RETHINK PRIORITIES

A review of GiveWell's discount rate

Global Health and
Development Department

October 2023 • Report

AUTHORS

Jenny Kudymowa — Senior Researcher, Rethink Priorities

James Hu — Research Analyst, Rethink Priorities

MANAGERS

Melanie Basnak — Senior Research Manager, Rethink Priorities

Tom Hird — Senior Research Manager, Rethink Priorities

A review of GiveWell's discount rate
Jenny Kudymowa — Senior Researcher, Rethink Priorities
James Hu — Research Analyst, Rethink Priorities
Melanie Basnak — Senior Research Manager, Rethink Priorities
Tom Hird — Senior Research Manager, Rethink Priorities

Suggested citation: Kudymowa, J., Hu, J., Basnak, M., & Hird, T. 2023. A review of GiveWell's discount rates. Rethink Priorities. https://rethinkpriorities.org/publications/a-review-of-givewells-discount-rate

Global Health and Development at Rethink Priorities

The Global Health and Development Department of Rethink Priorities is a multidisciplinary research and consulting group that seeks to disseminate actionable insights and drive positive change in the fields of global health, international development, and climate change.
https://rethinkpriorities.org/global-health-and-development

For all queries, please contact ghd@rethinkpriorities.org.

We thank GiveWell for commissioning and funding this research report. The views expressed herein are not necessarily endorsed by GiveWell.

GiveWell

RETHINK PRIORITIES
530 Divisadero St. PMB #796
San Francisco, California 94117
United States of America

About Rethink Priorities
Rethink Priorities is a research and implementation group that identifies pressing opportunities to make the world better. We act upon these opportunities by developing and implementing strategies, projects, and solutions to key issues. We do this work in close partnership with foundations and impact-focused nonprofits.
www.rethinkpriorities.org

Contents

Editorial note

This report was commissioned by GiveWell and produced by Rethink Priorities from June to July 2023. We revised this report for publication. GiveWell does not necessarily endorse our conclusions, nor do the organizations represented by those who were interviewed.

The primary focus of the report is to review GiveWell's current formulation of its discount rate by recommending improvements and reinforcing justifications for areas that do not require improvement. Our research involved reviewing the scientific and gray literature, and we spoke with 15 experts and stakeholders.

We don't intend this report to be Rethink Priorities' final word on discount rates, and we have tried to flag major sources of uncertainty in the report. We hope this report galvanizes a productive conversation within the global health and development community about discounting practices in cost-effectiveness analyses. We are open to revising our views as more information is uncovered.

Executive summary

Notes on the scope and process of this project

This project aims to serve the dual purposes of reviewing GiveWell's current approach to calculating its discount rate(s) to:

1. Provide recommendations to GiveWell on how its approach to discount rates could be improved.
2. Strengthen the justifications for its approach in cases where we do not recommend changes.

The direction of this project was mainly guided by our priors[1] that a prioritized investigation into three aspects could potentially make the biggest difference to GiveWell's discount rate:

1. A review of how other major organizations in the global health and development space (within and outside effective altruism) choose and justify their discount rates.
2. A review of GiveWell's overall approach to calculating discount rates to determine:
   a. Whether GiveWell should use a different overall calculation approach.
   b. Whether GiveWell should think differently about discounting consumption vs. health outcomes.
3. A review of the pure time preference component of GiveWell's discount rate.

We also reviewed several other components of the discount rate (consumption growth rate, compounding non-monetary benefits, temporal uncertainty), but decided to spend less time on those as we deemed it less likely to make major recommendations or expected it would be harder to make meaningful progress. Table 1 summarizes our recommendations for GiveWell's discounting practices.

The majority of this report focuses on the discount rate used for consumption benefits, as this appears to be the "main" discount rate used by GiveWell,[2] but we also discuss discounting of health benefits. We do not discuss discounting of costs in this report as (1) GiveWell's cost-effectiveness models rarely involve discounting costs, and (2) our general impression is that the typical approach across organizations is to discount monetary costs and benefits equally and we have seen very little discussion of alternative approaches.[3] A review of the shape of the utility functions[4] used is also out of scope for this review. Moreover, we focus exclusively

[1] Our priors are based on a combination of what we understood to be of high priority for GiveWell from speaking with Andrew Martin, and arguments and recommendations made in the "Change Our Minds" contest entries by SoGive and Julian Jamison.
[2] This is the discount rate that is discussed in GiveWell's 2020 discount rate write-up.
[3] The only case we've seen different discounting approaches being suggested for monetary benefits and costs is Dhaliwal et al. (2012, p. 38) (J-PAL): "The discounting of costs is representative of the choice a funder faces between incurring costs this year, or deferring expenditures to invest for a year and then incurring costs the next year. An organization or government's discount rate is usually calculated as the social opportunity cost of capital (SOC). [...] The discounting of benefits, on the other hand, represents how an end user of the program would trade off between the uses of the services this year versus next year. The appropriate discount rate for such a calculation is the social rate of time preference (SRTP) [...]." However, J-PAL itself does not use differential discounting of costs and benefits, but SOC for both outcomes.
[4] GiveWell uses three different utility functions for its cost-effectiveness analyses: an isoelastic utility function with $\eta = 1$ (also called log-utility) to model consumption benefits, an isoelastic utility function with $\eta = 1.59$ to calculate the "improving circumstances" or "wealth effect" component of the discount rate, and a linear utility function for health outcomes.

on temporal discounting.[5] If the time frame is not specified, all discount rates expressed as percentages are annual. Due to the variety of existing opinions and approaches with respect to discount rates and a relative lack of consensus, we opted to approach this project from a perspective of figuring out whether there are any compelling reasons to change GiveWell's current approach, rather than starting from scratch and coming up with a discount rate independently of the current approach.

[5] We are aware that GiveWell uses other types of discounting (e.g., generalizability/evidence discounting). We do not focus on those other types of discounting in this report.

Summary of recommendations

Table 1: Summary of Rethink Priorities' recommendations for GiveWell's discounting

Consideration	Current GiveWell choice	Rethink Priorities' recommendation	Comments
Overall annual discount rate	4.0%	4.3% (if current inconsistent choice of η is kept)	• The 0.3 percentage point increase is a result of an increased consumption growth rate estimate and a small change in the formula used to calculate the wealth effect. We recommend no other immediate changes. • However, as mentioned here, our recommended discount rate is also contingent on whether GiveWell decides to use a consistent utility curvature η across applications. For example, if $\eta = 1$, the wealth effect would be 0% and the resulting discount rate would decrease to 2.3%. • In several instances, our main reason for recommending to keep the current approach is that we could not find strong reasons to justify major changes. Thus, our recommendations do not always reflect a strong endorsement of a current approach, but can also reflect high uncertainty.
Overall approach to calculating the discount rate	Social rate of time preference (SRTP) approach	SRTP approach *[Confidence: Medium-high]*	• We argue that the following constitute compelling reasons to continue to use the SRTP approach: ○ In line with GiveWell's welfare-maximizing goals & focus on cross-intervention comparisons of cost- effectiveness; ○ Limited applicability of critiques in GiveWell's case; ○ Transparency of key assumptions; ○ Relative ease of use; ○ Commonly and increasingly used in practice. • However, we think the social opportunity cost of capital (SOC) approach could also be a reasonable choice.

GiveWell's current SRTP approach vs. Ramsey equation	• Determine individual components and sum them up • Calculate wealth effect implicitly in spreadsheet	• Use GiveWell equivalent of Ramsey equation, $r = \delta + (\eta - 1)g$. • This means that wealth effect is calculated explicitly via formula, $(\eta - 1)g$. • We recommend using a consistent η across applications. **Note that using η=1 in line with the implicit utility function would lead to a 0% wealth effect and thus a lower discount rate.** *[Confidence: High, though this is contingent on SRTP being the correct approach]*	• Current approach can already be considered a variant of the Ramsey equation that is more suited to GiveWell's modeling; we just think this should be more explicit. • Main difference to Ramsey: Ramsey is based on absolute increases in consumption, whereas GiveWell's model is based on percent increases, which yields a different wealth effect. • We found and fixed a small calculation error in GiveWell's calculations of the wealth effect. • We think that using a higher η for calculating the discount rate than for the underlying utility function to model income benefits (as is done currently) risks overdiscounting of benefits. However, we have not reviewed which assumption for η would be the best choice.
Discounting consumption vs. health benefits	Discount health benefits using only the temporal uncertainty component	• Tentatively keep current approach • Continue to discount health benefits at a lower rate than consumption *[Confidence: Low-medium]*	• We find that discounting health and monetary outcomes at equal rates is still dominant but do not find the arguments in favor of the dominant practice convincing. • Instead, we concur with the view, broadly and increasingly supported in the economic literature, that health outcomes should be discounted at a lower rate. • However, given the lack of a consensus view in the literature and various uncertainties such as those with respect to the shape of the utility function, we could not devise a superior approach to GiveWell's discount rate for health outcomes.
Consumption growth rate	• 3% consumption growth rate	• Raise consumption growth rate to 3.3%	• We recommend using published projections of real GDP per capita growth rates instead of eyeballing past GDP figures.

	• This results in a 1.7% discount rate from improving circumstances, or the wealth effect	• This results in a 2.0% discount rate from the wealth effect (according to the correct wealth effect formula) • We expect the growth rate to decline over time and recommend next revisiting this estimate in 2028 *[Confidence: Medium]*	• Our approach considers both sub-Saharan Africa and South Asia as part of a population-weighted composite. • We define time *windows* during which growth is being projected; our approach anchors the time interval to the longest-effect duration programs. • We define time *periods* during which the current wealth effect estimate applies; we suggest revisiting the estimate at the end of each period (5.2 years).
Pure time preference rate	0%	Keep current assumption of 0% *[Confidence: Medium-high]*	• Our impression is that the predominant opinion in the philosophical literature is that δ = 0%. Several arguments in favor of δ > 0% exist, though these seem highly context-dependent and somewhat controversial. • Empirical estimates of individuals' time preferences are extremely noisy and seem implausibly high to us; thus, not very useful. In a large expert survey on discount rates, δ = 0% is the modal response. • δ = 0% is rarely used in practice. In many (though not all) cases where researchers/organizations assume a positive δ, it actually reflects a positive temporal uncertainty component rather than what GiveWell considers pure time preferences. • Some theoretical articles show that δ = 0% leads to absurd predictions, but we do not think those specific predictions would apply in GiveWell's case.
Temporal uncertainty	1.4%	Tentatively keep current assumption of 1.4% *[Confidence: Low]*	• Temporal uncertainty has traditionally been defined more narrowly, as the risk of death or human extinction, than GiveWell does. GiveWell also considers changes in economic structure, catastrophe, and political instability.

			• The current GiveWell assumption is roughly in line with other estimates, e.g.: other organizations with links to effective altruism use a broad range of estimates for existential risk ranging from 0.1% to 2.3%.
Compounding non-monetary benefits	0.9%	• Tentatively keep current assumption of 0.9% • More reasoning transparency in public write-up *[Confidence: Low]*	• Including a component for non-monetary benefits sounds intuitively plausible to us, though we have not investigated the magnitude and the way of modeling. • Our main reason for keeping the current assumption is that we couldn't find strong counterarguments quickly. • The SPC approach to discounting is the only approach we know of that accounts for reinvestment of benefits, but it is highly impractical to use and covers only a part of what GiveWell tries to capture in this component. • We think a higher level of reasoning transparency in GiveWell's public write-up would facilitate more critical engagement with this component.

Review of discount rates in global health and development

We reviewed the literature on different theoretical and practical approaches to social discounting[6] to get a sense of whether there is any consensus on which approach is best suited and/or most commonly used in the philanthropic global health and development context. We first provide a brief theoretical overview of approaches (here), followed by a review of approaches used by major organizations in practice (here). This ultimately serves to answer the question of whether any particular approach seems preferable to GiveWell's current approach, which we discuss here.

Theoretical approaches to discounting: SRTP and SOC are the main theoretical approaches to constructing social discount rates

[Confidence: High. We are highly confident that we identified the most relevant and most commonly used approaches to social discounting, and we think it's unlikely that additional desk research would change our views. Additional literature on discounting from the perspective of philanthropies might change our views, but very little (if anything) has been published on this topic so far.[7]]

While many different approaches to social discounting exist, we are fairly confident that two approaches (and some variations thereof) are most relevant for GiveWell, which respectively represent consumer/beneficiary and producer/funder preferences:[8,9]

1. The **social rate of time preference (SRTP)** approach reflects the rate at which consumers — in GiveWell's case, donors or beneficiaries — prefer benefits[10] now rather than later, r_{SRTP}. There are various reasons for preferring benefits sooner rather than later, e.g., impatience or uncertainty about the future. **GiveWell currently uses this approach to discount its benefits.**
    a. Our impression is that the **Ramsey equation is the most commonly used SRTP method** (see here for how the Ramsey equation relates to GiveWell's approach). It is essentially a formula derived from a representative-agent growth model in economics.
2. The **social opportunity cost of capital (SOC)** approach reflects the rate at which the funds used for public — here, charitable — projects would otherwise yield returns in the private sector, r_{SOC}.

In a perfectly competitive market, SRTP and SOC approaches converge such that $r_{SRTP} = r_{SOC}$ (and both are equal to the market interest rate). In the real world, market distortions cause SOC

[6] According to Wikipedia, the "social discount rate (SDR) is the discount rate used in computing the value of funds spent on social projects" ("Social discount rate," 2022). Our impression is that this differs from private discount rates in that social discounting takes more social and ethical considerations into account, such as intergenerational and distributional equity, whereas private discounting is more concerned with profit maximization.
[7] In an email, Mark Moore (senior lecturer at Simon Fraser University's School of Business) said, "As far as we know, there is no published peer-reviewed literature on this topic. [...] Prof. Vining and I are contemplating writing something on this topic, and will keep you informed if we do."
[8] See Zhuang et al. (2007, pp. 2-14) for an excellent overview of the main social discounting approaches.
[9] Note that we've seen these approaches being discussed exclusively in the context of the public sector, not for the philanthropic sector.
[10] To be clear, the benefits are received by the beneficiaries. However, the preferences over the beneficiaries' benefits can be considered from the donors' or from the beneficiaries' standpoint.

approaches to generally yield higher discount rates than SRTP approaches, such that $r_{SRTP} < r_{SOC}$.[11] While some institutions have elected to use either r_{SRTP} or r_{SOC}, or a mixture of both for differential discounting of costs and benefits, others have adopted harmonizing approaches that select an intermediate value.[12]

Two harmonizing approaches are the "weighted average approach" (also known as the Harberger approach) and the shadow price of capital (SPC) approach. The weighted average approach proposes a weighted average of r_{SRTP} and r_{SOC}, with weights reflecting the proportions of public — or charitable — funding that constitutes forgone consumption and forgone investment, respectively.[13] The SPC approach builds upon the weighted average approach to reflect the value of reinvesting the social benefits from investment — or philanthropy — in the private sector.[14,15]

There is no clear theoretical consensus on which approach is generally most appropriate, as all of them have different advantages and drawbacks and are suitable for different contexts (see Appendix C for a brief general discussion of this). See here for an explanation of why we recommend SRTP for GiveWell's case.

Overview of institutional approaches: Other institutions' approaches do not provide useful lessons for GiveWell

[Confidence: Medium-high. We are confident that we have identified publicly available discount rates for most of the significant institutions that conduct cost-effectiveness analyses in global health and development. However, our survey of institutions was not exhaustive, and extra work, in particular direct outreach to personnel at relevant organizations such as the Copenhagen Consensus Center, which did not reply to our email query, could better illuminate their practices.]

See this spreadsheet for a non-exhaustive compilation of annual discount rates and approaches used by governmental and nongovernmental institutions in the global health and development sphere (~10 hours of desk research). We also surveyed several organizations with links to effective altruism and documented their policies (or lack thereof) in the spreadsheet.

---

[11] "In reality, the market is often distorted due to various imperfections. A typical example of imperfection is the taxes imposed on corporate incomes and individuals' interest earnings. Other examples are risks, information asymmetry, and externalities. These imperfections create a wedge between SRTP and SOC (with the former generally lower than the latter), and make both deviate from the market interest rate. Under such circumstances, the market interest rate will not reflect the marginal social opportunity cost of public funds, and the latter will vary depending on whether it is measured in terms of SRTP or SOC" (Zhuang et al., 2007, p. 3).

[12] By "mixture of both," we mean that some organizations use different discounting approaches for different outcomes, e.g., SRTP for benefits and SOC for costs. By "harmonizing approaches," we mean approaches that use some metric in between SRTP and SOC across different outcomes.

[13] According to the weighted average approach, the social discount rate can be calculated as: $r_{WA} = \alpha\, r_{SOC} + (1 - \alpha - \beta)\, i_f + \beta\, r_{SRTP}$, where "$i_f$ is the government's real long-term foreign borrowing rate, α is the proportion of funds for public investment obtained at the expense of private investment, β is the proportion of funds obtained at the expense of current consumption, and $(1 - \alpha - \beta)$ is the proportion of funds from foreign borrowing" (Zhuang et al., 2007, p. 12).

[14] A project's "shadow price of capital" is defined as the value of forgone future consumption caused by project spending's displacement of one unit of private investment, or the value of future consumption generated by reinvesting one unit of project benefits in the private sector (Zhuang et al., 2007, p. 13).

[15] We also came across the Epstein-Zin approach (Epstein & Zin, 1989). We don't think this approach is relevant for GiveWell, as it is mainly suited for the finance/asset pricing context. It provides a very flexible discount rate model, but at the expense of being very computation- and data-intensive.

Our survey of other institutions' approaches did not yield any major recommendations for GiveWell's discounting practices. In particular, we found:

- **Context differences.** Many institutions with well-justified, transparently determined discount rates operate in high-income countries, as opposed to low- and middle-income countries of the sort relevant to GiveWell-endorsed charities. Government expenditures also face distinct trade-offs that do not apply within GiveWell contexts.[16]
- **Lack of institutional positions.** Even institutions working in similar contexts — such as the Bill & Melinda Gates Foundation, Copenhagen Consensus, and Open Philanthropy — do not take (or publicize) clear, consistent positions on time discounting. It is therefore difficult to assess the extent to which their discounting practices apply to GiveWell's situation.
- **Frequent deference.** Particularly among non-government actors, we find that deference to other organizations' established practices is common. For example, within academic global health research, a 3% rate seems to have endured largely because of recommendations by the US Panels on Cost-Effectiveness in Health and Medicine (see Appendix A), while, within EA, organizations such as Charity Entrepreneurship and its incubatees rely on GiveWell's rate (more here).

Comparison of GiveWell's discount rate and approach with the rest of our sample:

- **There is considerable variation in discounting rates and approaches. GiveWell's 4% discount rate and SRTP approach[17] toward time discounting puts it in company with a sizable group within our sample.**[18]
    - A slim majority (54%) of the discount rates in our sample fall within the 3%-7% range, and a slim majority (53%) of the discount rates with clear justifications were devised using the SRTP approach. However, among discount rates that fall within the 3%-7% range, only a bare majority — 52% — of those were based on the SRTP approach.
    - Examples of institutions whose discount rates fall within the 3%-7% range and whose rates were derived using the SRTP approach include the UK Treasury's 3.5% rate[19] (HM Treasury, 2022, pp. 116-118) and the Spanish government's 4% and 6% rates for water and transport projects, respectively (Zhuang et al., 2007, p. 17).
    - GiveWell's 4% discount rate is also similar to — only slightly higher than — the 3% discount rate recommended by the First and Second Panels on Cost-Effectiveness in Health and Medicine (Lipscomb et al., 1996, p. 232; Basu & Ganiats, 2016, p. 278), which may have significantly shaped global health research norms (see also Appendix A). The First Panel — in contrast to GiveWell's use of the SRTP approach — justified its choice of rate using the SPC approach, but the

[16] For instance, it seems much more sensible to think of a high-income country's government expenditures — used for purposes other than foreign aid — as producing relatively high opportunity costs in private benefits (consumption and investment), in part due to the relatively broad remit of national governments, and in part because the majority of beneficiaries of fiscal policies in high-income countries already enjoy relatively high living standards. By contrast, GiveWell has a narrow remit to fund health and development interventions in some of the world's poorest countries using donations from the world's richest countries. These contrasts mean that high-income contexts have relatively few lessons to offer GiveWell.
[17] See more on GiveWell's specific approach here.
[18] Note that there are duplicate entries in our sample, because some institutions apply different rates in different contexts, and because our sample includes historical as well as current and proposed rates.
[19] Although health impacts are assessed at 1.5%, outside the 3%-7% range.

Second Panel — similarly to GiveWell — opted for the SRTP approach (Basu & Ganiats, 2016, pp. 284-286).[20]

- **GiveWell's 4% discount rate is lower than most LMIC-context rates in our sample. However, there is considerable variation in discounting practices across all institution categories, including among institutions focusing on LMIC contexts.** See also our spreadsheet as filtered by "LMIC" in the "Context" column.
    - GiveWell's 4% discount rate is lower than all those of multilateral development banks (range: 9%-12%) and all LMIC governments (range: 8.5%-15%) in our sample. LMIC governments also generally tend to set higher rates than HIC governments;[21] from our sample, GiveWell's discount rate appears to resemble HIC rates (range: 1%-10%) much more than LMICs'.
    - There is huge variation in aid agencies and research organizations that specialize in LMIC projects, with some reporting zero time preference and others using considerably higher discount rates than GiveWell. For example:
        - US aid agencies such as the International Development Finance Corporation and the Trade and Development Agency do not require discounting future benefits or costs, yet the Millennium Challenge Corporation uses a 10% discount rate (Kashi et al., 2022, p. 26).
        - USAID's practices vary significantly by program, with its education program using a zero discount rate (USAID, 2021, p. 62) and its general cost-benefit analysis guidelines recommending 12% (Kashi et al., 2022, p. 27).
- **GiveWell's 4% discount rate exerts moderate influence within the global health and development/wellbeing wing of effective altruism.** See also our spreadsheet as filtered by "EA" in the "Category" column.
    - Some organizations defer to GiveWell's discounting practices. For example, Charity Entrepreneurship stated in an email that it "defer[s] to GiveWell's 4%

[20] The Second Panel does not explicitly explain why it discarded the SPC approach endorsed by the First Panel. Instead, it lays out two competing theoretical conceptions of health discounting — "welfarist" and "extra-welfarist" approaches. Although the Second Panel ultimately does not reason from first principles toward a conclusive quantitative recommendation, both approaches it considers viable fall under the SRTP category. The welfarist approach takes the perspective of a society that is seeking to maximize the consumption value of health, and applies the Ramsey formula to both consumption benefits from health and costs. The extra-welfarist approach takes the perspective of a health-care sector operating within a fixed budget that is seeking to maximize health; it applies the social rate of time preference for health — rather than that for consumption — to health benefits and applies a modified version of the Ramsey formula to costs (Basu & Ganiats, 2016, p. 284). Ultimately, the Second Panel neither endorses one perspective over the other nor reconciles the two into a unified approach. Instead, it states: "Considerable uncertainty exists around each of the parameters that together comprise the appropriate discount rates from the societal and the healthcare sector perspectives. Because the goal of the Reference Cases is to promote comparability across studies, we recommend that a 3% interest rate be used for both costs and health effects in both the societal and the healthcare sector perspective analyses. [...] This preserves the recommendation made by the original Panel" (Basu & Ganiats, 2016, p. 286).

[21] There are two general explanations for why some discount rates are higher than others. First, in an analysis for the Asian Development Bank, Zhuang et al. (2007) attribute these differences to "the different analytical approaches followed" (p. 20). While this continues to hold true for our larger sample, we observe considerable overlap between the distributions of two of the main approaches followed, with SRTP-based rates in our compilation ranging from 1% to 9% and SOC-based rates ranging from 2.25% to 15%. Second, holding all else equal, the SRTP approach should yield generally higher discount rates in LMIC contexts than in HIC contexts, given that LMICs generally experience more rapid growth of consumption per head. Analysis of implemented discount rates above 0% shows that this generally holds true for our sample, although there is considerable overlap between the distributions of LMIC- and HIC-context rates, with HIC rates ranging from 1.5% to 10% and LMIC rates ranging from 3% to 15%.

[discount rate] in all cases."[22] As a funder, GiveWell's discount rates have directly influenced its commissioned research, including that by IDinsight.[23]

- Organizations with links to effective altruism that do not defer to GiveWell's discounting practices include Open Philanthropy and SoGive. Their discounting practices are summarized in Table 2. In addition, several organizations either do not currently have institutional positions on discounting or could not be reached.[24]

Table 2: Discounting practices of organizations with links to effective altruism

Organization	Discounting approach	More detail
Open Philanthropy (Global Health & Wellbeing focus areas)	No consistent approach, but generally SRTP	• Zero rate of pure time preference • Account for lower impacts due to improved circumstances (e.g. lower disease burdens or higher incomes in the future[25]) • Over longer time horizons (>25 years) consider existential risk or transformative artificial intelligence, 0.225%

SoGive	Tentatively recommend 4.8% for GiveWell	• In GiveWell review (Joshi, 2022), supposing Ramsey equation, do not recommend against GiveWell parameters of δ = 1.4% (consisting of zero pure time preference and 1.4% temporal uncertainty), η = 1.59, g = 3%
		  ○ This yields an overall rate of 6.2%
		• Supposing current GiveWell construction (at the time of its post, SoGive "would likely side with GiveWell's approach"):
		  ○ Raise 1.4% temporal uncertainty to 2.3% (relabeled "catastrophic risks")
		  ○ Lower overall rate by 0.1 percentage points due to uncertainty in discount rate
		  ○ This yields an overall rate of 4.8% and represents SoGive's tentative recommendation for GiveWell

[22] Consequently, nonprofits incubated by Charity Entrepreneurship, such as Lead Exposure Elimination Project (LEEP), also use 4% (LEEP, 2022).
[23] IDinsight does not have a "cohesive approach" to discounting according to its chief economist Dan Stein, but it used 4.2% in 2018, per GiveWell's staff median discount rate at the time, in its report on beneficiary preferences, which was funded by GiveWell (Redfern et al., 2018, footnote 158).
[24] The Happier Lives Institute does not currently take an institutional position on time discounting, according to Joel McGuire in an email. The Life You Can Save's cost-effectiveness analyses do not currently implement time discounting, which "remains an area of exploration," according to Katie Stanford in an email. Founders Pledge did not respond to our email query. Although Hoeijmakers's (2020) report "Investing to Give" refers to discount rates, we did not find any information that indicated Founders Pledge has taken any institutional positions on temporal discounting.
[25] Sam Donald said, "We value increases in log income, so a 10% increase in income is valued the same regardless of someone's income level, but a $10 increase in income to a high-income person would be valued less highly than a $10 increase to a low-income person. As such, if considering income impacts in absolute terms, we would adjust for growth in incomes (e.g if incomes will be 20% higher at time t+1, we will adjust any $ impact down by ~20% vs if it occurred now)."

Review of GiveWell's current discount rate

In the following, we first outline why we believe the SRTP approach is the most suitable for GiveWell (here). We then review GiveWell's current version of an SRTP approach vs. the most common SRTP method (the Ramsey equation) in the literature (here) and what the literature says about discounting health outcomes (here). We finally review several individual components of GiveWell's discount rate (here).

Given theoretical and practical considerations, we recommend that GiveWell keep using the SRTP approach

[Confidence: Medium-high. We have high confidence that the SRTP approach is a reasonable choice and has several advantages for GiveWell, but we also believe that a plausible case could be made to use the SOC approach instead. We do not think further desk research would be fruitful, but a conversation with one or several major experts on social discount rates (e.g., Mark Moore, Anthony Boardman, Aidan Vining) could help alleviate some remaining uncertainties. However, even these experts seem to have some disagreements on the different approaches.]

Based on theoretical considerations (here and in Appendix C) and evidence on discounting approaches used in practice (here), **we recommend that GiveWell keep using the SRTP approach for the following reasons:**

1. The SRTP approach reflects the goal of maximizing social welfare (i.e., the present value of the utility of a representative beneficiary)[26] and is suited for cross- intervention comparisons of cost-effectiveness,[27,28] whereas the SOC approach takes the perspective of a profit-maximizing investor. Thus, we think that the **SRTP is more aligned with GiveWell's philanthropic goals and focus on cross-intervention comparisons of cost-effectiveness.**

2. We think the **main critique of the SRTP approach of ignoring the opportunity cost of foregone private sector investment has limited applicability in GiveWell's case.**

    a. The SRTP approach has been criticized for ignoring the opportunity costs of foregone private sector investment, which is accounted for by SOC. However, SOC implicitly assumes that the next best use of public or philanthropic funds would be private sector investment. We expect (though we have not verified) that GiveWell's next best use of its funds would be other philanthropic initiatives rather than private sector investment. Moreover, GiveWell's funds are typically donated rather than borrowed, and thus do not have an (or have only a small) effect on interest rates that could crowd out private sector investment.

[26] "The STP method is based on the idea that the fundamental goal in welfare economics is to maximize the utility (or "happiness") of society (or of a representative individual), where utility depends on per capita consumption in present and future time periods" (Moore et al., 2013, p. 3).
[27] In an email, Mark Moore said, "Our first reaction is that, if a charity is concerned with the well being of the citizens of some country or the world, and is choosing between two interventions with different intra-generational time profiles, then our recommended approach would be to discount using a social rate of time preference-based social discount rate. [...] [I]f your concern is cost-effectiveness of two different interventions, then I would argue that the STP approach is the one to follow."
[28] When prompted with a question about the relative merits of SRTP and SOC in a philanthropic context, Anthony Boardman (professor emeritus at the University of British Columbia) responded that "SRTP is the correct approach" and said that "SOC had lost the debate." Boardman also referred us to the White House's proposed changes to Circular A-4 (see also Piper, 2023).

b. If GiveWell's main concern is spending its funds now vs. later (and investing the funds in the private sector in the meantime), then SRTP is still the recommended option, at least according to one major expert we've contacted.[29] However, another expert we contacted noted that some scholars would disagree with the former view and recommend SOC instead.[30] We don't have a clear stance on this point as the two experts shared their opinions without adding clear justifications. Nonetheless, given that GiveWell's main focus seems to be cross-intervention comparisons of cost-effectiveness rather than now vs. later comparisons of interventions, we recommend the SRTP.

3. We also like that the **SRTP approach makes assumptions and value judgments transparent**. SOC is sometimes considered the more objective social discounting approach as it reflects observable market returns. According to Creedy and Passi (2018), this is not the case, as "the very decision to select a SOC-based approach carries a number of implicit assumptions and value judgments. SRTP-based approaches require more transparent statements of the decision maker's value judgments."[31]

4. We think the **SRTP approach is relatively easy to use**, at least in comparison to the weighted average and the SPC approaches. While SRTP requires several parameters to be determined, our impression is that a lot of guidance on this exists in the literature. Moreover, SRTP is nowhere near as computation-and data-heavy as the weighted average and the SPC approaches.

5. **Lastly, the SRTP approach is commonly and increasingly used in practice.** As we found in our convenience sample of social discount rate approaches used in practice (see spreadsheet here; see text here), about half of the discount rates were devised using the SRTP approach. Moreover, the influential Panel on Cost-Effectiveness in Health and Medicine switched from using an SPC approach to SRTP in its most recent discount rate recommendation. Other researchers also found that SRTP is used more and more often in practice.[32,33]

Nonetheless, our impression is that all aforementioned theoretical approaches (i.e., SRTP, SOC, weighted average, and SPC) are well-respected methods for calculating social discount rates and could be reasonably justified. Given that the two harmonizing approaches (weighted average and SPC) are difficult to implement (as discussed in Appendix C), we think those are

[29] According to Anthony Boardman, "Your cash flows will depend on how well the investments will do. Your decision to start a project now or later should be based on discounting at the SRTP."
[30] Some scholars would take a different position if a charity has an endowment (e.g., BMGF), according to an interviewee. By spending now rather than later, such a charity would forgo the expected additional returns the endowment would have accrued. This would then suggest discounting at the endowment's risk-adjusted expected rate of return.
[31] "SOC-based approaches are traditionally regarded as being more objective, based as they are on observable market returns. However, the very decision to select a SOC-based approach carries a number of implicit assumptions. These include assumptions that market-based counterfactuals provide an appropriate counterfactual for public projects, public projects fully crowd-out private sector projects of equal magnitude, political decision makers should trade-off the future in the same way that individuals and businesses do when making decisions about their own personal consumption and investment—or all relevant preferences are measurable and accounted for in CBA cash flows. Also it is assumed that the way markets evaluate and price risk are also how governments evaluate and price risk" (Creedy & Passi, 2018, p. 154)
[32] "As the united opinion on the most suitable SDR approach does not exist, the analysis of scientists' researches shows that the priority is given to the SRTP approach more and more often" (Kazlauskienė, 2015, p. 464).
[33] SRTP is also the canonical approach used for discounting of intergenerational projects (Zhuang et al., 2007, p. 16). This usually involves timelines of more than 40 years, which seems less relevant for GiveWell.

less interesting for GiveWell. However, **we think a case could be made to use the SOC approach:**

- **The SOC approach could potentially be easier to use than SRTP**, as it is based only on one parameter (foregone private sector investment returns) whereas the SRTP approach requires determining several parameters (as we show here). SOC also requires fewer value judgments (e.g., concerning the wellbeing of future individuals).
- **SOC seems to be fairly commonly used by governments in LMICs** (see spreadsheet). It is also used by J-PAL who chose it mainly due to its practicality stating that "because of the high variance and scarce empirical data on time preferences in the developing world, the SRTP is not a practical option."[34]
- **SOC could potentially be more suitable if GiveWell's main concern was about spending its funds now vs. later** (rather than a comparison across different interventions), though not all experts agree with this view.

We recommend that GiveWell keep the current discount rate construction method over the Ramsey equation, but calculate it with a more explicit formula

[Confidence: High. We are fairly confident that we accurately represented the key difference between the Ramsey equation and the current GiveWell approach, and that the explicit formula we provide makes sense (though this is contingent on SRTP being the correct approach for GiveWell). We are also fairly confident that assuming a consistent utility curvature (η) would be preferable, though we have not considered what assumption of η would be preferable (out of scope). The highest uncertainty we currently have is how to conceptually think about the 'compounding non-monetary benefits' component and whether considering it a part of the 'pure time preference rate' component is the best way to think about it. Speaking with an expert on social discount rates (e.g., Mark Moore) might provide more clarity.]

We believe that GiveWell's current discount rate construction method is preferable to the Ramsey equation, as GiveWell's modeling works in a way which is not directly accommodated by the Ramsey equation (i.e., modeling consumption increases in % increases vs. absolute increases). Moreover, we believe that GiveWell's approach could already be considered a variant of the Ramsey equation adapted to GiveWell's specific modeling assumptions/context. Thus, we do not recommend a major change in the construction method (apart from fixing one small calculation error we found). However, we recommend calculating the discount rate (specifically the wealth effect component) with an explicit formula, rather than calculating the wealth effect implicitly in a spreadsheet. We also

[34] "The discounting of benefits, on the other hand, represents how an end user of the program would trade off between the uses of the services this year versus next year. The appropriate discount rate for such a calculation is the social rate of time preference (SRTP) [...]. There is relatively little information on the time preferences of people in poorer countries, and the fact that variations will depend upon the intended user of the program, rather than the implementer, makes it difficult to choose one rate which would be applicable in a variety of cases. If an organization were performing a cost-effectiveness analysis of programs that they run in particular countries, then it would be possible to use the SOC to discount its costs knowing its own cost of capital and use the SRTP of the country in which beneficiaries live to discount effects. However, in performing general cost-effectiveness analysis that is likely to be used by policymakers in different organizations and countries, one is unlikely to have such specific information about users, and so it is practical to choose a single discount rate. Because of the high variance and scarce empirical data on time preferences in the developing world, the SRTP is not a practical option. This suggests that the SOC may be the best available discount rate [...]." (Dhaliwal et al., 2012, pp. 38-39).

recommend that GiveWell reconsider the inconsistent use of utility functions and consider using a consistent η (though the choice of η is out of scope for this report).[35]

As explained here, we recommend that GiveWell keep using an SRTP approach for calculating its discount rate. As pointed out in SoGive's review of GiveWell's discount rates, GiveWell uses a non-standard SRTP approach to calculate its discount rate, which seems to stand in contrast with the Ramsey (1928) approach, the (by far) most commonly used SRTP approach. Using the Ramsey equation would yield a higher discount rate than GiveWell's approach does (Ramsey: 7.1%[36] vs. GiveWell: 4%) and the 'improving circumstances' component alone would almost triple (Ramsey: 4.77%[37] vs. GiveWell: 1.7%). As SoGive did not have a very firm conclusion on which approach is more suitable,[38] we decided to further investigate this topic. **Given the significant difference in discount rates from the two approaches, would it make more sense for GiveWell to adopt the Ramsey equation? We don't think so,** as we explain in the following.

GiveWell's current discount rate is 4% and is calculated as a sum of several components, as shown in Table 3.[39]

Table 3: GiveWell's discount rate and its components

Component	Value
Improving circumstances/wealth effect	1.7%
Temporal uncertainty	1.4%
Pure time preference (beneficiaries)	0%
Pure time preference (donors)	0%
Compounding non-monetary benefits	0.9%
Overall discount rate	4.0%

What is the Ramsey equation? According to Ramsey (1928), a social discount rate can be calculated as follows:

$$r_{Ramsey} = \delta + \eta g$$

where r is the discount rate, δ is the rate of pure time preference (also called the "utility discount rate," and ηg is a "wealth effect," which consists of η, the elasticity of marginal utility of consumption, and g, the consumption growth rate. How does this relate to GiveWell's

[35] Note that there have been some changes compared to a previous version of this report. We previously worked under the assumption that the unit to be discounted was utility, which we realized is not actually the case.
[36] According to SoGive, GiveWell's discount rate under the Ramsey equation would be 6.2%. However, we think it makes more sense to consider all discount rate components that are not part of the wealth effect as parts of the pure time preference rate component (δ), such that $r = (0\% + 1.4\% + 0.9\%) + 1.59 \times 3\% \cong 7.1\%$. See Table 3 for an overview of GiveWell's discount rate components.
[37] We calculated 4.77% by simply inserting GiveWell's parametric assumptions ($\eta = 1.59$; $g = 3\%$) into Ramsey's wealth effect of ηg (see here for GiveWell's assumptions).
[38] "If we were to select a view on this at this stage, we would likely side with GiveWell's approach, given its intuitiveness, but believe that it is unwise to assume that this is correct without further research." (SoGive, 2022, p. 6).
[39] An overview of GiveWell's current approach to calculate its discount rate can be found here. A bit more detail is available in older documents by James Snowden and Caitlin McGugan.

approach? The "wealth effect" (ηg) in the Ramsey equation corresponds to GiveWell's "improving circumstances" component. In the following, we use the terms "wealth effect" and "improving circumstances" interchangeably.

There are different interpretations of Ramsey's rate of pure time preference (δ). It is typically interpreted as including both what GiveWell calls pure time preference and temporal uncertainty (e.g., OECD, 2018, p. 5), though we think it makes sense to interpret it as the sum of all factors with which we discount utility (as, according to Ramsey (1928), r is used to discount consumption, and δ is used to discount utility). We consider GiveWell's "compounding non-monetary benefits" component to be a part of Ramsey's δ. Thus, we consider δ as comprising the rate of pure time preference, compounding non-monetary benefits, and temporal uncertainty, in GiveWell's terms.

The core difference between the Ramsey equation and GiveWell's approach then lies in how the wealth effect is calculated. Ramsey's wealth effect is calculated by ηg, which results from a growth model that was used to derive the optimal savings rate of a representative agent. GiveWell's wealth effect is not calculated with any explicit formula, but implicitly calculated in this spreadsheet by equating marginal utilities over time.

Fundamentally, both wealth effects are calculated very similarly: (1) Both approaches are based on the same assumed (isoelastic) utility function, and (2) both approaches derive utilities using an indifference equation that equates marginal utilities over time (see Appendix B for a short derivation of the Ramsey equation that shows this indifference equation). **GiveWell's approach to calculating the wealth effect can be considered a variant of the Ramsey equation, with one major difference: In GiveWell's CEAs, increases in consumption are modeled as percentage increases (+X%), whereas the Ramsey equation is based on increases in absolute terms (+$Y).** This distinction affects the resulting discount rate.[40] Thus, it would not make sense for GiveWell to calculate the wealth effect according to the Ramsey equation (ηg).

We think GiveWell's discount rate should be expressed as:

$$r_{GiveWell} = \delta + (\eta - 1)g$$

where $(\eta - 1)g$ is GiveWell's wealth effect, instead of ηg as in the Ramsey equation.

We derived this equation in two steps:

1. We did a small correction to GiveWell's spreadsheet calculations, as we think there may have been an error. Currently, the wealth effect is calculated using the following indifference equation: $U_{C0} * f = U_{C1}$ where $f = 1/(1 + r)$ is the discount factor due to the wealth effect, r is the corresponding discount rate and U_C represents marginal utility in years 0 and 1. Thus, GiveWell's wealth effect aims to discount marginal utility in year 0 such that it equals marginal utility in year 1. GiveWell calculates its discount rate as $r = 1 - \frac{U_{C1}}{U_{C0}} = 1 - f$. However, this is not what we get if we solve the above equation, $f = 1/(1 + r)$, for r. Solving the equation for r, we think the correct approach would be to calculate the discount rate as $r = \frac{U_{C0}}{U_{C1}} - 1$. We believe that GiveWell's wealth effect needs to be multiplied by $\frac{U_{C0}}{U_{C1}}$ to be correct. Thus, GiveWell's wealth effect ends up being

[40] This point was also made in SoGive's (2022) review (p. 5).

slightly underestimated, though the difference is small. If we recalculated GiveWell's wealth effect (keeping the parametric assumptions of g = 3% and η = 1.59), the resulting wealth effect would be 1.8% (instead of 1.7%).

2. After applying the aforementioned correction, we essentially used Tol's (2015) version of Ramsey's derivation of the discount rate (as shown in Appendix B), but replaced the marginal utilities of a unit increase in consumption by the marginal utility of a percentage increase in consumption. Note that the resulting discount rate formula we recommend GiveWell use, $\delta + (\eta - 1)g$, is an approximation of the precise wealth effect, which would be $(1 + \delta)(1 + g)^{\eta-1} - 1$. However, the same holds for the Ramsey discount rate, as $\delta + \eta g$ is also an approximation of the precise Ramsey discount rate $(1 + \delta)(1 + g)^{\eta} - 1$. See Appendix E for a more detailed derivation of this.

Another aspect which we think is important is the type of unit that is discounted. The Ramsey discount rate, $r_{Ramsey} = \delta + \eta g$, is meant to discount consumption, and δ is meant to discount utility (which is why δ is also called the "utility discount rate").[41] In GiveWell's case, however, neither of those units are discounted, but instead consumption growth (expressed as increases in log consumption) is discounted. Thus, if GiveWell were to change the unit used, a different discount rate may be applicable (see Table 4 below).

Table 4: Appropriate discount rate depending on the unit to be discounted

Benefit unit	Appropriate discount rate
Consumption flows (units of consumption)	$\delta + \eta g$
Consumption growth (ln consumption)	$\delta + (\eta - 1)g$
Utility flows (utils)	δ

Note. Consumption growth (ln consumption) is GiveWell's currently used unit.

There are several other points related to GiveWells discount rate construction method that we'd like to discuss:

- **We recommend that GiveWell reconsider its inconsistent use of utility functions for consumption and choose a consistent η.**
  - There is an inconsistency in how GiveWell uses utility functions for discounting as the curvature of the utility function for consumption (η) differs across applications: GiveWell assumes η = 1 for modeling consumption benefits, but assumes η = 1.59 for calculating the wealth effect component of the discount rate. This might seem inconsequential at first glance when looking at GiveWell's CEAs, as what is discounted is not utility but consumption growth.
  - However, our understanding is that the CEAs are implicitly based on a utility function with η = 1 (i.e., log utility). GiveWell uses changes in consumption growth as a means to calculate consumption/income doublings, which are then combined or compared with other intervention outcomes using moral weights. We have not seen GiveWell's utility function explicitly stated, but our understanding is that income doublings can only be meaningfully counted if a

[41] This probably gets clearer if you look at the derivation of the Ramsey discount rate in Appendix B.

doubling in income is worth the same increase in utility, no matter what the starting income is. This is, in our understanding, only the case if $\eta = 1$ in the isoelastic utility function. Thus, it seems to us that, at least implicitly, GiveWell assumes a utility function with $\eta = 1$, which is inconsistent with the utility function used to derive its wealth effect here that is based on $\eta = 1.59$.
   - We are aware that this inconsistency was a conscious decision, and $\eta = 1$ (i.e., log-utility) is used for convenience in GiveWell's CEAs.[42] **If GiveWell chose to consistently assume $\eta = 1$, the discount rate would be significantly lower, as the wealth effect would be zero (wealth effect: (1-1)*g=0%; resulting discount rate would drop from 4% to 2.3%).**[43]

- **We recommend including either pure time preferences of beneficiaries or of donors, but not both.** In the Ramsey approach, there is only one pure time preferences rate, which can either be interpreted as the beneficiaries' or the donors' time preferences. GiveWell's current approach sums up the pure time preferences of both beneficiaries and donors. This decision is currently inconsequential as GiveWell assumes both of these to be 0, but it becomes relevant in case GiveWell decides to increase them. We have not reviewed this aspect deeply, but it would intuitively seem more logical to us to (1) either focus on the beneficiaries' or the donors' perspective but not both, (2) or take a weighted average of both perspectives rather than adding them up, with the weights corresponding to the subjective emphasis on the beneficiaries' vs. the donors' perspectives.

We recommend that GiveWell continue discounting health at a lower rate than consumption, but we are uncertain about the precise discount rate

[Confidence: Low-medium. We are relatively confident that discounting health at a lower rate than consumption is in line with the economic consensus, but we have high uncertainty around the appropriate discount rate for health, as it depends on several factors (e.g., the utility function) that we have not been able to evaluate quickly. We are not sure whether further desk research would provide much more clarity, but we expect that speaking to experts about this topic would at least provide more clarity on the amount of progress possible on this question. Arthur Attema and Werner Brouwer are two key authors in this debate, so our recommended next steps would be to get in touch with them.]

The economic literature varies widely on how health should be discounted, ranging from suggestions to not discount health at all to discount health at the same rate as monetary outcomes. Our impression is that there has been a convergence away from equal discounting of health and monetary outcomes towards discounting health at a lower rate, though with little agreement on how precisely the discount rate for health should be constructed. This seems to depend mainly on the assumed shape of the utility function, but also partly on the types of health outcomes used.[44] Overall, we could not come up with a superior approach to constructing GiveWell's discount rate for health outcomes in the available time, and expect that progress on this question is difficult. However, GiveWell's current approach to

[42] According to an email exchange with James Snowden: "The motivation is that GW uses u=ln(c) for convenience in most of its model [sic]. But my best guess is that u diminishes more steeply than ln(c) so I wanted to account for that in my estimate of the discount rate. That means there's an inconsistency in how GW thinks about eta."
[43] GiveWell has also shown the effect of $\eta = 1$ on the wealth effect here.
[44] For example, QALYs already incorporate time preferences in the way they are constructed (MacKeigan et al., 2003) and should thus not be discounted again with respect to time preferences.

discounting health seems to be broadly in line with some of the literature.[45] Thus, we tentatively recommend keeping the current discounting approach for health outcomes until further progress on this question is made.

GiveWell's current approach to discounting health benefits is to only discount it with its temporal uncertainty component (currently 1.4%).

A relatively small and mainly theoretical literature speaks to the question of whether and how health consequences should be discounted relative to monetary consequences. This debate essentially consists of two camps: those who support equal discounting across domains and those who are in favor of differential discounting, with health being discounted at a lower rate than monetary costs/benefits, but varying opinions on the precise discount rates that should be used for health outcomes. We summarize the main arguments of both camps in the following subsections.

Equal discounting of health and monetary outcomes is still the dominant practice, but slowly falling out of favor

Equal discounting has historically been very common and is still the "dominant practice" (John et al., 2019, p. 2).[46] Our impression is that this was initially done mainly for practical reasons, but later supported by two influential theoretical arguments:

- The consistency argument: Weinstein and Stason (1977) argued that different discount rates for health benefits vs. costs would lead to inconsistencies. As Attema, Brouwer, et al. (2018) explain, "they illustrated this with two programs that are identical except for their timing. If one wants these identical programs to receive equal priority in decision making, this can only be accomplished by applying the same discount rate to costs and effects."

- The postponement paradox: Keeler and Cretin (1983) showed that if a lower discount rate is used for health effects than for costs, it theoretically becomes optimal to infinitely postpone an intervention because the cost-effectiveness ratio keeps improving over time.

We do not think those are strong arguments in favor of equal discounting. First of all, the theoretical arguments are only valid if the consumption value of health (typically denoted as v_H, the amount consumption regarded as equivalent to 1 unit of health,[47] also described as the "willingness to pay" for health[48]) is constant over time. It seems fairly widely accepted that v_H is growing over time.[49] Moreover, the arguments have limited practical relevance, as "infinite postponing was never observed in practice, and also not in countries recommending differential discounting" (Attema, Brouwer, et al., 2018, p. 747).

[45] See, e.g., Van Hout (1998); Klok et al. (2005); John et al. (2019).
[46] We couldn't figure out quickly what exactly is meant by "dominant practice." The typical examples given are European governments. For example, the National Institutes for Health and Care Excellence (NICE) in the UK stipulates a 3.5% discount rate for both health effects and costs (NICE, 2020, p. 29), with exceptions in specific cases. We are not sure whether this extends to non-governmental organizations or countries outside of Europe.
[47] Adapted from Clayton et al. (2019, p. 4).
[48] Attema, Brouwer, et al. (2018, p. 747).
[49] "In the health economic literature, it is often argued that v_H will grow over time with increases in income, but empirical estimates of g_V are lacking" (John et al., 2019, p. 2).

Discounting health outcomes at a lower rate than monetary outcomes is the approach preferred by much of the economic literature, but there is little agreement on what discount rates should be used

Our impression is that it seems pretty widely accepted in the economic literature that health outcomes should be discounted at a lower discount rate than monetary outcomes. There are several theoretical arguments that largely depend on the willingness to pay for health (v_H) growing over time (i.e., having a positive growth rate g_v).

These studies stipulate that $r_H = r_C - g_V$ with r being the respective discount rates for health and consumption. There seems to be a lot of disagreement on the magnitude of g_v, as the **discount rate essentially depends on the specific shape of the utility function and how the utilities of consumption and health interact**. According to John et al. (2019), "the health economic literature offers neither empirical evidence nor a strong theoretical a priori in support of the assumption that v_H will rise over time."

Some examples from the literature on how health should be discounted:

- Van Hout (1998) and Klok et al. (2005) argued that **health should, just like consumption, be discounted according to the Ramsey equation, but with health-specific parameters**. This means that, while consumption is discounted at $r_C = \delta + \eta_C g_C$, health should be discounted at $r_H = \delta + \eta_H g_H$, with only δ being constant across domains, and $g_V = r_C - r_H$. As the growth rate of health (often operationalized as the growth rate of life expectancy) is significantly smaller in practice than consumption growth, g_V can be considered positive, and r_H "can be expected to potentially grossly reduce to" $r_H = \delta$ (John et al., 2019, p. 3). However, according to John et al. (2019) this doesn't seem to be theoretically well-grounded, as the aforementioned equations would only hold under pretty strong and counterintuitive assumptions.[50]
- Gravelle and Smith (2001) provided a **theoretically grounded framework to help choose the appropriate discount rate based on two factors: (1) whether health affects consumption,[51] and (2) whether health has a direct effect on utility** (besides the indirect effect via consumption). We expect GiveWell's answer to both of those questions to be 'yes,' in which case the exact recommended discount rate is a complicated formula, but **converges to a simple discount rate formula as income becomes large:** $r_H = \delta - k$, where k represents the "rate of growth of [the] direct effect of health on utility" (p. 597), or phrased differently, the rate of growth of the direct marginal utility of health. This depends again very much on the assumed shape of the utility function, and whether there is increasing marginal utility of health. We have not investigated this further.

[50] John et al. (2019, p. 3): "However, it is not fully clear how Eqs. 1 and 2 were derived. Interpreting them as the results of a two goods extension of Ramsey's optimal growth model, as described in the next section, the first-order conditions for optimal growth are given by [...]. From this, it follows that Eqs. 1 and 2 hold only if the cross elasticities are zero, i.e. if the utility function is additively separable in health and consumption. This is a strong and counter-intuitive proposition which would need some justification in order to provide a plausible a priori."
[51] That is, whether a higher level of health can increase, e.g., labor market participation and therefore increase individual consumption.

- Some authors argue that **discounting "discriminates against well-accepted, once-off preventive and other programs that are characterized by early investment and late health outcome, including screening and pediatric vaccination, some authors argue that future benefits of such programs should not be discounted at all"** (Severens & Milne, 2004, p. 398). We have not followed this literature in detail, but some arguments we encountered are:
    - Temporal uncertainty (e.g., the risk of premature death) should not be modeled via discounting but more directly incorporated in a CEA (though we have not reviewed how this should be done) (Hillman & Kim, 1995, p. 199).
    - The effect of a decreasing marginal utility of consumption is unclear and could go both ways: "Marginal utility theory can be used to support both arguments. On the other hand, greater wealth in the future may make healthcare programmes easier to implement. so that the future welfare impact of ill health decreases. For example, the cost of producing a life-year is less, so a life-year saved in the future is also worth less. Conversely, economic growth could mean that a healthy life-year in the future will be more enjoyable and therefore worth more" (Hillman & Kim, 1995, p. 199).
    - There is also debate around the extent to which health is a good that can be traded and exchanged over time. While healthcare can be traded, health cannot be traded in the same way.
    - Some health metrics (e.g., QALYs) already incorporate time preferences in the way they are constructed. Discounting QALYs "may result in double discounting" (Tasset et al., 1999, p. S78).

There is also a growing strand of literature that measures individual time preferences for health outcomes vs. monetary outcomes (e.g., Attema, Bleichrodt, et al., 2018). We did not review this literature, as we expect it might be subject to similar empirical problems as explained in our section on pure time preference (though we have not checked whether this is the case).

Review of components of the discount rate

Consumption growth (g): We recommend that GiveWell estimate consumption growth as 3.3%, implying a 2.0% discount rate from improving circumstances, and next revisit this parameter in 2028

[Confidence: Medium. Long-term economic forecasts are unreliable, and economic growth is an imperfect proxy for consumption growth in countries, although we believe that these are likely the best available inputs for estimating g. We are also uncertain about whether population weights are the most appropriate for constructing a composite of sub-Saharan Africa and South Asia. Nonetheless, we expect that improving upon our current suggestion would be quite difficult.]

Summary of reasoning:
The consumption growth rate is a parameter for calculating the wealth effect or improving circumstances component, which we discuss in more detail here. As we explain earlier, we recommend using an explicit formula for the wealth effect, $(\eta - 1)g$. The wealth effect is intended to represent the reduction in marginal utility of consumption as consumption grows over time. The rate at which increases in consumption are valued less in the future than in the present is determined by (1) the curvature of the utility function and (2) the rate at which

consumption is expected to grow over time.[52] GiveWell's assumed parameters to calculate the wealth effect are shown in Table 5.

Table 5: Parameters used to calculate the discount rate from improving circumstances

Parameter	Interpretation	Current GiveWell assumption
η	Utility curvature (of the isoelastic utility function)	1.59
g	Consumption growth (proxied by real per capita GDP growth)	3%

We only review the consumption growth parameter, g, because the utility curvature parameter, η, is out of scope for this report. We propose using g = 3.3% and next revisiting this estimate in five years (2028). Our approach was motivated by three main considerations, outlined in Table 6. See this spreadsheet for our calculations.

Table 6: Considerations and recommendations for estimating the consumption growth rate

<table>
<tr><th colspan="2">Consideration</th><th>Aspect of recommended approach</th></tr>
<tr><td colspan="2">Projections exist, but GiveWell does not currently use them.</td><td>We use growth forecasts published by global institutions.</td></tr>
<tr><td colspan="2">Geography matters. GiveWell charities operate in both sub-Saharan Africa (SSA) and South Asia (SA), but the current estimate is mostly tailored for SSA.</td><td>We use a population-weighted composite of SSA and SA ("SSA+SA"). Instead of raising the compound average growth rate (CAGR) as might be expected based on historically stronger growth, inclusion of SA reduces the expected consumption growth rate. (In fact, our estimate for SSA+SA is lower than that for either SSA or SA considered individually.)</td></tr>
<tr><td rowspan="2">Time matters. The time window for which we anticipate consumption growth matters in two ways, but the current estimate does not specify any relevant temporal reference points</td><td>First, the width of the time window is defined by the longest-duration interventions.</td><td>We set the window width as t = 40 years, which is GiveWell's analytic duration for deworming programs. Supposing we set our "start year" as the current year y, then 2063 is the "end year" (more here).</td></tr>
<tr><td>Second, the time window shifts forward in time as time elapses, resulting in changes to g, but</td><td>We define the period length, τ, as the expected number of years elapsed before the implicit discount rate from consumption growth decreases[53] by 0.1 percentage points.[54] We calculate τ as 5.2 years, then shift the t-year window by the half-period length, $\tau/2$ = 2.6 years,</td></tr>
</table>

[52] Say we have a utility function such that an individual experiences diminishing marginal returns to consumption. If this individual experiences a 10% growth in consumption in year 2, her utility would increase by less than the same 10% growth in consumption in year 1, because of both (1) a higher baseline consumption due to the increase in period 1, and (2) the assumption that there is a positive consumption growth rate g (corresponding to GDP growth per capital) regardless of any individual income shocks.
[53] According to Goldman Sachs (2022), SSA real per capita growth rates are expected to peak in the 2030s, while SA rates are expected to have already peaked (p. 39).
[54] This step is performed in the interest of reducing the frequency at which GiveWell needs to adjust its discount rate as time elapses and the t-year window shifts (assuming constant t, from 2023-2063 to 2024-2064, and so forth). See the footnote here for an explanation of our calculation.

or an update schedule.	GiveWell prefers not to adjust its discount rate at an excessive frequency.	**forward.** Adding $\tau/2$ to the current year, we define the first t-year window as starting in 2026 and ending in 2066. • **We also suggest that GiveWell recalculate the wealth effect every five years** (more here).

More detail:

Projections exist for all relevant metrics. For near- and long-term real GDP growth and population growth, we rely on the following forecasts by major global institutions:

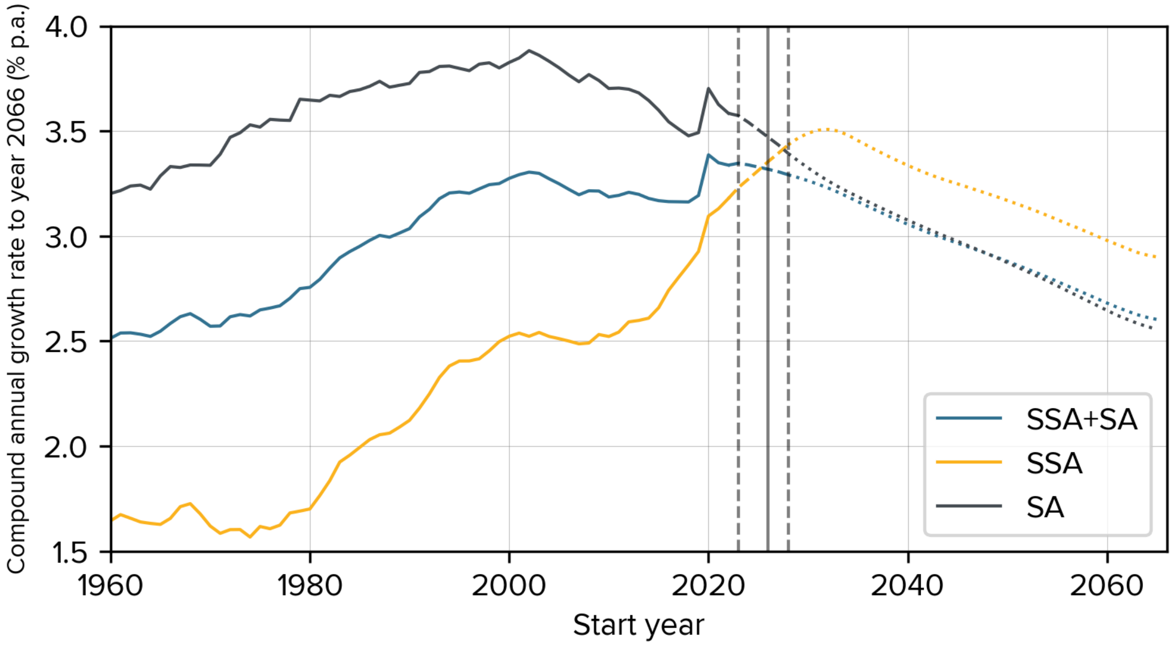
- *World Economic Outlook* (IMF): future real GDP growth, 2023-2028;
- *The Path to 2075* (Goldman Sachs): future real GDP growth, 2030s-2070s;[55]
- *World Population Prospects* (UN): future population growth.

Geography matters, but not how one might expect. Figure 1 shows that, although SA has historically experienced much faster growth in real GDP per capita than SSA, considering both SSA and SA as part of a population-weighted composite in fact results in a *lower* expected growth rate than considering either separately.[56]

[55] We use linear interpolation — while preserving decadal averages — to smoothen real GDP per capita growth rates so it does not appear that, e.g., SSA grows at a constant rate of 6.0% per year from 2030 to 2039 then abruptly drops to 5.5% growth in 2040s. Goldman Sachs also projects expected rates of growth from 2024 to 2029, but given that we trust the IMF's projections until 2028 more, we use the average of the IMF's 2028 projection and our interpolated value for 2030 to calculate the growth rate for 2029.

[56] With ~1.5 hours' research time, we failed to formulate an intuitive, satisfying explanation for why the composite entity's CAGR is lower than each of its components' CAGRs during the period 2026-2048. We also find it peculiar that the composite's CAGR tracks the SA curve so much more closely than it does the SSA curve from ~2030 onward. We hypothesize that the phenomenon can be explained in part by the differential timing of the peak growth rates for the components (i.e., SSA's rate of growth is expected to peak in the future while SA's is expected to have already peaked), and in part by the nature of exponential growth (e.g., insofar as smaller base numbers will yield higher growth percentages). Ben Snodin at Rethink Priorities said that the phenomenon reminded him of Simpson's paradox, although he was not sure that they were related. We consider the question unresolved and *cannot* rule out the possibility of errors in our calculations. However, absent knowledge of errors, we stand by our findings as it does not seem obviously implausible that two entities with different growth trajectories could have, in aggregate, a CAGR that does not fall between their individual CAGRs at any given point in time.

Figure 1: Projected compound annual growth rate (CAGR) of real GDP per capita to an "end year" of 2066, for SSA, SA, and their composite

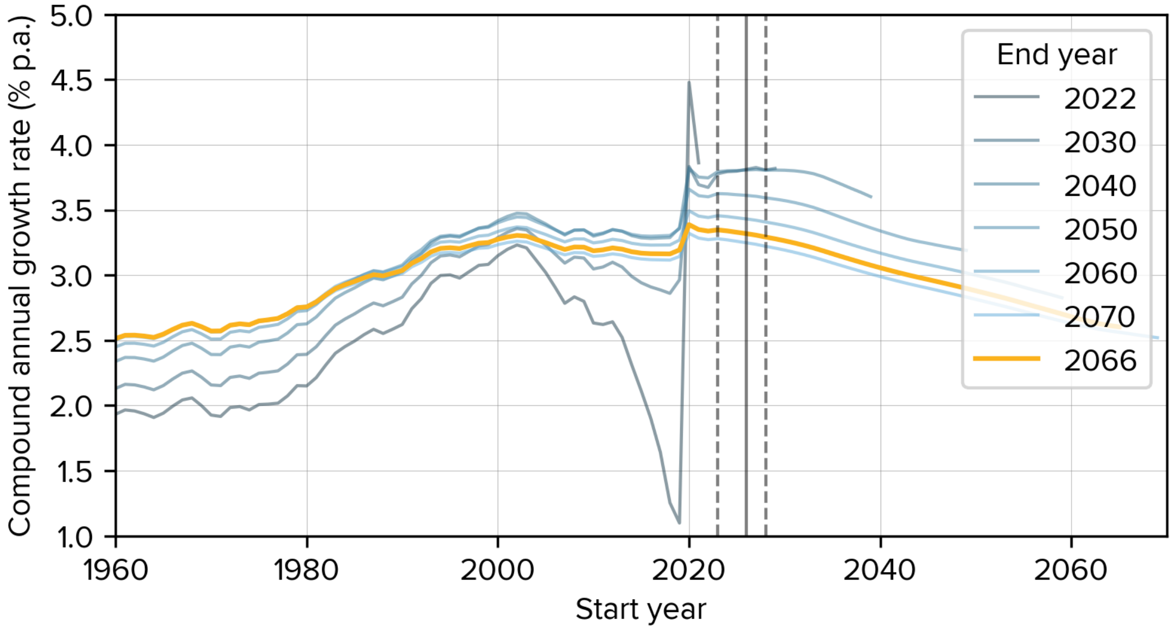


Note. SSA: sub-Saharan Africa; SA: South Asia; SSA+SA: population-weighted composite of SSA and SA. The dashed vertical lines bound the τ-length period; the dotted vertical line represents the period midpoint.

Alternatively, GiveWell could develop a composite of SSA and SA that is weighted by expected funding allocation. As we do not have access to data that would allow us to do so, we use population weights only.

Time matters in two ways. First, in Figure 2, looking at the right half of the plot, holding the horizontal-axis value constant, the gray-blue-shaded curves are arranged in sequential order of CAGR with respect to "end year" — as "end year" increases, the curves are lower in the plot. This means that, as the end year selected for calculating CAGR is extended into the future (from 2030 to 2040 and 2050, and so forth), CAGR is lower for any given start year. Therefore, as the t-year window is widened, CAGR decreases. **If GiveWell begins to evaluate interventions with t > 40 years, i.e., those with even more durable effects than those of deworming programs, it should consider lowering g accordingly, and vice versa.**

Figure 2: Projected CAGR of real GDP per capita, for SSA and SA



Note. Horizontal axis represents "start year" for calculating CAGR. Curves represent various "end years"; the yellow curve is the recommended end year, 2066. The dashed vertical lines bound the τ-length period; the solid vertical line represents the period midpoint.

Second, in Figure 2, the three bluest curves slope downward to the right. In other words, provided the end point is ~2050 or later,[57] as the start year selected for calculating CAGR is extended into the future, CAGR decreases monotonically from the current year. Therefore, as time elapses and the t-year window shifts rightward, assuming constant t = 40 years, CAGR will continuously decrease. **If GiveWell sets its discount rate in 2023 and does not substantially change its current evaluation durations, it should consider recalculating the wealth effect component of its discount rate in τ years.** (We suggest that GiveWell use the expected g in the middle of each τ-length period, in order to smooth out discrepancies within each period.[58])

Our calculations indicate that τ = 5.2 years is an appropriate period length.[59] (Our calculations assume that it makes sense to update g when a 0.1 percentage point change to the discount rate is expected.) Adding the half-period length as $\tau/2$ = 2.6 years to the current year, we therefore recommend 2026-2066 as a reference window to determine that g = 3.3% and a wealth effect of 2.0% are the most appropriate values for the first period, 2023-2028.[60] Assuming it does not

[57] For end points before 2050, refer to the two grayest curves, which show that, with end years of 2030 and 2040, CAGR increases slightly before decreasing after a start year of ~2030.
[58] We use $\tau/2$ in defining the start and end years of the first t-year time *window*, because we conceive of the first τ-year *period* as starting in year y (2023) and ending in year $y + \tau$ (2028). To smoothen discrepancies within each τ-year period, we select the g associated with the t-year window that starts in the middle of the first τ-year period, i.e., the *window* that starts in $y + \tau/2$ (2026) and ends in $y + t + \tau/2$ (2066).
[59] The calculation goes roughly as follows: (a) we expect that, every year, the 40-year g will decrease by ~0.032 percentage points; (b) we consequently expect that, every year, the implicit discount rate from consumption growth will decrease by ~0.019 percentage points; (c) we posit that a −0.1 percentage point change to the discount rate merits an update on GiveWell's part; (d) we expect a −0.1 percentage point effect on the discount rate every ~5.2 years and therefore propose 5.2 years as an appropriate period length.
[60] Note that the current τ-year period runs from 2023 to 2028; assuming that economic projections do not change significantly, the next period will run from 2028 to 2033. The current t-year window runs from 2026 to 2066; assuming that economic projections do not change significantly *and* that GiveWell's interventions remain of the same effect duration, the next window will run 2031 to 2071.

substantially change its evaluation durations, we recommend that GiveWell next recalculate the wealth effect component of its discount rate in 2028. Based on current economic projections, we expect GiveWell will lower the discount rate by 0.1 percentage points upon recalculation in 2028.

Pure time preference (δ): We recommend that GiveWell continue to use a zero rate of pure time preference

[Confidence: Medium-high. We are highly confident that we found the most important arguments in favor of and against a positive pure time preference rate, and fairly confident that GiveWell's current 0% pure time preference rate is a reasonable choice. Our main remaining uncertainty is around the philosophical debate, as we have not been able to evaluate whether Mogensen's and Purves' arguments in favor of a positive pure time preference rate are valid in GiveWell's specific circumstances. We recommend getting directly in touch with both to evaluate whether those circumstances apply.]

Summary of reasoning:
We reviewed the philosophical and economic (empirical and theoretical) literature on pure time preferences in discounting. Overall, we think that GiveWell's current choice of δ = 0% is reasonable, and we have not encountered very strong reasons for why δ should definitely be above zero:[61]

- **Philosophical considerations**: Many (if not most) philosophers argue that δ = 0%, though there are also some strands of literature arguing in favor of δ > 0%, which seem highly context-dependent and somewhat controversial. We have not been able to evaluate whether GiveWell's specific circumstances are such that the arguments in favor of δ > 0% are applicable, but conversations with Andreas Mogensen and/or Duncan Purves could be helpful to investigate this further.

- **Empirical considerations**: Empirical estimates of individuals' time preferences (which are typically positive) are so noisy that we don't think they can immediately be used as an estimate of δ. In a survey of >200 experts on discount rates, δ = 0% is the single most popular recommended time preference rate (though the median is δ = 0.5%). In many (though not all) cases where researchers/organizations assume a positive δ in practice, it actually reflects a positive temporal uncertainty component rather than what GiveWell considers pure time preferences.

- **Theoretical considerations**: Some theoretical articles show that δ = 0% leads to extreme predictions that would imply an absurdly high savings rate in exchange for a tiny benefit to future generations. However, those extreme predictions occur only if utility is not discounted at all, i.e., if there is also no discounting for, e.g., temporal uncertainty, which is not currently the case for GiveWell.

More detail:
The pure time preference rate (δ) is also called the "utility discount rate" and represents the "proportional rate of decline in the weight placed on a unit of utility in the future compared with an equal unit of utility experienced today" (Beckerman & Hepburn, 2007, pp. 191-192). The exact interpretation of δ varies. For example, many consider it to consist of two components: pure time preferences (also called myopia or impatience), and a component

[61] For full transparency, we would like to mention that we argued in favor of δ = 0.1% in a previous draft of this report. After further thinking, we conclude that this choice was not well-justified and the specific value for δ was somewhat arbitrary. Also, we gained a better understanding of the extreme theoretical predictions of δ = 0% and realized that these do not apply to GiveWell's specific case.

reflecting uncertainty about the future (e.g., risk of death or catastrophic risk) (e.g., OECD, 2018, p. 5). We interpret δ as only the first component (i.e., myopia[62]), as GiveWell estimates those two components separately. We discuss the temporal uncertainty component separately here. These two components are often not disentangled in the literature, so it is not always clear which interpretation of δ is used.

Our general impression from reviewing the literature is that most experts consider a δ of 0-1% reasonable, but there is fierce and long-standing debate in the literature on whether δ should be exactly zero or larger than zero.[63] This debate has recently been re-ignited in environmental economics, as recommended actions related to climate change are highly sensitive to the choice of the pure time preference rate.[64]

Typically, researchers clash on the question of whether this should be a prescriptive/ethical issue (debating whether it's ethically justified to place more value on current rather than future welfare) or a descriptive issue (reflecting people's actual behavior). We summarize the main philosophical and economic (empirical and theoretical) considerations in the following and provide our own takes on the debate:

<u>Philosophical considerations:</u>
Many moral philosophers are opposed to a positive pure time preference rate (Mogensen, 2019), which is due to the basic argument that "a person's place in time" has no implications for that person's moral status (Caney, 2014).[65] This view is also held by many eminent economists, such as Pigou, Ramsey,[66] Sen, and Solow[67] (Stern, 2006, postscript). Moreover, Stern's influential *Review on the Economics of Climate Change* reignited the debate on discount rates as it states that "we treat the welfare of future generations on a par with our own" (p. 76), which also implies a zero rate of pure time preference.[68]

There are two claims behind this view. One of them is that future people are just as morally important as present people. The second is that some temporally impartial moral principle is correct, such as utilitarianism's principle of utility (Beckerman & Hepburn, 2007, p. 196). One

[62] We use the term "myopia" as it is commonly used to describe reasons for temporal discounting that are not related to temporal uncertainty. We do not mean to imply that individuals with myopic preferences are necessarily irrational.
[63] This might seem like a small disagreement, but, for example, a δ = 0.5% implies that "the welfare of people alive in 139 years is worth half that of people alive today" (Shepherd et al., 2020, p. 2).
[64] This debate resurfaced after the publication of Stern's (2006) influential *Review on the Economics of Climate Change* which was criticized by Nordhaus (2007).
[65] "A person's place in time is not, in itself, the right kind of feature of a person to affect his/her entitlements. For example, it does not make someone more or less deserving or meritorious. Similarly, it does not, in itself, make anyone's needs more or less pressing. … It is not the right kind of property to confer on people extra or reduced moral status" (Caney, 2014, pp. 323-324).
[66] Ramsey (1928) described a positive pure time preference rate as "ethically indefensible and [arising] merely from the weakness of the imagination."
[67] According to Solow (1974), "we ought to act as if the social rate of time preference were zero (though we would simultaneously discount future consumption if we expected the future to be richer than the present)" (p. 9).
[68] According to Stern (2006), "it is, of course, possible that people actually do place less value on the welfare of future generations, simply on the grounds that they are more distant in time. But it is hard to see any ethical justification for this" (p. 31).

can endorse the first claim without the second.[69] Many people believe that *physically* distant strangers are just as important (in the abstract sense that's relevant to an individual's moral status) as those who are near and dear, but don't therefore think that distant strangers figure into their obligations in the same way as those near and dear. Likewise, one can believe that *temporally* distant strangers are just as important as those in the current generation, but not therefore think that future individuals are relevant to your obligations in the way that current individuals are. Thus, a positive pure time preference does *not* necessarily imply that future individuals have less moral status, as there are other morally relevant considerations pertaining to acting on behalf of future people.

According to Bob Fischer, a philosopher and senior research manager at Rethink Priorities, there are many possible philosophical justifications for a positive pure time preference:

> For instance, there are epistemic considerations: sometimes, we know more about how to benefit present people than future people. There are practical considerations: sometimes, we can do more for present people than future people. There are considerations of partiality: we can be in relationships with present people that command our moral attention (Mogensen, 2019). There are anti-aggregative considerations: we might be skeptical that relatively weak moral interests can be summed to outweigh relatively strong moral interests, and so deny that the weaker interests of future people can outweigh the strong interests of the worst off present people (Curran, 2022). Even though most philosophers accept moral views that are favorable to some positive time preference rate, the literature on social discount rates tends to skew toward the temporally egalitarian perspective — influenced, no doubt, by Parfit (1984), a prominent opponent of positive pure time preference on social discount rates. That being said, many authors suggest that discount- supporting considerations deserve more attention.[70]

We also asked Andreas Mogensen, a philosopher at the Global Priorities Institute, how strong he believes the philosophical case is for $\delta > 0\%$, as he recently published an article with the title "The Only Ethical Argument for Positive δ? Partiality and Pure Time Preference" (Mogensen, 2022). He replied:

> In spite of the questions arising from my paper that remain to be answered, I do think discounting for kinship is a plausible enough idea that it justifies some degree of increased concern for nearer generations over further future generations in some contexts. But I also think it's important to keep in mind that exactly how it does so is

[69] According to a conversation with Bob Fischer at Rethink Priorities: "Philosophers typically distinguish between axiology, on the one hand, and normative ethics, on the other. Axiology is your theory of value; normative ethics is your theory about what you ought to do in light of what's valuable. Utilitarianism basically blurs this distinction because it implies that what you ought to do is straightforwardly determined by summing what's valuable. But almost no other normative theory works like this. So, on most theories, it's perfectly consistent to say that some distant stranger matters as much as my kid, but that I have special duties to my kid that explain why it would be wrong for me to help that person over my kid, even if the stranger has more pressing needs. Utilitarianism, by contrast, does not necessarily imply that I can prioritize my kid."

[70] For example, according to Kelleher (2017), "we need further substantive discussion of intergenerational ethics to decide whether pure [agent-relative] discounting should be part of the current generation's response to the situation it faces" (p. 469). Moreover, according to Beckerman and Hepburn (2007), "we do not presume here to adjudicate between various ethical systems. The point is that, whatever the 'right' answer, climate policy cannot properly be conducted without considering a range of ethical perspectives, including those that attach a lower value to a unit of welfare accruing to a distant generation as to one accruing today" (p. 201).

> contextual and that only in some special circumstances is this approach likely to give rise to some kind of unitary discount rate applied to the welfare of all future generations, considered as a whole. As I say in the paper: "Precisely because reasons for discounting are presumed to be agent-relative, there will be no shared rate of pure time preference such that each currently existing individual ought to discount the welfare of future people at a certain rate per period, no matter which future people they might be." While I suggest that something like the standard economist's practice of discounting the welfare of future people in general at the same rate may be justifiable from the perspective of international decision-making bodies addressing global problems like climate change, the same might not be true for the decision-making contexts you're considering. Furthermore, I don't think this approach can justify a constant rate of pure time preference; the rate of pure time preference has to be diminishing and the discount factor should level out at some non-zero level.

Moreover, Mogensen recommended reviewing Purves (2016), as the study identifies a justification for positive delta not based on considerations of partiality, but instead on the growing significance of the non-identity problem over time. We did not have time to review this paper.

Our take on the philosophical debate around pure time preferences in discounting is the following: **On the one hand, it seems to be a pretty widely held view among philosophers (and economists) that δ should be zero. On the other hand, several arguments exist in favor of a positive time preference rate, though these arguments are not uncontroversial, and they also do not seem to be valid in all circumstances.[71] We have not been able to evaluate whether these circumstances hold in GiveWell's case and we are not confident that our team is well-positioned to answer this question due to our lack of philosophical training. Overall, our impression is that many philosophers would support GiveWell's current choice of a zero pure time preference rate, and we have not encountered definitive reasons for why GiveWell's δ cannot be zero.** If GiveWell wants to explore further whether its specific circumstances give rise to Mogensen's "discounting for kinship" or Purves' "non-identity effect,"[72] we recommend getting in touch with Andreas Mogensen and/or Duncan Purves directly.

Empirical considerations:
In this section, we discuss three types of empirical estimates: (1) estimates of individuals' time preferences elicited through surveys or experiments, (2) experts' opinions on what time preference rate should be used for social discounting, and (3) estimates of pure time preferences used in practice.

1. *Estimates of individuals' time preferences:*
   In the empirical economic literature, a positive pure time preference of individuals is a very well-documented phenomenon. Time preferences have been measured in many ways, including actual observed behavior in the "real world," self-reported measures,

[71] Here we are specifically referring to Mogensen's thinking around "discounting for kinship," which, he noted, is very contextual and which "only in some special circumstances is [...] likely to give rise to some kind of unitary discount rate applied to the welfare of all future generations."
[72] We learned about Purves' (2016) discussion of the "non-identity effect" at a late stage of this report, and we did not have sufficient time to understand what it means nor to form any views on the validity of his arguments.

and hypothetical or real experimental elicitation of preferences[73] (Frederick et al., 2002, p. 377). Frederick et al. (2002) reviewed more than 40 studies on empirical estimates of individual discount rates (i.e., time preference rates) based on various timelines and found that there is a very high variation in estimates (ranging from -6% to infinity; frequently above 100%). The vast majority of estimates are positive, though in many cases they seem implausibly high to us. Moreover, the estimates are extremely noisy and seem highly sensitive to the specific research design. In recent decades, there have been several methodological improvements in measuring time preferences, but estimated discount rates still seem too high to be plausibly usable for GiveWell's CEAs (typically double-digit figures).[74] While we do not feel fully up to date with the empirical literature on pure time preference rates, we would be fairly surprised if it could yield reliable estimates for GiveWell's purposes.[75]

A major caveat to those estimates is that they do not only include what we consider pure time preferences (i.e., myopia), but basically any reasons for time discounting at the individual level (e.g., uncertainty about the future), which we consider separate components. Thus, we cannot disentangle how much of a discount rate is due to δ vs. other components. Another issue with those and most other preference estimates is that they are predominantly focused on individuals from high-income countries and may not hold for GiveWell's typical beneficiaries.[76]

2. *Survey of experts on their recommended time preference rate:*
Due to a lack of consensus among various experts on discount rates, Drupp et al. (2018) surveyed over 200 experts[77] on their estimates of the long-term (>100 years) global social discount rate and its components, including pure time preferences. For the pure time preference rate, they found a modal value of 0% (chosen by 38% of experts), a median of 0.5% and a mean of 1.1% (see Figure 3):

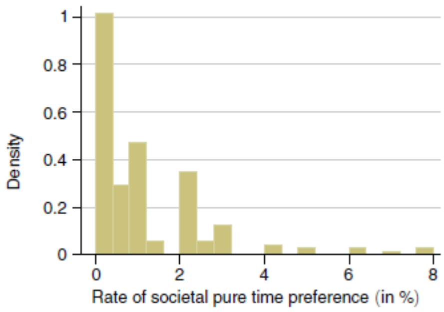
The elicited parameters refer explicitly to projects "with intergenerational consequences," which may not be relevant for a typical GiveWell-supported organization.[78] It is noteworthy that the UK government opted to use δ = 0.5% based on Drupp et al.'s (2018) findings.

[73] This often involves asking individuals questions along the lines of "Which would you prefer: $100 today or $150 one year from today?" either with real money involved or in hypothetical terms (Frederick et al., 2002, p. 377).
[74] One major methodological improvement in the economic literature that led to more plausible estimates was the development of preference elicitation methods that allow one to account for confounding factors, such as risk preferences. Since then, individual discount rate estimates tend to be smaller, but still larger than anything we'd consider reasonable for GiveWell's specific use case. For example, Andreoni et al. (2013) estimate an annual discount rate of 47%-74%, depending on the model (p. 14). Intuitively, this would mean that individuals would be willing to give up 47%-74% of the value of a reward for receiving it immediately instead of receiving it in the future.
[75] This intuition is mainly based on Jenny's own experience of eliciting time preferences during her PhD.
[76] A famous preference elicitation study carried out across many countries found that Africans have a significantly stronger preference for receiving money in the present than in the future relative to any other cultural group in the study (Wang et al., 2016), though the authors did not translate their estimates explicitly into time preference or discount rates.
[77] "For the purposes of this paper, an individual is deemed to be a potential "expert" if he or she is a (co-)author of at least one pertinent publication in the field of (social) discounting in a leading economics journal" (Drupp et al., 2018, p. 116).
[78] According to Andrew Martin, GiveWell's CEAs don't use timelines longer than 40 years.

Drupp et al. (2018) also asked the experts whether they recommend focusing on the normative or the descriptive perspective when determining the social discount rate. They found that 80% of experts think both perspectives are relevant, though a majority (~62%) recommend a stronger focus on normative issues. To our knowledge, Drupp et al. (2018) is the only such study that investigated expert opinions on different components of the social discount rate.

It is again noteworthy that what Drupp et al. (2018) call the "rate of societal pure time preference" is open to interpretation and the study did not measure how exactly experts interpreted this parameter. Our guess is that at least some of the experts consider this parameter to consist both of myopia and temporal uncertainty, which could potentially explain why many experts chose $\delta > 0\%$.

Figure 3: Rate of societal pure time preference based on expert survey.



Note. From Drupp et al. (2018, p. 119), *American Economic Journal: Economic Policy*.

3. *Estimates of the pure time preference rate used in practice:*
   See Appendix D for an overview of some of the pure time preference rates used in practice, compiled by Zhuang et al. (2007). The estimates range from 0% to 3% and reflect both myopia and temporal uncertainty, though we don't think this compilation is representative. In only three out of 13 cases, an explicitly positive rate due to myopia is used (> 0% - 0.5%), with a positive temporal uncertainty component on top. In two cases, the interpretation is not very clear, and in the remaining eight cases, the time preference rate reflects purely temporal uncertainty (e.g., probability of extinction or probability of death), ranging from 0.1% to 2.2%.[79] This shows that in many cases where a positive pure time preference is used in practice, it actually reflects the risk of extinction or death rather than what GiveWell considers pure time preferences.

We think the empirical literature measuring individuals' preferences clearly points to a positive pure time preference rate, but the elicited preferences are so noisy and sensitive to the study design that we don't think they are directly usable as such. The expert survey showed that the majority of experts believe time preference rates between 0% and 1% are acceptable, but zero is the single most popular answer. Moreover, in many cases where a positive pure time preference rate is used in practice, it actually reflects temporal uncertainty rather than what GiveWell considers pure time preferences. Overall, we do not think that the empirical considerations provide definite reasons for using $\delta > 0\%$.

[79] For example, Stern (2006, postscript) argued that "the primary justification for a positive rate of pure time preference in assessing the impacts of climate change is the possibility that the human race may be extinguished. As the possibility of this happening appears to be low, we assume a low rate of pure time preference of 0.1%, which corresponds with a 90% probability of humanity surviving a 100-year period."

Theoretical considerations:
Let's assume that δ = 0% is a reasonable choice. What are the theoretical implications of using a pure time preference rate of zero? According to [Arrow (1999)](), δ = 0% would yield an implausibly high optimal savings rate using the Ramsey model. For example, if η = 1.5, the optimal savings rate would be 67%.[80,81] Similarly, Nordhaus (2007), a critical review of Stern's (2006) famous review on climate change, investigated the theoretical implications of using Stern's (2006) assumption of δ = 0%. He found that this assumption would lead to some model predictions he called "completely absurd"[82] and explained:

> The bizarre result arises because the value of the future consumption stream is so high with near-zero discounting that we would trade off a large fraction of today's income to increase a far-future income stream by a very tiny fraction. This bizarre implication reminds us of Koopmans's warning quoted above to proceed cautiously to accept theoretical assumptions about discounting before examining their full consequences. (Nordhaus, 2007, p. 12)

We initially found these theoretical arguments in favor of δ > 0% convincing. However, we later realized that these extreme results only occur if not only what we consider the pure time preference rate is zero but also the temporal uncertainty component is zero (which is usually considered a component of the pure time preference rate). If a positive temporal uncertainty component is part of the discount rate (which is currently the case for GiveWell), there will be no such "completely absurd" model predictions. Thus, we don't see these theoretical considerations as strong arguments against δ = 0%.

Temporal uncertainty: We tentatively recommend keeping the current assumption of 1.4% as it seems to be roughly in line with other existing estimates

[Confidence: Low. We spent little time reviewing this component and have not thought about it deeply. However, the assumption of 1.4% seems roughly aligned with other estimates we've seen and we could not find strong reasons for why its value should definitely be higher or lower. We would like to note that we also have low confidence with respect to our overview of other existing estimates, partly because of our shallow and time-limited engagement with the subject and partly because of inherent difficulties in estimating the risk of human extinction and major civilizational catastrophes.]

Summary of reasoning:
We find that temporal uncertainty has traditionally been more narrowly defined as the risk of death or human extinction, whereas GiveWell uses a broader scope that also includes other risks that would prevent the realization of intervention benefits, such as "major changes in economic structure, catastrophe, or political instability." We also provide an overview of existential risk estimates — ranging from 0.1% to 2.3% — used by other organizations with links to effective altruism.

---

[80] Arrow therefore concludes that "the strong ethical requirement that all generations be treated alike, itself reasonable, contradicts a very strong intuition that it is not morally acceptable to demand excessively high savings rates of any one generation, or even of every generation. We must accept that the pure rate of time preference is positive" (pp. 97-98).
[81] Stern (2006, p. 47) argues that Arrow's suggestion of assuming δ = 1% to prevent such an implausibly high savings rate is "very ad hoc" and "not convincing."
[82] "Suppose that scientists discover that a wrinkle in the climatic system will cause damages equal to 0.01 percent of output starting in 2200 and continuing at that rate thereafter. How large a one-time investment would be justified today to remove the wrinkle starting after two centuries? The answer is that a payment of 15 percent of world consumption today (approximately $7 trillion) would pass the Review's cost-benefit test. This seems completely absurd" (Nordhaus, 2007, p. 12).

More detail:
A thorough investigation of GiveWell's temporal uncertainty component is out of scope for this report. However, we attempted to at least provide a superficial overview of this topic to provide a starting point for GiveWell.

In the literature, what GiveWell calls "temporal uncertainty" is typically modeled as one of the components of the pure time preference rate δ. Many studies do not clearly distinguish between what GiveWell calls pure time preferences (also called myopia) and temporal uncertainty, so it is not always straightforward to see what assumptions other researchers or organizations use.

A shallow overview of how others model temporal uncertainty in practice:

- *"Traditional" global health and development/climate change:*
    - Our general impression is that most (non-EA) researchers/organizations that explicitly consider temporal uncertainty as a component in their discount rate only consider "the risk of death or human race extinction" (Zhuang et al., 2007, p. 4). Thus, **many other researchers/organizations (at least outside of EA) seem to use a much narrower definition of temporal uncertainty than GiveWell**, which considers any reasons for intervention benefits not to materialize in the future, such as "major changes in economic structure, catastrophe, or political instability" (Snowden, 2020). If we consider others' estimates of risk of death/human extinction alone, many choose rates between 0.1% and 2.2% (see a compilation of different estimates used in practice in Appendix D, taken from Zhuang et al., 2007):
    - Individual risks of death range between 1% and 2.2%, but most of these are based on annual survival probabilities based on the UK in the 20th century. Thus, they are not applicable to GiveWell's typical beneficiaries. The only somewhat relevant figure for GiveWell is 1.3%, which is based on annual death rates in India during the latter half of the 20th century (Kula, 2004). This may not be useful for GiveWell as "the probability that a person will die before realizing the full benefits of the intervention [...] is captured elsewhere in [GiveWell's] cost-effectiveness analysis" (Snowden, 2020).
    - Risks of human extinction or the total destruction of a society range between 0.1% and 1.5% with only the 1.5% estimate being based on non-EU countries (see Appendix D, taken from Zhuang et al., 2007).
- *Organizations with links to effective altruism:*
    - CEARCH uses ~0.003% as an existential risk discount[83] and add a "broad uncertainty discount" of 0.1%[84] (p. 8).
    - A paper from the Happier Lives Institute (Donaldson et al., 2020, footnotes 35-36) suggests discounting for "risks from global, regional or national

[83] "Existential risks (e.g., nuclear winter), with this meaning that we are all dead, and neither the maximum potential benefit or cost, or indeed the solution itself, persists. CEARCH calculates the expected proportion of global population dead from large scale nuclear war at 0.003415436936% per annum, and uses this as our discount" (CEARCH, p. 8).
[84] "Beyond specific variables applied in the above categories, a broad uncertainty discount of 0.1% is also applied to take into account the fact that there is a non-zero chance that in the future, the benefits/costs or the intervention itself do not persist for factors we do not and cannot identify in the present" (CEARCH, p. 8).

catastrophic events" with rates between 0.18%[85] and 0.4%.[86] However, this is not the institutional position.

- For Open Philanthropy (Global Health & Wellbeing), Sam Donald stated that "over longer time horizons (>25 years) we would generally account for extinction risk via applying a 0.225% annual hazard rate (based on Toby Ord's estimate of a 1/6 chance of extinction by then) - though we plan to do more work on this in the near future."
- SoGive (pp. 11-13) outlines an alternative view on catastrophic risks in its review of GiveWell's discount rate and suggested that a higher value than GiveWell's assumed 1.4% could potentially be plausible. The review decomposes the catastrophic risks into several components (nuclear weapons usage, geopolitical instability, pandemics, AI risk, and other) and estimates a rate of 2.3%, though with very high uncertainty. We have not reviewed this in detail.
- Michael Aird at Rethink Priorities also previously compiled a database of existential risk estimates. Note that these estimates are not immediately comparable (e.g., as assumed timeframes and definitions of existential risks differ). Thus, we were not able to provide a summary overview of this database quickly. Nonetheless, **we think this database provides a good starting point to think about existential risk estimates.**

Compounding non-monetary benefits: We recommend that GiveWell explain its reasoning on this parameter more transparently, and we tentatively recommend keeping the current assumption of 0.9%, as we did not find good reasons to change it

[Confidence: Low. We find the reasons for including this component intuitively plausible, but we have not investigated its magnitude, nor alternative ways of modeling it. We learned more detail about the reasoning behind this component at a late stage of this report and did not have sufficient time for any deeper engagement with it. Speaking with an expert on how this component could be modeled might be a useful way forward.]

Summary of reasoning:

Our tentative recommendation of keeping the current assumptions regarding this component reflects more **a lack of counter-arguments we could find quickly rather than an active endorsement of the current approach**. The SPC approach to discounting is the only approach we know of that explicitly accounts for reinvestment of benefits, but it is highly impractical to use and covers only a part of what GiveWell tries to capture in this component. Regardless, we believe that a **higher level of reasoning transparency in GiveWell's public write-up** would facilitate more critical engagement with this component.

More detail:

GiveWell's current discount rates write-up is somewhat vague on what precisely the "compounding non-monetary benefits" component entails and how the 0.9% figure was determined:

> There are non-monetary returns not captured in our cost-effectiveness analysis which likely compound over time and are causally intertwined with consumption. These

[85] "This is calculated by taking Toby Ord's estimate of a 1 in 6 chance of humanity not making it through the next century and assuming the risk is constant in that period" (Donaldson et al., 2020, footnote 35).
[86] "This is very speculative, but is calculated based on additional risks to the relevant population (e.g., regional catastrophes) being equally as likely as existential risks" (Donaldson et al., 2020, footnote 36).

include reduced stress and improved nutrition. We chose a rate of 0.9% to account for this based on discussion.

We got more insight into the reasoning behind this component at a relatively late stage of this report from James Snowden via email:

> Conceptually, it's trying to capture the dynamics behind poverty traps. e.g., an argument for increasing consumption now being more valuable than in the future is that it releases credit constraints, allows additional spending on things like nutrition, and those investments pay off down the line.
>
> This parameter was originally 1.9%, and meant to capture the benefits of being able to get returns on capital. But then I think that argument got weaker based on long term fade out evidence from GiveDirectly's program. I'd wanted to cut it to 0, but Caitlin made the case that (a) there may still be other unmeasured non-monetary benefits that compound over time (b) getting money earlier might still allow consumption smoothing. So we halved it as a kind of compromise position.

Caitlin McGugan responded:

> Yes, I agree with James's summary of our reasoning. I also want to plug that while I think including compounding benefits makes sense conceptually, we're really uncertain about the quantitative value assigned to it.

We did not have sufficient time to think deeply about whether including this component makes sense and whether its magnitude is reasonable, but there are a few points we'd like to raise that came to mind quickly:

- **Our gut reaction to this component is that we find the reasoning behind including it intuitively plausible**, though we only spent ~10 minutes thinking about this and might have missed potential counterarguments. We are less sure about its magnitude, and we are also unsure about whether the current approach is the best way to model it.
- **The shadow price of capital (SPC) approach is, to our knowledge, the only formal[87] approach to discounting that explicitly accounts for reinvestment of benefits, but is unlikely to be a good option for GiveWell**. First of all, it is very data- and math-heavy, and thus not practical to use. Second, it only accounts for re- investment of monetary benefits in the private sector and thus ignores any potential non-monetary benefits. Therefore, it captures only part of what is relevant to GiveWell. We have not seen any alternative efforts to include such a component, but we have not looked for it specifically. **We currently favor GiveWell's approach (i.e., including it as an additive component to the discount rate) over the fairly complicated SPC approach, even though GiveWell's approach is a bit more "hacky" — though we have a lot of uncertainty.**
- **We recommend that GiveWell be more explicit in its reasoning regarding this component in its public write-up**. This could make it easier for readers to critically engage with this component. Our guess is that the current phrasing might have

[87] By "formal approach" we mean an approach that is theoretically grounded, such as all theoretical discount rate approaches we mention here.

contributed to previous critical reviews (e.g., [SoGive, 2022]) not having engaged with this component.

Other relevant considerations that we would review with more time

- Is it possible to make meaningful progress with respect to how health should be discounted?
  - As there does not appear to be much consensus in the economic literature on discounting health, we are unsure whether meaningful progress can be made by further desk research. Overall, we are not highly confident in our understanding of the literature and we would speak with experts about this.
- Is the compounding non-monetary benefits component of 0.9% reasonable? Should this be larger/smaller or modeled entirely differently?
- Should GiveWell use hyperbolic (i.e., discount rates getting smaller over time) rather than constant discounting?
  - We've seen several reasons mentioned in the literature on why hyperbolic discounting may be better suited than constant discounting; for example:
    - Individual hyperbolic discounting seems to be a well-established empirical phenomenon in behavioral economics
    - In the SRTP approach, discount rates depend on GDP growth rates, which might decline in the future
- Should donor preferences be incorporated in GiveWell's discount rate, and if so, how?
  - This point was brought up in a discussion with SoGive, which stated that incorporating donor preferences may potentially increase donor engagement.
- Are there any reasons to discount morbidity differently than mortality?
  - Due to time constraints, we have not reviewed this question at all, but can imagine that different health outcomes should not necessarily be discounted in the same way.
- Are we calculating the consumption growth rate using the best available inputs and correct methods?
  - Are there other projections of long-term real GDP growth that should be considered alongside (or instead of) Goldman Sachs' figures?

Contributions and acknowledgments

Jenny Kudymowa and James Hu jointly researched and wrote this report. Jenny Kudymowa also served as the project lead. Melanie Basnak and Tom Hird supervised the report. Special thanks to Bob Fischer, Sagar Shah, and Ben Snodin for their generous assistance in specific sections of the report, and further thanks to Andrew Martin (GiveWell), Ruby Dickson, and Aisling Leow for helpful comments on drafts.

Thanks also to Anthony Boardman (University of British Columbia), Vicky Cox (Charity Entrepreneurship), Sam Donald (Open Philanthropy), Spencer Ericson (SoGive), Sanjay Joshi (SoGive), James Snowden (Open Philanthropy), Joel McGuire (Happier Lives Institute), Caitlin McGugan (GiveWell), Andreas Mogensen (Global Priorities Institute), Mark Moore (Simon Fraser University), Chris Smith (Open Philanthropy), Katie Stanford (The Life You Can Save), Dan Stein (IDinsight), Aidan Vining (Simon Fraser University), and Damian Walker (Management Sciences for Health) for taking the time to speak with us.

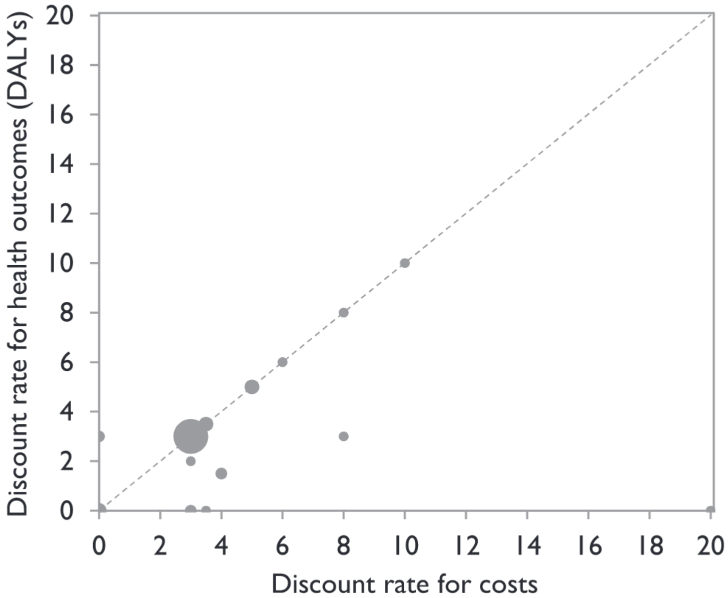
Appendices

Appendix A: A partial history of discount rates in global health research

Within global health research, a discount rate of 3% for both costs and benefits appears to have taken hold as a standard over the course of the 1990s-2010s, in line with the contemporaneous convergence toward the 3% rate in the broader health research community.

The US Public Health Service-convened **Panels on Cost-Effectiveness in Health and Medicine — held in 1996 and 2016 — have consistently recommended 3% as a standard for discounting** in health research (Lipscomb et al., 1996, p. 233; Sanders et al., 2016, p. 1098).[88] Before 1996, a 5% discount rate appears to have been so common in health studies[89] that the First Panel recommended that (in addition to adopting the new 3% recommendation) researchers still calculate results using a 5% discount to facilitate cross-comparability. The lack of an explicit recommendation[90] by the Second Panel to include a 5% discount rate suggests that the higher rate fell out of favor in the intervening years. This is consistent with Haacker et al.'s (2019) finding — which they partly attribute to the panels' influence — that the "practice on discounting in economic evaluations in global health overwhelmingly is aligned with the recommendation of applying a discount rate of 3% to both costs and health outcomes," with 85% of relevant studies they reviewed using the rate (p. 110; see Figure A1).

Figure A1: Discount rates used in 188 global health cost-effectiveness studies with time horizon ≥3 years



Note. Bubble area size is proportional to the root of the number of studies. From Haacker et al. (2019, p. 110), *Health Policy and Planning*.

[88] See here for a fuller synthesis of the US Panels' reasoning behind these recommendations. In short, the First Panel's recommendation was based on a shadow price of capital approach, while the Second Panel's equivalent recommendation was justified on the grounds of cross comparability — although its write-up heavily implies an endorsement of the social rate of time preference approach.
[89] According to Lipscomb et al. (1996), "scores of existing CEAs have adopted 5%" (p. 232).
[90] Instead, the Second Panel "recommends conducting sensitivity analyses that allow for a reasonable range of rates, along with more research on the topic of using different discount rates for costs and health effects in cost-effectiveness analyses" (Sanders et al., 2016, p. 1098).

Haacker et al. (2019) criticize the blanket application of a 3% discount rate in global health, which was originally determined within the US economic context (Lipscomb et al., 1996, p. 232), arguing that LMICs require rates of "at least" 4%-5%[91] owing to their higher average rates of growth in real GDP per capita. We are not aware how influential Haacker et al.'s (2019) advice has been, nor to what extent observed developments can be traced back to their article; regardless, global health researchers and philanthropists may be raising their discount rates. Damian Walker, formerly Deputy Director of Data and Analytics at the Bill and Melinda Gates Foundation, expressed a similar view in an email, saying that "for LICs, a higher rate is justified, but 3% has somehow become the de facto standard."

The Bill & Melinda Gates Foundation (BMGF), the only major foundation for which we found public thinking on discount rates, suggested[92] that its grantees apply a 3% annual discount rate for costs and benefits as recently as 2014 (Wilkinson et al., 2014) and 2016 (Wilkinson et al., 2016), then subsequently adopted a higher rate of 5% around 2019 for its funding of the "three Gs" consisting of Gavi, the Global Fund and the Global Polio Eradication Initiative (Walker, 2019). However, the 3% guideline appears to have been justified based on its popularity — in order "to facilitate comparability" (Wilkinson et al., 2014, p. 50) — and does not constitute an independent analysis.

According to a 2019 blog post on the BMGF website, "Here at the foundation, we don't have a consistent viewpoint on this topic" (Walker, 2019). Walker confirmed in an email that "BMGF never had – still doesn't, I think, have – an explicit discount rate, as it didn't want to be hemmed in." However, for its funding of the three Gs, BMGF "decided to go with a 5 percent discount rate" (Walker, 2019).

According to Walker in his email, the decision to adopt a 5% discount rate was likely made by Bjorn Lomberg of the Copenhagen Consensus Center, but Walker could not recall the precise justification for the approach.[93] A recent BMGF-funded paper by Copenhagen Consensus Center that we were able to review uses a 5% central discount rate (Copenhagen Consensus Center, 2021, p. 8). Earlier research by the think tank appears to have used both 3% and 5% discount rates (see Anderljung et al., 2015; McGreevey et al., 2011), suggesting that it may have shifted from 3% to 5% as its recommended rate between 2015 and 2021.

[91] As a lower bar, 5% for low- and lower-middle-income countries, and 4% for upper-middle-income countries.
[92] In an email, Walker said, "I often debated whether we should make the reference case guidance a requirement of our funding. But we never did. However, even with such soft guidance, I think many grantees sought to adhere."
[93] We approached Lomberg, but he did not respond to our email query.

Appendix B: Derivation of the Ramsey equation for discounting

The Ramsey equation for discounting is based on a growth model which Ramsey (1928) used to derive the optimal savings rate of a representative agent. We do not recommend reviewing Ramsey's (1928) original publication, as the math is fairly technical and lengthy.

Several easier and more intuitive derivations are available, e.g., Tol (2015).[94] According to Tol (2015, p. 3), ε denotes the present value of $1 at time t, and r is the consumption discount rate:

$$\varepsilon = e^{-rt}$$

ε also represents the discount factor.[95] We can then derive the Ramsey rule from equating marginal utilities across time:

$$U_{C0}\,\varepsilon = e^{-\delta t}\,U_{Ct}$$

where U is the utility function, U_C is the the first partial derivative to consumption which represents the marginal utilities at times 0 and t, δ is the pure time preference rate (also called utility discount rate), and ε is the consumption discount factor. We then assume an isoelastic utility function ($U(C_t) = \frac{C_t^{1-\eta}}{1-\eta}$) and express consumption at time t in terms of consumption at time 0 ($C_{t+1} = C_0(1+g)^t$):

$$\varepsilon = \frac{e^{-\delta t}U_{Ct}}{U_{C0}} = \frac{e^{-\delta t}C_t^{-\eta}}{C_0^{-\eta}} = \frac{e^{-\delta t}C_0^{-\eta}e^{-\eta g t}}{C_0^{-\eta}} = e^{-(\delta+\eta g)t}$$

Combining this with the first equation defining that $\varepsilon = e^{-rt}$, we can then solve the equation for r, which yields the Ramsey equation:

$$r = \delta + \eta g$$

An alternative, fairly straightforward derivation of the Ramsey equation in continuous time can be found in Zhuang et al. (2007, p. 8).

[94] See here for a video by Richard Tol in which he derives the Ramsey equation step by step.
[95] Note that the above equation represents discounting in continuous time, not in discrete time (as done by GiveWell). For discounting in discrete time, we would use: $\varepsilon = 1/(1+r)^t$. However, the math needed for the Ramsey equation is easier to show in continuous time, and the discrete form converges to the continuous form if t converges to 0.

Appendix C: A brief discussion of advantages and drawbacks of the main social discounting approaches

- The SRTP approach has the advantage that it accommodates and makes transparent several parameters (e.g., time preferences, shape of the utility function). However, it has been criticized for ignoring potential market effects of public projects on private sector investment. When public and private investments compete for financial resources, funding for public projects may crowd out private investment. This opportunity cost is not accounted for in the SRTP approach.[96]

- The SOC approach accounts for the opportunity cost of forgone private sector investment, but ignores the opportunity cost of forgone current consumption (which is captured by the SRTP approach). Moreover, it implicitly assumes that the next best use of public or philanthropic funds would be private sector investment. This is often not, or only partially, the case.[97] If there is no or only partial crowding out of private sector investment, the SOC would yield an overestimate of the opportunity cost and thus the social discount rate. We think this "crowding out" effect is less relevant for the philanthropic sector, as funds are typically donated rather than borrowed or raised from investors.[98]

- The weighted average approach is an attempt to reconcile both approaches (SOC and SRTP). While it has the advantage that the opportunity cost of forgone private sector investment is taken into account, "it assumes that benefits will be consumed immediately and ignores the fact that they could also be reinvested in the private sector, generate future consumption, and bring more social value than if they were consumed immediately" (Zhuang et al., 2007, p. 11). This would lead to overdiscounting of social benefits. Moreover, estimating the weights in this approach can be difficult and data-intensive, as it requires estimating various "elasticities of demand and supply of funds with respect to changes in interest rates" (ibid, p. 24).[99]

- The SPC approach is similar to the weighted average approach, but additionally accounts for the fact that benefits can be reinvested by the private sector. While this approach offers a lot of flexibility and seems theoretically attractive, it is very

---

[96] According to Zhuang et al. (2007, p. 9), "a major criticism on using SRTP as the social discount rate is that it is purely a measure of the social opportunity cost in terms of foregone consumption and ignores the fact that public projects could displace or crowd out private sector investment if they cause the market interest rate to rise (Baumol 1968 and Harberger 1972). If additional public investment is made at the cost of displacing private investment, its marginal social opportunity cost should also reflect what the displaced private investment would otherwise bring to the society, which can be measured by the marginal social rate of return on private sector investment (SOC)."

[97] "However, this approach to defining the SOC assumes that, in the absence of undertaking a public sector project, the next-best use of the funds would be to invest in a private sector project of equal magnitude and risk. In other words, it assumes that a public sector project fully crowds-out or displaces a private sector investment of the same cost and risk profile. This is a key judgement. It is clearly the appropriate counterfactual for private sector investment decisions. However, in most cases the appropriate counterfactual for setting the government's opportunity cost is likely to be one of the following two other possibilities" (Creedy & Passi, 2018, p. 148).

[98] Adapted from Jansen and McKatz (2002, p. 3).

[99] See Zhuang et al. (2007, p. 24) for an example of the weighted average approach used for Papua New Guinea. The discount rate was estimated using "elasticities of savings, supply of foreign capital, and private investment with respect to the interest rate" and "shares to the total savings by various groups of domestic savers and foreign savers" as well as "the investment share of various business sectors."

complicated to implement and highly sensitive to some key assumptions such as the length of a project.[100]

[100] "A different shadow price of capital has to be estimated for every project according to the length of life of the project" (Zhuang et al., 2007, p. 14), which does not seem very practical.

Appendix D: Empirical estimates of the pure time preference rate in the literature

Table D1: Empirical estimates of the utility discount rate.

SOURCE	EMPIRICAL ESTIMATES	THEORETICAL BASIS
Scott (1977)	1.5%	Component reflecting myopia is 0.5%, and that reflecting the changing life chance due to the risk of total destruction of a society is 1.0%
Kula (1985)	2.2%	Reflecting average annual survival probability in the UK during 1900–1975
Kula (1987)	1.2%	Reflecting average annual probability of death in the UK in 1975
Scott (1989)	1.3%	Component reflecting myopia is 0.3%, and that reflecting the changing life chance due to the risk of total destruction of a society is 1.0%
Newbery (1992)	1.0%	Perceived risk of the end of mankind in 100 years
Dynamic Integrated Model of Climate and the Economy (DICE) model (Nordhaus 1993)	3% per year	Utility discount rate reflecting pure social time preference, determined by calibrating the DICE model to match actual data
Pearce and Ulph (1995)	1.1%	Reflecting the average annual probability of death in the UK in 1991
Arrow (1995)	1%	Utility discount rate reflecting pure social time preference, and matching the observed savings behavior
OXERA (2002)	Myopia = 0-0.5% Risk of death = 1.1% with a projected change in the near future to 1.0%	Based on previous studies and projected and recent average annual death rates in the UK
Evans and Sezer (2004)	1.0–1.5%	1% for EU countries and 1.5% for non-EU countries, reflecting catastrophe risks
Kula (2004)	1.3%	Reflecting the average annual death rate in India during 1965–1995
Evans (2006)	1%	Based on the approximate average annual death rate in 2002–2004 in 15 countries of the European Union
Stern Review (2006)	0.1%	Probability of human race extinction per year

Sources: Compiled by authors.

Note. From Zhuang et al. (2007, p. 6), Asian Development Bank.

Appendix E: Derivation of GiveWell's discount rate formula

GiveWell and Ramsey both assume isoelastic utility where $u(c) = \frac{c^{1-\eta}-1}{1-\eta}$. As we explain in Appendix B, to obtain the discount rate according to the Ramsey rule, we need to equate marginal utilities across time. The marginal utility of an additional unit (or USD) of consumption is given by $u'(c) = c^{-\eta}$.

As GiveWell models consumption benefits not in absolute but in percentage terms, GiveWell equates marginal utilities of an additional percent of consumption over time. These marginal utilities are approximated by $\phi(c) = zc\,u'(c) = zc^{1-\eta}$, where z is small.[101]

Let f represent the discount factor and r represent the discount rate, then $f = \frac{1}{1+r} \approx e^{-r}$. Thus, we equate GiveWell's marginal utilities of a percentage change of consumption over periods 0 and 1 as follows:

$$f\phi\left(c_0\right) = e^{-\delta}\phi\left(c_1\right)$$

As $\phi(c) = zc^{1-\eta}$ and $c_1 = e^{g}.c_0$, we can rewrite the above equation as: $fzc_0^{1-\eta} = e^{-\delta}z\left(e^{g}c_0\right)^{1-\eta} = e^{-\delta+g(1-\eta)}zc_0^{1-\eta}$. Solving this for f yields $f = e^{-\delta+g(\eta-1)}$. Since $f = e^{-r}$, we get:

$$r_{GiveWell} \cong \delta + g(\eta - 1)$$

Note that if we expressed the indifference equations in terms of discrete time, rather than continuous time, we would obtain the exact GiveWell discount rate expressed as:

$$r_{GiveWell} = (1 + \delta)(1 + g)^{\eta-1} - 1$$

This difference between the exact and the approximate discount rate is small for small parameter values,[102] so we recommend that GiveWell use the approximate version to calculate its discount rate.

[101] We noticed at a later stage that an alternative expression of the marginal utility of a percentage increase in consumption would be $du/d\ln(c) = c^{1-\eta}$. This can be shown to hold via the chain rule.
[102] The approximation is more accurate at small parameter values because the similarity of the simplified expression to the original depends on how well $1/(1 + x)$ approximates e^{-x}; the expressions are most similar when x is small.