Discounting for uncertainty in health

Owen Cotton-Barratt

Book chapter draft

1. Introduction

Discount functions are very common in economic modelling. They help people to compare options whose costs and benefits are spread over time. Several factors can affect how we make such trade-offs, and these are often combined into a single discount function. This function can then be used as a black box, without the users needing to understand everything that went into its creation. This is a great strength of the method, because it is simple and accessible, and can be easily and accurately applied by non-experts.

Nonetheless, in some contexts the behaviour of the different drivers of discounting can come apart. In order to choose an appropriate discount function it is then necessary to unpack and analyse these separately: arguments that shed light on one aspect of discounting may be inapplicable to another.

This is relevant when choosing a discount function in a health context. Much of the discussion about discounting health has not differentiated between different components of discounting, or has focused on discounting for time preference rather than for uncertainty.

Discounting for uncertainty has an important role to play in modelling health outcomes. In this chapter I will explain how it differs in principle from other types of discounting (Section 2), explore the implications in discounting health (Section 3), and show how it offers a resolution to some apparent paradoxes (Section 4).

2. Discounting for uncertainty

2.1 Reasons for discounting

Discounting is the practice, in building economic models, of including an adjustment so that a benefit or cost later in time is weighted less than the same benefit or cost earlier in time. There are three common reasons for including such an adjustment.

Discounting for time preference is based on the observed preferences people show for receiving benefits earlier or costs later. *Discounting for opportunity costs* is based on resources being valued more earlier because they can be put to work in the meantime. *Discounting for uncertainty* is based on effects further in the future being adjusted for the possibility that the model does not describe them properly.

Often these are combined into a single rate for simplicity, because they affect the model in very similar ways. However, the factors driving them can behave differently in different contexts. So when we're thinking about how to adjust discount rates for new contexts it can be important to consider the role of these components separately.

Discounting for time preference amounts to a claim that earlier goods are intrinsically more valuable. Discounting for opportunity costs amounts to a claim that earlier goods are instrumentally more valuable. Discounting for uncertainty amounts to a claim not that earlier goods are more valuable, but that projections of future goods are more likely correct when they concern the near future than the long-term future. It is most frequently seen as a 'catastrophe rate', discounting the future for the possibility that a large catastrophe renders predictions useless.

All of these may in theory be relevant in a health context (although there is agreement that discounting for opportunity costs should be much smaller in the case of health than for capital[123]). To explain how they can be relevant, we will examine an example from the literature:

> *"As an example, let us assume that about £70 000 is available for hip fracture prevention in 100 women, and there are two strategies under consideration: 10 years of hormone replacement therapy (given to 50 year old women), which prevents 50% of fractures in 30 years' time; or 10 years of calcium and vitamin D (given to 70 year old women), which prevents 30% of hip fractures in 10 years' time. [...] Without discounting, hormone replacement therapy produces a lower cost per avoided hip fracture than vitamin D and calcium. On the other hand if the hip fracture benefits are discounted [at 6% per annum], then the reverse is true."*[4]

This example was originally used just to demonstrate that the choice of how to discount can be decision-relevant. We will see further that each of the three types of discounting could be applicable in this scenario.

First, people typically have preferences for receiving health benefits earlier rather than later.[5] If we think this is appropriate to apply at a social choice level, we may wish to apply a discount rate for time preference. This discount rate would be applied to the direct health benefit of avoiding hip fractures.

Second, there are social costs associated with the treatment of hip fractures (after they occur). The resources used in treatment might be used for other valuable purposes, so it is appropriate to apply a discount rate for opportunity cost to the projected savings in future treatment costs.

Third, something might happen which would render the longer-term intervention partially or wholly irrelevant before all of the benefits were realised. Perhaps we will discover a cheaper and more effective alternative to the current hormone replacement therapy that can be administered to 70 year-olds. Perhaps a rise in robotic mobility assistance will halve the number of falls that could result in fractures within the next 25 years. This uncertainty could lead us to discount the entire projection of future hip fractures averted.

2.2 Discount functions for uncertainty

Because they are applied for different reasons, they sometimes behave differently. As an example, when discounting for uncertainty there is a difference between looking forwards and looking backwards which is not present when discounting for time preference or opportunity cost.

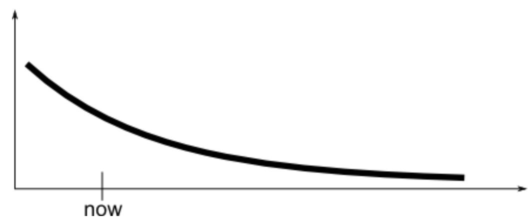

Figure 1: Value of resources over time when discounting for time preference

Compare Figure 1 and Figure 2. In both cases the value of resources in the future is discounted. In Figure 1 this continues smoothly into the past, so that past resources are judged more valuable than present resources. In Figure 2, it is assumed there is no uncertainty about past resources, so there is no discount factor applied to the past.

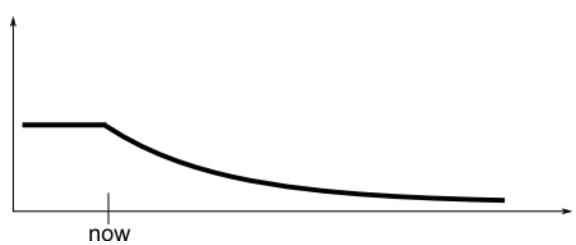

Figure 2: Value of resources over time when discounting for uncertainty

Where does the curve in Figure 2 come from? There are many possible futures in which the value of the resources in question could differ. While *ex post* only one of these will be correct, and the most likely single one is usually that which caused no change in the value of resources, we face uncertainty *ex ante*. The value of resources might stay the same. It might drop suddenly (if they are obsoleted, or the system they are in collapses), and a drop like this could happen at many different possible times. There may be multiple drops, or a smooth period of change, or an increase. It is only in expectation that these are aggregated and give something like the graph of Figure 2. This is illustrated in Figure 3.

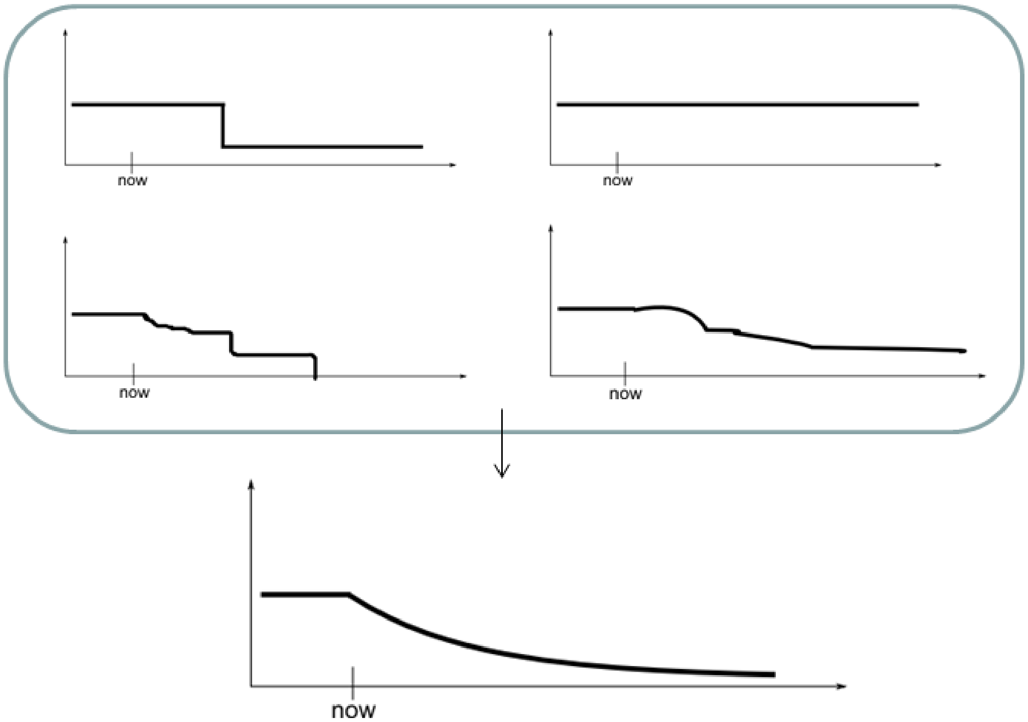

Figure 3: possible *ex post* behaviours at the top combining to give an *ex ante* discount function

It might look like we've introduced a time-inconsistency here: if the curve, as drawn in the future, will be undiscounted between then and now, why do we discount in the curve as drawn now? Indeed it is sometimes argued that discount functions should be exponential, as non-exponential functions can lead to predictable preference reversals.[6] This argument doesn't apply in the case of discounting for uncertainty, as we will gain new information in the future. This means that while preference reversals are common, they are not predictable.

However, if the drivers of the uncertainty are approximately constant over time, the *ex ante* discount function will be approximately exponential in the future. This seems to be a reasonable assumption over short and medium timescales. Exponential discount functions have the additional advantage that they are particularly easy to work with, which means that as a practical matter they will be more frequently and consistently applied. We will sometimes talk about the *discount rate*, meaning the annual decay under the assumption that the function is exponential.

2.3 Including uncertainty in models versus in discount functions

One question is whether it’s appropriate to include uncertainty in the discount function at all. It could instead be built in as a component of whatever model we are using to project future effects.

There is something to be said for preferring to keep uncertainty in the model. It concerns what will or won’t happen in the world, rather than how we should trade off between different known goods. Conceptually this puts it closer to modelling, which concerns facts about the world. If we wish to carve the world at its joints, it would seem odd to count this as part of discounting.

Moreover, the degree of uncertainty depends on the context. Suppose a worker in her 20s faces danger of an accident which could cause paralysis or death. We are much more uncertain about whether we will have an effective treatment for paralysis in twenty years time than whether we will have an effective treatment for death, so we should apply a higher discount rate to the projected years of future paralysis. Discount rates are usually applied without too much consideration across a range of contexts, so work poorly if the rate is too context-sensitive.

So in theory it seems preferable to keep uncertainty separate from the discount function. On the other hand as a matter of pragmatism there is a lot to be said for including it. First, because in practice many models omit consideration of background uncertainties, so the choice may be between including it in the discount rate or not at all. Second, because although it is somewhat context-sensitive, a lot of uncertainty is relatively context-insensitive. For different interventions we may need to account for the same unlikely-but-possible disruptions to the wider system, such as government collapse, or obsolescence of the project.

This matters because these events are often complex and hard to assess. That means that evaluating them will not lie within the expertise of everyone wishing to model the effects of interventions. Even if the modellers are able to estimate the effects of such broad uncertainties, it is an unnecessary burden to place upon them, which will slow them down. We can increase the efficiency of modelling by grouping uncertainties together, and looking at these carefully a few times rather than every time. This has the advantage of consistency: if different teams are using the same estimates for systemic uncertainty, this increases the comparability of their model outputs.

It is worth noting that these practical reasons have led to an uncertainty component of discount rates being recommended as best practice in some domains. The UK Treasury Green Book includes a term for ‘catastrophe rate’; that is discounting future effects for “the likelihood that there will be some event so devastating that all returns from policies, programmes or projects are eliminated, or at least radically and unpredictably altered."[7]

Of course when we talk about discounting for uncertainty we don't mean to include all uncertainty. If we are implementing a new programme with a 10% chance of collapse due to specific circumstances, it is appropriate to include that in our model rather than a discount rate. It is the broad and shared uncertainties which are often hardest to model that benefit most from being included in the discount rate.

3. Discounting for uncertainty in a health context

3.1 Discounting health

The Global Burden of Disease project measures the health losses (i.e. loss of capacity along certain dimensions such as mobility) associated with different health conditions worldwide. It then converts the multi-dimensional health loss into a single-dimensional 'burden of disease', which measures how bad the health losses are. It is assumed that the conversion between health losses and burden of disease is constant over time. The original Global Burden of Disease project included a discount rate on health losses. Following extensive discussion, the methodology was changed in the 2010 update, to not discount health losses at all.[8] This change was based principally on ethical arguments that it was indefensible to weight health earlier more than health later. I think these arguments are entirely correct with respect to the intrinsic value of health, which determines that the rate of time preference included in a discount rate for health should be zero.

It is also not clear that health earlier is more valuable instrumentally than health later. A year of healthy life lived earlier may lead to someone doing more productive work and helping the welfare of those around them. These benefits may in turn compound. On the other hand productivity has been increasing with time, so a year of healthy life lived later may be able to achieve more than one lived earlier. So in principle it seems that there could be a discount rate for opportunity costs of health, but it is unclear even whether it is positive or negative, and in any case is likely to be small.

What about uncertainty? The ethical case against discounting is not applicable to discounting for uncertainty, since such discounting does not involve the claim that health earlier is more valuable. There is certainly some uncertainty around projections of future health losses. While much of it is of the type that slots easily into the modelling, I will claim that at least some of it is most naturally incorporated into a discount rate.

3.2 Historical trends

In the history of modern medicine, the chief trend is both familiar and encouraging: we have learned how to prevent, cure, or treat increasing numbers of conditions, and how to treat those increasingly effectively. Mortality from syphilis fell dramatically in the years following the discovery of penicillin.[9] Mass vaccination led to the eradication of smallpox.[10] Near-skin accommodations such as glasses can reduce the burden of some conditions.

This trend has continued more recently, too. Life expectancies have consistently risen in developed countries by around two years every decade over the past fifty years. Of course part of this gain is attributable to lifestyle changes, but various estimates have suggested that improved medicine is responsible for around half of the gain[11], or between one-quarter and one-half of the gain[12]. Survival rates for many kinds of cancer have risen. Modern prosthetics are much more versatile than earlier versions. Handheld devices with optical character recognition and text-to-speech software on smartphones allow blind people to read printed texts.

On the other hand, it is very rare for us to lose the ability to treat anything. Much of our progress has come from research, and from codifying evidence-based best-practice. Barring catastrophe this knowledge is safe. The rise of antibiotic-resistant strains of some diseases is a very small flow compared to the huge sweep of gains we have made.

Going forward, we should expect this trend to continue. That is, there are some conditions that we will be able to more effectively treat or prevent in the coming years. However, we cannot say with confidence now which these will be. For this reason it is hard to directly model, and will usually better be accounted for by discounting.

3.3 Implications for uncertainty discounting of health

Our expectation of finding new cures and treatments matters when we are projecting future health losses. If we always ignore the possibility of future developments, we will systematically overestimate future health losses; this could lead to us predictably paying too little attention to averting current health losses relative to future health losses. If instead we discount future health losses for this uncertainty, then we will sometimes underestimate the future health losses (when no better treatment or prevention is discovered), and sometimes overestimate them (if a cure is discovered shortly afterwards), but we will be closer to correct on average. This discount rate could be thought of as a catastrophe rate for diseases, where the ‘catastrophe’ is actually the good event that the disease no longer represents a problem.

It’s important to note that this is relevant only when evaluating projections of future health losses; not for example when comparing measured health loss at different times. Uncertainty discounting is therefore not needed for the current Global Burden of Disease project, which estimates the current health burden of different conditions. On the other hand it is relevant when choosing between health interventions, a question commonly faced by health agencies and funders. It may also be relevant for the Global Burden of Disease project, if they follow their stated intention to start estimating not just what role historical risk factors have played in the current distribution of health states, but also predicting what future health states will be produced by current risk factors.

If we decide that we want to discount health for uncertainty, there's still a question of what discount rate to use. The most appropriate approach here is probably to look to the data. We could examine historical data on the burden of diseases to estimate the average rate at which the health burden of conditions decays. The Global Burden of Disease fell from 47,104 DALYs/100,000 people in 1990 to 36,027 DALYs/100,000 people in 2010[13]. This corresponds to an annualised drop of 1.35%, of which only a fraction should be attributed to new medical discoveries. If that fraction is between one-quarter and one-half, this suggests a discount rate of between 0.3% and 0.7%, much lower than the discount rates typically used on capital. This estimate is however quite crude: it might be better to look carefully at the data and estimate what reduction has come from new discoveries.

Even after determining an appropriate rate for health as a whole, we should be aware that our uncertainty is sensitive to context. In the context of health, there is an important distinction to be drawn between years lived with disability (YLDs) and years of life lost (YLLs).

In the case of YLDs, we discount for uncertainty about whether the condition can be prevented or treated. The relevant time by which this must have happened is usually the year in which the health loss would occur. So we should discount each YLD to the year in which it is projected to occur, much as we would discount any other resource.

For YLLs the story is a little different. After a death has occurred, we are much more confident that no future discovery will undo this. It therefore makes sense to discount years of life lost to the year of projected death, but to stop the discounting at that year*. This stands in contrast to discounting health for its intrinsic or instrumental value: in those cases we would continue to discount so that each year of foregone life is discounted to the year in which it is projected to occur.

---

* Or perhaps to include only a very small discount rate beyond that point. For example we might include a catastrophe rate of 0.1% to account for the possibility of human extinction, as Stern does in his report on climate change.

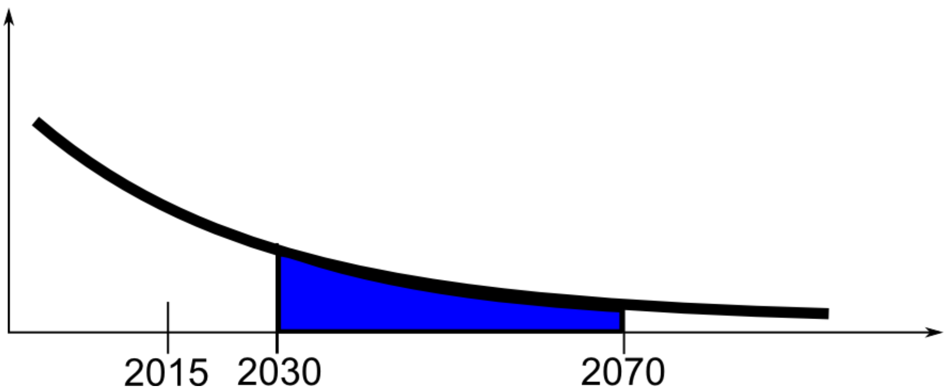

Figure 4: discounting years of life lost following a death in 2030, discounted for intrinsic value

Figures 4 and 5 illustrate this difference. Both show a death projected to occur in 2030, from the vantage point of 2015. Without this death the person was expected to live until 2070. In Figure 4 the blue area shows the YLLs that would be counted if we discount them for intrinsic or instrumental value. The burden drops every year according to the discount rate. In Figure 5, we look at the same projection but discount only for uncertainty. Now the only effect of the discounting is to reduce the chance that the death actually occurs, so we reduce the height of the blue area uniformly from its undiscounted value.

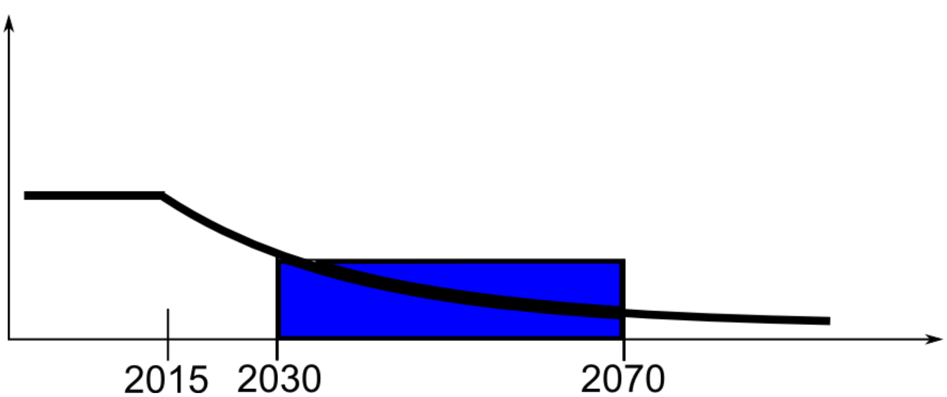

Figure 5: discounting years of life lost following a death in 2030, discounted for uncertainty

This distinction is important in the case of infant mortality. Discounting health has traditionally been seen as reducing the burden associated with infant mortality, since many of the YLLs are many years in the future. If we are only discounting for uncertainty, however, then we should take full account of future YLLs after the death is known for certain.

There may be other times where it is helpful to know how much of the discount rate is reducing the projected incidence of a disease rather than its impact. For example if we discount future projections of deaths from heart disease, this might increase the projected burden of other conditions such as Alzheimer's disease, as more people live into old age and are susceptible.

3.4 Granularity

The considerations which could lead us to choose an uncertainty discount rate in health will be different from those in other domains, so we shouldn't be surprised if the discount rate is different. Similarly, health is not homogeneous, and it may be that it is appropriate to use different discount rates for different parts of health. Perhaps there are different forces driving our uncertainty for communicable and non-communicable diseases, or for chronic and acute conditions. In the next section I discuss one reason why we might expect the degree of uncertainty around YLDs and YLLs to come apart.

There are, however, disadvantages to separating out discount rates too finely by area. First, we are less able to draw upon historical data to predict future behaviour. While the trend of treatments for more conditions seems robust for medicine as a whole, it would be absurd to take a very narrow view and project no new treatments for conditions where we have not yet found a treatment, and better new treatments for every condition that we recently improved our treatment for. While there are different trends in different narrow areas, some of this difference represents a genuine difference that we should expect to continue going forward, and some is just noise.

Second, it is usually the case that a range of discount rates are defensible. By establishing a standard used across the field, comparisons of different health interventions will more properly be comparing like-with-like. This limits the ability of advocates for a particular course of action to nudge an analysis by choosing a favourable discount rate.

3.5 Extra considerations in discounting YLDs

So far we've discussed discounting future health losses because of uncertainty about the possibility that new discoveries will allow us to avert or treat the condition; that is, uncertainty about the degree of health loss. Of course the burden of disease is not a function just of the health state, but also of the environment in which people live. That environment varies geographically, but it also varies with time. Our trend towards increasing technology often provides increasing numbers of options and substitutes, and in many cases this may reduce the burden of diseases. For example the rise of online grocery shopping may make mobility problems less burdensome. Because of this, we should discount our projections of the burden caused by a given health loss as well as of the amount of health loss there will be.

On occasion perhaps technology increases the burden associated with some health states. For example the rise of personal computers may have made it more burdensome to have a condition which restricts the use of a mouse or keyboard. It appears to me that on balance that the trend is towards reduction of the burden associated with some health states, but further analysis of this would be helpful. If we could estimate the rate at which this occurs, this could be used as an extra term in the discount rate used for discounting future YLDs. It would not be used to discount YLLs.

4. Application: the research and eradication paradoxes

As an example of the value of discounting for uncertainty, we'll consider the research paradox and the eradication paradox. These have been presented as problems for the practice of not discounting health.

The research paradox says that if we don't discount health, then we should devote our entire healthcare budgets to research (so long as it has a non-zero chance of success), since it will produce an infinite stream of benefits. This is taken as a paradox since nobody thinks this is a reasonable conclusion. The eradication paradox applies the same reasoning to the project of eradicating a communicable disease.

In removing discounting, the GBD 2010 team considered these paradoxes. Their conclusion was to admit that research or eradication would produce more health in total than any direct intervention, but contest that we were not obligated to redirect current health budgets for reasons of intergenerational equity.[14]

I think that the conclusions of these two paradoxes are wrong as a matter of fact. If we are considering the health benefits of funding research or eradication programmes, we should compare this to the scenario in which we don't fund the research or eradication. In that counterfactual scenario it would be naive to assume that the research or eradication will never happen. Instead we should expect that research on the problem at hand or eradication of the disease in question will be accomplished at some later point. It is hard to predict exactly when that will occur, but we could build probabilistic models to estimate the total expected health benefit. The expected benefits, net of these additional future activities, would probably be finite.[15]

As an alternative to detailed modelling of these counterfactual scenarios, we could discount the future health benefits of research or eradication by our uncertainty discount rate. Since this would be a positive discount rate, the infinity – and hence the paradoxes – would disappear.

Here discounting for uncertainty is adjusting the future health stream for the chance that the problem would be solved anyway. This is essentially the same consideration that would lead you to complex counterfactual models. The answer produced may be slightly different if it doesn't account for extra facts about the particular disease, but it will be approximately correct. Incorporating this discount rate thus makes us more robust to extra considerations we didn't think to model.

5. Conclusions

Uncertainty about the future is certainly important when making decisions about health programmes. There are many possible sources of uncertainty, and each could be accounted for in a model or in the discount rate. Some are certainly most appropriate to include in the model, but I've argued that others, in particular the chance of unexpected new treatment methods, are hard to model and sufficiently context-insensitive that they are often better included in the discount rate.

If these factors always made it into our models we could be happy to leave them there, and omit them from discounting. But in practice they are not always included in the models — witness the discussion over the research and eradication paradoxes. They are also complicated enough that it would be an unreasonable burden to demand from all modellers. Given this, I believe that the appropriate route is to include them by default in a discount rate. This should be applied to projections of future YLDs, and projections of future deaths. This component should be excluded from the discount rate only in cases where it is replaced with something more accurate as part of the model.

Doing this would increase the relative value of direct health interventions relative to research projects or preventative measures whose effects may not be felt for many years, but because the discount rate is likely small the effect would be modest. This would help to correct a systematic bias in unadjusted projections, which consistently assume too little change in future knowledge.

---

[1] Parsonage, M. & Neuburger, H. (1992). Discounting and health benefits. *Health Econ*. Apr;1(1):71-6
[2] van Hout, B.A. (1998). Discounting costs and effects: a reconsideration. *Health Econ*. Nov;7(7):581-94
[3] Gravelle, H. & Smith, D. (2001). Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Econ*. Oct;10(7):587-99
[4] Torgerson, D. J. & Raftery, J. (1999). Discounting. *BMJ : British Medical Journal*, *319*(7214), 914–915
[5] Cairns, J.A. (1994). Valuing future benefits. *Health Econ*. Jul-Aug;3(4):221-9.
[6] Mulligan, C. (1996). A logical economist's argument against hyperbolic discounting. *University of Chicago mimeo*, February.

[7] The Green Book: Appraisal and Evaluation in Central Government (2003). *HM Treasury, London*

[8] Wang, H., Dwyer-Lindgren, L., Lofgren, K. T., Rajaratnam, J. K., Marcus, J. R., Levin-Rector, A., ... & Murray, C. J. (2013). Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, *380*(9859), 2071-2094.

[9] Douglas, J.M. (2009). Penicillin Treatment of Syphilis. *JAMA*. 301(7):769-771

[10] Henderson, D.A. (2011). The eradication of smallpox--an overview of the past, present, and future. *Vaccine*. Dec 30;29 Suppl 4:D7-9

[11] Bunker, J.P. (2001). The role of medical care in contributing to health improvements within societies, *Int. J. Epidemiol*. 30 (6): 1260-1263

[12] Mackenbach, J.P. & Looman, C.W. (2013). Life expectancy and national income in Europe, 1900-2008: an update of Preston's analysis. *Int J Epidemiol*. Aug;42(4):1100-10

[13] Data taken from http://vizhub.healthdata.org/gbd-compare/, February 2015

[14] Murray, Christopher JL, et al. "Comprehensive Systematic Analysis of Global Epidemiology: Definitions, Methods, Simplification of DALYs, and Comparative Results from the Global Burden of Disease 2010 Study."

[15] Cotton-Barratt, O. & Chandaria, S., On the marginal benefits of research, paper in preparation