

What should NICE's discount rate on future health-related well-being be?

Rob Pepper

Abstract

I argue that the National Institute for Health and Care Excellence (NICE) should use a lower discount rate on future health-related well-being¹ than its current 3.5% rate. A positive discount rate, applied to cost-effectiveness analysis (CEA), means the value of a given unit of health is considered greater the closer to the present it is realised. I recognise the discount rate should be higher than 0% because of uncertainty about the future and the instrumental benefits of earlier investment in health. Nonetheless, the current discount rate leads NICE to implausibly value healthcare interventions that treat existing conditions at dramatically higher levels than interventions that prevent future conditions. Though I do not specify what the actual discount rate should be, I argue that a rate lower than 3.5% would enable NICE to make more defensible trade-offs between present and future health.

1.

In this paper, I discuss how CEA should be applied to healthcare interventions with long-term effects. NICE, a UK government agency that undertakes CEA to guide health policy, down-weights future health using a discount rate. Current guidelines state that the costs and health effects of interventions should be

¹ For conciseness, I refer to “health-related well-being” as “health”.

discounted together at an annual rate of 3.5% (NICE, 2012). In practice, this means that 1 quality-adjusted life year² (QALY) or £100 next year has a value of 0.965 QALYs or £96.50 today. I argue that discounting future health is justifiable insofar as health is instrumentally valuable though not insofar as it is intrinsically valuable. To realise this, NICE's CEA discount rate should be lower than 3.5%.

In Section 2, I explain the rationale behind the 3.5% discount rate, distinguishing the instrumental and intrinsic justifications within it. I argue in Section 3 that this rate leads an implausibly low CEA ratio to be attached to interventions with long-term effects. In Section 4, I consider and accept Hilary Greaves' (2017) argument that there should at least some discount rate on the value of future health, given instrumental benefits of earlier investment in health and uncertainty about the future. In Sections 5 and 6, I consider and reject arguments presented by Joseph Heath (2017) that discounting the intrinsic future value of goods like health (even at a very low rate) is justified. I conclude in Section 7.

2.

Applying a positive discount rate to CEA means the value of a given unit of health (e.g. 1 QALY) is considered greater the closer to the present it is realised. In practice, if we were calculating the cost-effectiveness of an intervention whose health effects were realised over a period of z years, the CEA ratio would

² "A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life. One QALY is equal to 1 year of life in perfect health" (NICE, 2019).

be modified taking account of the discount rate, the value of z and the number of intermediate years between z and the present. Imagine, for example, a preventive intervention saved a person's life today and gave them 60 years of healthy life. Applying a 3.5% annual discount rate, as NICE currently does, would lead us to calculate that approximately 25 QALYs had been gained.

In NICE's CEA, future costs and benefits of healthcare interventions are, in general, discounted together at a rate of 3.5%. (As I explain in Section 3, however, NICE's guidelines do allow differential discounting of costs and benefit in exceptional situations). The reasons for discounting future costs and future health effects should, nonetheless, be differentiated. In explaining this, the distinction between intrinsic and instrumental value is useful. An intrinsically valuable good is valuable in and of itself, whereas an instrumentally valuable good is only valuable to the extent that it is a means to realise intrinsic value (by obtaining intrinsically valuable goods). Discounting instrumentally valuable goods may reflect the fact that their potential to realise intrinsic value declines over time. By contrast, discounting intrinsically valuable goods can only be the result of a pure time preference – in other words, a preference, which may or may not be justified, for realising intrinsic value sooner rather than later. Importantly, it is consistent to discount instrumentally valuable goods while not discounting intrinsically valuable goods.

Whereas money is only instrumentally valuable, health is arguably of both instrumental *and* intrinsic value. As I explain further in Section 4, health is instrumentally valuable because it enables the production of socially beneficial goods (which may even include *further* health improvements) and intrinsically

valuable in that persons living healthy lives is plausibly good in and of itself. Consequently, discounting future costs and future health should be justified in different ways. Discounting future costs is justified since the same amount of money spent on a healthcare intervention in the future will plausibly deliver less health (measured in QALYs) than it would if it were spent on an intervention in the present. This is because, even apart from inflationary reasons, economic growth means that future societies will be wealthier than present societies and more able to bear costs. Discounting future health, however, is not necessarily justified. It is not clear, insofar as health is intrinsically valuable, why it would be any less valuable if it were realised in the future rather than the present.

NICE's guidelines do not explain exactly why a discount rate of 3.5% for costs and health effects in CEA has been selected. Yet if, as is plausible, they are following guidance from HM Treasury's Green Book, then there are two reasons for it: (1) "time preference" and (2) "wealth effects", which correspond with intrinsic and instrumental reasons for discounting respectively (HM Treasury, 2018: 27). This corresponds with Frank Ramsey's equation for determining social discount rates (1928). Formally, this equation is usually presented in the following way:

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

On the left, δ refers to pure time preference. On the right, η and g mean elasticity of marginal utility of consumption and economic growth rate respectively. Multiplied together, η and g capture the idea that, as a society gets richer, each additional unit of resources will have a smaller and smaller effect on welfare. In other words, the left side of the equation corresponds to discounting

for intrinsic reasons, while the right side corresponds to discounting for instrumental reasons. To justify a given rate r , therefore, both intrinsic and instrumental reasons for discounting costs and health effects should be given.

3.

The need for a convincing justification of NICE's 3.5% discount rate is particularly acute given the intuitively implausible results to which it gives rise. Practically, the rate means that health interventions whose effects are realised immediately (treatment of existing health conditions), relative to interventions whose effects are realised over a longer period (preventive or public health interventions), are regarded as more valuable – perhaps substantially so.

Imagine that, in the present day, NICE were comparing the value of treating current patients with kidney failure, through renal dialysis, against the value of preventing citizens in the future from being afflicted with kidney failure in the first place, through a public health initiative aiming to improve diets. Idealising somewhat, let us assume that, before discounting, the costs per QALY gained through renal dialysis and the public health initiative are £20,000 and £7,000 respectively. The public health initiative clearly seems more cost-effective. However, the QALYs gained through renal dialysis will be realised immediately, while those gained through the public health initiative will be realised in 30 years. Applying a 3.5% discount rate, 1 QALY gained 30 years from now has a value of approximately 0.34 QALYs. The cost per QALY of the public health initiative, on this basis, is £20,588, making it less cost-effective than renal dialysis.

Intuitively, such results seem implausible. This intuition appears to have been shared by a Citizens Council, convened in 2011 to reflect on NICE's assessment of future costs and health benefits (2011b). The Council was "surprised and puzzled" by the implications of the discount rate, suggesting that the valuing of "treatment over prevention" it entailed was "paradoxical" (NICE, 2011b). Given that prevention ultimately leads to a diminished need for treatment, though the contrary is not necessarily true, the application of a 3.5% discount rate seems to conflict with the underlying rationale of CEA.

NICE do seem to have recognised this as an issue. In 2011, a document clarifying their guidance on technology appraisals stated that, in "special circumstances", discount rates of 1.5% and 3.5% should be applied to future health and future costs respectively (NICE, 2011a). Applying such differential discounting is intended to avoid the kinds of intuitively implausible results outlined above. However, this does not fully resolve the problem. NICE define such "special circumstances" as ones in which treatment effects are "substantial in restoring health and sustained over a very long period" (NICE, 2011a). As James F. O'Mahoney and Mike Paulden argue, this definition is "ambiguous" and "inconsistent" (2014: 493 – 5). It is not clear what a "substantial" effect is, nor how long a "sustained" effect lasts. Even if these were defined, moreover, they would give rise to arbitrary thresholds. Imagine that the threshold of a "sustained" effect were set at thirty years. On this rule, two interventions that deliver the same annual health benefit, but one for twenty-nine years and the other for thirty years would be judged differently; a lower, differential discount rate would be applied to the one with a thirty-year effect than to the one with a

twenty-nine-year effect. The result could be significant, with one intervention qualifying and the other failing to qualify for funding.

Granted, any form of CEA is prone to generate results that are – or at least appear to be – somewhat arbitrary, because thresholds of acceptable costs given desirable benefits must be set. Yet the arbitrariness of these thresholds is arguably justified in that it is the only feasible way to make effective decisions about the distribution of healthcare resources under constraints. This justification, by contrast, is not open to arbitrary temporal thresholds. They seem to be underpinned by the rationale that the intrinsic value of health does not depend on when it is realised, though this rationale is inconsistently implemented. This view is supported by the surprise of the Citizen Council at the implications of NICE's 3.5% discount rate. The Council's comments are indicative of an underlying assumption: that future health should not be valued at a significantly lower level than present health. To defend the 3.5% discount rate, convincing arguments against this assumption must be offered.

4.

A discount rate may be applied to future health without necessarily entailing that present and future health are not of equal intrinsic value. Greaves provides two persuasive reasons for discounting future health that do not appeal to a positive pure time preference.

Firstly, earlier investment in healthcare may be a beneficial decision for both present and future society because it delivers instrumental gains by increasing economic productivity (Greaves, 2017: 5). Improved health in a

population means that *more* workers are able to spend *more* time being economically productive. "Economic productivity is [...] more valuable the earlier it occurs", Greaves writes, "because its outputs can [...] be put to work" (2017: 5). Earlier investment in healthcare might lead to a virtuous circle, as rising levels of economic productivity enable higher economic growth, and more government revenues which can be reinvested in improving health.

Secondly, uncertainty about the future means that policymakers cannot necessarily be sure that health interventions will produce their expected effects (Greaves, 2017: 3 - 4). The risk of human extinction represents one dramatic way in which this could be the case. However, a health intervention may also fail to bring about its expected benefit because of some change in the environment in which it is being delivered (for instance, subsequent policymakers may decide to cancel the intervention), or simply because concurrent improvements in healthcare mean the benefit has already been delivered by the time that the relevant intervention would have done so. A discount rate enables such uncertainty to be factored into CEA.

Discounting, however, is not the only way to capture uncertainty in CEA. For instance, an intervention's expected benefit could be calculated using probabilistic methods. An advantage of this approach would be that these probabilities could be varied according to the contexts of different interventions, with different chances of success. However, the kinds of uncertainty to which Greaves refers may be severe, such that it might not be possible to assign precise probabilities to them. In this case, a uniformly applied discount rate may be preferable to *ad hoc* probabilistic methods for pragmatic reasons. For one

thing, this would mean greater transparency in the way in which uncertainty is accounted for in CEA. Moreover, if some interventions are more and others less effective than anticipated, the effects of uncertainty may cancel out in the long-run with a uniform way of capturing it.

I want to flag here that it is not abundantly clear whether uncertainty should count as an instrumental or intrinsic reason for discounting. On the one hand, it does not seem strictly instrumental as it does not clearly relate to either the elasticity of marginal utility of consumption or the economic growth rate. On the other hand, it does not seem strictly intrinsic either. The preference for realising health in the present rather than the future it entails does not necessarily imply future health is in and of itself less valuable than present health (as a pure time preference does), only that there is a risk of foregoing some health gains due to uncertainty about the future that should be factored into calculations. Though somewhat ambivalent, I think uncertainty is, on balance, more an instrumental reason for discounting and assume this for the remainder of the paper.

Given the two reasons Greaves presents, I would argue that at least some discount rate on future health is justified. This is not to say, however, that NICE's 3.5% discount rate necessarily is, since, as I noted in Section 2, this specific rate is justified for intrinsic as well as instrumental reasons. To the extent that it is motivated by a positive pure time preference for health realised closer to the present, NICE's 3.5% discount rate requires further justification.

5.

Very few philosophers, however, defend a positive pure time preference (which, following Heath, I will treat as synonymous with a positive social time preference). Heath's attempt to justify a positive but low social time preference in the context of climate ethics stands out. If successful, Heath's argument may provide a justification for the pure time preference that underlies NICE's 3.5% discount rate.

Before presenting an argument for a positive social time preference, Heath motivates his case by responding to two major arguments against one (2017: 442 – 7). The first is that a positive social time preference is tantamount to discrimination against future persons, because it means their welfare is considered less valuable than that of present persons. The second is that, since when something happens is just as morally irrelevant as where it happens (assuming we accept where something happens is a morally irrelevant consideration), a positive social time preference violates a kind of neutrality with respect to time we should have in moral considerations. Derek Parfit, who advances the second argument, even claims that a "temporal discount rate" is "as little justified" as a "spatial discount rate" (1984: 486).

Heath claims both arguments, as they stand, do not convincingly show a positive but low social time preference to be unjustified. He argues the first depends on a confused understanding of how temporal discounting works in practice. Provided it is instantiated in policy that is consistent over time, temporal discounting means that everybody's welfare is considered less valuable the further into the future it is likely to be realised. With the passing of time,

moreover, the welfare of future persons is considered increasingly valuable, unlike familiar cases of discrimination, in which the welfare of oppressed groups is undervalued in a systematic and unchanging way.

In response to the second argument, Heath contests the analogy that Parfit draws between spatial and temporal proximity. Space and time, Heath suggests, may be different in a morally relevant way. Everyone who is alive today, regardless of where they are, are actual persons, while those who are not yet alive are merely possible persons. Actual persons, moreover, may have rights that merely possible persons do not. It is plausible, for instance, that an individual's "right to life" comes "only sometime *after* conception", so, while actual persons have this right, merely possible persons do not (Heath, 2017: 446). A positive social time preference could reflect the fact that these additional rights of actual persons need to be factored into decision-making about policies affecting the future.

Having motivated his case, Heath presents an argument for a positive social time preference (2017: 456 – 61). In a rule utilitarian sense, Heath argues that the most effective way to realise our "abstract moral commitments" at society level is through a shared system of rules that he calls "institutional morality". For example, the most effective way to realise an imperfect duty to aid others at society level could be by institutionalising certain rules about helping friends and relations, rather than demanding each individual to consider whom among billions of people they should help on each occasion they can.

Heath suggests that temporal limitations, in the form of temporal discounting, could be among such rules. Therefore, while temporal limitations

need not have any intrinsic moral significance, they may have an instrumental moral significance as “the system of institutional morality makes use of [them] [...] as a way of creating determinate moral obligations” (Heath, 2017: 457). Though such temporal limitations would likely give rise to some apparently arbitrary results, the point is that they would (purportedly) allow us to realise our abstract moral commitments more effectively than we could were they lifted.

This justification of a positive social time preference, I would suggest, is a sophisticated form of a demandingness objection. Without a positive social time preference, Heath implies, it may be necessary that we sacrifice practically all present consumption in order to pursue long-term benefits. Assuming that this is unfeasible and could even result in paralysis, it would be practically better to adopt a positive social time preference. For instance, in the context of an urgent need to address climate change but huge problems in agreeing how much present generations should curb their carbon emissions, applying a positive social time preference to the social cost of carbon may be a way to ensure the maximum possible reductions in emissions.

Though Heath’s argument is formulated in the context of climate change, it could theoretically apply to priority-setting in healthcare too. In other words, a pure time preference incorporated within NICE’s discount rate on future health may be justified if it leads to the maximisation of QALYs. Yet, as I argue in Section 6, even aside from the issue of whether Heath’s proposal could plausibly apply in a health as well as an environmental context, there are considerable problems with his argument.

6.

Heath's argument, somewhat counter-intuitively, is an instrumental justification for discounting intrinsically valuable goods. Yet this justification, at least in the context of discounting future health, arguably confuses instrumental and intrinsic reasons for discounting. As I argued in Section 4, there are sound reasons for discounting health insofar as it is instrumentally valuable or, in the case of uncertainty, when we are not sure that healthcare interventions will be as effective as anticipated. However, it is not clear that additionally treating future health as if it were intrinsically less valuable than present health through a pure time preference would result in the instrumentally better outcome Heath suggests. In fact, adopting a pure time preference for instrumental reasons may simply obscure our underlying reason for doing so in the first place.

Heath's argument for a pure time preference from "institutional morality" requires, like demandingness objections, empirical justification. Otherwise, it is not a sound argument. Institutional morality, as Heath calls it, is a political rather than a moral concept; it provides guidance about what is feasible, rather than what is right. Therefore, if temporal limitations do not in fact make it easier for abstract moral commitments to be met, then they are unjustified. I accept that, with respect to priority-setting in healthcare, removing all temporal limitations (meaning a 0% discount rate on future health) may have implications that are too demanding to be effectively realised. Since prevention is usually substantially cheaper than treatment, though its effects are felt over a longer period, a 0% discount rate may entail the radical defunding of treatment to shift resources to prevention. Yet this could be avoided without a pure time

preference with respect to future health, by discounting health only insofar as it is instrumentally valuable. The urgency of Heath's demandingness objection is therefore blunted.

Heath's contention that temporal proximity may be a morally relevant factor is more challenging to address. It is significant, however, that this argument assumes some form of deontology. The potential moral significance of temporal proximity stems from the fact that actual persons have certain rights, which merely possible persons do not. By contrast, CEA assumes an essentially consequentialist view. It involves seeing health benefits as valuable not because they enable the rights of individuals to be respected, but because they are good consequences in and of themselves. While this observation does not respond directly to Heath's argument, it does suggest that it would be strange to justify a discount rate to CEA for deontological reasons. On its own, a discount rate would be a crude way of capturing deontological concerns. It would affect a very young but actual person and a merely possible but soon-to-be-born person in similar ways, even though the distinction between actual and merely possible persons is the one it is supposed to track. Further consideration of deontological objections to CEA is beyond the scope of this paper, but it suffices to say that CEA as NICE employs it would likely not exist if underpinned by a deontological view.

If we are to recognise that there are practical reasons against a discount rate on future health, we should also accept that there may be practical reasons for it. John Broome highlights problems that arise over time with a positive discount rate on future health. Prioritising present health, Broome argues, could "impl[y] that a government may rightly select a policy in the full knowledge that

in a few years time it will rightly judge that policy to have been wrong" (2016: 909). Granted, exponential discounting, like NICE's 3.5% discount rate, does not lead to the kind of dynamic inconsistency that hyperbolic discounting does. Decision-makers applying hyperbolic discounting are prone to make decisions in the present that, even with exactly the same information, their future selves would not have made. Nonetheless, in interpersonal cases exponential discounting does produce seemingly inconsistent results. To illustrate this problem, consider again the example of treating or preventing kidney failure that I introduced in Section 3.

Imagine that, in the present day, NICE were comparing the value of treating current patients with kidney failure, through renal dialysis, against the value of preventing citizens in the future from being afflicted with kidney failure in the first place, through a public health initiative aiming to improve diets. Discounting the value of future health at a rate of 3.5%, the institution may judge that the treatment is more cost-effective than the prevention. Yet, if the institution were comparing the same interventions in the future, they may make the *opposite* judgement that the prevention would have been more cost-effective than the treatment. Because the health benefits of the prevention would be realised closer in time, the value of these health benefits, even after discounting, may, at this future point, be greater than the value of the health benefits from the "treatment".

Discounting the value of future health could thus lead to a situation in which an evaluating institution would have to accept that their past judgement was wrong. This would not be, in Broome's words, "consistent with rational

agency” (2016: 909). In intrapersonal cases, in which an individual agent makes judgements about their preferences over time, this violation of rational choice may not be such a problem, because the individual agent *themselves* reaps the earlier reward, as well as paying the later costs. If I were choosing what kind of health-care I want for myself, my decision to opt for immediate treatment rather than longer-term prevention would be less problematic. However, in complex, interpersonal cases, such as priority-setting in healthcare, those who benefit from the earlier reward are different from those who pay the later costs. This situation is more clearly problematic, because those who pay the later costs may reasonably claim that they have been treated unfairly – and the evaluating institution may have reason to agree. While any discount rate on the value of future health may be susceptible to this problem, the relatively higher discount rate of 3.5% would be particularly liable to generate it.

7.

In this paper, I have argued that NICE’s current discount rate of 3.5% on costs and health effects is too high, leading to implausibly low cost-effectiveness ratios to be assigned to interventions with long-term effects. I recognise that some form of discounting is necessary on future health is necessary because of uncertainty about the future and the instrumental benefits of earlier investment in healthcare but maintain that discounting on the basis of a pure time preference with respect to health is not clearly justified. Since NICE’s current 3.5% discount rate is seemingly justified for both instrumental and intrinsic reasons, it should be reduced to a level at which it is justified only for instrumental reasons.

A lower discount rate in CEA, I suggest, would enable NICE to make more defensible trade-offs between the value of present and future health. I close by observing that, however, that the usefulness of CEA itself, even with a lower discount rate, is likely to decrease when it is used to evaluate interventions which have extremely long-term and uncertain effects. CEA is a particularly useful tool when it comes to precisely weighing costs against QALYs. When QALYs will be realised very far in the future, it is very difficult to estimate their values. Using CEA in this context may be like using a magnifying glass to look at the moon.

Bibliography

- Broome, John, (2016) 'The Well-Being of Future Generations' in *The Oxford Handbook of Well-Being and Public Policy*, eds. Matthew D. Adler and Marc Fleurbaey, (Oxford: Oxford University Press), pp. 901 – 928
- Greaves, Hilary, (2017) 'Discounting Future Health', Draft v8 (2017), *Global Health Priority-Setting: Cost-Effectiveness and Beyond*, eds. Ezekiel Emanuel et al., (Oxford: Oxford University Press), forthcoming, pp. 1 – 11
http://users.ox.ac.uk/~mert2255/papers/discounting_health.pdf
[Accessed 18th Nov 2018]
- Heath, Joseph, (2017) 'Climate Ethics: Justifying a Positive Social Time Preference', *Journal of Moral Philosophy*, 14, pp. 435 – 462
- HM Treasury, (2018) 'The Green Book: Appraisal and Evaluation in Central Government', <https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government> [Accessed 15th April 2019]
- National Institute for Health and Care Excellence, (2011a) 'Discounting of health benefits in special circumstances',
http://www.nice.org.uk/media/955/4F/Clarification_to_section_5.6_of_the_Guide_to_Methods_of_Technology_Appraisals.pdf [Accessed 18th Nov 2018]
- National Institute for Health and Care Excellence, (2011b) 'How Should NICE Assess Future Costs and Health Benefits?', NICE Citizen Council Reports,

No. 15, <https://www.ncbi.nlm.nih.gov/pubmed/28230942> [Accessed 18th Nov 2018]

National Institute for Health and Care Excellence, (2012) *The Guidelines Manual*, London <https://www.nice.org.uk/process/pmg6/chapter/assessing-cost-effectiveness> [Accessed 18th Nov 2018]

National Institute for Health and Care Excellence, (2019) 'Quality-Adjusted Life Year', *Glossary*, <https://www.nice.org.uk/glossary?letter=q> [Accessed 15th April 2019]

O'Mahoney, James F. and Paulden, Mike, (2014) 'NICE's Selective Application of Differential Discounting: Ambiguous, Inconsistent and Unjustified', *Value in Health*, 16 (3), pp. 493 – 6

Parfit, Derek, (1984) *Reasons and Persons* (Oxford: Oxford University Press)

Ramsey, Frank, (1928) 'A Mathematical Theory of Saving', *The Economic Journal*, 38 (152), pp. 543 - 559