Value choices in summary measures of population health

S. Andrew Schroeder
Claremont McKenna College
aschroeder@cmc.edu

Abstract. Summary measures of health, such as the QALY and DALY, have long been known to incorporate a number of value choices. In this paper, though, I show that the value choices in the construction of such measures extend far beyond what is generally recognized. In showing this, I hope both to improve the understanding of those measures by epidemiologists, health economists, and policy-makers, and also to contribute to the general debate about the extent to which such measures should be adjusted to reflect ethical values.

Introduction

The quality-adjusted life year (QALY) and disability-adjusted life year (DALY), as well as

related measures such as health-adjusted life expectancy (HALE), are typically described as measures of

health.[1] Accordingly, they are often treated by scientists and policy-makers as "objective" indicators, of a

kind with measures like mortality rates and disease prevalences, which aim to quantify features of the

natural world. At the same time, though, QALYs and DALYs have long been known to incorporate a

number of value choices. Perhaps the clearest example of this is the practice of *age-weighting*, a

prominent part of several iterations of the Global Burden of Disease Study (GBD). In the 1990 GBD,

health losses were given differential weight, based on the age of the person suffering the loss. The

weighting factor climbed quickly from zero at birth to a maximum at age 25, then gradually declined into

old age. The justification given for age-weighting in this way was an ethical one: "The well-being of

some age groups, we argue, is instrumental in making society flourish; therefore collectively we may be

more concerned with improving health status for individuals in these age groups" (Murray and Acharya

1997). So, when the 1990 GBD attributed 28 million years of life lost to malaria, it did not mean that 28

---

[1] See e.g. Murray *et al.* (2002), which begins, "Summary measures of population health…combine information on mortality and non-fatal health outcomes *to represent the health of a particular population* as a single numerical index" (*emphasis added*).

million actual years of life were lost to malaria. Instead, it meant that some number of years were lost to malaria which, *when weighted according to the given function*, yielded 28 million. Therefore, the measurement "28 million years of life lost" represented both an empirical fact (that a number of people were killed by malaria at particular ages) *and* an ethical claim (that health at different ages should be of different social concern).

This departure from objectivity, and others like it, has sparked a significant debate. Some have endorsed value-based practices like age-weighting, arguing that since QALYs and DALYs often play an important role in decision-making, we should want them to incorporate whatever values are relevant to those decisions. Here, for example, is Alan Williams:

> If the nature and implications of particular positions are to be clarified in a policy-relevant way, this discussion has to move on to seek quantification of what are otherwise merely vaguely appealing but ambiguous [ethical] slogans... Only with some quantification will it be possible to devise rules that can be applied in a consistent manner with a reasonable chance of checking on performance. (Williams 1997: 120; *cf.* Menzel *et al.* (1999), Nord (1999; 2014), Schokkaert (2015))

Others, however, have argued that QALYs and DALYs should remain as free of values as possible. A variety of justifications are offered for this view: that health economists are not experts when it comes to questions of value, that issues connected to value are the province of decision-makers and not scientists, that including values may tend to confuse decision-makers, and so forth.[2] And, setting aside these specific worries, there does seem to be something important about having a "pure" measure of health. It would seem odd to weight mortality rates or disease prevalences to account for ethical judgments about the importance of different people's lives. So perhaps there is good reason to preserve QALYs and DALYs as objective (or largely objective) measures of health.[3]

---

[2] For samples of these arguments, see Arnesen and Kapiriri (2004), Nygaard (2000), Barker and Green (1996), and Lyttkens (2003). There is also, of course, a long history of treating science as a discipline free of (non-epistemic) values. See Kincaid, Dupré, and Wylie (2007) for a collection of essays on that ideal, which, despite recently receiving more critical attention, is still probably the dominant view among scientists. See also note 9, below.

[3] The GBD team has recently offered an argument of this sort to justify the removal of age-weighting and discounting from its primary results, claiming that DALYs are best viewed "as a strict summary measure of population health"(Murray *et al.* 2012: 16).

My sympathies on this issue lie closer to the former, value-friendly point of view, but I won't argue for that here. Instead, I would like to lay some necessary but overlooked groundwork for the debate. Thus far, the academic and policy literature has focused almost exclusively on a relatively small number of value choices: age-weighting, temporal discounting, health state valuation, and (occasionally) choice of life expectancy. Many commentators implicitly or explicitly assert that these are the only important value choices commonly built into measures of health. In their well-known paper on value choices in the DALY, for example, Anand and Hanson (1998) mention only disability weights, age-weighting and discounting; Arnesen and Kapiriri (2004) do the same, while briefly mentioning life expectancy as an additional "arbitrary choice." Empirical studies are generally worse, often explicitly stating that DALYs incorporate only a limited number of value choices. Chapman *et al.* (2006) note more value choices than most when they say,

> Four value choices are associated with the DALY health measure. First, a premature death is defined as a death occurring before age- and sex-specific limits based on a model life table… Secondly, the standard DALY indicator employs a non-uniform age weighting function… As a third value choice, the future health events are discounted at a rate of 3% per year. Finally, a set of disability weights has been developed to reflect the relative value of time lived in various compromised health states.[4]

This widely-accepted picture is wrong. Many more value choices are standardly built into QALYs and DALYs. In this paper, I will demonstrate this by looking at a particular case: the use of DALYs in burden of disease studies, and in particular in the Global Burden of Disease Study. I choose this example because the value choices in the DALY have been subjected to greater academic scrutiny than the value choices in other measures of health. Nevertheless, I will show that the value choices in the DALY extend far beyond the few examples that are typically acknowledged.

---

[4] For studies that strongly suggest there are less than four value choices involved, see Michaud *et al.* 2006 (life expectancy, discounting, age-weighting); Robberstad *et al.* 2007 (discounting and age-weighting); and Plass *et al.* 2013 (discounting and age-weighting).

In showing this, I aim to accomplish two things. First, it is important for producers and consumers of studies like the GBD to be aware of the many value choices that go into its construction. Although many (but not all) of the value choices I describe have been identified somewhere in the literature, no source lays out these value choices in a systematic way, and discussions of any value choices beyond discounting, age-weighting, and health state valuation are hard to find.[5] With this paper, then, I hope to contribute to the ability of epidemiologists, health economists, and decision-makers to properly interpret GBD results and accordingly to make informed decisions on their basis. Second, I believe that greater clarity about the extent to which QALYs and DALYs are structured by values will lead to a more informed debate about values in measures of health. Both advocates and critics of incorporating additional values into measures of health — values relating to inequality, severity of illness, and aggregation, for example — should begin with a sound understanding of the extent to which those measures are already permeated with values. No existing source provides this.

A preliminary matter[6]

Before looking at the specific value choices built in to the DALY, we should briefly explore what it means to say that a value has been incorporated into a scientific measure, or (for short) for a measure to be *value-laden*. Surprisingly, though there is much discussion of value choices in summary measures of health, and similar discussions of values in other composite measures (such as price indices),[7] I am not aware of any explicit discussion of what it means for a value to be incorporated into a measure. The issue is, I think, more complicated than it may appear. Take the example of discounting future health benefits. Does this involve a value choice? Many, including the authors of the GBD, think that it does, arguing that discounting amounts to saying that health in the future is (all else equal) less important than health in

---

[5] The most comprehensive source to date is still Murray (1996), but it includes discussions of both ethical and non-ethical methodological issues, in ways that are not always easy to distinguish. Other important, broad discussions include Voigt (2012), Murray *et al.* (2002), and sections of Murray *et al.* (2012).
[6] I thank two anonymous reviewers for pressing me on many aspects of this section; it is greatly improved as a result.
[7] For a good discussion of value choices in price indices, see Reiss (2008, ch.2-3).

the present (Anand and Hanson 1998). But others claim that this isn't the case. Weinstein and Stason, for example, say that discounting need not "assume that life years in the future are less valuable than life years today, in any absolute utilitarian sense" (1977: 720, *cf.* Sunstein and Rowell 2007). So who is right? Does discounting involve a value choice, or doesn't it?

The answer, I think, is that it depends. Consider the following five (caricatured) health economists, who each discount the future health benefits from a vaccine program at a 3% rate:

> • *Alice discounts because a recent, well-regarded study showed that there is a 3% chance per year of a technological advance that would inexpensively cure the illness in question, rendering the vaccine unnecessary.*
> • *Bob discounts because he believes that there is a 3% chance each year that the world will end.*
> • *Carol discounts because she believes, following Weinstein, Stason, and others, that not discounting yields nonsensical or inconsistent results.*
> • *David discounts because the equation he mindlessly copied from his textbook included a 3% discount rate.*
> • *Ellen discounts because she believes we have stronger political obligations to care for our contemporaries.*

My sense is that Alice's discounting clearly does not involve a value choice and Ellen's clearly does. The middle three cases are harder. Bob's case looks a lot like Alice's, in that he discounts on the basis of an empirical belief, and so we might be inclined to say that his discounting does not involve a value choice. At the same time, though, his empirical belief is an absurd one. A better justification for Bob's discounting might therefore appeal to considerations like Ellen's. And if the best or most charitable justification of Bob's discounting appeals to values, then it might make sense to treat it as value-laden.[8] With Carol, we might want to know more about the sense in which non-discounted results are supposedly "nonsensical." If Carol's claim is that not discounting leads to straightforward inconsistencies (e.g. that $7=$1), that wouldn't seem to involve any ethical value. On the other hand, if she believes that not discounting leads to the "nonsensical" conclusion that we should indefinitely postpone spending on health, that seems to be based on the ethical claim that we ought not indefinitely postpone spending.

---

[8] Another way of putting this is that, to a consumer of Bob's study, it is best treated as equivalent to Ellen's, and therefore value-laden.

Finally turning to David, we might want to know more about the history of his textbook. If Ellen was the textbook author, we might be inclined to think David's study "inherits" Ellen's values, and accordingly should be considered value-laden. If, on the other hand, the author had no explicitly ethical motivations or if the textbook was written in a way consistent with disciplinary norms, then I think it becomes less clear what to say.

Bob, Carol, and David, then, appear to represent complicated cases — and they are only the tip of a very large iceberg. So what should we say about them? The solution, it seems to me, is to acknowledge that there are a variety of different respects in which a measure can be value-laden. The most conservative account would focus only on cases like Ellen's, saying that a measure is value-laden when scientists adopt it or structure it in a particular way for value-based reasons. More liberal accounts might also include cases like David's (where a measure can ultimately be traced back to values), some versions of Carol's (where a measure was structured based on implicit or unrecognized values), Bob's (where the best justification for structuring a measure in a particular way is grounded in values), or others we haven't discussed here. For the purposes of this paper, I will adopt the most conservative account, calling a measure value-laden if it is adopted or structured in a particular way for value-based reasons. An obvious (and, I think, sufficient) reason for choosing that standard is that it will reinforce how value-laden the DALY is: I will show that *even on the most restrictive account of value-ladenness*, the DALY incorporates many more values than are generally recognized. (Given a more liberal account, of course, the DALY would likely turn out to be value-laden in many additional ways.)

I think there is also, however, a substantive reason to be interested in this sort of value-ladenness. There is a large philosophical and scientific literature on the relationship between social values and science. Though it has recently come under growing attack, the dominant view has been that social values should play, at most, a limited role in science. Many of the strongest and most influential arguments offered for this view directly or indirectly target the value-based *motives* of scientists. Commentators, for example, frequently worry about the politicization of science, about the lack of trust that may ensue when scientists are viewed as advocates, and that researchers may pre-determine the

outcome of their research to ensure that it coheres with their values. Another set of arguments claims that scientists generally step beyond their expertise when they venture into the domain of values, and argues that questions of value are appropriately resolved by judges, policy-makers, or other political bodies, not scientists.[9]

Set aside the question of whether these arguments are good ones. (I think that several of them are, but others are not.) Note that Ellen's use of discounting arguably raises all of these worries: she adopts and (implicitly) advocates for a particular view about intergenerational justice, and for that reason her results may reasonably be viewed skeptically by those with different ethical views. In making claims about intergenerational justice, she could be taken to step outside her professional expertise, and in some sense to insert herself into a decision-making process which is appropriately resolved through democratic or judicial processes. Alice's, Bob's, Carol's and David's actions, however, aren't like this. None of them means to adopt or advocate for any ethical position through their use of a 3% discount rate. For that reason, there seems to be no special worry that they have manipulated their research to support any particular conclusion. None of them claims expertise in matters of value. And therefore none of them is attempting to play a role in a decision-making process, apart from providing empirical information. If, then, many critics of incorporating values into science are, justifiably or not, especially worried about matters connected to the reasons for which scientists make certain decisions, I think it is important to pay attention to those reasons.

This analysis also enables us to distinguish the issue of when measures are value-laden from a distinct issue, with which it is frequently confused. QALYs are sometimes described as incorporating the assumption that any given year of life in full health is of equal moral value.[10] This is related to the claim that un-age-weighted DALYs are just as value-laden as age-weighted DALYs, since un-age-weighted DALYs are based on the ethical assumption that a year of healthy life at age 20 is of the same value as a

---

[9] For more philosophical versions of these arguments, see Betz (2013), Haack (1993), Pielke (2007), and Anderson (2004). For expressions of these views by scientists from a variety of disciplines, see Lackey (2007), Bray (2010), Bergman *et al.* (2013), and Rykiel (2001).

[10] This view is commonly attributed, perhaps unfairly, to Harris (1987).

year of healthy life at age 90. We can see why these claims are mistaken by looking at an analogous situation. Suppose I am looking for a new apartment and measure the square footage of each potential dwelling. The resulting measurements need not be based on the assumption that each square foot is of equal value — that, for example, an extra square foot of closet space is of equal value to an extra square foot of kitchen space. In measuring square footage, I can reasonably claim to be agnostic concerning the relative value of each square foot, and so I see no reason to declare those measurements value-laden. Of course, if I go on to choose an apartment solely by comparing square footage (or price per square foot), then it may be correct to say that my decision-making process implicitly presupposes that each square foot is of equal value. But the value judgment there lies in my *use* of the measure, not in the measure itself.

Analogously, we can see that QALYs and DALYs need not incorporate a presupposition that each year of healthy life is of equal moral value. It is true that certain uses of them involve this commitment, but that value need not be part of the measure itself. In what follows, I will not be interested in the question of what values are presupposed in the *use* of summary measures of health. (To answer that question, we would need to investigate the many different ways that QALYs and DALYs can be used.) Instead, I will focus on the question of when the measures themselves are value-laden.

Values frequently in DALYs

According to a very restrictive account, then, we can call a measure value-laden if it was adopted or structured as it was for value-based reasons. Let us now turn to the Global Burden of Disease Study, which purports to use DALYs to measure the "health loss" or "disease burden" associated with various diseases, injuries, and risk factors. In this section, I will identify at least eight respects in which the GBD's understanding of what constitutes health loss or disease burden is guided by ethical values. In accordance with our conservative account of value-ladenness, in each case I will highlight the reasons given for the choice of a measure or method of calculation over alternatives, showing that the choice was made on the basis of social or ethical values.

8

Age-weighting

As we saw above, DALYs in the 1990 GBD Study — as well as a number of other studies that adopted the same methodology (Polinder *et al.* 2012) — applied a weighting factor to health losses, based on the idea that the social value of a health loss varied, based on the age of the patient. (Additional arguments for age-weighting have been grounded in considerations of economic productivity, benefit to others, concern for the worst-off, fairness, and prudential allocation across a lifespan (Bognar 2008, Tsuchiya 2000).) This justification is clearly an ethical one, and thus age-weighting is perhaps the least controversial sense in which DALYs are frequently value-laden.

Recent updates to the GBD have abandoned the use of unequal age-weights. Whether this equal weighting constitutes a value choice, on our restrictive account, depends on why it was adopted. If it was justified by an assertion that health at all ages should be of equal concern, then that would count as value-laden. If, on the other hand, it was adopted out of a desire to simply measure health losses, while remaining agnostic about the ethical value of those losses, then it need not be value-laden. The authors of the GBD aren't entirely clear, but it appears that they intend to take the value-free route.[11] It remains to be seen whether other burden of disease studies will follow the GBD in abandoning unequal age-weights.

Discounting

Along with age-weighting, the most commonly-discussed value incorporated into DALYs is that involved with temporal discounting, the standard economic practice of counting effects that come in the future as less significant than otherwise comparable effects in the present. As noted above, there are non-value-based justifications for discounting — though, arguably, they are typically not good justifications.[12] Nevertheless, the authors of early versions of the GBD are explicit that they discount for value-based

---

[11] They justify their decision to remove age-weights by describing DALYs as "a strict summary measure of population health" — though later they do note that "one can argue that a year of health life regardless of age should be treated equally." (Murray *et al.* 2012: 16).

[12] At least for discounting at any substantial rate. See Nord (2011) and Menzel (2011) for discussions.

reasons. In particular, Murray is most worried that in the absence of a discount rate, present generations

may be asked to make excessive sacrifices on behalf of future generations (Murray 1996: 53). In recent

versions of the GBD, Murray and colleagues have rejected their earlier argument and accordingly do not

discount future health benefits (Murray *et al.* 2012: 15-6). Discounting, however, is standard practice in

most other burden of disease studies (Polinder *et al.* 2012), and it remains to be seen whether future

studies will follow the GBD's lead in abandoning discounting.

Life expectancy (length)

Though most commentators focus on age-weighting and discounting as the primary value choices

in DALYs, the authors of the GBD are clear about a third important value choice. In calculating the years

of life lost (YLLs) to some disease or injury, a life expectancy must be assigned to each person to enable

us to decide how many years premature a given death was. The natural and arguably value-free way to do

this would be to use a *counterfactual* definition: to determine how many years of life were lost in, say, a

fatal car accident, we could simply ask how many more years the victim would likely have lived, had she

not died in the car accident. But, as Murray points out, to do this

> would lead us to conclude that the death of a 40 year-old woman in Kigali contributes less to the
> global burden of disease than the death of a 40 year-old woman in Paris because the expectation
> of life at age 40 is lower in Rwanda than in France. Equivalent health outcomes would be a
> greater burden in richer communities than in poorer communities. (Murray 1996: 14)

This seems unfair. Certain populations are already disadvantaged by low life expectancy, and on this

proposal deaths in those populations would register fewer DALYs because of that disadvantage. If we're

trying to calculate the health loss experienced by populations, or the burden of ill health on populations,

then it seems perverse to count a death for less because the victim would otherwise have died prematurely

of another health problem.

To avoid this result, Murray proposes that we apply the same life expectancy to all deaths at a

given age, making the death of every 40 year-old woman count the same, regardless of her personal

characteristics. He adopts this proposal because of its "egalitarian nature" (1996: 15). This is, clearly, a

value choice: this proposal was adopted because the obvious alternative -- to simply count how many

years were actually lost due to the death -- yielded (by Murray's lights) morally unacceptable results.[13]

Life expectancy (gender)

In early versions of the GBD, women were given a slightly longer life expectancy (2.5 years at

birth) than men at each age. This was done to represent "a biological difference in survival potential for

males and females" (Murray 1996: 17). If we assume that 2.5 years represented Murray's best estimate of

the biological difference in survival potential for men and women, then this doesn't involve a value

choice, and hence the resulting YLL calculations are not value-laden. In the most recent GBD, however,

the difference in life expectancy by gender was removed, perhaps because it was in tension with the

fairness-based considerations that led Murray to use a standard global life expectancy:

> Several arguments contributed to the view that the same reference standard should be used for
> males and females. First, the empirically observed gap in life expectancy between males and
> females continues to narrow… Second, there is no reason that society should have lower
> aspirations for health for males than females. (Murray *et al.* 2012: 14)

The assertion that society *should* have the same aspiration for men and women is of course an ethical one.

Since this seems to be a critical part of the argument (there is no claim that the empirically observed gap

has or will completely disappear), the use of the same life expectancy for men and women in recent

versions of the GBD is based on a value judgment.

Health state valuations

The quality/disability weights used to compare different states of ill health are typically arrived at

through surveys. Because health states are assigned values on a 0-1 scale, many commentators seem to

---

[13] This section argues that a value choice is involved in using a single standard life expectancy across the globe. I suspect a value choice is also involved in choosing precisely what life expectancy to use as that standard. My argument for that claim would require a fair amount of background information and would be somewhat speculative, though, so I won't explore it here.

assume that this embodies a value judgment. Things, however, aren't that simple. The authors of recent

iterations of the GBD maintain that their quality/disability weights measure a *quantity* of health, and the

wording from their most recent surveys reflects this (Salomon *et al.* 2003; 2012). Their idea is that,

intuitively, we have no trouble making judgments of relative health. We say, for example, that someone

with a mild cold is *more healthy* than someone bedridden with pneumonia. The GBD team intends for

their surveys to quantify just how much more healthy the person with the mild cold is. The resulting

quality/disability weights, therefore, represent *quantities* of health, not the *value* of different health states.

Now, Hausman (2015) has persuasively argued that this view of health is mistaken, and that there is no

sense in which health can be quantified in the way the GBD team supposes. Even if Hausman is correct,

though, it may still be true that the disability weights used in the GBD should not count as value-laden, at

least on our restrictive account of value-ladenness. As we saw above, the justification given for a

measure matters. If the GBD team is trying to measure quantity of health, then, even if such a thing

doesn't exist, their results needn't be considered value-laden.[14]

      That said, few other studies take this approach. Most, including earlier versions of the GBD,

acknowledge that quality/disability weights *value* health, measuring how *good* or *bad* it is (Murray 1996).

Now, since these valuations of health are arrived at through surveys, it might be argued that although the

resulting measurements are value-laden, they are not laden with the values of the scientists. Instead, they

are laden with the values of the survey participants.[15] Since, as noted earlier, many concerns about

incorporating social values into science are grounded in concerns about *scientists'* motives, this could be

an important distinction, since it might absolve scientists of stepping beyond their professional expertise,

inappropriately inserting themselves into political processes, and so forth. Even if that is correct, though,

---

[14] If Hausman is right, then there is of course a serious flaw in the GBD, since it purports to measure something unmeasurable. It therefore seems similar to the case of Bob, above, who discounted on the basis of an absurd empirical belief. As I noted there, more liberal accounts of value-ladenness might describe such cases as value-laden.

[15] See Broome (2008) for a critical discussion of the general reluctance of economists to acknowledge making assertions of value, and health economists' attempts to avoid making value judgments by instead claiming only to report the value judgments of the general population.

many choices need to be made about how the surveys will be constructed. And in making these choices, the values of the scientists will typically be incorporated into the results.

First, there are a number of different question types that are used to value health states, including the standard gamble (SG), time trade-off (TTO), and person trade-off (PTO) methods.[16] Each of these asks a different question of survey-takers, and accordingly measures a different thing. The SG, for example, is affected by the subject's attitudes towards uncertainty and risk; the TTO is affected by the subject's time preferences and attitudes towards duration; and the PTO is affected by the subject's attitudes towards distribution (Brazier *et al.* 2007: ch. 5). In making an informed choice for one method over another, therefore, an economist will likely appeal to her beliefs about which of these factors is most important for healthcare decision-making, and/or which are mere "biases". Nord (1995), for example, argues that because QALYs are often used to guide resource allocation across groups of people, it is critical that they consider distributional factors. Since the PTO (unlike the other methods) does account for distribution, he argues it is superior on this score. Voigt (2012: 144) argues that the PTO was used in early versions of the GBD for precisely this reason, making its choice value-laden.

Second, a choice needs to be made about how health states are to be defined. Should paraplegia, for example, be evaluated as a single health state, assigned a uniform value everywhere in the world? Or should we consider paraplegia in a mountainous country with no wheelchair accommodations to have a different value than paraplegia in an urban area full of ramps, elevators, and wheelchair-accessible public transportation? (Reidpath *et al.*, 2003). Murray defends the former answer by appealing to the *ethical* proposition that "the non-health characteristics of the individual…that should be considered in calculating the associated burden of disease should be restricted to age and sex," which he notes gives DALYs an "egalitarian flavor" (Murray 1996: 6-7; but *cf.* Voigt 2012: 149-50).

Third, a decision needs to be made about whom to survey. This decision is sometimes made by an appeal to accuracy — e.g. that people who have experienced a given condition are likely to have more

---

[16] For a survey of these and other methods, see Brazier *et al.* (2007), especially chapters 5-7.

information about what living with the condition is like, or (alternately) that people without the condition are less likely to be biased. This does not obviously involve a value judgment. But the decision can also be made through an appeal to ethical factors. For example, if a study is to be used to allocate public resources, then perhaps the general public has a right that its values carry the day, and accordingly that it be surveyed. (Brazier *et al.* 2007: 114-117; *cf.* Gold *et al.* 1996: 99).

Fourth, and relatedly, a decision needs to be made about how to handle adaptation. We know that, over time, people with a given health condition typically adapt to it, finding it less bad than they did initially.[17] In valuing health, should we seek to measure pre-adapted values or post-adapted values? Sometimes, the former seems preferable. Suppose a talented and ambitious violinist, who dreams of becoming an international soloist, acquires a hearing impairment and accordingly changes her aspirations, seeking only to scrape by as a member of a regional orchestra. To many, it may seem that something bad has happened, and that the badness of her impairment is in some way reflected in her changed expectations. This speaks in favor of using pre-adapted values, since adaptation may disguise a legitimately important loss. On the other hand, suppose that a newly blinded person initially believes that her condition is horrible, since she can barely navigate her apartment without assistance. Over time, however, she learns to accomplish most tasks without help, and accordingly decides that being blind does not significantly impact her quality of life. In this case, the new skills she acquires seem wholly desirable, and accordingly it may seem appropriate to use her post-adapted values when deciding how good or bad blindness is. It is therefore not clear whether to use pre- or post-adapted values, or whether to combine them in some way when assigning values to health states. To resolve this issue, it seems that we must decide which forms of adaptation are wholly good, and which are, at least in some sense, regrettable. As Murray states, this seems to be a "vexing moral problem," (1996: 32) and so adequately addressing adaptation will likely require an appeal to values (Menzel *et al.* 2002).

---

[17] This is related to the question of whom to survey, because surveying people who have a given condition will typically deliver post-adapted values, whereas surveying the general public will typically elicit pre-adapted values.

To sum up, then, even if we acknowledge that survey results report the values of the survey-takers, rather than the values of the scientists, quality/disability weights will nevertheless typically be laden with the values of the researchers. In choosing what type of question to ask, how to define or differentiate health states, whom to survey, and how to deal with the discrepancy between pre- and post-adapted values, thoughtful researchers will probably make decisions based on their values.[18]

Co-morbidity

How should the DALY handle cases of co-morbidity? In some cases two health problems may combine in a greater-than-additive way, making things especially bad. (It seems likely that blindness-plus-deafness constitutes a far greater impairment than blindness or deafness alone.) In other cases, a co-morbidity may not noticeably make things worse. (If I am already confined to a wheelchair, a foot amputation may not constitute a significant additional burden.) In principle, then, it seems that co-morbidities should be assessed individually. That is, we should conduct a survey to determine the significance of blindness, a second to determine the significance of deafness, and then a third to determine the significance of blindness-plus-deafness. In practice, of course, this is not feasible, and so a different approach is needed. Many studies, including the 2010 GBD (Murray *et al.* 2012: 11) have assumed that a multiplicative model is the most accurate approximation: two co-morbid conditions, each with a disability weight of 0.2 result in a total disability weight of 0.36. But in the 1990 GBD, Murray rejects this approach:

> For co-disability due solely to chance, there is the suggestion that the disability weights for such co-disability should remain the sum of the individual disability weights… If we "correct" disability weights for co-disability—reduce them such that the combination of two conditions is less than the sum of the disability weights for each condition separately—becoming blind in a population with a high prevalence of disability would lead to a smaller incremental increase in

---

[18] There are a range of other value choices that are likely to be involved in conducting quality/disability weight surveys. For example, a health economist may constrain survey outcomes on grounds of "consistency". (See Hausman 2015: 51-57 and Murray 1996: 92 for different senses in which survey responses may be thought inconsistent.) Such judgments of consistency are likely based on values. (See Nord 1999: 121-3 for an illustration of this in the GBD.) For reasons of space, I will not discuss such additional value choices connected to quality/disability weight surveys, in large part because they will vary a great deal from study to study.

burden than becoming blind in a population with low disability prevalence. Correcting for independent co-disability would introduce a bias in favor of saying that the same event caused more burden in the better off than in the worse off. As a result, adjustments have not been made for [independent] co-disability. (Murray 1996: 42)

Murray's argument here seems to echo the earlier argument regarding life expectancy. Note that he does not argue that a simple additive model more accurately reflects the burden on people suffering from multiple disabilities. (In fact, he later acknowledges that an additive model could lead to "nonsensical" results, with individuals having a total disability weight greater than 1.) He instead argues that it is *unfair* to count a disability as less significant, just because it happens to affect someone already suffering from a disability. This is a value-based justification.

Incidence vs. prevalence approach

The examples above are the clearest cases of values being incorporated into DALYs, and they have each at least been mentioned in the literature. But a closer look reveals several additional value choices, which have not yet been discussed. When calculating the health lost to disability in a particular time period, two different calculation methods are sometimes used, each of which has intuitive support. The first option is to measure the amount of disability *experienced* in the time period in question. Since this amounts to measuring the prevalence of the condition, Murray (1996) calls this a *prevalence perspective*. The second option is to measure the amount of disability *caused* in the time period in question, regardless of when the disability will be experienced. Murray calls this an *incidence perspective*. As I have argued elsewhere (2012), these are simply measures of different quantities, and accordingly are useful for different purposes.

For the 1990 GBD, Murray chose to use an incidence perspective, citing its utility in highlighting "current epidemiological trends" (1996: 9). The 2010 GBD switched to a prevalence perspective. In justifying the switch, Murray and colleagues appealed to the utility of a prevalence perspective: "[I]ndividuals living with reductions in health functioning who need health and other social services may not be reflected in burden estimates if incidence has declined substantially… For most planning and

decision-making purposes assigning the burden to the age at which health loss is experienced is more useful" (Murray *et al.* 2012: 15). In choosing one calculation method over another based on its utility for policy choices, Murray was making a value judgment about the relative importance of those policy decisions. For the 1990 GBD, Murray apparently held that highlighting current epidemiological trends was more important than reflecting the functional abilities of those in need of health services. For the 2010 GBD, he made the opposite determination. In each case, therefore, the choice of one calculation method over the other was value-based.

Grouping of causes, risk factors, etc.

Some of the most visible aspects of the GBD are its "top ten" lists of the most significant causes of health loss. Producing such lists, though, requires grouping causes together, and there is no obvious way to decide how coarse or fine-grained the groupings should be. Suppose Jane has a malignancy, caused by smoking, that began in her left lung. For the purposes of GBD ranking, it could be grouped only with left-lung-cancers-caused-by-smoking, with all left-lung cancers (regardless of cause), with all lung cancers, with all respiratory cancers, with all cancers, and so forth. Some of these groupings may seem silly (why distinguish left- from right-lung cancers?), but it is easy to see that there are multiple groupings here that could be reasonable, and which may be useful for different purposes. Murray and colleagues recognize this, and note that the choice is important:

> For presentation of the leading causes of DALYs, we need to choose the level in the cause hierarchy at which we rank conditions. Because the leading causes of burden tend to have some influence on the perception of disease control priorities, the choice of aggregation is at once important and subject to debate. (Murray *et al.* 2012: 6)

The causal groupings that the GBD ultimately settled on don't seem to be entirely uniform. Their "top 10" list includes apparently broad groupings like lower respiratory infections, low back pain, and preterm birth complications, and also narrow-looking groupings like HIV/AIDS and malaria. (Cancers were divided by site, so that the highest-ranking cancer was lung cancer, at 22nd. Had cancers been

17

grouped together, they would have ranked 1st.) The only explanation given for this was that the causal

groupings were chosen "to distinguish and cluster conditions together that may have programmatic or

public health significance" (Murray *et al.* 2012: 6). Without more information, it is hard to say

definitively whether value choices were involved here, but it seems highly likely that they were. First, in

deciding to group causes together based on "programmatic or public health significance" — as opposed,

say, to grouping them based on similarity of cause, resultant functional limitation, etc. — the GBD team

apparently decided that this grouping provided more important information, or that grouping causes in

this way would have better consequences than alternative groupings. These judgments reflect ethical

values. Second, there is no single scale of "programmatic or public health significance." One person, for

example, might think that the primary aim of public health programs should be to improve health

outcomes, a second might think it should be to give people the chance to make informed decisions

concerning their health, while a third might think it should be to reduce health-related inequality. In

choosing one of these conceptions over the others, it seems likely that the GBD team implicitly appealed

to their vision of what public health programs should aim at. For all of these reasons, then, it seems

highly likely that the resultant ranking lists should be considered value-laden.[19]

Conclusion

        Above, I've outlined at least eight ways in which the DALYs used in different versions of the

GBD, as well as in many other burden of disease studies, are appropriately considered value-laden on

even an extremely conservative interpretation. I don't think those are the only eight,[20] but I think they are

the clearest examples, and they should be sufficient to show that values permeate the DALY in many

more ways than are typically recognized. This is, of course, of direct importance, since epidemiologists

---

[19] The problem of aggregation comes up in many other parts of the GBD. Geographical regions, age groups, and risk factors, for example, can each be grouped in different ways. I suspect that ethical values are involved there, as well.

[20] I think values are likely also implicated in the GBD's definition of health, its analysis of causation (in particular, how it handles jointly caused events and different levels of a causal hierarchy), and in the baseline it uses in defining counterfactual scenarios for risk factor analysis. For hints of some of these issues, see Salomon *et al.* (2003), Murray *et al.* (2002: 28-31), and Murray and Lopez (1999), respectively. Adopting a more liberal account of value-ladenness would, of course, expand things even farther.

and health economists conducting similar studies should understand the ways in which ethical values are a part of the (standard) DALY framework, modifying and criticizing them as appropriate. And decision-makers making use of DALYs need to be aware of the ways in which the measures they use already incorporate certain ethical assumptions, so that they are able to make more informed decisions on the basis of GBD results. (I haven't explicitly discussed QALYs, but most of the value choices described here apply directly to QALYs, and others have QALY-analogues.[21])

At a more general level, I think what I've shown here has an important role to play in debates about the extent to which QALYs, DALYs, and other related measures *should* incorporate values. Many have argued that QALYs and DALYs should be modified to reflect additional ethical factors. Jamison and colleagues (2006), for example, propose that DALYs account for fetal health, treating fetal deaths continuously with deaths at very young ages. Many have argued that we owe special concern to those with very serious afflictions, that it is morally more urgent to treat one person with a serious disability (weight=0.4) than four people with much less serious disabilities (weight=0.1). And there has been a steady push to adjust QALYs to account for inequalities in health.[22] It is tempting to respond to these calls by retreating to the idea that these measures should somehow be "pure" measures of health, or that we should keep science separate from ethics whenever possible. This position may be correct, but its advocates should be aware of the extent to which values already pervade QALYs and DALYs. If we're serious about keeping values out of the DALY, we need to do more than reject age-weighting and discounting. We would also need to rethink our approach to life expectancy, disability weights, co-morbidity, and a host of other factors. If we're not willing to do this, or if we think this can't be done, then we need to abandon the simplistic view that the DALY should be purged of values and instead ask

---

[21] The case of life expectancy may appear to be the main outlier. Since QALY calculations typically either add up health experienced, or calculate the health that would be gained through interventions, no life expectancy needs to be specified. That said, Murray's argument regarding life expectancy does point out a potential ethical concern with QALYs: under standard QALY methodology, a lifesaving intervention may have a better cost-per-QALY ratio is a wealthier community than in a poorer one, if those in the wealthier community have a longer life expectancy. For the reasons Murray identifies, we might regard this as unfair, and so stipulate that the QALY gain from a treatment will be considered the same in richer and poorer communities. This choice would make the resulting QALY calculations value-laden, in a similar way to the DALY's handling of life expectancy.

[22] For severity, see Nord (1999: 32-8). On inequality, see Eyal *et al.* (2013).

the more complicated question: given that values should sometimes be built into measures of health,

when should they be built in, and whose values should they incorporate? These, I think, are very difficult

questions that have not yet received anything near an adequate treatment in the literature. But they will

have to wait for another occasion.[23]

Works Cited

Anand, S. and Hanson, K. 1997. "Disability-adjusted life years: a critical review," *Journal of Health Economics* 16: 685-702.

Anand, S., and Hanson, K. 1998. "DALYs: Efficiency Versus Equity," *World Development* 26: 307-10.

Anderson, E. 2004. "Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce," *Hypatia* 19: 1-24.

Arnesen, T. and Kapiriri, L. 2004. "Can the value choices in DALYs influence global priority-setting?" *Health Policy* 70: 137-49.

Barker, C. and Green, A. 1996. "Opening the debate on DALYS," *Health Policy and Planning* 11: 179-83.

Bergman, A. and 40 colleagues. 2013. "Science and policy on endocrine disrupters must not be mixed: a reply to a 'common sense' intervention by toxicology journal editors," *Environmental Health* 12:69.

Betz, G. 2013. "In defense of the value free ideal," *European Journal for the Philosophy of Science* 3: 207-20.

Bognar, G. 2008. "Age-Weighting," *Economics and Philosophy* 24: 167-89.

Bray, D. 2010. "The Scientific Consensus of Climate Change Revisited," *Environmental Science and Policy* 13: 340-50.

Broome, J. 2008. "Why Economics Needs Ethical Theory," in *Arguments for a Better World: Essays in Honor of Amartya Sen: Volume 1: Ethics, Welfare, and Measurement* (eds. Basu and Kanbur), Oxford University Press: 7-14.

Brazier, J., Ratcliffe, J., Salomon, J.A., Tsuchiya, A. 2007. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press.

Brock, D. 2006. "Ethical Issues in the Use of Cost Effectiveness Analysis for the Prioritization of Health Care Resources," in *Public Health, Ethics, and Equity* (eds. Anand, Peter, and Sen), Oxford University Press.

Chapman, G., Hansen, K.S., Jelsma, J., Ndhlovu, C., Piotti, B., Byskov, J., and Vos, T. 2006. "The burden of disease in Zimbabwe in 1997 as measured by disability-adjusted life years lost," *Tropical Medicine & International Health* 11: 660-671.

Eyal, N., Hurst, S.A., Norheim, O.F., and Wikler, D. (eds). 2013. *Inequalities in Health: Concepts, Measures, and Ethics*. Oxford University Press.

Gold, M., Siegel, J., Russell, L., and Weinstein, M. 1996. *Cost-Effectiveness in Health and Medicine*. Oxford University Press.

Haack, S. 1993. "Epistemological Reflections of an Old Feminist," *Reason Papers* 18: 31-43.

Harris, J. 1987. "QALYfying the value of life," *Journal of Medical Ethics* 13: 117-23.

Hausman, D. 2015. *Valuing Health: Well-Being, Freedom, and Suffering*. Oxford University Press.

Kincaid, H., Dupré, J., and Wylie, A. (eds). 2007. *Value-Free Science? Ideals and Illusions*. Oxford University Press.

Jamison, D., Shahid-Salles, S.A., Jamison, J., Lawn, J.E., and Zupan, J. 2006. "Incorporating Deaths Near the Time of Birth Into Estimates of the Global Burden of Disease," in *Global Burden of Disease and Risk Factors* (eds. Lopez, Mathers, Ezzati, Jamison, and Murray). The World Bank: 427-63.

Lackey, R.T. 2007. "Science, scientists, and policy advocacy," *Conservation Biology* 21: 12-17.

Lyttkens, C. 2003. "Time to disable DALYs?" *The European Journal of Health Economics* 4: 195-202.

Menzel, P. 2011. "Should the Value of Future Health Benefits Be Time-Discounted?" in *Prevention vs. Treatment: What's the Right Balance?* (ed. Faust and Menzel), Oxford University Press.

Menzel, P., Dolan, P., Richardson, J. and Olsen, J.A. 2002. "The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis," *Social Science & Medicine* 55: 2149-2158.

Menzel, P., Gold, M.R., Nord, E., Pinto-Prades, J., Richardson, J., and Ubel, P. 1999. "Toward a Broader View of Values in Cost-Effectiveness Analyses of Health," *The Hastings Center Report* 29: 7-15.

Michaud, C.M. and 11 colleagues. 2006. "The burden of disease and injury in the United States 1996," *Population Health Metrics* 4:11.

Murray, C.J.L. 1996. "Rethinking DALYs," in *The Global Burden of Disease* (ed. Murray and Lopez). World Health Organization: 1-89.

Murray, C. and Acharya, A. 1997. "Understanding DALYs," *Journal of Health Economics* 16: 703-730.

Murray, C.J.L, Salomon, J.A., and Mathers, C.D. 2002. "A Critical Examination of Summary Measures of Population Health," in *Summary measures of population health: concepts, ethics, measurement and applications* (eds. Murray, Salomon, Mathers, and Lopez)*.* World Health Organization: 13-40.

Murray, C.J.L. and 11 colleagues. 2012. "Comprehensive Systematic Analysis of Global Epidemiology: Definitions, Methods, Simplification of DALYs, and Comparative Results from the Global Burden of Disease Study 2010," *Lancet*, online supplement to vol. 380.

Murray, C.J.L. and Lopez, A. 1999. "On the Comparable Quantification of Health Risks: Lessons from the Global Burden of Disease Study," *Epidemiology* 10: 594-605.

Nord, E. 2014. "Cost-Value Analysis of Health Interventions: Introduction and Update on Methods and Preference Data," *PharmacoEconomics* 33:89-95.

Nord, E. 2011. "Discounting future health benefits: the poverty of consistency arguments," *Health Economics* 20: 16-26.

Nord, E. 1999. *Cost-Value Analysis in Health Care*. Cambridge University Press.

Nord, E. 1995. "The Person-trade-off Approach to Valuing Health Care Programs," *Medical Decision Making* 15: 201-8.

Nygaard, E. 2000. "Is It Feasible or Desirable to Measure Burdens of Disease as a Single Number?" *Reproductive Health Matters* 8: 117-25.

Pielke, R.A. 2007. *The Honest Broker*. Cambridge University Press.

Plass, D., Chau, P.Y.K., Thach, T.Q.T., Jahn, H.J., Lai, P.C., Wong, C.M. and Kraemer, A. 2013. "Quantifying the burden of disease due to premature mortality in Hong Kong using standard expected years of life lost," *BMC Public Health* 13:863.

Polinder, S, Haagsma, J.A., Stein, C., and Havelaar, A.H. 2012. "Systematic review of general burden of disease studies using disability-adjusted life years," *Population Health Metrics* 10:21.

Reidpath, D.D., Allotey, P.A., Kouame, A., and Cummins, R.A. 2003. "Measuring health in a vacuum: examining the disability weight of the DALY," *Health Policy and Planning* 18: 351-6..

Reiss, J. 2008. *Error in Economics: Towards a More Evidenced-Based Methodology.* Routledge.

Robberstad, B., Hermed, Y., and Norheim, O.F. 2007. "Cost-effectiveness of medical interventions to prevent cardiovascular disease in a sub-Saharan African country – the case of Tanzania," *Cost Effectiveness and Resource Allocation* 5:3.

Rykiel, E.J. 2001. "Scientific Objectivity, Value Systems, and Policymaking," *BioScience* 51: 433-6.

Salomon, J.A., Mathers, C.D., Chaterji, S., Sadana, R., Üstün, T.B., and Murray, C.J.L. "Quantifying Individual Levels of Health: Definitions, Concepts, and Measurement Issues," in *Health Systems Performance Assessment: Debates, Methods, and Empiricism* (eds. Murray and Evans), World Health Organization: 301-318.

Salomon, J.A. and 127 co-authors. 2012. "Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010," *Lancet* 380: 2129-43.

Schokkaert, E. 2015. "How to introduce more (or better) ethical arguments into HTA?" *International Journal of Technology Assessment in Health Care* 31: 1-2.

Schroeder, S.A. 2012. "Incidence, prevalence, and hybrid approaches to calculating disability-adjusted life years," *Population Health Metrics* 10.

Sunstein, C.R. and Rowell, A. 2007. "On Discounting Regulatory Benefits: Risk, Money, and Intergenerational Equity," *The University of Chicago Law Review* 74: 171-208.

Tsuchiya, A. 2000. "QALYs and Ageism: Philosophical Theories and Age Weighting," *Health Economics* 9: 57-68.

Voigt, K. 2012. "Measuring global health," in *Global Justice and Health Inequalities* (eds. Lenard and Straehle), Edinburgh University Press.

Weinstein, M.C. and Stason, W.B. 1977. "Foundations of Cost-Effectiveness Analysis for Health and Medical Practices," *New England Journal of Medicine* 296: 716-21.

Williams, A. 1997. "Intergenerational equity: An exploration of the 'fair innings' argument," *Health Economics* 6: 117-32.