






The Devils in the DALY: Prevailing Evaluative Assumptions


Carl Tollef Solberg , * Bergen Centre for Ethics and Priority Setting—BCEPS, Department of Global Public Health and Primary Care, Faculty of Medicine, University of Bergen

Preben Sørheim , Department of Philosophy, Faculty of Humanities, University of Bergen

Karl Erik Müller , The Gade Research Group for Infection and Immunity, Department of Clinical Science, Faculty of Medicine, University of Bergen; Institute of Biosciences, São Paulo State University; Department of Internal Medicine, Drammen Hospital, Vestre Viken Hospital Trust

Espen Gamlund , Department of Philosophy, Faculty of Humanities, University of Bergen

Ole Frithjof Norheim , Bergen Centre for Ethics and Priority Setting—BCEPS, Department of Global Public Health and Primary Care, Faculty of Medicine, University of Bergen; Harvard TH Chan School of Public Health, Harvard University

Mathias Barra , The Health Services Research Unit—HØKH, Akershus University Hospital HF

*Corresponding author: Carl Tollef Solberg, Bloksbergstien 54, 1394 Nesbru, Norway. Tel.: +47 93833641; Email: carl.solberg@uib.no and catoffel@gmail.com

In recent years, it has become commonplace among the Global Burden of Disease (GBD) study authors to regard the *disability-adjusted life year* (DALY) primarily as a *descriptive* health metric. During the first phase of the GBD (1990–1996), it was widely acknowledged that the DALY had built-in *evaluative* assumptions. However, from the publication of the 2010 GBD and onwards, two central evaluative practices—*time discounting* and *age-weighting*—have been omitted from the DALY model. After this substantial revision, the emerging view now appears to be that the DALY is primarily a descriptive measure. Our aim in this article is to argue that the DALY, despite changes, remains largely evaluative. Our analysis focuses on the understanding of the DALY by comparing the DALY as a measure of disease burden in the two most significant phases of GBD publications, from their beginning (1990–1996) to the most recent releases (2010–2017). We identify numerous assumptions underlying the DALY and group them as descriptive or evaluative. We conclude that while the DALY model arguably has become more descriptive, it remains, by necessity, largely evaluative.

Introduction

The Global Burden of Disease (GBD) study is ‘a systematic, scientific effort to quantify the comparative magnitude of health loss due to diseases, injuries, and risk

factors by age, sex and geographies for specific points in time’ (Murray *et al.* 2012a: 1). The GBD study quantifies disease burden via a measure called *disability-adjusted life years* (DALYs).

doi:10.1093/phe/phaa030

Advance Access publication on 29 October 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

During the first phase of the GBD study (1990–1996), it was widely acknowledged that the DALY model included *evaluative* assumptions (Murray, 1994, 1996; Murray and Lopez, 1996; Murray and Acharya, 1997).¹ However, with the publication of the 2010 GBD study, two central evaluative practices—time discounting and age-weighting—were discontinued. After a substantial revision, the emerging view now appears to be that the DALY is primarily a *descriptive* measure of overall disease burden (Murray *et al.*, 2012a,b; Salomon *et al.*, 2012; Murray and Lopez, 2013; Knudsen *et al.*, 2016).²

In brief, the DALY has seemingly undergone a transition from being a measure of burden ‘based on explicit and transparent *value* choices’ (Murray and Lopez, 1996: 8) to ‘a major step toward a replicable scientific approach to global *descriptive* epidemiology’ (Murray *et al.*, 2012b: 2065—our italics).³ This emerging terminology obscures the fact that several critical evaluative assumptions remain embedded within the DALY model. As we will see, both the magnitude and distribution of disease burden rely heavily on these evaluative assumptions.

The GBD study was conceived to inform policy-makers (Murray, 1994; Murray and Lopez, 1996). Since then, the DALY model has become increasingly popular in the global health community. Major organizations and institutions, as well as national health authorities, use this measure,⁴ and DALY-publications regularly appear in high-ranking academic journals. By using DALY data from the GBD study, epidemiological trends are monitored, disability groups are ranked according to their disease burden, and health inequity is quantified.

Our aim in this article is to argue that the DALY model, despite the recent modifications, remains largely *evaluative*. We will restrict our analysis to the assumptions underlying the DALY model rather than the assumptions underlying the GBD study as such. Moreover, we will focus on the practices and corresponding rationales that have been suggested by the DALY architects themselves (e.g. Murray, 1994, 1996; Murray and Lopez, 2013).

This article proceeds as follows: first, we motivate and review the basics of the DALY model. Next, we examine and contrast the core DALY assumptions during the two most significant phases of the GBD: 1990–1996 and 2010–2017. For our purposes, we classify these assumptions into two categories: *descriptive* and *evaluative*.⁵ We conclude that evaluative assumptions are ubiquitous and that while the DALY model has arguably become somewhat more descriptive, it necessarily remains largely evaluative.

The Disability-Adjusted Life Year

The DALY model can be viewed as a natural extension of earlier efforts (1940–1950) to evaluate instead of simply counting deaths. The initial ambition was to analyze disease burden beyond the descriptive *crude death rates* (CDRs) by instead evaluating deaths according to the prematurity of their occurrence (see e.g. Dempsey, 1947; Robinson, 1948; Haenszel, 1950). Two decades later, *health-* and *quality-adjusted life years* (HALYs and QALYs, respectively), and similar models⁶—undertook to include the burden morbidity confer on people while they still are alive, in addition to the disease burden resulting directly from death. The DALY model was launched in 1993 with the first GBD study, and its *raison d’être* was to obtain a universally applicable measure integrating morbidity *and* mortality.

According to the DALY model, morbidity is measured by assigning *disability weights* (DWs) to health conditions, where 0 represents the absence of disability, $0 \leq DW \leq 1$ quantifies the burden that a particular health condition incurs and 1 is the highest possible DW, defined as a loss ‘equivalent to death’ (Salomon *et al.*, 2015: 712). After a condition has been assigned its DW, the *years lived with disability* (YLDs) is calculated as the product of the condition’s duration and its DW, which account for morbidity. *Years of life lost* (YLLs), relative to a reference life expectancy, account for mortality. Finally, $YLDs + YLLs = DALYs =$ disease burden.

A simplified example of the calculation of lifetime DALYs for an individual may be helpful. Imagine a person who suffers from $A =$ severe anxiety disorder ($DW_A = 0.5$) during 8 years of her young adulthood (see Figure 1).⁷ This suffering generates duration \times DW = 8 years \times 0.5 YLDs/year = 4.0 YLDs. Later in life, the person endures $B =$ chronic neck pain ($DW_B = 0.3$) which lasts for 15 years, generating 15 years \times 0.3 YLDs/year = 4.5 YLDs, until she dies at the age of 70. For 70-year-olds, the DALY reference life expectancy is 88.9 years, meaning that the person lost 88.9 years $-$ 70 years = 18.9 years, which are the YLLs generated by her (premature) death (Murray *et al.*, 2012a). Accordingly, the aggregate lifetime DALY amount is 4.0 YLDs + 4.5 YLDs + 18.9 YLLs = 27.4 DALYs. See Figure 1 for an illustration.

The GBD employs this type of calculation to arrive at estimates of the overall disease burden in a population. However, in the second phase of the GBD (2010–2017), the DALY estimates are (usually) given as annual figures and quantify the total amount of disease burden generated during that year.

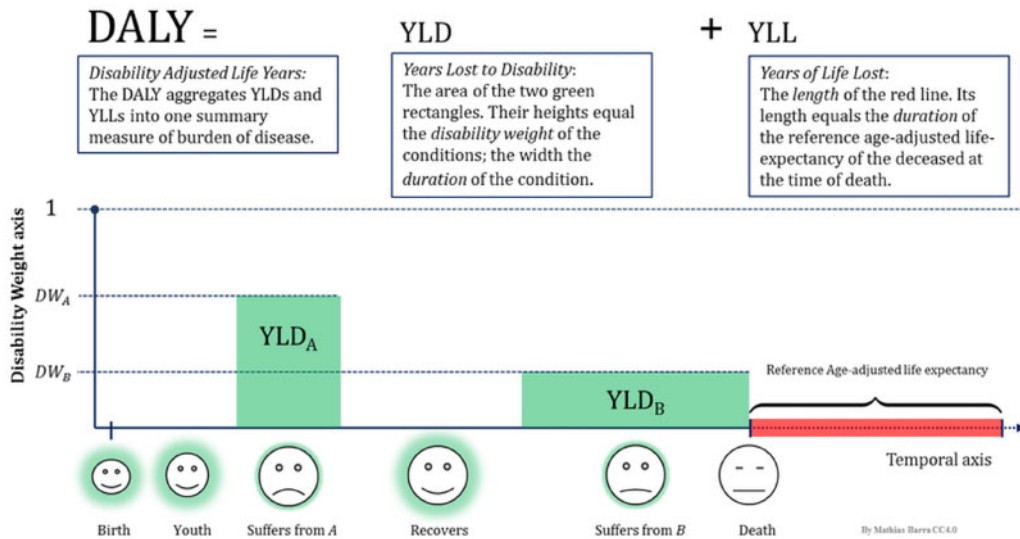


Figure 1. An illustration of the DALY.

Descriptive versus Evaluative Epidemiology

Historically, epidemiologists were chiefly concerned with the mortality in a population, and mortality rates were the standard measures of disease burden (in addition to incidences and prevalences). This initial approach included *the CDR* and *age-specific death rates*. The CDR is the number of fatalities per year per 1000 people. Age-specific death rates are CDRs that are restricted to a predefined age bracket. The child mortality rate, for example, reports the CDR for those between 1 and 5 years of age (Porta, 2014). Such descriptive mortality measures have several virtues. In general, they are simple, transparent, and inherently universal (Haenszel, 1950). Evaluative measures, in contrast, will, by necessity, rely on contested value assumptions.⁸

No measure of mortality, health or well-being serves equally well for all purposes. For instance, descriptive mortality measures do not highlight the fact that individual deaths are postponed rather than prevented. Moreover, they say nothing about whether some deaths are worse than others. For some purposes, we need evaluative measures instead (Gamlund and Solberg, 2019).

In the philosophy of science, several frameworks have been suggested for how values influence the scientific process (Longino, 1990; Lacey 1999; Douglas, 2009; 2016). Moreover, there is a growing literature on the philosophy of measurement (see e.g. Cartwright and Bradburn, 2011; Cartwright and Runhardt, 2014; Tal,

2017). Furthermore, for philosophers of science, it is common to regard most scientific disciplines as more value-laden than what is acknowledged by the practitioners within those disciplines. Social and ethical values permeate scientific endeavors at many different levels (Chang and Cartwright, 2014).

With regard to the DALY, one framework suggested by Hausman and McPherson (2006) could serve as an illustration of the intuitions we draw on. Admittedly, Hausman and McPherson (2006: 9, Figure 1.2.1) describe their framework as displaying ‘exaggerated contrasts between facts and values’. According to their caricatured framework, factual claims are characterized by disagreement that can easily be resolved, hypotheses that can be determined as true or false, and hypotheses that are independent of evaluative claims. In contrast, evaluative claims are characterized by little agreement and are not easily resolved, hypotheses that cannot easily be determined as true or false, and hypotheses that are dependent on factual claims (Hausman and McPherson, 2006: 9).

However, it is difficult, or even impossible, to tease apart every evaluative aspect from every descriptive aspect of the assumptions within the DALY. While this might sometimes make it difficult to classify an assumption as ‘clearly evaluative’ or ‘predominantly descriptive’ in a manner that would be universally acceptable, we believe one can often rely on intuitions to discern that some assumptions contain largely evaluative, as opposed to descriptive, aspects.

A rigorous account of descriptive assumptions, as opposed to evaluative ones, remains elusive, and we offer

no formal criteria to distinguish descriptive from evaluative assumptions. Instead, we rely on what we regard as shared intuitions concerning the classification of the various aspects of the DALY discussed below. Even if some readers disagree with some of our intuitions, we believe that the overall case we present will be compelling.

The Devils in the DALY

Certain assumptions in the DALY model have been discussed frequently, such as *discounting*, *age-weighting*, *DWs* and *choice of reference life expectancy* (Anand and Hanson, 1997; Murray and Acharya, 1997; Fox-Rushby and Hanson, 2001; Lyttkens, 2003; Arnesen and Nord, 1999; Arnesen and Kapiriri, 2004; Voigt, 2012). Aside from the assumptions above, Murray (1994, 1996) also addressed the *incidence versus prevalence* approach and *comorbidity* at an early stage. That the DALY is assumed to be a measure of *health* rather than *well-being* has also been subject to debate (Broome, 2002; Hausman, 2015). The perhaps most systematic and rigorous work on the assumptions behind the DALY to date has been conducted by the DALY-architect himself Christopher J. L. Murray and by the philosopher S. Andrew Schroeder (Schroeder, 2012, 2017, 2019). In addition to the assumptions mentioned above, Schroeder adds *grouping of disabilities* to the list above (2016).⁹

In the present exposition, we expand on this list with a discussion of four further assumptions: *intrapersonal and interpersonal aggregation*, *individual versus societal burden*, *gradualism versus non-gradualism* and *commensurability*.¹⁰

Health versus Well-Being

The *health versus well-being* debate lies at the core of the descriptive versus evaluative question. One of the strengths under which the current DALY is marketed is that it measures *health simpliciter* (Voigt and King, 2014). The discussion of this controversial issue began when the first GBD was launched (Anand and Hanson, 1997; Murray and Acharya, 1997); to date, it has not been satisfactorily resolved (Schroeder, 2016, 2019). Still, the DALY authors continue to describe the DALY as an objective and descriptive measure of health. Indeed, without health simpliciter, much of the mission behind the DALY may be undermined: one of the great advantages of the DALY is that it can be *universally applied* (conditional on measuring *health simpliciter*). In contrast, most health jurisdictions mandate that *national tariffs* for QALYs are used for health technology assessment

analyses since it is empirically recognized that preferences do vary between cultures and within populations.

Since the architects of the DALY strive to measure health descriptively, their methodology has (as previously mentioned) been revised to remove evaluative components (Salomon *et al.*, 2012). Whether the DALY should measure health or well-being is already thoroughly discussed in the literature (e.g. Broome, 2002; Hausman, 2015). It is unclear whether it is desirable to ignore the impact of health, let alone if *ill health* is a robust and meaningful construct when considered in isolation of the individuals who are burdened by it (Arnesen and Nord, 1999; Broome, 2002; Voigt and King, 2014). Lastly, the *health versus well-being* debate also illustrates the need for a careful analysis of that which we tentatively call the *descriptive versus evaluative* distinction in health measurement and epidemiology.

Disability Weights

The quantification of the burden through the so-called *DWs* is central to the DALY model. The DW is the device for modeling the assumption that some conditions are *worse* than others. The measuring of the burden of morbidity is carried out by assigning DWs to different conditions: 0 represents the absence of disability, 1 represents the maximal possible disability and intermediate values represent degrees of disability. For example, for the DWs set in 2017, *severe multiple sclerosis* was assigned a DW of 0.72, while a *symptomatic tension-type headache* was given a DW of 0.04 (Global Burden of Disease Collaborative Network, 2018). These numbers mean that the burden associated with severe multiple sclerosis is considered to be 18 times greater than the burden associated with a tension-type headache, per unit time.

All work on the DWs in the DALY has relied on one or more of the standard preference-based evaluation methods: *the person trade-off (PTO)*, *the standard gamble* and *the time trade-off*.¹¹ In the early GBD studies, the PTO method was central (Arnesen and Nord, 1999). The current practice (as of 2020) is based on utility theory and draws on discrete choice methodology combined with the PTO method. Nevertheless, the DALY is currently referred to as a non-preference-based measure: the explanation provided is that respondents are asked to set DWs *not* based on their own preferences, but instead ‘to state which of the two individuals they would deem as being healthier than the other’ (Salomon *et al.*, 2015: 713). However, there are reasons to believe that a change of wording alone cannot transform the inherent

evaluative nature of such choice tasks into an objective measurement of health *qua* health (Voigt and King, 2014).

A preference for something over something else implies the ranking of one alternative above another. Some preferences are simple and deeply rooted in our biology, such as preferring pleasurable sensations over pain or the taste of healthy nutrients over poisons. Such preferences tend to appear instinctive and require no conscious choice of our own. They may reasonably be referred to as descriptive.¹² However, we should not confuse such physiological preferences with deliberative preferences. The rank order ‘healthier than’ encodes a high degree of complexity, since this ranking is (supposedly) not an automated input-output procedure. Rather, it relies on deliberative thinking informed by both intuitions and arguments, usually a long-term perspective, and the processing of information, experience, and context. Thus, it is hard to escape the view that the DWs are evaluative by their very construction.¹³

Second, expert panels set the DWs initially, but, since the 2010 GBD, the public’s involvement in establishing DWs has been imperative. The choice of involving the public can be based on at least two rationales: *accuracy* and *legitimacy*. The DALY architects were arguably mostly concerned with legitimacy (Salomon *et al.*, 2012). Hence, the move from the expert toward the public view implied the prioritization of public over technocratic legitimacy.¹⁴ To add to this, we can easily imagine that if ordinary people (instead of technocrats) are asked about their views regarding different conditions, this may be informative as to what is important for them. That being said, the YLL construct—which generates the largest share of the total disease burden worldwide—remains completely uninformed by the public’s views.

Third, the question remains of whether some conditions are more burdensome than death. Recall that a DW of 1 has been interpreted as the maximum possible burden. No ‘worse than death’ values—that is, DWs above 1—have ever been used in the DALY model.¹⁵ Comorbidity is also accounted for in such a way that no compound health condition can ever be worse than death (Burstein *et al.*, 2015). This characteristic suggests the evaluative assumption that life is better than death—no matter what.

In summary, the DWs remain evaluative because people are asked how good or bad different conditions seem to them. We doubt that any system for setting DWs can avoid being evaluative (at least, such a system would no longer measure something that matters). Consequently, the YLD-component of the DALY is inherently evaluative as well.

Discounting

During the first phase of the GBD (1990–1996), future DALYs were discounted by a fixed rate of 3% per year. This practice reduced the YLLs attributed to premature deaths. Because future years were discounted (combined with age-weighting), death at birth was not counted as 86 YLLs but rather as approximately 32 YLLs. This discounting practice also implied that if two individuals each lost 20 YLLs and one person lost 40 YLLs, the latter loss would carry less weight since these YLLs were more distant. Three reasons were offered for this discounting practice: (i) the future is shrouded in uncertainty; (ii) health interventions are likely to improve in the future; and (iii) people tend to prefer goods in the near rather than the far future (Murray, 1996; Murray and Acharya, 1997).

Reason (i) can be seen as descriptive as it is based on an epistemic concern about how to forecast the future. In isolation, this issue does not concern any value-theoretical questions. That is to say, it is simply true and indisputable that the future is shrouded by epistemic uncertainty. Reason (ii) can also be seen as descriptive. On the one hand, one may argue that the very concept of *improvement* itself relates to something evaluative as the improvement implies that something is *better than* it was before. Moreover, that *something* resembling health interventions is likely to improve may involve value concerns. Such an improvement may include ‘better health’, ‘better quality of life’, or ‘better lives’—all evaluative concepts.

On the other hand, there is a correct answer as to whether health interventions will improve in the future, and reasonable agreement on the assumption that health interventions will improve. Thus, given a narrow definition of evaluative, we may determine reason (ii) to be descriptive. Reason (iii), however, seems to be evaluative, even in a narrow sense. We can indeed describe people’s preferences—doing so is a descriptive endeavor. Moreover, we can provide true empirical answers to questions about what preferences people have. However, recall that deliberative preferences have an evaluative rather than a descriptive nature. More specifically, reason (iii) implies the judgment that goods in the near future are evaluated and ranked as *better than* goods in the far future, which is evaluative.

Scholars have heavily debated such discounting (Anand and Hanson, 1997), and from 2010 GBD and onwards, it was omitted to make the DALY more descriptive. Nevertheless, this omission was grounded in yet another evaluative rationale, namely that every life year should count equally, independently of when in life it occurs (Murray *et al.*, 2012a).

Age-Weighting

Another characteristic of the first phase of the GBD studies (GBD 1990–1996) was age-weighting, which meant that less weight was attributed to years lived at very young and very old ages. In this marginal age-weighting, life years between the ages of 15 and 40 were given the highest relative value, while the years in the very first phase of life, as well as in the final period, were given the least relative value. Three main rationales were offered for this age-weighting: (i) *Well-being*; people themselves may value life years differently at different life-stages, (ii) *Productivity*; one may attribute a higher value to the most productive life years of an individual's life, (iii) *Well-being interdependence*; the belief that some people play a unique role in providing well-being for others, such as children and elderly parents (Murray and Acharya, 1997). We see that (i–iii) are arguably evaluative: (i) concerns preferences for prudential value (see the DW discussion below), while (ii–iii) highlight instrumental value assumptions relating to particular people. From the 2010 GBD and onwards, this age-weighting was omitted based on the same evaluative premise that omitted the discounting practice: that every life year for every person around the world should count equally (Murray et al., 2012a).¹⁶

Choice of Reference Life Expectancy

The largest part of the total disease burden is generated by deaths rather than morbidity. Hence, it is crucial to know how these deaths are evaluated. Several of the assumptions in the YLL component of the DALY will have high *elasticity*: specifically, small changes in the assumptions may incur large changes in the overall disease burden. The YLL component of the DALY is based on the concept of *potential years of life lost* (PYLL), a mortality measure that originated in the late 1940s (Dempsey, 1947; Robinson, 1948; Haenszel, 1950).

In the second phase of the GBD, YLLs are calculated in the following way. If an individual dies at age 20, her YLLs are calculated as the temporal distance between her age of death and a reference life expectancy for her age group. According to the life table used in the 2010 GBD, this death would generate $86.4 - 20 = 66.4$ YLLs. Similarly, a 10-year-old would lose $86.3 - 10 = 76.3$ YLLs, an 80-year-old $91.0 - 80.0 = 11.0$ YLLs, and a still-born child $0 - 0 = 0$ YLLs (Murray et al., 2012a). This practice implies that an individual generates more YLLs the younger she is at the time of death, reaching a maximum immediately after birth. The YLLs have been counted from the time of birth throughout the history of the GBD, but the choice of reference life expectancy has

varied (see e.g. Murray, 1994 versus Murray et al., 2012a).

Several questions need to be answered in order to create a reference life expectancy. To begin with, when do individuals begin to accrue YLLs? This issue was debated when the PYLL—the precursor of the YLL—was developed in the 1940s. Most authors suggested counting from birth (Dempsey, 1947; Haenszel, 1950), but opinions ranged from including stillbirths (Robinson, 1948) to counting from age one (Romed and McWhinnie, 1977). The DALY model starts counting YLLs at birth, thus excluding all stillbirths. Discussion of this issue, however, is largely absent in the GBD literature.

From the 1990 GBD until the 2017 GBD, stillbirths were excluded (i.e. generating 0 DALYs).¹⁷ Until the 2010 GBD, the death of 10-year-olds was measured as the greatest possible amount of DALYs lost due to the combined effect of age-weighting and discounting. Since the 2010 GBD, however, neonatal deaths have been attributed the maximal possible burden (approximately 86 DALYs). In the 2017 GBD, stillbirths were counted in the mortality statistics but not in the DALY count that generates disease burden (Wang et al., 2016). As of 2020, stillbirths do not generate any disease burden in the GBD study. The question of the age at which we should begin to count DALYs is an important aspect of the GBD, and merely small changes in the lower age limit would result in a large difference to the total disease burden.

Since a discussion of reasons for setting a lower age limit is lacking in the DALY literature, we are led to imagine candidate reasons for this practice. This issue ultimately relates to the following question: what is disease burden? Providing an answer to this question is not trivial. One approach is to argue that disease burden is usually something we experience, and since embryos and fetuses (usually) cannot have experiences, they cannot be subjected to any disease burden. However, such an approach seems unreasonable given that the majority of the total disease burden is a direct result of YLLs, which occur when individuals have died and therefore cannot experience at all. Another strategy could be to argue that individuals can be harmed by their own death only after they are born. Such an approach would involve value-theoretical considerations regarding the harm of death. A third approach is to argue that the lower age limit is set at birth because this is in line with ordinary norms and sensible in a practical sense. However, even this third approach involves choices that are value-laden and subject to reasonable disagreement. In relation to this, Murray's claim that every life year should count equally for everyone disregards the discontinuous jump in disease burden between fetuses and neonates.

Furthermore, the question remains as to what the upper limit for YLLs should be. In the first GBD, the reference life expectancy used for measuring YLLs was 80.0 years at birth for men and 82.5 years at birth for women (Murray, 1994). This female reference life expectancy was based on Japan's, which had the highest life expectancy at the time (Murray and Acharya, 1997). The decrement of 2.5 years for males was arbitrarily estimated. This sex difference was later omitted. From the 2010 GBD, age-specific reference life expectancies are synthesized by choosing the lowest national age-specific death rate recorded (Murray *et al.*, 2012a). The use of these so-called *synthetic reference life expectancies* means that the longevity used in YLL-computations is estimated so that individuals always enjoy the lowest recorded national age-specific death rate.

Another important issue concerns the question of whether the reference life expectancy should be local or universal. The initial approach suggested in the PYLL used national life expectancies (e.g. Dempsey, 1947). Throughout all of the GBD studies, however, the reference life expectancy has been universal. This means that death at, say, age 60 is attributed the same number of DALYs regardless of the nationality of that individual. Murray explicitly mentions that the universal application of synthetic life expectancy is grounded in an 'egalitarian nature'. That is to say, if one were to count YLLs from local life expectancy, then preventing the death of a 40-year old woman in a high-income country (with higher life expectancy) would lead to a larger reduction in the global burden of disease than preventing the death of 40-year-old in a developing country (with low life expectancy; Murray, 1996).

Finally, should the reference life expectancy be fixed or progressive? Some of the very first PYLL measures were fixed, which meant that one calculated the PYLL for all individuals based on life expectancy at birth (Dempsey, 1947). In all of the GBD studies, however, the reference life expectancy has been progressive. This means that one uses statistical tables, which show life expectancy for each age and make age-adjustments so that the older a person becomes, the higher her life expectancy will be. This progressive approach is reflected in the YLL so that, while life expectancy falls for each year a person ages, it does not fall with a full year. The use of such progressive instead of fixed life-expectancy models implies that more YLLs are attributed to the elderly, more YLLs are generated in total, and the DALY acquires a slightly less egalitarian flavor.

There are a few things to note about the four assumptions above: they have competing alternatives, and there is reasonable disagreement on the alternatives. Taken

together, we see that even though life expectancy is descriptive in an empirical sense, the choice of one system of reference life expectancy over another as a way of calculating YLLs is evaluative. There are reasons to believe that we cannot choose systems of reference life expectancies in a value-neutral way (see Anand and Reddy, 2019).

The Incidence versus Prevalence Approach

Theoretically, the two components of the DALY—the YLD and the YLL—can be measured both as *incidence* and as *prevalence* parameters. Because death rates are incidence rates, the YLL has been accounted for by incidence rate by default (Murray, 1994, 1996), but this default is not self-evident.¹⁸ In contrast, both incidence and prevalence rates make immediate sense for the YLD. The GBD in 1990 and onwards used an incidence perspective for both the YLD and the YLL. Recall that DALYs are calculated for 1 year at a time. The issue of the incidence versus prevalence approach has to do with the year in which morbidity is assigned. Under an incidence approach, all DALYs associated with a diagnostic incident in a given year, including expected future DALYs, are assigned to that year. This incidence practice means that if a person is diagnosed with a chronic obstructive pulmonary disease in 2005 (moderate DW 0.23), and she is expected to live for a further 10 years, then $10 \times 0.23 = 2.30$ DALYs are attributed to the disease burden for the year 2005. Three reasons were initially given for the practice of having a pure incidence perspective. First, quantifying incidence YLDs is more consistent with incidence YLLs. Second, an incidence perspective for YLDs identifies the impact of health interventions more rapidly. Third, with the prevalence YLD alternative, there is a risk of uncritical reading (Murray, 1994, 1996).

The second phase of the GBD (2010–2017) saw the use of a prevalence perspective of the YLD. This prevalence perspective implies that YLDs are accounted for one year at a time, instead of all at once. In the chronic obstructive pulmonary disease example above, this practice means that only $1 \times 0.23 = 0.23$ DALYs were accounted for in 2015, and the same will be the case for the next 9 years. There are at least two rationales given for the switch from incidence to prevalence YLDs. First, incidence YLDs rely on strong assumptions about an uncertain future. Second, under a falling incidence, future years may come out better than they should (if DALYs should describe health care needs) because the need for health care services might still be high (Murray *et al.*, 2012a; Schroeder, 2016). In this case, the rationales behind

the choice between an incidence and prevalence YLD can be seen as descriptive as they concern epistemic forecasting rather than value theory.

Individual versus Societal Burden

The question remains of whether the DALY is concerned with morbidity and mortality for those who are sick and dying, for their dependents, or society itself. The DALY has always been primarily concerned with an *individual* burden. However, in the first phase of the GBD, the DALY also referred to a *societal* burden. Murray suggested two possible rationales for this concern: first, the *human capital approach*, where the value of the time at each age should be proportional to the productivity at that age. Second, social values were attributed to age groups that normally act as caregivers for their children and parents. As mentioned, these two rationales were supposed to favor age-weighting (Murray, 1994, 1996). When the GBD also measured societal burden, there seemed to be no principled reasons for a sharp cutoff in the count's lower limit. A concern for societal burden may imply a gradual increase with regard to disease burden generated. This is because societal burden refers to how all but the deceased are affected by a person's premature death or disability. According to this line of thinking, it is not unreasonable to claim that stillbirths also incur at least a minimal degree of societal burden.

From the GBD 2010 and onwards, societal burden was excluded. In the words of Murray *et al.* (2012a, 14) 'Burden should be assessed individual by individual'. This omission led some authors to conclude that the GBD is now almost value-free (Murray *et al.*, 2012a; Salomon *et al.*, 2012; Salomon *et al.*, 2015; Knudsen *et al.*, 2016). The choice to omit societal burden seems reasonable and represents a step toward a slightly more descriptive DALY. However, even the individual *burden* itself implies something of disvalue to the individual, and it is hard to see that such a concern does not rely on value theory or other evaluative approaches. Even if the omission of societal burden can be seen as non-evaluative, individual disease burden is itself an evaluative concept.

Gradualism versus Non-Gradualism

From the 1990 GBD until the 2010 GBD, the combined effect of age-weighting and time discounting gave a gradual curve for disease burden throughout individual lives. This combined effect implied that the highest possible number of YLLs was incurred when 10-year-olds

died (Murray *et al.*, 2012a). Importantly, this gradual function was a result of the combined effect of individual and societal burden. It is hard to say whether this precise implication—that the death of 10-year-olds incurred the greatest number of DALYs—was intended or not.

It is hard to see how a human capital approach would only matter from birth onwards. When age-weighting and time discounting were omitted, the GBD-curve showed a non-gradual function with a sharp discontinuous boundary at birth. Moreover, this non-gradual curve represents an individual burden only. This latter view implies that the age at which death becomes a burden is also when death generates the greatest possible burden. In the current GBD, the burden of death is assumed to be the greatest at birth, where neonatal deaths incur around 86 DALYs each (Murray *et al.*, 2012a). Rationales for gradualism versus non-gradualism are lacking in the GBD literature but can be found elsewhere (McMahan, 2002; Millum, 2015; Solberg and Gamlund, 2016). The choice of whether gradualism or non-gradualism should apply to the DALY may, arguably, be classified as both descriptive and evaluative. The choice is descriptive if we assume that there is a true answer as to when the worst time to die is. However, it is also a value-theoretical question and therefore evaluative in this sense. Moreover, if there is no true answer as to when the worst time to die is, how we evaluate deaths in the DALY remains an open question that requires reasons that are directly value-laden. Thus, we hold that the challenge of gradualism versus non-gradualism in the DALY is an evaluative concern.

Aggregation (Intra- and Interpersonal)

There are two forms of aggregation in the DALY. First, *intrapersonal aggregation*—that is, an aggregation of burden across time within an individual's life. In the GBD study, the individual is the fundamental unit for the disease burden (Murray *et al.*, 2012a). Several assumptions need to be in place for intrapersonal aggregation to make sense, and this has been discussed elsewhere (Broome, 2004; Hirose, 2015).

The current GBD indirectly assumes that we begin to exist from the moment of birth. Additionally, it is assumed that burdens can accumulate within the lives of individuals (as illustrated in Figure 1). This kind of intrapersonal aggregation entails the idea that some relevant property (e.g. the brain, our bodies or sentience) grounds an individual's identity throughout her life. In philosophy, this property is called personal identity. There is little consensus among philosophers about what constitutes the grounds for personal identity

(Parfit, 1984; McMahan, 2002; Olson, 2007). Still, to make sense, there are reasons to believe that the DALY must rely on the assumption that personal identity is acquired at birth and continues until the current definition of death occurs.¹⁹

Second, the DALY presupposes *interpersonal aggregation*—that is, an aggregation of burden across people, at least for estimating DWs. Interpersonal aggregation is an assumption of the DALY that is seldom articulated, even though the founders of the DALY are probably aware of it. An important aspect of interpersonal aggregation is *additive aggregation*. Regardless of what position one adopts on the issue of distribution, the choice of employing a straightforward additive aggregation formula is a value choice.

Commensurability

Closely related to the aggregation of burden is the issue of *commensurability*. There are at least two assumptions regarding commensurability in the GBD.²⁰ First and foremost, YLDs quantify disease burden—an inherently multidimensional construct. More precisely, in the 2015 GBD, the YLD quantifies the burden of 235 distinct conditions. The underlying assumption here is that these conditions are commensurable as individual burdens. However, several of these conditions are not intuitively comparable. Compare, for instance, mild low back pain, moderate hearing loss and amputation of one arm or severe dementia. We should ask what these conditions have in common. To answer this question, we will need an account that unifies these conditions.

A second concern is the paramount assumption that it makes sense to aggregate YLDs and YLLs. The idea is that YLDs and YLLs can be measured on a cardinal ratio scale, which captures the assumption that YLDs and YLLs are commensurable as an individual burden (Murray *et al.*, 2000; Murray and Evans, 2003). However, it is not entirely clear that the YLL measures an individual burden in the first place, as we do not experience—or even exist—while ‘being dead’. If the YLL is, in fact, not an individual burden, then the YLD and the YLL will be incommensurable qua individual burden. This may be a severe problem for the DALY as the YLL sets the reference frame for the YLD and because the majority of the total disease burden consists of YLLs.

The best candidate for a justification of commensurability is perhaps the fact that the DALY presupposes a *counterfactual account of harm*.²¹ Reference to counterfactual harm is probably the best candidate for explaining how the YLD is a multidimensional concept, as well as how the YLD and the YLL are commensurable. Note that harms and benefits directly concern how our well-

being is affected, and so the concern about commensurability is strongly related to evaluative concerns (Solberg *et al.*, 2018).

There are reasons to question whether purely descriptive concepts of morbidity and mortality can be commensurable at all. If the YLD and the YLL in the DALY are to be seen as commensurable, then one will have to admit that there are evaluative aspects involved in this measure. However, this concern about commensurability is absent in the GBD literature. Schroeder has, however, responded to this concern. He argues that if we grant that the DALY is best understood as an index, then, the concern for YLD-YLL commensurability may matter less (Schroeder, 2018). Whether or not Schroeder is right is an open question, but the very assumption of equivalence in value between YLD-YLL is, in our view, an evaluative matter.

Summary of Assumptions

In summary, even if we grant a narrow definition of evaluative assumptions, most of the assumptions that we have discussed are evaluative. We have provided strong reasons in support of the view that the DALY measure should still be regarded as an evaluative endeavor. See Table 1 for a summary of the DALY assumptions that we have explored.

Why the DALY is Primarily Evaluative

All measures of morbidity known to us, such as QALYs and DALYs, erect that scale on evaluative judgments. It is hard to imagine any way to circumvent evaluative judgments, and the burden of justification lies with those who claim that this can be done. Thus, there are reasons to believe that descriptive mortality measures are unsuitable for direct comparison with evaluative morbidity measures without further evaluative adjustments.²² According to this line of reasoning, the DALY is *not*, in this publicly accessible sense, *descriptive* since it is erected on a scaffold of evaluative assumptions.

Moreover, the DALY measures *disease burden*, and *burden* is a normative term—it connotes something negative that one wants to discard. This is another sense in which the DALY is evaluative. The closeness between *disease burden* and the monitoring of global health, discussions on inequalities in health, and the aim of prioritization between major health programs should at least remind us that motivation behind the construction of the DALY is inherently ethical. If the DALY was

Table 1. DALY assumptions in the GBD

Assumption	GBD 1990–1996	GBD 2010–2017	Descriptive
1. Health versus well-being	Practice <ul style="list-style-type: none"> • Proxy of well-being with an attempted rewording to ameliorate the situation 	Practice <ul style="list-style-type: none"> • Health 	No
2. Disability weights	Practice <ul style="list-style-type: none"> • No condition worse than death • Expert panels • The PTO method 	Practice <ul style="list-style-type: none"> • No condition worse than death • Public involvement • Utility like-choice tasks 	No
3. Discounting	Practice <ul style="list-style-type: none"> • Yes, 3% p.a. Reasons <ul style="list-style-type: none"> • Uncertainty • Improvement • Time preferences 	Practice <ul style="list-style-type: none"> • No, 0% p.a. Reasons <ul style="list-style-type: none"> • Every life year should count equally 	No
4. Age-weighting	Practice <ul style="list-style-type: none"> • Yes Reasons <ul style="list-style-type: none"> • Well-being • Productivity • Well-being interdependence 	Practice <ul style="list-style-type: none"> • No Reason <ul style="list-style-type: none"> • Every life year should count equally 	No
5. Choice of reference life expectancy	Practice <ul style="list-style-type: none"> • Birth as lower limit. • 80 years for men, and 82 years for women as upper limit. • Age-adjusted (life table) • Universal (except sex difference) 	Practice <ul style="list-style-type: none"> • Birth as lower limit. • Synthetic life-tables • Age-adjusted (life table) • Universal • No sex difference 	No
6. The incidence versus prevalence approach	Practice <ul style="list-style-type: none"> • Incidence YLD • Incidence YLL Reasons <ul style="list-style-type: none"> • Consistency • Rapidity • Uncritical reading 	Practice <ul style="list-style-type: none"> • Prevalence YLD • Incidence YLL Reasons <ul style="list-style-type: none"> • Incidence YLDs: strong future assumptions, if falling incidence, future 	Yes

Table 1. (continued)

Assumption	GBD 1990–1996	GBD 2010–2017	Descriptive
		years may come out too good.	
7. Individual versus societal burden	Practice <ul style="list-style-type: none"> • Individual (majority) and societal burden 	Practice <ul style="list-style-type: none"> • Individual burden 	No
8. Gradualism versus non-gradualism	Practice <ul style="list-style-type: none"> • Gradualism 	Practice <ul style="list-style-type: none"> • Non-gradualism 	No
9. Aggregation (inter- and intrapersonal)	Practice <ul style="list-style-type: none"> • Both 	Practice <ul style="list-style-type: none"> • Both 	No

Our categorization and contrasting of the assumptions of the first phase of the GBD (1990–1996), against its second phase (2010–2017). We believe that most underlying assumptions and their corresponding rationales are evaluative rather than descriptive.

descriptive in this stronger sense, then to learn that, say, a virus-pandemic generated n DALYs would be less interesting. However, DALYs are not intended to be seen as purely descriptive observations but rather information that motivates policymakers to act.

It is important to remark that we see nothing wrong with using an evaluative measure such as the DALY to measure disease burden. What we take issue with is the air of authority that comes with claiming objectivity and a descriptive model. The assumptions underlying the DALY that we have found are not neutral and demand scrutiny and continuous reassessment. Accordingly, the DALY should not be able to evade the ongoing critical discourse regarding its axiological foundation.²³

Conclusion

We acknowledge that the revisions made in conjunction with the second phase of the GBD (2010–2017) have made the DALY slightly more descriptive of individual disease burden. Notwithstanding—as we have argued—the DALY is still dissimilar from descriptive endeavors such as the crude prevalence or incidence metrics. Many evaluative assumptions will, by necessity, remain embedded in the DALY construct. Modifying these assumptions may affect both the size and the distribution of disease burden across the globe. Our exposition has been a call for more transparency as well as continued scholarly and public scrutiny of the DALY. We conclude that the DALY is primarily evaluative and encourage scholars to continue to seek a firmer ethical foundation of this influential measure.

Conflict of Interest

O.F.N. has been a co-author on several GBD-articles since 2010.

Notes

1. Christopher J. L. Murray, the leading architect of the DALY, dedicated 80 pages to the evaluative choices in the DALY in his article ‘Rethinking DALYs’ (Murray, 1996).
2. The view that the DALY is now a descriptive rather than evaluative measure is evident from the title of a book by the leading architects of the DALY: *An integrative metaregression framework for descriptive epidemiology* (Flaxman *et al.*, 2015).
3. The last quote describes the GBD effort broadly.
4. For a few examples of DALY publications in high-ranking academic journals, see *The Lancet* (Salomon *et al.*, 2015) and *The New England Journal of Medicine* (Murray and Lopez, 2013). For major organizations and institutions, see the [World Health Organization \(2018\)](#) and the [World Bank \(1993\)](#). Furthermore, the DALY is expected to play a significant role in monitoring the United Nations’ 2030 Agenda for Sustainable Development (see [Reddy, 2016](#); IHME, 2018).
5. Please note that our aim is not to construct a novel and rigorous philosophical distinction between evaluative and descriptive assumptions. Instead, this is a first step, in line with previous work on value assumptions in the DALY measure.

6. For more on these methods, see, e.g. Gold et al. (2002) and Weinstein et al. (2009).
7. In the second phase of the GBD (after the publication of the 2010 GBD), a slightly different method from life time DALYs is used. The YLD is now a prevalence rather than an indices measure.
8. In line with the current terminology within epidemiology, we do believe that it is constructive to continue to use and reserve the term descriptive for certain mortality measures (such as the CDRs), even if they must all invoke certain assumptions. Most notably, all mortality measures must rely on a specific definition of death. In the last fifty years, the brain-death criterion has been the dominant view.
9. In contrast to Schroeder, we consider the issue of grouping conditions and disabilities as something outside the DALY-paradigm.
10. We focus on these additional assumptions because they are scarcely discussed in the DALY literature, or not discussed at all, while relating in a direct way to the DALY construct, and small changes in these assumptions will have significant implications for the total amount of disease burden. There is also a parallel discussion about what purposes the DALY model is suitable (Arnesen and Kapiriri, 2004; Voigt and King, 2014). Since this discussion is beyond the DALY model itself, we do not discuss it here.
11. The standard gamble and the time trade-off are similar methods to the person trade-off, but instead of the number of people, they establish a relative burden.
12. One may, of course, refer to such deeply biological preferences as non-choices and pseudo-preferences in this sense.
13. For the utility theoretical foundation of the DW, see Neumann and Morgenstern (2004).
14. As Ubel et al. (1996) show, priority-setting policies that do not survey the public views on trade-offs and reflect these are often disbanded. There is also a further concern about who to survey in the trade-offs. This concern involves balancing many variables, such as age, sex, culture, personal experiences and expertise. For more on this issue, see e.g. Weinstein et al. (2009) and Schroeder (2016).
15. In contrast, negative value has been practiced in the QALY.
16. This latter point is also a value judgment, as scholars reasonably disagree about whether every life year lost has the same value regardless of age.
17. There are several definitions of stillbirths. In this article, we accord with the WHO's definition where fetal death between the 28th gestational week and birth count as stillbirths.
18. First, the DALY measure seems to presume that being dead has a value of 1. Second, one could have applied a prevalence YLL, where one YLL was attributed for each year. See, e.g. Bradley (2009) for more on this issue.
19. The 'counting from birth' practice is probably chosen for political reasons. Our point is that an underlying account of personal identity is necessary for the practice to make sense.
20. Commensurability, in this context, means that the things we are measuring (such as A, B and C) have something relevant F in common that allows us to measure them on the same scale.
21. For recent developments on the counterfactual account of harm, see, e.g. Klocksien (2012) and Feit (2016).
22. Perhaps the best example is the QALY, which is acknowledged as an evaluative measure. See, e.g. Weinstein et al. (2009).
23. CTS had the initial idea and drafted the first manuscript. All authors circulated and revised the manuscript versions and held working sessions; MB made the figures and formulae; and all authors revised and approved the final manuscript.

Acknowledgements

This article was presented at the Department of Philosophy, University of Oslo (2016), the international CSMN-closing conference in Oslo (2017), an internal seminar at BCEPS in Bergen (2019), an international practical philosophy workshop in Bergen (2019), as well as the 33rd European Conference on Philosophy of Medicine and Health Care, in Oslo (2019). We would especially like to thank Jakob Elster, Kristine Kirkebø Giske, Axel Gosseries, David Miller, S. Andrew Schroeder, Anders Strand, Sveinung S. Sivertsen, Jesse Tomalty, and Joost Jacob Vecht, as well as the associate editor and one anonymous reviewer for their valuable comments on earlier drafts of this manuscript.

References

- Anand, S. and Hanson, K. (1997). Disability-Adjusted Life Years: A Critical Review. *Journal of Health Economics*, 16, 685–702.
- Anand, S. and Reddy, S. G. (2019). *The Construction of the DALY: Implications and Anomalies*. Department of Economics Discussion Paper Series. University of Oxford. Number 877, ISSN 1471–0498.

- Arnesen, T. and Kapiriri, L. (2004). Can the Value Choices in DALYs Influence Global Priority-Setting? *Health Policy*, **70**, 137–149.
- Arnesen, T. and Nord, E. (1999). The value of DALY life: problems with ethics and validity of disability adjusted life years. *BMJ*, **319**, 7222, 1423–1425.
- Bradley, B. (2009). *Well-Being and Death*. New York: Oxford University Press.
- Broome, J. (2002). Measuring the Burden of Disease by Aggregating Well-Being. In C. J. L. Murray, J. A. Salomon, C. D. Mathers, and A. D. Lopez. (Eds.) *Summary Measures of Population Health. Concepts, Ethics, Measurement and Applications*. Chapter 3, 91–113. Geneva: World Health Organization.
- Broome, J. (2004). *Weighing Lives*. Oxford: Oxford University Press.
- Burstein, R., Fleming, T., Haagsma, J., Salomon, J. A., Vos, T., and Murray, C. J. L. (2015). Estimating Distributions of Health State Severity for the Global Burden of Disease Study. *Population Health Metrics*, **13**, 31.
- Cartwright, N. and Bradburn, N. (2011). A Theory of Measurement. In *The Importance of Common Metrics for Advancing Social Science Theory and Research: A Workshop Summary*. Washington: National Academies Press, pp. 53–56.
- Cartwright, N. and Runhardt, R. (2014). Measurement. In N., Cartwright and E., Montuschi (eds.), *Philosophy of Social Science. A New Introduction*, Chapter 14. New York: Oxford University Press, pp. 265–288.
- Chang, H. and Cartwright, N. (2014). Measurement. In M., Curd and S., Psillos (eds.), *The Routledge Companion to Philosophy of Science*, 2nd edn. London: Routledge, pp. 411–419.
- Dempsey, M. (1947). Decline in Tuberculosis: The Death Rate Fails to Tell the Entire Story. *American Review of Tuberculosis*, **56**, 157–164.
- Douglas, H. (2009). *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.
- Douglas, H. (2016). Values in Science. In P. Humphreys (ed.), *The Oxford Handbook of Philosophy of Science*. New York: Oxford University Press, pp. 609–32.
- Feit, N. (2016). Comparative Harm, Creation and Death. *Utilitas*, **28**, 136–163.
- Flaxman, A. D., Vos, T., and Murray, C. J. L. (2015). *An Integrative MetaRegression Framework for Descriptive Epidemiology*. Seattle: University of Washington Press.
- Fox-Rushby, J. and Hanson, K. (2001). Calculating and Presenting Disability Adjusted Life Years (DALYs) in Cost-Effectiveness Analysis. *Health Policy and Planning*, **16**, 326–331.
- Gamlund, E. and Solberg, C. T. (eds) (2019). *Saving People from the Harm of Death*. New York: Oxford University Press.
- Global Burden of Disease Collaborative Network (2018). *Global Burden of Disease Study 2017 (GBD 2017)*. Seattle, United States: Institute for Health and Evaluation (IHME).
- Gold, M. R., Stevenson, D., and Fryback, D. G. (2002). HALYs and QALYs and DALYs, Oh My: Similarities and Differences in Summary Measures of Population Health”. *Annual Review of Public Health*, **23**, 115–134.
- Haenszel, W. (1950). A Standardized Rate for Mortality Defined in Units of Lost Years of Life. *American Journal of Public Health*, **40**, 17–26.
- Hausman, D. (2015). *Valuing Health: Well-Being, Freedom, and Suffering*. New York: Oxford University Press.
- Hausman, D. and McPherson, M. S. (2006). *Economic Analysis, Moral Philosophy, and Public Policy*. Cambridge: Cambridge University Press.
- Hirose, I. (2015). *Moral Aggregation*. Oxford: Oxford University Press.
- Klocksien, J. A. (2012). A Defense of the Counterfactual Comparative account of Harm. *American Philosophical Quarterly*, **49**, 285–300.
- Knudsen, A. K., Kinge, J. M., Skirbekk, V., and Vollset, S. E. (2016). *Sykdomsbyrde i Norge 1990:2013: Resultater Fra Global Burden of Disease, Injuries, and Risk Factors Study 2013. (GBD 2013)*. Oslo: Nasjonalt Folkehelseinstitutt.
- Lacey, H. (1999). *Is Science Value-Free?* London: Routledge.
- Longino, H. (1990). *Science as Social Knowledge*. Princeton: Princeton University Press.
- Lytckens, C. H. (2003). Time to Disable DALYs? on the Use of Disability-Adjusted Life-Years in Health Policy. *European Journal of Health Economics*, **4**, 195–202.
- McMahan, J. (2002). *The Ethics of Killing: Problems at the Margins of Life*. New York: Oxford University Press.
- Millum, J. (2015). Age and Death: A Defence of Gradualism”. *Utilitas*, **27**, 279–297.
- Murray, C. J. L. (1994). Quantifying the Burden of Disease: The Technical Basis for Disability-Adjusted Life Years. *Bulletin of the World Health Organization*, **72**, 429–445.
- Murray, C. J. L. (1996). Rethinking DALYs. In C. J. L., Murray and A. D., Lopez *The Global Burden of Disease*. Cambridge, MA: Harvard University Press, pp. 1–98.
- Murray, C. J. L. and Acharya, A. K. (1997). Understanding DALYs. *Journal of Health Economics*, **16**, 703–730.

- Murray, C. J. L. and Evans, D. B. (eds) (2003). *Health Systems Performance Assessment: Debates, Methods and Empiricism*. Geneva: World Health Organization.
- Murray, C. J. L. and Lopez, A. (1996). *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020: Summary*. Cambridge: World Health Organization.
- Murray, C. J. L. and Lopez, A. (2013). Measuring the Global Burden of Disease. *The New England Journal of Medicine*, **369**, 448–457.
- Murray, C. J. L., Ezzati, M., Flaxman, A. D., Lim, S., Lozano, R., Michaud, C., Naghavi, M., Salomon, J. A., Shibuya, K., Vos, T., Wikler, D., and Lopez, A. D. (2012a). Comprehensive Systematic Analysis of Global Epidemiology: Definitions, Methods, Simplification of DALYs, and Comparative Results from the Global Burden of Disease Study. *The Lancet*, **380**, 1–140.
- Murray, C. J. L., Ezzati, M., Flaxman, A. D., Lim, S., Lozano, R., Michaud, C., Naghavi, M., Salomon, J. A., Shibuya, K., Vos, T., Wikler, D., and Lopez, A. D. (2012b). GBD 2010: Design, Definitions, and Metrics. *The Lancet*, **380**, 2063–2066.
- Murray, C. J. L., Salomon, J. A., and Mathers, C. A. (2000). A Critical Examination of Summary Measures of Population Health. *Bulletin of the World Health Organization*, **78**, 981–994.
- Neumann, J. J. and Morgenstern, O. (2004 [1944]) *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- Olson, E. T. (2007). *What Are We? A Study in Personal Ontology*. New York: Oxford University Press.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Porta, M. (ed) (2014). *A Dictionary of Epidemiology: Sixth Edition*. Oxford: Oxford University Press.
- Reddy, K. S. (2016). Global Burden of Disease Study 2015 Provides GPS for Global Health 2030. *The Lancet*, **388**, 1448–1449.
- Robinson, H. L. (1948). Mortality Trends and Public Health in Canada. *Canadian Journal of Public Health*, **39**, 60–70.
- Romeder, J. M. and McWhinnie, J. R. (1977). Potential Years of Life Lost between Ages 1 and 70: An Indicator of Premature Mortality for Health Planning. *International Journal of Epidemiology*, **6**, 143–151.
- Salomon, J. A., Haagsma, J. A., Davis, A., de Noordhout, C. M., Polinder, S., Havelaar, A. H., Cassini, A., Devleeschauwer, B., Kretschmar, M., Speybroeck, N., Murray, C. J. L., and Vos, T. (2015). Disability Weights for the Global Burden of Disease 2013 Study. *The Lancet Global Health*, **3**, e712.
- Salomon, J. A., Vos, T., Hogan, D. R., Gagnon, M., Naghavi, M., Mokdad, A., Begum, N., Shah, R., Karyana, M., Kosen, S., Farje, M. R., Moncada, G., Dutta, A., Sazawal, S., Dyer, A., Seiler, J., Aboyans, V., Baker, L., Baxter, A., Benjamin, E. J., Bhalla, K., Abdulkhak, A. B., Blyth, F., Bourne, R., Braithwaite, T., Brooks, P., Brugha, T. S., Bryan-Hancock, C., Buchbinder, R., Burney, P., Calabria, B., Chen, H., Chugh, S. S., Cooley, R., Criqui, M. H., Cross, M., Dabhadkar, K. C., Dahodwala, N., Davis, A., Degenhardt, L., Díaz-Torné, C., Dorsey, E. R., Driscoll, T., Edmond, K., Elbaz, A., Ezzati, M., Feigin, V., Ferri, C. P., Flaxman, A. D., Flood, L., Fransen, M., Fuse, K., Gabbe, B. J., Gillum, R. F., Haagsma, J., Harrison, J. E., Havmoeller, R., Hay, R. J., Hel-Baqui, A., Hoek, H. W., Hoffman, H., Hogeland, E., Hoy, D., Jarvis, D., Jonas, J. B., Karthikeyan, G., Knowlton, L. M., Lathlean, T., Leasher, J. L., Lim, S. S., Lipshultz, S. E., Lopez, A. D., Lozano, R., Lyons, R., Malekzadeh, R., Marcenes, W., March, L., Margolis, D. J., McGill, N., McGrath, J., Mensah, G. A., Meyer, A.-C., Michaud, C., Moran, A., Mori, R., Murdoch, M. E., Naldi, L., Newton, C. R., Norman, R., Omer, S. B., Osborne, R., Pearce, N., Perez-Ruiz, F., Perico, N., Pesudovs, K., Phillips, D., Pourmalek, F., Prince, M., Rehm, J. T., Remuzzi, G., Richardson, K., Room, R., Saha, S., Sampson, U., Sanchez-Riera, L., Segui-Gomez, M., Shahraz, S., Shibuya, K., Singh, D., Sliwa, K., Smith, E., Soerjomataram, I., Steiner, T., Stolk, W. A., Stovner, L. J., Sudfeld, C., Taylor, H. R., Tleyjeh, I. M., van der Werf, M. J., Watson, W. L., Weatherall, D. J., Weintraub, R., Weisskopf, M. G., Whiteford, H., Wilkinson, J. D., Woolf, A. D., Zheng, Z.-J., and Murray, C. J. (2012). Common Values in Assessing Health Outcomes from Disease and Injury: Disability Weights Measurement Study for the Global Burden of Disease Study 2010. *The Lancet*, **380**, 2129–2143.
- Schroeder, S. A. (2012). Incidence, Prevalence, and Hybrid Approaches to Calculating Disability-Adjusted Life Years. *Population Health Metrics*, **10**, 19.
- Schroeder, S. A. (2017). *Value Choices in Summary Measures of Population Health*. *Public Health Ethics*, **10**, 2, 176–187.
- Schroeder, S. A. (2018). Is Consistency Overrated? *Journal of Medical Ethics*, **44**, 199–200.
- Schroeder, S. A. (2019). Which Values Should Be Built into Economic Measures? *Economics & Philosophy*, **35**, 3, 521–536. doi:10.1017/S0266267118000317.

- Solberg, C. T. and Gamlund, E. (2016). The Badness of Death and Priorities in Health. *BMC Medical Ethics*, 17, 21.
- Solberg, C. T., Norheim, O. F., and Barra, M. (2018). The Disvalue of Death in the Global Burden of Disease. *Journal of Medical Ethics*, 44, 192–198.
- Tal, E. (2017). Measurement in Science. In E. N., Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), available from: <https://plato.stanford.edu/entries/measurement-science/>.
- Ubel, P. A., DeKay, M. L., Baron, J., and Asch, D. A. (1996). Cost-effectiveness Analysis in a Setting of Budget Constraints—Is It Equitable? 334, 18, 1174–11777.
- Voigt, K. (2012) Measuring Global Health. In P. T., Lenard and C., Straehle (eds.), *Health Inequalities and Global Justice*. Edinburgh: Edinburgh University Press, pp. 139–156.
- Voigt, K. and King, N. B. (2014). Disability Weights in the Global Burden of Disease 2010 Study: Two Steps Forward, One Step Back? *Bulletin of the World Health Organization*, 92, 226–228.
- Wang, H., Bhutta, Z. A., Coates, M. M., Coggeshall, M., Dandona, L., Diallo, K., Franca, E. B., Fraser, M., Fullman, N., Gething, P. W., Hay, S. I., Kinfu, Y., Kita, M., Kulikoff, X. R., Larson, H. J., Liang, J., Liang, X., Lim, S. S., Lind, M., Lopez, A. D., Lozano, R., Mensah, G. A., Mikesell, J. B., Mokdad, A. H., Mooney, M. D., Naghavi, M., Nguyen, G., Rakovac, I., Salomon, J. A., Silpakit, N., Sligar, A., Sorensen, R. J. D., Vos, T., Zhu, J., Abajobir, A. A., Abate, K. H., Abbas, K. M., Abd-Allah, F., Abdulle, A. M., Abera, S. F., Aboyans, V., Abraham, B., Abubakar, I., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Abyu, G. Y., Achoki, T., Adebisi, A. O., Adedeji, I. A., Adelekan, A. L., Adou, A. K., Agarwal, A., Ajala, O. N., Akinyemiju, T. F., Akseer, N., Alam, K., Alam, N. K. M., Alasfoor, D., Aldridge, R. W., Alegretti, M. A., Alemu, Z. A., Ali, R., Alkerwi, A., Alla, F., Al-Raddadi, R., Alsharif, U., Altirkawi, K. A., Martin, E. A., Alvis-Guzman, N., Amare, A. T., Amberbir, A., Amegah, A. K., Ameh, E. A., Ammar, W., Amrock, S. M., Andersen, H. H., Anderson, G. M., Antonio, C. A. T., Ärnlöv, J., Artaman, A., Asayesh, H., Asghar, R. J., Assadi, R., Atique, S., Avokpaho, E. F. G. A., Awasthi, A., Quintanilla, B. P. A., Bacha, U., Badawi, A., Balakrishnan, K., Banerjee, A., Banigbe, B. F., Barac, A., Barber, R. M., Barker-Collo, S. L., Bärnighausen, T., Barrero, L. H., Bayou, T. A., Bayou, Y. T., Bazargan-Hejazi, S., Beardsley, J., Bedi, N., Bekele, T., Bell, M. L., Bello, A. K., Bennett, D. A., Bensenor, I. M., Berhane, A., Bernabé, E., Betsu, B. D., Beyene, A. S., Bhatt, S., Biadgilign, S., Bikbov, B., Birlik, S. M., Bisanzio, D., Bjertness, E., Blore, J. D., Bourne, R. R. A., Brainin, M., Brazinova, A., Breitborde, N. J. K., Brown, A., Buckle, G. C., Burch, M., Butt, Z. A., Campos-Nonato, I. R., Campuzano, J. C., Cárdenas, R., Carpenter, D. O., Carrero, J. J., Carter, A., Casey, D. C., Castañeda-Orjuela, C. A., Rivas, J. C., Castro, R. E., Catalá-López, F., Cercy, K., Chang, H.-Y., Chang, J.-C., Chibueze, C. E., Chisumpa, V. H., Choi, J.-Y. J., Chowdhury, R., Christopher, D. J., Ciobanu, L. G., Colquhoun, S. M., Cooper, C., Cornaby, L., Damtew, S. A., Danawi, H., Dandona, R., das Neves, J., Davis, A. C., de Jager, P., De Leo, D., Degenhardt, L., Deribe, K., Deribew, A., Jarlais, D. C. D., deVeber, G. A., Dharmaratne, S. D., Dhillon, P. K., Ding, E. L., Doshi, P. P., Doyle, K. E., Duan, L., Dubey, M., Ebrahimi, H., Ellingsen, C. L., Elyazar, I., Endries, A. Y., Ermakov, S. P., Eshrati, B., Esteghamati, A., Faraon, E. J. A., Farid, T. A., Farinha, C. S. e S., Faro, A., Farvid, M. S., Farzadfar, F., Fereshtehnejad, S.-M., Fernandes, J. C., Fischer, F., Fitchett, J. R. A., Foigt, N., Franklin, R. C., Friedman, J., Fürst, T., Gambashidze, K., Gamkrelidze, A., Ganguly, P., Gebre, T., Gebrehiwot, T. T., Gebremedhin, A. T., Gebru, A. A., Geleijnse, J. M., Gessner, B. D., Ginawi, I. A. M., Giref, A. Z., Gishu, M. D., Gomez-Dantes, H., Gona, P., Goodridge, A., Gopalani, S. V., Goto, A., Gouda, H. N., Gugnani, H. C., Guo, Y., Gupta, R., Gupta, R., Gupta, V., Gyawali, B., Haagsma, J. A., Hafezi-Nejad, N., Haile, D., Hailu, A. D., Hailu, G. B., Hamadeh, R. R., Hamidi, S., Hancock, J., Handal, A. J., Hankey, G. J., Harb, H. L., Harikrishnan, S., Harun, K. M., Havmoeller, R., Hay, R. J., Heredia-Pi, I. B., Hoek, H. W., Horino, M., Horita, N., Hosgood, H. D., Hotez, P. J., Hoy, D. G., Hsairi, M., Hu, G., Huang, C., Huang, J. J., Huang, H., Huiart, L., Huynh, C., Iburg, K. M., Idrisov, B. T., Innos, K., Jacobsen, K. H., Jahanmehr, N., Javanbakht, M., Jayatilleke, A. U., Jee, S. H., Jeemon, P., Jha, V., Jiang, G., Jiang, Y., Jibat, T., Jin, Y., Jonas, J. B., Kabir, Z., Kalkonde, Y., Kamal, R., Kan, H., Kang, G., Karch, A., Karema, C. K., Kasaeian, A., Kaul, A., Kawakami, N., Kayibanda, J. F., Kazanjan, K., Keiyoro, P. N., Kemp, A. H., Kengne, A. P., Keren, A., Kereselidze, M., Kesavachandran, C. N., Khader, Y. S., Khalil, I. A., Khan, A. R., Khan, E. A., Khang, Y.-H., Khonelidze, I., Khubchandani, J., Kim, C.-i., Kim, D., Kim, Y. J., Kissoon, N., Kivipelto, M., Knibbs, L. D., Kokubo, Y., Kosen, S., Koul, P. A., Koyanagi, A., Defo, B. K., Bicer, B. K., Kudom, A. A., Kumar, G. A., Kutz, M. J., Kyu, H. H., Lal, D. K., Lalloo, R., Lam, H., Lam, J. O., Lansingh, V. C.,

- Larsson, A., Leigh, J., Leung, R., Li, Y., Li, Y., Lindsay, M. P., Liu, P. Y., Liu, S., Lloyd, B. K., Lo, W. D., Logroscino, G., Low, N., Lunevicius, R., Lyons, R. A., Ma, S., Razek, H. M. A. E., Razek, M. M. A. E., Mahdavi, M., Majdan, M., Majeed, A., Malekzadeh, R., Mapoma, C. C., Marcenes, W., Martinez-Raga, J., Marzan, M. B., Masiye, F., McGrath, J. J., Meaney, P. A., Mehari, A., Mehndiratta, M. M., Mekonnen, A. B., Melaku, Y. A., Memiah, P., Memish, Z. A., Mendoza, W., Meretoja, A., Meretoja, T. J., Mhimbira, F. A., Miller, T. R., Mills, E. J., Mirarefin, M., Misganaw, A., Mock, C. N., Mohammad, K. A., Mohammadi, A., Mohammed, S., Monasta, L., Hernandez, J. C. M., Montico, M., Moore, A. R., Moradi-Lakeh, M., Morawska, L., Mori, R., Mueller, U. O., Murphy, G. A. V., Murthy, S., Nachega, J. B., Naheed, A., Naidoo, K. S., Naldi, L., Nand, D., Nangia, V., Neupane, S., Newton, C. R., Newton, J. N., Ng, M., Ngalesoni, F. N., Nguhiu, P., Nguyen, Q. L., Nisar, M. I., Pete, P. M. N., Norheim, O. F., Norman, R. E., Ogbo, F. A., Oh, I.-H., Ojelabi, F. A., Olivares, P. R., Olusanya, B. O., Olusanya, J. O., Oren, E., Ota, E., PA, M., Park, E.-K., Park, H.-Y., Parsaeian, M., Caicedo, A. J. P., Patten, S. B., Pedro, J. M., Pereira, D. M., Perico, N., Pesudovs, K., Petzold, M., Phillips, M. R., Pillay, J. D., Pishgar, F., Polinder, S., Pope, D., Popova, S., Pourmalek, F., Qorbani, M., Rabiee, R. H. S., Rafay, A., Rahimi-Movaghar, V., Rahman, M., Rahman, M. H. U., Rahman, S. U., Rai, R. K., Raju, M., Ram, U., Rana, S. M., Ranabhat, C. L., Rao, P., Refaat, A. H., Remuzzi, G., Resnikoff, S., Reynolds, A., Rojas-Rueda, D., Ronfani, L., Roshandel, G., Roth, G. A., Roy, A., Ruhago, G. M., Sagar, R., Saleh, M. M., Sanabria, J. R., Sanchez-Niño, M. D., Santos, I. S., Santos, J. V., Sarmiento-Suarez, R., Sartorius, B., Satpathy, M., Savic, M., Sawhney, M., Schneider, I. J. C., Schöttker, B., Schwebel, D. C., Seedat, S., Sepanlou, S. G., Servan-Mori, E. E., Setegn, T., Shahrzad, S., Shaikh, M. A., Shakh-Nazarova, M., Sharma, R., She, J., Sheikhbahaei, S., Shen, J., Sheth, K. N., Shibuya, K., Shin, H. H., Shin, M.-J., Shiri, R., Shuie, I., Sigfusdottir, I. D., Silva, D. A. S., Silverberg, J., Simard, E. P., Sindi, S., Singh, A., Singh, J. A., Singh, O. P., Singh, P. K., Singh, V., Soriano, J. B., Soshnikov, S., Sposato, L. A., Sreeramareddy, C. T., Stathopoulou, V., Steel, N., Stroumpoulis, K., Sturua, L., Sunguya, B. F., Swaminathan, S., Sykes, B. L., Szoeker, C. E. I., Tabarés-Seisdedos, R., Tavakkoli, M., Taye, B., Tedla, B. A., Tefera, W. M., Tekle, T., Shifa, G. T., Terkawi, A. S., Tesfay, F. H., Tessema, G. A., Thapa, K., Thomson, A. J., Thorne-Lyman, A. L., Tobe-Gai, R., Tonelli, M., Topor-Madry, R., Topouzis, F., Tran, B. X., Troeger, C., Truelsen, T., Dimbuene, Z. T., Tura, A. K., Tyrovolas, S., Ukwaja, K. N., Uneke, C. J., Uthman, O. A., Vaezghasemi, M., Vasankari, T., Vasconcelos, A. M. N., Venketasubramanian, N., Verma, R. K., Violante, F. S., Vladimirov, S. K., Vlassov, V. V., Vollset, S. E., Wang, L., Wang, Y., Weichenthal, S., Weiderpass, E., Weintraub, R. G., Weiss, D. J., Werdecker, A., Westerman, R., Widdowson, M.-A., Wijeratne, T., Williams, T. N., Wiysonge, C. S., Wolfe, C. D. A., Wolfe, I., Won, S., Wubshet, M., Xiao, Q., Xu, G., Yadav, A. K., Yakob, B., Yano, Y., Yaseri, M., Ye, P., Yebyo, H. G., Yip, P., Yonemoto, N., Yoon, S.-J., Younis, M. Z., Yu, C., Zaidi, Z., Zaki, M. E. S., Zeeb, H., Zhang, H., Zhao, Y., Zheng, Y., Zhou, M., Zodpey, S., and Murray, C. J. L. (2016). Global, Regional, National, and Selected Subnational Levels of Stillbirths, Neonatal, Infant, and under-5-Mortality, 1980–2015: A Systematic Analysis for the Global Burden of Disease Study 2015. *The Lancet*, **388**, 1725–1774.
- Weinstein, M. C., Torrance, G., and McGuire, A. (2009). QALYs: The Basics. *Value in Health*, **12**, S5–S9.
- World Bank (1993). *World Development Report 1993. Investing in Health*. New York: Oxford University Press.
- World Health Organization (2018). The top 10 causes of death—Leading causes of death by economy income group. Available from: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> [Accessed 25 August 2020].