Article

Universal Metrics for Climate Change Adaptation Finance? A Cautionary Tale

Andrea Klonschinski

Faculty of Philosophy, Christian-Albrechts-Universität Kiel, 24118 Kiel, Germany; klonschinski@philsem.uni-kiel.de

Abstract: Climate change adaptation is receiving ever more attention in the literature and in practice. Since available funds are not meeting adaptation needs, the question of how to allocate scarce resources becomes pressing. Universal adaptation metrics promise to facilitate the allocation process ex ante and the evaluation of projects ex post. Two such metrics have been proposed recently: Saved Wealth (SW), measured in terms of money, and Saved Health (SH), gauged in terms of disability adjusted life-years (DALYs). The paper analyzes this SWSH approach and shows that it is replete with unresolved conceptual and normative-ethical problems, which are exemplary for universal metrics seeking to combine concerns for equity and efficiency at once. The paper's aim is to uncover these issues, and its conclusion is modest: universal metrics such as SW and the DALY have to be designed and used with great caution and further research is necessary.

Keywords: climate change adaptation; resource allocation; efficiency; equity; health; DALY

1. Introduction

As climate change adaptation is getting ever more attention in the literature and in practice and since available funds are not meeting adaptation needs, the question of how to allocate scarce resources becomes pressing. Hence, a central issue in the current debate is the question of how to measure, aggregate, and compare the effects of climate change adaptation projects ([1] (p. 7)). If these effects could be measured and aggregated into one or a few general metrics, the latter could be applied within cost-effectiveness analysis (CEA) to identify and prioritize the most efficient adaptation projects ex ante and to evaluate their success ex post ([2] (p. 29)). Yet, since adaptation is an inherently multidimensional and context-dependent phenomenon, a universally accepted definition of adaptation is still lacking; it is therefore contested in the literature whether universal metrics for climate change adaptation are feasible at all ([1] (p. 8), [2] (p. 31), [3] (p. 152)).

Michaelowa and colleagues have recently proposed two such metrics, though: Saved Wealth (SW) and Saved Health (SH) [4–7] (For reasons of brevity, I will refer to these papers as "Michaelowa et al." or "Michaelowa and colleagues" in the following text, since Axel Michaelowa is the only author involved in all the papers considered. Although the papers differ in their specific content, I take it that there is sufficient common ground to warrant this way of speaking). They define the aim of adaptation in rather general terms as improving the "well-being of persons and countries" ([5] (p. 110) and operationalize it by means of the three indicators (I will use the terms "metric" and "indicator" interchangeably): absolute economic savings and relative economic savings on the one hand (SW) and human health and lives saved on the other (SH). (Stadelmann et al. [5] only tackle human lives saved. Yet, for this purpose, counting DALYs prohibited by some policy is neither necessary nor sufficient. I will return to this issue in Section 4.) Taking into account the relative economic savings is supposed to ameliorate some ethically troublesome implications of cost–benefit analysis (CBA). While the latter only focuses on the sum of monetary benefits saved and is thus higher the wealthier a community is, the same amount of benefits constitutes a larger saving relative to the income of poorer communities. Hence, SW is supposed to

incorporate concerns for vulnerability ([6] (p. 2150)). SH, in turn, consists in the number of disability-adjusted life-years (DALYs) prohibited by a project. Stadelmann and colleagues endorse the DALY, especially because a non-monetary measure of health benefits allows to circumvent the "endless political debates about an equitable valuation of human life and health" ([6] (p. 2149)), which are "fraught with ethical and political challenges" ([7] (p. 67); see also [4], (p. 9), [6], (p. 2152)). In general, Michaelowa and colleagues seem anxious to be as normatively modest as possible in formulating their approach and stress several times that they do not seek to tackle ethical questions, which, they argue, call for political decisions ([5] (p. 113), [6] (p. 2146)). SW and SH are thus intended to serve as *effectiveness*, not as *equity* measures ([5] (p. 101)).

　　Yet, their approach does not live up to its task, since neither the SW nor the SH metric is as normatively innocuous as Michaelowa et al. seem to assume, as the present paper demonstrates. The basic problem of both measures is that they compound *descriptive elements* of money or health saved on the one hand (effectiveness) with distributional considerations as to how the units of value saved should be distributed among the persons concerned when it comes to adaptation funding decisions on the other (equity). In effect, SW and SH may end ethical and political debates, albeit not by solving them, but by hiding critical normative assumptions in highly aggregated numbers. As the use of these indicators does not enhance but diminishes transparency of decision making, they should at best be used with great caution as adaptation metrics.

　　To support these claims, the argument proceeds as follows. The ensuing section deals with the SW indicator and analyzes Michaelowa and colleagues' arguments for using relative and absolute economic savings as measures for adaptation benefit (Section 2). In doing so, I will argue, first, that they reject a "purely" economic approach, focusing only on the economic benefits of adaptation projects, for the wrong reasons; second, that the measure of relative economic savings does not accomplish what the authors intend it to; and third, that the justification of combining absolute and relative economic savings is conceptually unsound and ultimately unconvincing. While, in general, (philosophical) problems of CBA are quite well-known in environmental ethics and the climate change debate [8,9], the same cannot be said for the discussion on the use of summary measures of population health within CEA; Michaelowa and colleagues hardly discuss the DALY at all. To fill this lacuna in the literature on climate change, the better part of this paper critically examines the DALY (Section 3). After outlining the theoretical basics of the measure, three important steps of the DALY's theoretical development since its inception in 1990 will be sketched to demonstrate that the DALY is not ethically innocuous. What is especially troublesome is the amalgamation of concerns for the descriptive measurement of health on the one hand and of distributional concerns of how health care *should* be allocated on the other hand. As serious conceptual and ethical issues still remain unsolved, the DALY does not meet the demands made by Michaelowa et al. Note that by undertaking a close analysis of the specific arguments presented by Michaelowa et al., the paper makes a general contribution to the debate on climate change adaption finance and adaptation metrics, since the authors' defense of the SWSH approach raises some general problems of economic evaluations, summary measures of population health, and resource allocation. The paper concludes that SW and SH should only be used with great caution as adaptation metrics (Section 4).

　　Beforehand, three clarificatory remarks are in order, the first two of which concern the aim and scope of the paper (I am grateful to two anonymous reviewers for pressing me on these points). First, it is beyond this paper's scope to propose alternative solutions to the difficult questions of how adaptation should be measured and how adaptation funding should be allocated. It contributes to these debates, though, by proposing an elaborate analysis of universal metrics and, thus, providing a basis for further research on the matter. Second, I appreciate the fact that Michaelowa et al.'s paper serves the same function of enhancing the debate and I am sure they themselves were aware of the problems their SWSH approach is facing. By criticizing it, I am taking up the ball and take their

contribution seriously. The final remark concerns the concept of "adaptation finance." I will use this term broadly and will not differentiate between different sources or levels of funding, since these differences do not seem to impact the fundamental normative analysis of summary measures undertaken here. However, they certainly will make a difference when it comes to the need for and the normative consequences of using universal metrics. This is an important topic for further research.

2. Saved Wealth: Absolute and Relative Savings

2.1. Absolute Wealth and Cost–Benefit Analysis

The first metric for measuring the effectiveness of climate adaptation projects proposed by Michaelowa et al., SW, is itself a product of two indicators, saved absolute and saved relative wealth. The former measures the absolute wealth of a community saved by the adaptation project, adjusted by purchasing power parity when it comes to international comparisons ([4] (p. 10), [5], (p. 2150)). This is the classical approach of CBA, which defines all benefits of a certain project in terms of money. If the benefits in question cannot be assigned a market value, a method called contingent valuation is used to mimic the market valuation process ([10], (p. 2f.)). Thereby, respondents are asked for their willingness to pay (WTP) for some hypothetical alternative presented to them (such as saving an endangered species) or the sum of money they would regard as compensation for forgoing the option (e.g., for letting the species become extinct). If the accumulated WTP values outweigh a project's costs, the respective policy is regarded as efficient.

Michaelowa et al. reject CBA's measure of value as *sole* aim of adaption projects for three reasons. First, they claim that CBA cannot account for "non-monetized" benefits, such as some social and environmental benefits ([5] (p. 111)). Second, they criticize contingent valuation by arguing that the elicited WTP depends on the person's ability to pay, which, in turn, hangs on a country's wealth, so resource allocation in climate adaptation based on CBA would systematically disfavor the most vulnerable countries ([5] (p. 111)). Third, they regard the matter of assigning a monetary value to human life as too contentious ([6] (p. 2149)). In the following text, each of these objections will be considered in some detail. The present Section 2.1 deals with the first, the following section deals with the second objection (Section 2.2), and the third will be examined comprehensively in Section 3.

Regarding the first criticism, i.e., CBA's restriction to "economic benefits" ([6] (p. 2148) and its disregard of "non-monetized" benefits ([5] (p. 111)), it remains unclear what exactly the authors have in mind. That CBA simply looks at "economic benefits" is certainly true, but then again, from the perspective of CBA, anything of value can be considered an economic benefit since anything can be assigned a monetary value. Hence, Michaelowa et al. either must consider "absolute economic savings" only in terms of actual money and goods that already "come with" a market value, or their criticism is not that CBA *cannot* incorporate non-monetized benefits, but rather, that it *should* not. The first option would require some specification of which values should enter the economic benefit in question. After all, the value of real estate, for instance, is also an estimate of what others would be willing to pay for it, if it were sold. In any case, this option seems not to be what Michaelowa et al. mean, as absolute savings are supposed to include public infrastructure, "which can include natural resources and services" ([4], (p. 10)). Thus, they seem to regard some environmental impacts, such as the loss of biodiversity, as too "difficult to measure in terms of monetary wealth" ([4] (p. 9)). However, since they do not spell out what "too difficult" means, it seems to be an ad hoc decision which benefits to express in monetary terms and which to consider qualitatively.

It needs to be stressed at this point that the question of which benefits—and which costs, for that matter—to include in CBA is no mere technical issue, but a normative-ethical one ([11] (pp. 166–176)). Theoretically, CBA is supposed to be "as all-encompassing as possible" when it comes to the valuation of costs and benefits, i.e., it is meant to adopt a social perspective ([11] (p. 168); see also [12] (pp. 11, 15). Yet, since not each and every cost and benefit, including future ones, can possibly be considered, some limitation has

to be made. In practice, this restriction often is not only pragmatically but also ethically motivated. When it comes to health care, for instance, applied CBA usually disregards "the indirect benefits of increased production" stemming from improved health ([13] (p. 794). As to the choice of which costs to consider, the following example taken from Garber et al. is illustrative of its ethical dimension: "Suppose, for example, that we contemplate instituting a suicide prevention program in high school. It is highly effective and reduces teenage suicides by 50%. Students who would otherwise have died now lead lives of average length and have medical care utilization comparable to those of average persons of their age. Should the future cost of health care that they consume be counted as costs of the intervention?" ([14] (p. 45)).

I guess most of us would answer in the negative. Obviously, by excluding some values ad hoc, the value-maximizing logic of CBA is not strictly adhered to in praxis, which suggests that something may be wrong with this logic in the first place. Coming back to Michaelowa and colleagues, they seem hesitant to incorporate all environmental impacts into the CBA of adaptation projects as well. I surmise that the respective problem does not consist in the technical difficulty of measurement, though, but rather points to normative-ethical issues of CBA. Put differently, the real question is not whether all benefits *can* be assigned a monetary value but whether they *should*. In effect, Michaelowa et al. mask the normative decision not to include some adaptation benefits as a technical one.

As a second objection leveled against CBA, the authors criticize that allocating resources with the aim to maximize absolute wealth saved not necessarily addresses vulnerability. (The authors do not define "vulnerability" in more detail and neither will I. For the complex connection between wealth and vulnerability in the context of prioritizing adaptation projects, see [15] (p. 604f).) In fact, it is likely that the opposite is true, since the richer a country is, the less vulnerable to climate change its inhabitants tend to be, but the more absolute savings are possible ([5], (p. 115))— "Where more valuable objects [. . .] exist, more can be lost" [16], (p. 82)). Note that in contrast to the aforementioned criticism, this objection against CBA does not refer to its monetary measure of benefit but to the ethically undesirable consequences of allocating adaptation funding with the aim of maximizing this benefit. That is to say, it stems from concerns for distributional equity.

Before discussing Michaelowa et al.'s proposed remedy, another question presents itself: In light of their objections leveled against CBA, why do the authors not generally reject absolute wealth saved and replace it with another measure? Somewhat surprisingly, Stadelmann et al. state that "absolute economic value has its own merits: it is the usual way to measure macro-economic effectiveness, it is a standard indicator for evaluating the well-being of societies, it reflects overall utility as stated by market participants, and, finally, vulnerable people may benefit from absolute economic assets via redistribution" ([5] (p. 111)). We are thus presented with four reasons for not dropping absolute economic value. The first and the second reason given boil down to the claim that economic value is commonly used to assess aggregate well-being. Yet, the mere fact that some measure's use is widespread does not mean that using it is warranted ([16] (p. 82)). As to the third supposed merit of economic value, it remains unclear in which respect it differs from the second, for within applied welfare economics, "overall utility" is usually equated with well-being. That being said, regardless of whether "utility" is understood in terms of (some kind of substantive) well-being or of preference satisfaction, it is well established in the literature by now that there is neither a proportional relationship between GDP and well-being nor one between the satisfaction of market preferences and any kind of substantive well-being ([17], (pp. 77–87)). Michaelowa et al.'s reasons two and three thus do not stand critical scrutiny. Finally, more economic assets do indeed potentially allow for more redistribution so that the poor could indirectly benefit from absolute economic savings. Then again, this argument is reminiscent of the so-called "trickle-down theory," popular in the 1990s, according to which the wealth of a few will ultimately trickle down to the less fortunate in society—a hope that proved unwarranted. Beyond that, if the merit of considering absolute savings ultimately consists in its (*potential*) contribution

to alleviate vulnerability via (*hypothetical*) redistribution, why not focus on vulnerability directly or, at least, on more objective measures of well-being and resilience, bypassing the problems associated with CBA altogether? All in all, the arguments for considering aggregate absolute economic value even as a part of the equation remain unconvincing.

2.2. Relative Wealth

Instead of dropping the metric of absolute saved wealth, Michaelowa et al. argue that the trade-off between saving wealth and tackling vulnerability can be eased by combining absolute with *relative* wealth saved, which is the absolute wealth saved by a project divided by the total wealth of the respective community ([4] (p. 10)). Since the same amount of monetary benefits saved constitutes a larger fraction of income for the poor than for the rich, aiming at saving relative wealth favors poorer regions, they argue, so that it indirectly considers vulnerability. Stadelmann et al. differentiate between an individual-level and an aggregate assessment of relative wealth saved ([6] (p. 2150)). While applying the former would be preferable, lack of data and privacy protection render this option infeasible, so they propose using the relative wealth concept on an aggregate basis.

However, once an aggregate assessment of relative wealth saved takes place, most of the asserted advantages of this concept vanish. The claims that "the relative savings indicator measures the number of personal livelihoods that can be saved" ([5] (p. 114)) or that relative wealth saved addresses inequalities in wealth and vulnerability within communities ([4], (p. 10)), for instance, only hold if a project's impact on each individual is considered separately. Aggregated data, by contrast, do not tell anything about the distribution of the respective losses *within* the community. This is highly relevant, because "losing, say, 10% of one's income might impose serious hardship on a poor person while it does not [do so] for a very wealthy member of society" ([16] (p. 83)). Therefore, saved wealth fails to tackle social inequalities within one community.

Beyond that, the aggregated relative saved wealth indicator may lose its significance when it comes to the comparison *between* communities as well. To see this, consider the example given by Köhler and Michaelowa of a community of 1000 people and a total wealth—or GDP—of USD 1 million ([4] (p. 2)). If an adaptation measure saves this community USD 0.2 million, the relative saving amounts to 20%. To be sure, if USD 0.2 million is spared in a community with the same number of inhabitants with a GDP of USD 10 million, the relative saving is only 2% and the indicator would favor the poorer area. If, however, the richer community also stands to save more absolute wealth in virtue of the adaptation project, let us say USD 2 million, the relative savings are the same and vulnerability is not tackled at all. Aggregated sums cannot, in principle, take distribution into account.

Finally, even if an allocation of resources focusing on maximizing saved relative wealth would favor poorer communities, it is unclear to which extent alleviating poverty lessens vulnerability. After all, poverty or lack of income cannot be equated with vulnerability, since irrespective of its specific definition, the latter certainly also depends on other factors, such as "the extent of people's dependency on risky activities and sources of income such as agriculture and fishing" ([15], (p. 605)). Thus, whether SW accounts for vulnerability at all depends on a project's specific context, a fact that renders its usefulness as a *universal* adaptation metric questionable. (Note that the fact that the climate change vulnerability of a certain region is by definition geo-localized does not in general undermine the search for universal metrics. For once a measure for vulnerability was defined, this could be applied to different adaptation projects across different regions. I am grateful to an anonymous reviewer for raising this point.)

In sum, the SW indicator is not successful in addressing the first two objections leveled against CBA by Michaelowa et al. The following section analyzes whether the DALY provides for a satisfying solution to their third issue with CBA: the controversial matter of assigning a monetary value to human life ([6] (p. 2149)).

3. Saved Health: The DALY

3.1. Saved Health and Cost-Effectiveness Analysis

Incorporating the prevention of ill-health and premature death into a climate change adaptation measure is certainly plausible (for the critical relevance of public health data for the (public) debate on climate change, see [9]). The question is how to define and measure ill-health, though. In this respect, Michaelowa and colleagues endorse the DALY. The justification given for this theoretical choice mainly refers to the drawbacks of other approaches and the public rejection of attaching monetary values to human life. Traditional methods to do so make the value of a human life contingent upon the respective person's income. The human capital approach, for instance, determines "the pay-back" on the "investment" in human capital as the present value of the individual's future earnings, using market wage rates ([18] (p. 215f.)). This is especially troublesome when it comes to the comparison of data for industrialized and developing countries, which is common practice in the discourse on climate change ([4] (p. 3)). Michaelowa and colleagues reckon that since monetary valuation of life is "fraught with ethical and political challenges" ([7] (p. 67)), it "should be avoided, especially if comparing industrialized and developing countries" ([6] (p. 2152)). To escape the "endless political debates about an equitable valuation of human life and death" ([6] (p. 2149) it would be critical "to have a non-monetary indicator that addresses the health benefits of adaptation projects" ([7] (p. 67); see also [4] (p. 9)).

Considering non-monetary benefits, Michaelowa and colleagues shift from CBA to a form of CEA, which defines the benefit measure in terms of some unit other than money, usually some physical target ([12] (p. 16)). In health care, this could be reduction in blood pressure in mmHg. Although such physical measures appear straightforward and easy to understand, they obviously do not allow for comparing the efficiency of projects with different goals. At this point, *generic* measures of health, such as DALYs or quality-adjusted life years (QALYs), enter the scene, which allow to capture both premature mortality and the reduced quality of life due to ill-health. The DALY does so by combing the number of *years of life lost* (YLL) with the number of *years lived with disability* (YLD). The term "disability" is used in a broad sense here and denotes any short- or long-term loss of health ([19] (p. 2198), [20] (p. 2130)). To measure the YLDs, each year lived with a disability is attached a disability weight, where 0 represents perfect health and 1 expresses death or a condition equivalent to death. Being a measure of *ill*-health and life *lost*, the DALY is not a "good" to be saved or maximized but a "bad" to be minimized ([21] (p. 307)). (The terminology "disability-adjusted life-year" is misleading, since, to speak with Anand/Hanson [21] (p. 310), "more of a 'life-year' (even adjusted for disability) is normally understood as a 'good', which should be maximized and not minimized." Accordingly, there exists some confusion on this matter in the literature on climate change as well, for instance, when Köhler and Michaelowa speak of the "the concept of Disability Adjusted Life Years (DALYs) saved" ([4] (p. 3); see also [5] (p. 111)). In an otherwise illuminating paper, Nolt describes the DALY consistently wrong when he claims that "[o]ne DALY is one year of healthy life. [. . .] [E]ach disability is assigned a value between 0 (death) and health (1), lower numbers representing greater severity" and so forth ([9] (p. 351)). What he says is right—though not with respect to the DALY but for its close relative, the QALY.).

Michaelowa and colleagues hardly discuss the DALY, but their reasons for using it quoted above suggest that the authors consider the DALY as *not* being fraught with ethical challenges and as *not* prompting political debate. Due to the DALY's lack of transparency, the latter claim may indeed be correct ([9] (p. 351)), but the former is far from true. To see this, a closer examination of the DALY and its measurement is illuminating.

The DALY was developed during the first *Global Burden of Disease* (GBD) study 1990, launched inter alia by the World Health Organization (WHO), with the purpose of estimating the global burden of health loss due to diseases, injuries, and risk factors (such as tobacco use or high blood pressure) differentiated by age, sex, and geographical region. Adjacent to providing a unit of measurement for monitoring the global burden of disease, the DALY is also intended to serve as an outcome measure within CEA ([22], (p. 704)).

It seems fair to say that the DALY is for public health what a universal metric would be for climate change adaptation. Since the beginning, however, the DALY has ignited a critical debate on its theoretical and methodological foundations and has undergone crucial modifications [21–29]. In the following text, the debates on the central issue of what the DALY is supposed to measure in the first place will be illustrated by considering three important steps of the DALY's development within the GBD framework. The purpose of this endeavor is twofold: for one thing, the analysis supports the thesis that the DALY incorporates lots of controversial normative assumptions. For another, it will become clear that the construction of the DALY as a descriptive measure of ill-health is already heavily influenced by distributional considerations—just as the SW metric proposed by Michaelowa et al. is.

For the measurement of disability weights, preliminary considerations on the DALY proposed six disability classes describing some condition in general terms, an example being the following: "[L]imited ability to perform activities in two or more of the following areas: recreation, education, procreation or occupation" (class 3) or "[n]eeds assistance with activities of daily living such as eating, personal hygiene or toilet use" (class 6) ([23] (p. 438, Table 2)). These classes were evaluated by a group of medical experts by means of a method called magnitude estimation ([23], (p. 439)), which asks the respondents a question of the form "How many times worse is one state than another [reference] state?" ([30] (p. 16)). In this way, class 3 was assigned a weight of 0.400 and class 6 a weight of 0.920. Each disability class was supposed to represent "a greater loss of welfare or increased severity than the class before," and the weights were intended to measure a disability's "impact on the individual" ([23] (p. 438)). However, the disability weights were not supposed to measure the myriad ways in which a health state affects individual well-being (the intricate relationship between "health" and "well-being" and the question of what should be tackled by means of summary measures of population health has been discussed extensively by Hausman [17]), for such an account would have to take into account the social context of the respective individual ([23] (p. 437f.)). Instead of tackling the specific *disadvantage* caused by a health state, the disability classes and, hence, the weights were supposed to represent the *disability* in terms of human functioning ([23] (p. 438)).

This individualist perspective was thwarted by two other normative choices, though. First, to determine the YLL, the designers of the DALY applied a maximum life expectancy of 82.5 years for women, which equaled the life expectancy for women in Japan, the country with the highest life expectancy worldwide ([22] (p. 711)). For men, the maximum life expectancy was set to 80 years, where the divergence was supposed to mirror different life expectancies due to biological factors, not lifestyle choices ([22] (p. 711)). A standardized life expectancy was applied globally because it should not be considered "more important to save the life of a person in a rich country (with greater life expectancy) than to save the life of someone in a poor country" ([28] (p. 201)). Both choices can be criticized. For one thing, and as pointed out by Lyttkens (ibid.), if "biologically" different life expectancies are taken into account in the case of sex, why not also consider such differences between other social groups, once they are detected? As to the application of a universal life expectancy, it can be argued that it is beside the point to claim that a man living in Sierra Leone and dying at the age of 30 due to a certain disease loses 50 DALYs *attributable to the respective disease*, while his life expectancy had only been 38 years anyway ([26] (p. 5)). (The 38 refers to 1997, the year Murray and Acharya's paper was published. In 2018, according to the World Bank, the life expectancy in Sierra Leone was 54.309 years. See https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=SL, accessed on 24 August 2020.) While driven by plausible distributional considerations, the use of a universal life expectancy is thus problematic when it comes to the descriptive measurement of the burden of disease due to specific diseases or risk factors.

The second normative choice of concern here refers to differential valuations of the DALY prohibited at different ages, the value being highest for young adults around the age of 25 and lowest for both children under 10 and the elderly over 60 years ([23] (p. 436)).

The reason for attaching age weights to DALYs is stated by Murray as follows: "Higher weights at a particular age does [sic] not mean that the time lived at that age is per se more important to the individual, but because of social roles the social value of that time may be greater" ([23] (p. 435)). Young adults, so the idea, play an important role "in providing for the well-being of others" (ibid., see also [22] (p. 718)). That is to say, the well-being of the most vulnerable is supposed to be accounted for by attaching a higher value to the health of those who care for them. (Since I suppose that discussions about discounting future benefits are well known in the adaptation literature, I do not elaborate upon discounting future DALYs here. See [26] (pp. 695–98) and [29].) This is highly problematic, though, especially for two reasons. First, this reasoning opens Pandora's box, as Lyttkens ([28], (p. 200)) put it, since it implies that a higher value would have to be attached to the health of doctors, nurses, or other persons providing care and crucial services for others as well, whereas the health of persons without children or elderly dependents to take care of would have to be assigned a lower value ([25], (p. 692)). The same would be true for chronically ill or disabled adults who require care. This seems to be an ethically unacceptable consequence. Second, age weighting based on social considerations, along with using universal life expectancies, introduces distributive considerations as to who should receive priority when it comes to treatment into the DALY, so the resulting number cannot be regarded a purely *descriptive* measure of ill-health.

For the final GBD 1990, some aspects of the DALY were reconsidered ([22,24]). While age weighting and a universally high life expectancy were adhered to, the question of what to measure was now answered unambiguously regarding a *social* evaluation of health: "It can be argued that for burden of disease [. . .] and cost-effectiveness analyses that are intended to inform social choices, a method that directly measures social preferences for health states would be more appropriate than one that measures individual preferences" ([22], p. 713). In this context, social preferences are preferences individuals have not with regard to their own health but concerning the distribution of health or health care on other persons, not including themselves ([31], p. 26).

These social preferences were elicited by means of the person trade-off (PTO) method from a group of health care providers from each region of the world convened at the WHO in 1995 ([22] (pp. 713, 715)). The respondents were confronted with two version of the PTO. In the first version, PTO1, they were asked to compare life extension for a healthy individual with life extension for a person with a disability: "[W]ould you as decision maker prefer to purchase, through a health intervention, 1 year of life extension for 1,000 perfectly healthy individuals or 2000 blind individuals?" ([22] (p. 714)). In a series of such questions, it was elicited how large the number of blind persons must be so that the respondent is indifferent between the two scenarios. If the number is, say, 8000, the disability weight would amount to 1 minus 1000 divided by 8000, that is, 0.875 ([32] (p. 1424)). In the second version of PTO, PTO2, the respondents were asked to compare the value of curing a certain number of individuals with disabilities on the one hand with life extension for a certain number of healthy persons on the other: "[H]ow many people cured of blindness do you consider equal to prolonging the lives of 1000 healthy people?" ([32] (p. 1424)) If the disability weights derived from PTO1 were inconsistent with those inferred from PTO2, the respondents were instructed to reconcile their answers ([24] (p. 36)).

However, it is questionable whether different weights resulting from the two PTOs present an inconsistency in the first place, since they ask for quite different things. Consider PTO1 first: the question assumes that a life-year gained for (or better: in) a disabled person is of less value than a life-year gained in an otherwise healthy person. Yet, the respondents might believe that both groups of persons have equal claims to the life-saving procedure ([32] (p. 1424)). In terms of value, this means that they regard prolonging the life of 1000 healthy persons and prolonging the life of 1000 disabled persons as equivalent. This, however, implies a disability weight of zero, which means that a blind person counts as perfectly healthy and has no claim to health care when it comes to resource allocation. This is why such results were considered irrational. The problem here, again, is that the

task of descriptively evaluating how bad a health state is *for* the respective individual is meshed up with distributional considerations about resource allocation. Note that at this point, we are back to the very problem of valuing human lives, which Michaelowa and colleagues were so eager to evade by using the DALY in the first place.

PTO2, by contrast, does not entail any devaluation of life-years gained by a certain group of patients and might therefore lead to a different disability weight for the health state at stake. In effect, the respondents might interpret the PTO1 scenario as an issue of distributive justice, that is, social value, whereas they might read PTO2 as asking for their evaluation of the health state in question, i.e., individual value (see also [28] (p. 197)). Asking the respondents to resolve the apparent inconsistency not only forces them to attach less weight to life-years gained for the disabled but also renders the resulting disability weights meaningless ([32] (p. 1424), [27] (p. 121)).

In the course of the GBD 2010, the DALY's conceptual, normative, and methodological foundations have been substantively revisited ([19]). To begin with, the GBD now decidedly seeks "to quantify health loss rather than welfare loss" ([20] (p. 2130)) so that the disability weights are taken to "reflect the general population judgment about the 'healthfulness' of defined states, not any judgments of quality of life or the worth of persons or the social undesirability or stigma of health states" ([33] (p. 12)). Accordingly, the age weights have been dropped, whereas the application of a universal life expectancy is still adhered to ([19] (p. 2199)). (Life expectancy is no longer differentiated between men and women, though, and YLL are calculated using a new reference-standard life expectancy at each age.) For measuring the "healthfulness" of different health states, two new methods were developed, and the respective questions were answered by a sample of the public from all over the world. To be able to discuss what the methods elicit, the questions need to be considered at length. The first method, a paired comparison, looks as follows:

> "Now, we want to learn how people compare different health problems. [. . .] I will ask you to tell me which person you think is healthier overall, in terms of having fewer physical or mental limitations on what they can do in life. [. . .] There are no right or wrong answers to these questions. Instead, we are interested in finding out your personal views.
>
> The first person [[. . .] has mild tremors and moves a little slowly, but is able to walk and do daily activities without assistance.
>
> The second person [[. . .] has some trouble remembering recent events, and finds it hard to concentrate and make decisions and plans.
>
> Who do you think is healthier overall, the first person or the second person?" ([34] (p. 2)).

While this type of question focuses on chronic disabilities, the following, so-called population health equivalence method, draws on the PTO and seeks to elicit trade-offs between mortality and non-fatal health states:

> "The last questions will ask you to compare the overall health benefits produced by two different programs. Imagine there were two different health programs.
>
> The first program prevented 1000 people from getting an illness that causes rapid death.
>
> "The second program prevented [*Number selected randomly from {1500, 2000, 3000, 5000, 10 000}]* people from getting an illness that is not fatal but causes the following lifelong health problems: [[. . .] Some difficulty in moving around, and in using the hands for lifting and holding things, dressing and grooming.
>
> Which program would you say produced the greater overall population health benefit?" ([34] (p. 3))

Whereas the first question asks the respondents whom of two patients they consider "healthier overall," the latter asks them for the program producing "the greater overall pop-

ulation health benefit" (ibid.). This way of eliciting disability weights is highly problematic, though, and there are three reasons why.

First, regarding the population health equivalent, the authors stress that they do not want to mesh up the measurement of health on the one hand with distributional issues on the other when they state, "In keeping with the focus in the GBD 2010 on the construct of health loss, we explicitly avoided framing of the question in terms of resource allocation decisions, as this framing may evoke distributional concerns that are orthogonal to the health construct" (ibid.). While it is true that they do not explicitly ask the respondents to make a (hypothetically) distributional decision, neither do they ask them to imagine themselves in the respective health state. Instead, they choose a consequentialist account that asks for some abstract sum of health aggregated above all persons concerned. In doing so, the question presumes that the respondents are able and willing to evaluate abstract units of health, add these values up, and draw a balance sheet across the patients concerned. Yet, this consequentialist manner does not seem to be the most natural way to understand the population health equivalent. Most likely, the respondents do not think of the "amount" of health produced by each program, but rather regard it as a distributional task, which makes them consider which program should be realized and, thereby, which group of patients should be treated.

Second, without considering the impact of health on well-being or quality of life, it is conceptually unclear what it means to say that one of two persons is "healthier overall." Health is a multidimensional concept, and different health states imply quite different limitations on what one can do. These differences cannot simply be measured in terms of "more" or "less." To see this, try to figure out who is healthier in terms of facing fewer limitations on what he or she can do: (i) a person sitting in a wheelchair, (ii) a blind person, (iii) a person suffering severe migraine attacks two or three times each month, or (iv) a person with arthrosis and constant pain in the joints. Considering this question, two conclusions can be drawn. First, without balancing the different dimensions of health somehow, it is impossible to tell who is healthier as such. In addition, while it is possible to say whether one person is healthier than another as long as only one dimension is at stake (e.g., a person blind in one eye is "healthier" than a person blind in two eyes), the relevance of such comparison is rather limited. This is because, second, the severity of the limitations associated with a health state depends on the latter's effect on the person's quality of life and socially available activities and on an assessment of the relative importance of these activities ([17] (pp. 54f.)). Hence, health as such cannot be *quantified* but must be *valued* ([17], p. 42).

Third, and intricately connected with the aforementioned issue, it is impossible to separate the valuation of health from the social context. This claim can be substantiated by considering an example taken from Voigt and King [35]: Imagine two countries with equal numbers of persons with impaired vision. In one country, corrective lenses are available, so the persons concerned hardly suffer any negative consequences from their condition at all. In the other country, by contrast, there are no corrective lenses and persons with impaired vision have difficulties finding jobs, gaining a sufficient income, and so forth. The designer of the GBD 2010 in effect argue that although the impact of impaired vision on well-being may vary between those countries, the health state as such does not, and this is the invariant construct they want to measure. However, even if one is willing to consider the persons in both countries as equally healthy, when it comes to allocating resources, it seems reasonable to say that it should make a difference of how their health state actually affects their life ([35] (p. 227)).

To conclude, measuring units of health as such is impossible and asking the respondents to compare programs in terms of the overall population health benefit presupposes a framework they are not accustomed with. While I argued that the respondents probably understand the population health equivalent as a distributional choice, the meaning of their answers to the paired comparisons and, thus, the respective disability weights remains obscure ([36], [17] (p. 42)).

3.2. Implications for SH

To sum it up until this point, the previous subsection illustrated that the DALY's development has been accompanied by critical debates on and modifications of its conceptual and normative foundations. The designers of the GBD 2010 tried to react to criticism leveled against the DALY, but their proposals remain unconvincing. The contentious normative choices incorporated in the DALY and the problematic methods to elicit disability weights not only call into question what the DALY measures, especially whether it tackles the individual value of health or distributional goals, but also reveal that it is by no means an ethically innocuous measure of health. Instead of enhancing the transparency of the decision-making process, such a heterogeneous and highly aggregated measure in fact diminishes intelligibility by hiding controversial value judgments ([29]). This is a highly relevant insight when it comes to the SH indicator, for the DALY obviously does not live up to the task of avoiding ethical challenges, as Michaelowa and colleagues assumed it would. In addition to this fundamental issue, I want to point out some further problems of their use of the DALY as an SH indicator and, in doing so, draw some general conclusions as to the DALY's usefulness.

First, Stadelmann et al. take the DALY to measure human lives savings due to some adaptation project ([5] (p. 111)). As we saw, however, the DALY is a product of YLL and YLD, and it is impossible to know whether saving some number of DALYs means saving *any* life at all. The number could as well represent the prevention of a lot of minor disabilities. This ambiguity of the DALY raises the question of whether death and non-fatal health states should be compounded within one number in the first place ([25] (p. 689)), for it arguably makes a difference whether a project saves lives or prevents deteriorations in health.

For a similar reason, second, the claim that SH is a proxy for reduced vulnerability ([5] (p. 111) is not necessarily true. Around 1,300,000 DALYs could be saved by curing cellulitis ([19] (p. 2208, Table 1)), but would this reduce vulnerability? Allocating adaptation money strictly so as to maximize the sum of DALYs prohibited may well lead to systematically disadvantaging those who are already worse off ([15]). These first two aspects together call for a disaggregation of the data and the value judgments incorporated in the DALY in order to enhance the transparency of the decision-making process ([29]).

Third, Michaelowa and colleagues search for a measure that does not put different values on the lives of rich and poor persons (see e.g., [5] (p. 111)). When it comes to calculating the YLL, Stadelmann et al. point out that the choice of reference region is "an ethical issue": "From an equity viewpoint, the average global life expectancy should be chosen, whereas from a comparative economic view, the locally applicable life expectancy would be preferred" ([6] (p. 2153)). Why the first alternative would be a demand of equity and the second one of economics remains unclear, though. In their case study in Vietnam, Köhler and Michaelowa [4] choose the local life expectancy of Vietnam (74.2 years) to calculate the YLL. Since the respective CEA concerns two different projects in the same region, this choice does not really matter, but it would be interesting to know which life expectancies the authors would propose for international comparisons.

However, even if the same life expectancy was used, fourth, the question arises as to whether the DALY with its focus on health instead of well-being actually provides a *universal* metric at all, for using the DALY as a benefit measure within CEA may hamper the equal consideration of persons living in poor and those living in rich countries in another respect. As we saw by considering the example of corrective lenses, the social context of a person crucially influences what some health state means for the patient. The context in question incorporates not only technological possibilities and medical infrastructure but also "softer" social factors such as the support ill or disabled persons receive from their community. Hence, it could be argued that a DALY saved in Germany does not represent the same *burden* of disease as a DALY saved in the Democratic Republic of the Congo. All in all, the considerations presented here show that it may be doubted whether Michaelowa et al. indeed gain much by avoiding to put a price tag on human life, because relative

evaluations of life and health happen anyway once the value-maximizing perspective of CEA is adopted (for a similar point, see [11] (pp. 176–78)).

4. Conclusions

This paper analyzed the SW and SH indicators that were recently proposed by Michaelowa et al. as universal adaptation metrics. While SW incorporates saved absolute and saved relative weights, SH is represented by the number of DALYs prohibited due to a particular adaptation project. As both measures are general and highly aggregated, they are supposed to serve both as descriptive measures to monitor and evaluate the success of adaptation projects ex post and as input into CEA to determine cost-effective projects ex ante. I argued that neither of these measures lives up to its task. As to SW, relative wealth saved would only tackle the poor if the effect of an adaptation project would be estimated for each individual separately. Once aggregate data are used, the sum of relative wealth saved is just as distributional insensitive as the sum of absolute wealth is. Beyond that, the attempt to account for vulnerability by modifying the benefit measure within CBA introduces distributional considerations of justice into the definition of the units of value. This does not only run counter to Michaelowa et al.'s asserted ethical modesty but also diminishes transparency of decision making. Finally, the arguments presented by the authors for complementing as opposed to substitute absolute savings with relative savings is not convincing. I have raised similar concerns regarding the DALY as a measure of ill-health. As the sketch of the DALY's theoretical development illustrates, it is still contested in the literature what the disability weights are supposed to represent at all. The debates around the use of a certain life expectancy and the different weightings of DALYs according to the age of the recipient reveal that the DALY also suffers from an insufficient separation between the measurement of (the value of) health on the one hand and the question of how health and health care should be prioritized and distributed on the other hand.

Hence, although highly aggregated summary measures such as the DALY are usually praised as allowing for transparent and fair decision making, they pose the danger of hiding controversial normative assumptions, treating them as mere technical issues, and thus withdrawing them from public discourse. The sum of economic savings or DALYs generated by a project, in turn, also disguises how DALYs are distributed on the persons concerned and how individuals are treated. The "ethical and political challenges" associated with measuring what is valuable for human beings in different societies and with distributing resources fairly cannot be evaded and must not be concealed.

Where does this leave us? The upshot of this paper is modest but important: designing universal adaptation metrics incorporating different adaptation aims is tremendously difficult and may not be feasible at all. To determine whether it is, further research on fundamental concepts, such as vulnerability, and on the role efficiency considerations can and should play in priority setting is warranted.

Institutional Review Board Statement: The study did not involve humans or animals.

Informed Consent Statement: The study did not involve humans. Written informed consent for publication must be obtained from participating patients who can be identified (in-cluding by the patients themselves). Please state "Written informed consent has been obtained from the patient(s) to publish this paper" if applicable.

Data Availability Statement: The study did not report any data.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Christiansen, L.; Martinez, G. Editorial. In *Adaptation Metrics: Perspectives on Measuring, Aggregating and Comparing Adaptation Results*; Christiansen, L., Martinez, G., Naswa, P., Eds.; UNEP DTU Partnership: Copenhagen, Denmark, 2018; pp. 7–13.
2. Leiter, T.; Pringle, P. Pitfalls and potential of measuring climate change adaptation through adaptation metrics. In *Adaptation metrics: Perspectives on Measuring, Aggregating and Comparing Adaptation Results*; Christiansen, L., Martinez, G., Naswa, P., Eds.; UNEP DTU Partnership: Copenhagen, Denmark, 2018; pp. 29–48.
3. Leiter, T. Assessing results of climate change adaptation projects in practice: Learning from German Technical Development Cooperation. In *Adaptation Metrics: Perspectives on Measuring, Aggregating and Comparing Adaptation Results*; Christiansen, L., Martinez, G., Naswa, P., Eds.; UNEP DTU Partnership: Copenhagen, Denmark, 2018; pp. 139–155.
4. Köhler, M.; Michaelowa, A. Saved Wealth, Saved Health: An Approach to Quantifying the Benefits of Climate Change Adaptation. GIZ, Eschenbach. 2013. Available online: https://www.adaptationcommunity.net/?wpfb_dl=139 (accessed on 20 August 2020).
5. Stadelmann, M.; Persson, A.; Ratajczak-Juszko, I.; Michaelowa, A. Equity and cost-effectiveness of multilateral adaptation finance: Are they friends or foes? *Int. Environ. Agreem.* **2014**, *14*, 101–120. [CrossRef]
6. Stadelmann, M.; Michaelowa, A.; Butzengeiger-Geyer, S.; Köhler, M. Universal metrics to compare the effectiveness of climate chance adaptation projects. In *Handbook of Climate Change Adaptation*; Leal Filho, W., Ed.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 2143–2160.
7. Michaelowa, A.; Stadelmann, P. Development of universal metrics for adaptation effectiveness. In *Adaptation Metrics: Perspectives on Measuring, Aggregating and Comparing Adaptation Results*; Christiansen, L., Martinez, G., Naswa, P., Eds.; UNEP DTU Partnership: Copenhagen, Denmark, 2018; pp. 63–72.
8. Sagoff, M. *Price, Principle, and the Environment*; Cambridge University Press: Cambridge, UK, 2004.
9. Nolt, J. Casualties as a moral measure of climate change. *Clim. Chang.* **2015**, *130*, 347–358. [CrossRef]
10. Mitchell, R.C.; Carson, R.T. *Using Surveys to Value Public Goods: The Contingent Valuation Method*; Resources for the Future: Washington, DC, USA, 1989.
11. Hansson, S.O. Philosophical problems in cost-benefit analysis. *Econ. Philos.* **2007**, *23*, 163–183. [CrossRef]
12. UNFCCC. Potential Costs and Benefits of Adaptation Options: A Review of Existing Literature. Bonn. 2010. Available online: https://unfccc.int/sites/default/files/resource/docs/2009/tp/02r01.pdf (accessed on 7 August 2020).
13. Richardson, J.; McKie, J. Economic evaluation in the context of a national health scheme: The case for an equity-based framework. *J. Health Econ.* **2007**, *26*, 785–799. [CrossRef] [PubMed]
14. Garber, A.M.; Weinstein, M.C.; Torrance, G.W.; Kamlet, M.S. Theoretical foundations of cost-effectiveness analysis. In *Cost-Effectiveness in Health and Medicine*; Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C., Eds.; Oxford University Press: Oxford, UK, 1996; pp. 25–53.
15. Paavola, J.; Adger, W.N. Fair adaptation and climate change. *Ecol. Econ.* **2006**, *56*, 594–609. [CrossRef]
16. Baatz, C.; Bourban, M. Distributing scarce adaptation finance across SIDS: Effectiveness, not efficiency. In *Dealing with Climate Change on Small Islands: Towards Effective and Sustainable Adaptation?* Klöck, C., Fink, M., Eds.; Göttingen University Press: Göttingen, Germany, 2019; pp. 77–99.
17. Hausman, D. *Valuing Health: Well-Being, Freedom, and Suffering*; Oxford University Press: Oxford, UK, 2015.
18. Drummond, M.F.; Sculpher, M.J.; Torrance, G.W.; O´Brien, B.J.; Stoddart, G.L. *Methods for the Economic Evaluation of Health Care Programmes*; Oxford University Press: Oxford, UK, 2005.
19. Murray, C.J.L.; Vos, T.; Lozano, R.; Naghavi, M.; Flaxman, A.D.; Michaud, C.; Ezzati, M.; Shibuya, K.; A Salomon, J.; Abdalla, S.; et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2013**, *380*, 2197–2223. [CrossRef]
20. Salomon, J.A.; Vos, T.; Hogan, D.R.; Gagnon, M.; Naghavi, M.; Mokdad, A.; Begum, N.; Shah, R.; Karyana, M.; Kosen, S.; et al. Common values in assessing health outcomes from disease and injury: Disability weights measurement study for the Global Burden of Disease Study. *Lancet* **2012**. [CrossRef]
21. Anand, S.; Hanson, K. DALYs: Efficiency versus equity. *World Dev.* **1998**, *26*, 307–310. [CrossRef]
22. Murray, C.J.L.; Acharya, A.K. Understanding DALYs. *J. Health Econ.* **2017**, *16*, 703–730. [CrossRef]
23. Murray, C.J.L. Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bull. World Heal. Organ.* **1994**, *72*, 429–445.
24. Murray, C.J.L. Rethinking DALYs. In *The Global Burden of Disease. A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020*; Murray, C.J., Lopez, L., Alan, D., Eds.; Harvard University Press: Cambridge, UK, 1996; pp. 1–89.
25. Anand, S.; Hanson, K. Disability-adjusted life years: A critical review. *J. Health Econ.* **1997**, *16*, 685–702. [CrossRef]
26. Williams, A. Calculating the global burden of disease: Time for a strategic reappraisal? *Health Econ.* **1999**, *8*, 1–8. [CrossRef]
27. Nygaard, E. Is it feasible or desirable to measure burdens of disease as a single number? *Reprod. Health Matters* **2000**, *8*, 117–125. [CrossRef]
28. Lyttkens, C.H. Time to disable DALYs? On the use of disability-adjusted life years in health policy. *Eur. J. Health Econ.* **2003**, *4*, 195–202. [CrossRef]
29. Arnesen, T.; Kapiriri, L. Can the value choices in DALYs influence global priority setting? *Health Policy* **2004**, *70*, 137–149. [CrossRef]

30. Richardson, J. Cost utility analysis: What should be measured? *Soc. Sci. Med.* **1994**, *39*, 7–21. [CrossRef]
31. Stouthard, M.; Essing-Bot, M.-L.; Bonsel, G.J. Disability weights for diseases. A modified protocol and results for a Western European region. *Eur. J. Public Health* **2000**, *10*, 24–30. [CrossRef]
32. Arnesen, T.; Nord, E. The value of DALY life: Problems with ethics and validity of disability adjusted life years. *BMJ* **1999**, *319*, 1423–1425. [CrossRef] [PubMed]
33. WHO. WHO Methods and Data Sources for Global Burden of Disease Estimates 2000–2015. 2017. Available online: https://www.who.int/healthinfo/global_burden_disease/GlobalDALYmethods_2000_2015.pdf (accessed on 1 March 2019).
34. Salomon, J.A.; Vos, T.; Hogan, D.R.; Gagnon, M.; Naghavi, M.; Mokdad, A.; Begum, N.; Shah, R.; Karyana, M.; Kosen, S.; et al. Supplementary appendix to: Common values in assessing health outcomes from disease and injury: Disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380*, 2129–2143. [CrossRef]
35. Voigt, K.; King, N.B. Disability weights in the global burden of disease 2010 study: Two steps forward, one step back? *Bull. World Health Organ.* **2014**, *92*, 226–228. [CrossRef]
36. Nord, E. Disability weights in the Global Burden of Disease 2010: Unclear meaning and overstatement of international agreement. *Health Policy* **2013**, *111*, 99–104. [CrossRef] [PubMed]