Environmental decision-support tools often predict a multitude of different human health effects due to environmental stressors. The accounting and aggregating of these morbidity and mortality outcomes is key to support decision making and can be accomplished by different methods that we call human health metrics. This article attempts to answer two questions: Does it matter which metric is chosen? and What are the relevant characteristics of these metrics in environmental applications? Three metrics (quality adjusted life years (QALYs), disability adjusted life years (DALYs), and willingness to pay (WTP)) have been applied to the same diverse set of health effects due to environmental impacts. In this example, the choice of metric mattered for the ranking of these environmental impacts and it was found for this example that WTP was dominated by mortality outcomes. Further, QALYs and DALYs are sensitive to mild illnesses that affect large numbers of people and the severity of these mild illnesses are difficult to assess. Eight guiding questions are provided in order to help select human health metrics for environmental decision-support tools. Since health metrics tend to follow the paradigm of utility maximization, these metrics may be supplemented with a semi-quantitative discussion of distributional and ethical aspects. Finally, the magnitude of age-dependent disutility due to mortality for both monetary and nonmonetary metrics may bear the largest practical relevance for future research.

KEY WORDS: WTP; DALYs; QALYs; life cycle impact assessment; comparative risk assessment; VSL

1. INTRODUCTION

Environmental impacts on human health are (1) significant compared to other health impacts,2 (2) considered as important as damages to ecosystems,1,2 and (3) often trigger a change of behavior and regulations.3 However, environmental impacts cause a myriad of different health effects for different durations4,5 and in environmental decisions these impacts often have to be compared with a different set of health impacts caused by competing decision alternatives (comparative risk assessment or life cycle assessment), with regulation costs (benefit-cost analysis), or impacts on ecosystems and resources (life cycle assessment). Therefore, a common metric for health outcomes that allows aggregating a wide range of different health outcomes would enable decisions that are more informed.4

In many environmental studies where human health metrics were used, the metrics have been used...

4 This is especially true if the human health endpoint, survival and cure rate, age of onset, and the duration of disability are known, i.e., a complete prediction of health profiles is possible. Epidemiological studies provide most of this data but are often unavailable for environmental risk factors. Metrics that need less data have been suggested6–8 but are not discussed here.
2. INTRODUCTION TO HUMAN HEALTH METRICS

The term “human health metrics” is used here as a label for methods that quantify and make comparable morbidity or/mortality outcomes. Because the human health impacts caused by environmental exposures may include many types of illnesses (with different symptoms and durations) and because mortality may also occur, we are interested in health metrics that allow one to combine morbidity and mortality outcomes and account for differences in duration and severity of symptoms.

Experience with health metrics in environmental decision-support tools is limited to the use of lives lost or years of life lost in early energy studies,(21) monetized estimates of health damages;(13,14,22,23) and, more recently, quality adjusted life years (QALYs)\(^{(15,24,25,26)}\) and disability adjusted life years (DALYs).\(^{(1,5,16,18,27,28)}\) All these studies used human health metrics in order to aggregate different health outcomes in one dimension so as to make them more comparable and interpretable within the respective decision-support systems. For this same reason, we focus here on one-dimensional approaches.\(^9\) For pragmatic reasons we will focus on and introduce those metrics that are already used and readily applicable.

2.1. Health Adjusted Life Years (QALYs, DALYs)

The term health adjusted life years (HALYs) is used here to indicate health metrics that transform any type of morbidity or mortality into an equivalent number of life years. Fig. 1 shows a hypothetical health profile\(^{10}\) of an individual and illustrates the two most common HALY approaches. The gray and black areas represent the quality adjusted life years (QALYs) and disability adjusted life years (DALYs), respectively. While QALYs measures the actual health quality integrated over time, DALYs measure the loss compared

---

5 Most methods used in life cycle impact assessment for the assessment of human toxicity have their roots in chemical risk assessment.\(^{(9–11)}\) The chosen metric to compare impacts from different pollutants is a derivate of the margin of safety concept (ratio of exposure increase and no-effect exposure limit). The nonoccurrence of health impacts is in this case the anchor of the health metrics scale.

6 Externality studies used methods to assign monetary values to different health outcomes because their aim was to value environmental damages in monetary units.\(^{(12–14)}\)

7 A recent example may be the comparative risk assessment study performed by U.S. EPA,\(^{(15)}\) where QALYs have been chosen to express human health impacts due to microbiological water pollution and the effects of chlorination and its side-products while another group at the RIVM, the Netherlands\(^{(16)}\) chose to use DALYs for a similar case study.

8 A broader and more detailed discussion of these issues, including a review of the literature in medical decision making and health economy, is available in Hofstetter and Hammitt.\(^{(20)}\)

9 We recognize that in many decision-making settings it may be preferable to carry several dimensions and to make the reduction to a single dimension at the latest possible time. We assume here that either the methods are used at this latest possible time or that carrying several dimensions is impractical because the analysis includes too many other dimensions such as biodiversity, ecosystem functions, and resource depletion.

10 An illustration would be jaundice at birth, a broken leg due to a skiing accident at the age of 12, a major accident with a motor bike at the age of 18, burn-out syndromes at the age of 35, heart attack at the age of 45 with almost full recovery, typical age-related morbidities between the age of 50 and 70 with a skin cancer surgery at the age of 58. Lung cancer at the age of 70 leads to death at age 72.
Fig. 1. Graphical illustration of a health profile and its measurement by quality adjusted life years (QALY, gray area) and disability adjusted life years (DALYs, black area).

to a hypothetical profile of perfect health, i.e., DALYs are the natural complement of QALYs.

The simplest description of QALY refers to an integration of health quality over time or:

\[
\text{QALY} = \Sigma (H(Q) \times \Delta t) \quad [a]
\]

where \( Q \) is a constant health state during the period \( \Delta t \). \( H(Q) \) refers to the value function of quality (we will call it quality weight). It is also common practice to include a discount factor that adjusts for individuals’ or societies’ time preference, i.e., the empirical finding that most people put a higher value on good health in the coming year than on good health 10 years from now (see also Section 4.5). Further variations and the theoretical foundations can be found in References 29 and 30.

DALYs are the sum of the years of life lost (YLL) and the years lived with disability (YLD):\(^{(31)}\)

\[
\text{DALY} = \text{YLL} + \text{YLD}
\]

\[
= \text{discounting} \times \text{age-weighting} \times (\text{SEYLL} + \text{disability weight} \times \text{disability duration}) \quad [a]
\]

The YLLs lost are calculated with the standard expected years of life lost (SEYLL). For both YLL and YLD, a continuously falling discounting function of the form of \( e^{-rt} \) is used, where \( r \) is the discount rate and \( t \) the time, in order to adjust for the time preference (Section 4.5). Age-weighting is used to adjust for age-dependent differences of how a society values life (see Section 4.4). Age-weighting in DALYs uses the expression \( C \cdot a \cdot e^{-\beta a} \) where \( C \) and \( \beta \) are constants and set equal to 0.1658 and 0.04, respectively, and \( a \) is the age in years. For YLD, similar to QALYs above, the disability weight is multiplied by the disability duration. Detailed equations have been developed for continuous\(^{(31:64ff)}\) and for discrete age of onset\(^{(32)}\). Murray et al\(^{(31)}\) allow DALYs that both use or do not use discounting and age weighting.

There are only a few key differences between QALYs and DALYs.

1. They are complements and changes in their values are inverses (\( \Delta \text{QALY} \sim - \Delta \text{DALY} \)).
2. Methods used to derive quality weights and disability weights differ (see Sections 4.2 and 4.3).
3. DALYs assume as a reference perfectly healthy individuals who would have died at their standard expected age while QALYs consider typical age-related disabilities and actual life expectancies.
4. DALYs may use an age-weighting function that is not used in QALYs (see Section 4.4).

Therefore, we use the term health adjusted life years (HALYs) as an umbrella term for both metrics.

2.2. Willingness to Pay (WTP)

Both QALYs and DALYs make the restrictive assumption of time-proportionality and use as unit of measurement a constructed index. Willingness to pay (WTP) is embedded in welfare economics and values loss in life quality in monetary units that have an external reference. Further, it makes no restrictive assumption like time-proportionality and can be written as:

\[
\text{WTP}(t) = V(\Delta Q, \Delta t) \quad [\$]
\]

where \( V \) is the value function of the health state change \( \Delta Q \) during the time interval \( \Delta t \). WTP should be understood as the rate of substitution between health and wealth. It is typically used to evaluate small changes in health states, such as those commonly anticipated to occur from small changes in environmental exposure, rather than to construct a total burden of disease (see Hammitt\(^{(30)}\) for a more detailed elaboration of the nature of WTP for mortality risk).

3. COMPARISON OF DALYS, QALYS, AND WTP BASED ON AN EXAMPLE

Before discussing the relevant characteristics of different health metrics in detail (Section 4), we consider the question of whether the choice of method for aggregation matters in practice. For this purpose,
we chose an example that covers a large set of different types of human health impacts and applied all three of the most widely used summary health metrics (DALYs, QALYs, and WTP).

Our example relies on estimates of human health impacts of environmental exposure in the Netherlands and uses data from de Hollander et al.\(^{(5)}\) For pragmatic reasons we excluded risk factors that are not strictly caused by (external) environmental pollution such as accidents, environmental tobacco smoke, or damp houses, and also exclude a large number of carcinogens that contribute little to the total health effects. The remaining five risk factors\(^{11}\) are therefore neither a complete set of all environmental health effects in the Netherlands nor necessarily the most important ones. For acute morbidity cases, incidence data with estimates for the duration of diseases have been used. We have estimated the life years lost by premature death using Dutch life tables. For chronic morbidity, prevalence data have been used (see Columns 3 and 4 in Table I). A detailed list of assumptions and references for all data used in Table I are provided in Hofstetter and Hammitt.\(^{(20)}\) Some major assumptions include:

- We provide only the central estimates, without additional information on uncertainty and variability. Therefore, no probabilistic analysis on significant differences can be made.
- Although we use incidence and prevalence data, it is assumed that all health effects occur in the same year. Therefore, time discounting was not included because it would not alter the presented results.
- The disability weights for the calculation of DALYs (Columns 5–7) have been taken from de Hollander et al.\(^{(5)}\) where 0 stands for perfect health and 1 for death.
- Age weighting, as suggested by Murray et al.\(^{(31)}\) for one version of DALYs (see Equation (2)), has not been applied in de Hollander et al.\(^{(5)}\)
- QALYs are calculated in Columns 8–10 using quality weights from various sources\(^{12}\) (perfect health = 1, death = 0). For noise effects, the effective health state of “severe annoyance” has been approximated by “anxiety” and the “sleep disturbance” approximated by “sleep disorders.” These are obviously different severity levels but are the most similar quality weights available in the literature.
- We did not account for the decreased utility of life years lost at higher ages due to comorbidities. To do so one would need the information on the age distribution of the premature deaths, which was not presented by de Hollander et al.\(^{(5)}\)
- The WTP values are effectively a mixture of WTP (based on contingent valuation or labor market studies and hedonic price methods for noise) and cost of illness (COI). This inconsistency is slightly reduced by heavily relying on one compilation of values.\(^{(23)}\) Since U.S. EPA\(^{(23)}\) uses in the baseline scenario the value of statistical life (VSL) approach\(^{(13)}\) without adjustment for age, this assumption has been adapted.

The following insights are important.

- The resulting DALYs and loss of QALYs can be compared to about 15 million Dutch inhabitants who live 15 million life years per year. Between 1.3% (DALYs) and 2.7% (QALYs) of these 15 million life years are lost due to the five environmental risk factors studied above. Melse et al. (2000) estimate the Dutch burden of disease in 1994 due to 48 selected diseases at 2.584 million DALYs. 7.3% of these DALYs can be attributed to the five risk factors analyzed in Table I. The health risk costs of 95 billion U.S.$ (almost completely intangible costs) amount to about 30% of the Dutch GDP in 1990. The magnitude of this amount suggests either that these substantial externalities are acceptable or that there are errors in the valuation of lives lost or in estimation of health impacts of exposure to particles.
- The assessment of the relative importance of noise compared to the total impact of all five environmental risk factors is very different

\(^{11}\) Long-term effects from particles smaller 10 µm (PM10), short-term effects from increased tropospheric ozone levels, impacts due to lead from drinking water pipes, traffic-related noise, and health effects due to increased UV-A and UV-B exposure caused by ozone-layer degradation.

\(^{12}\) See Hofstetter and Hammitt\(^{(20)}\) for detailed references to the used weights.

\(^{13}\) The value of a statistical life (VSL) can be calculated by workers’ willingness to accept increased job risks (wage-risk approach) or by consumers’ willingness to pay for reducing individual risks (market approach). Accepting an increase in mortality risk of 0.0001 for a compensation of $500 results in a VSL of $5 million. However, the fact that environmental risks are perceived as very different from job risks may limit the usefulness of wage-risk estimates in environmental decision-support tools.\(^{(33)}\)
<table>
<thead>
<tr>
<th>Risk Factors</th>
<th>Health Effects</th>
<th>Incidence or Prevalence</th>
<th>Cases per Year</th>
<th>Duration</th>
<th>Disability Weight [a]</th>
<th>DALYs [%]</th>
<th>DALYs [%]</th>
<th>Quality Weight</th>
<th>ΔQALY [%]</th>
<th>WTP or COI [1990$] per Case</th>
<th>WTP or COI [1990$]</th>
<th>WTP/COI [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM long term</td>
<td>Mortality total</td>
<td></td>
<td>15594</td>
<td>9.6</td>
<td>1</td>
<td>149186</td>
<td>78.76%</td>
<td>0</td>
<td>149186</td>
<td>37.10%</td>
<td>4800000</td>
<td>74851</td>
</tr>
<tr>
<td></td>
<td>Chronic respiratory</td>
<td></td>
<td>10138</td>
<td>1</td>
<td>0.17</td>
<td>1723</td>
<td>0.91%</td>
<td>0.86</td>
<td>1419</td>
<td>0.35%</td>
<td>28946</td>
<td>793</td>
</tr>
<tr>
<td></td>
<td>symptoms, children</td>
<td></td>
<td>4085</td>
<td>1</td>
<td>0.31</td>
<td>1266</td>
<td>0.67%</td>
<td>0.86</td>
<td>572</td>
<td>0.14%</td>
<td>28946</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>152176</td>
<td>80.35%</td>
<td>0.86</td>
<td>151177</td>
<td>37.59%</td>
<td>0.86</td>
<td>75260</td>
<td>79.32%</td>
<td>75263</td>
<td></td>
</tr>
<tr>
<td>O3</td>
<td>Mortality, total</td>
<td></td>
<td>3840</td>
<td>0.25</td>
<td>0.7</td>
<td>672</td>
<td>0.36%</td>
<td>0.56</td>
<td>75</td>
<td>0.02%</td>
<td>4800000</td>
<td>18432</td>
</tr>
<tr>
<td></td>
<td>Hospital admission,</td>
<td></td>
<td>4490</td>
<td>0.038</td>
<td>0.64</td>
<td>109</td>
<td>0.06%</td>
<td>0.56</td>
<td>75</td>
<td>0.02%</td>
<td>6000</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>respiratory</td>
<td></td>
<td>1300</td>
<td>0.69%</td>
<td>0.49</td>
<td>519</td>
<td>0.27%</td>
<td>0.49</td>
<td>519</td>
<td>0.13%</td>
<td>194</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>1764</td>
<td>70</td>
<td>0.06</td>
<td>7409</td>
<td>3.91%</td>
<td>0.94</td>
<td>7409</td>
<td>1.84%</td>
<td>10005</td>
<td>18</td>
</tr>
<tr>
<td>Lead (*)</td>
<td>Neurocognitive development</td>
<td></td>
<td>1767000</td>
<td>1</td>
<td>0.01</td>
<td>17670</td>
<td>9.33%</td>
<td>0.91</td>
<td>159030</td>
<td>39.54%</td>
<td>265</td>
<td>468</td>
</tr>
<tr>
<td>Noise</td>
<td>Severe annoyance</td>
<td></td>
<td>1030000</td>
<td>1</td>
<td>0.01</td>
<td>10300</td>
<td>5.44%</td>
<td>0.92</td>
<td>82400</td>
<td>20.49%</td>
<td>265</td>
<td>273</td>
</tr>
<tr>
<td></td>
<td>Sleep disturbance</td>
<td></td>
<td>3830</td>
<td>0.038</td>
<td>0.35</td>
<td>51</td>
<td>0.03%</td>
<td>0.56</td>
<td>64</td>
<td>0.02%</td>
<td>9000</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Hospital admissions IHD</td>
<td></td>
<td>40</td>
<td>0.25</td>
<td>0.7</td>
<td>7</td>
<td>0.00%</td>
<td>0</td>
<td>10</td>
<td>0.00%</td>
<td>4800000</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>Mortality IHD</td>
<td></td>
<td>28028</td>
<td>14.80%</td>
<td>241504</td>
<td>60.05%</td>
<td></td>
<td></td>
<td>968</td>
<td>1.02%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>478</td>
<td>0.25%</td>
<td>511</td>
<td>0.13%</td>
<td></td>
<td></td>
<td>109</td>
<td>0.11%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ozone depletion</td>
<td>Total (morbidity and</td>
<td></td>
<td>189390</td>
<td>1</td>
<td>402155</td>
<td>1</td>
<td></td>
<td></td>
<td>94888</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mortality)</td>
<td></td>
<td>79.35%</td>
<td></td>
<td>37.44%</td>
<td></td>
<td></td>
<td></td>
<td>98.61%</td>
<td>1.39%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Only best estimates are shown, number of given digits does not suggest that these are significant digits.

*From drinking water pipes.
between the three metrics (DALYs 15%, QALYs 60%, WTP/COI 1%). The reason is that environmental noise induces very few cases of mortality (which dominate the WTP) and that the number of concerned Dutch inhabitants is so huge that small absolute differences in quality weights (that are actually large relative differences) have major effects on the morbidity burden. A probabilistic analysis would show less difference between DALYs and QALYs because Müller-Wenk (27) reports disability weights for sleep disturbance and severe annoyance that are about five times higher than those used by de Hollander et al. (5)

Further, we used for QALY quality weights that most probably apply to more severe disturbances than those from noise. Since we neither considered age weighting for DALYs nor age-related co-morbidities for QALYs, the only difference between QALYs and DALYs originates in the difference between quality and disability weights. This means that the HALY approaches are sensitive to these weights, especially for mild illnesses.

- The share of the total health burden due to mortality impacts varies from 37% for QALYs, 79% for DALYs to 98.6% for WTP/COI. The difference between QALYs and DALYs may be biased by our assumptions on the quality weights for noise and the DALYs value may be more appropriate. Therefore, we can conclude that all health metrics are heavily influenced by mortality outcomes and that in this application WTP/COI makes the morbidity assessment irrelevant (last column Table II).
- While the rank order is stable between DALYs and QALYs (only noise gets different ranking, which may be an artifact, see above), the WTP suggests that tropospheric ozone should get high attention, whereas lead exposure from drinking water pipes appears to be a very minor problem (see Table II). This rank-order reversal is due to the dominance of mortality rates in the WTP approach.

Let us assume that the presented analysis will be used, for example, in decisions on public health investment priorities or in a trade-off analysis where a new high-efficiency particle filter increases noise levels of trucks. For such cases, we have shown that the results of the analysis may be influenced substantially by the choice of health metric. However, we have also seen that differences are largely due to the insensitivity of WTP to morbidity outcomes and the large effect of uncertainties in the assessment of mild diseases.

### 4. METHODOLOGICAL CHOICES RELEVANT FOR ENVIRONMENTAL APPLICATIONS

The three major issues that must be resolved in developing simple measures of aggregate public health impacts are:

1. How to combine morbidity and mortality impacts.
2. How to aggregate across time.
3. How to aggregate across individuals.

To answer Question 1 beyond the introduction given in Section 2 we will address the following questions: Which consequences do we include? Whose values matter? How should values and utilities be elicited? And how should premature death be measured and valued? The aggregation across time is usually accomplished by discounting future impacts. Finally, the aggregation across individuals requires tackling distributional and ethical questions.

### 4.1. Which Consequences Should We Include?

Morbidity and mortality outcomes impact the concerned individuals through reduced social,
Table III. Overview on the Costs of Morbidity

<table>
<thead>
<tr>
<th>Cost of Illness (Medical)</th>
<th>Cost of Illness (Production)</th>
<th>Cost of Averting Behavior</th>
<th>Intangible Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collectively borne</td>
<td>Treatment cost (health care, infrastructure, medication etc.)</td>
<td>Loss of production (GDP)</td>
<td>Averting expenditures (noise protection barriers, water treatment plants etc.)</td>
</tr>
<tr>
<td>Individually borne</td>
<td>Treatment cost (health insurance, medication etc.)</td>
<td>Loss of production (household income)</td>
<td>Averting expenditures (water and air filters in private homes, no outdoor sport during high ozone periods etc.)</td>
</tr>
</tbody>
</table>

Note: Adapted from Seethaler. Dark shaded indicates “included in nonmonetary health metrics”; light shaded indicates “market prices are available”; no shading indicates “usually neglected.”

4.2. Whose Values Matter?

Before we turn to the description of methods to elicit values for quality weights needed in the QALYs approach, disability weights needed in the DALYs approach, and for WTP (see Equations (1)–(3)), we need to ask whose values should be considered in those elicitation procedures? This question can be split into three subquestions: Do weights differ for different groups of individuals? Which groups should be considered because of their stakes in the decision? and Which groups are actually able to state values?

A review of 38 studies, which included groups of patients and nonpatients, to elicit quality weights found that 11 of these studies show no statistically significant differences between different groups (in many cases due to small sample sizes). Twenty-two studies reported higher patient values, two studies showed lower patient values, and three studies found contradictory results. Therefore, it may matter which study population is selected. In the course of the Global Burden of Disease Study, it was questioned whether globally universal disability weights make sense due to cultural differences in health perception and the very different consequences of disabilities. An empirical study performed in 14 different countries suggests a fairly stable rank ordering among 17 selected health conditions, with the big exception of HIV infection. These authors also find that the differences in ranking of mental versus physical conditions are larger between different groups of physicians and caregivers than between countries.

15 See Gold et al. and Weinstein et al. for guidance on which cost factors are included in the nominator and denominator.

16 According to ISO, life cycle assessment includes effects on human health, ecosystems, and natural resources. Therefore, only intangible costs would be included directly. However, environmental impacts due to treatment, production loss, and averting behavior would be considered if relevant.

17 Since the study looked at quality weights where 0 stands for dead and 1 for perfect health, higher refers here to less severe symptoms.
What are the reasons for higher quality weights of patients compared to health professionals or the public? Several explanations have been offered, including:

- The description of the disease or health condition given to the general public did not correspond with what patients actually suffer.\(^\text{40}\)
- Human beings are very flexible in adapting to new situations.\(^\text{41}\)
- Human beings tend to state relative preferences that probably compare to people of similar age or fate.\(^\text{42,43}\)
- Aversion against disability only plays in \textit{ex ante} situations; patients are in \textit{ex post} situations.
- Aversion against death (which is often used as scale end in elicitation methods) may be higher for patients because death is more real or closer.\(^\text{45}\)
- The whole meaning of quality of life is redefined for ill or disabled persons.\(^\text{19}\)

For patients’ decisions on treatment methods, most of the stated reasons for higher quality weights of patients are not just plausible but also valid, i.e., not distortions to be controlled for. In environmental decision making, the number of “cases” can be influenced, that is, how many people get asthma attacks or die prematurely. This means that aversion against the disability as shown by the public may make sense and adaptation by comparing just with people of similar age or fate may not. Health professionals, on the other hand, may have a good idea what patients are actually suffering but may have systematic biases related to their training, social status, and work experience.\(^\text{47}\) Practically speaking, reasonable weights for use in environmental decision making may lie somewhere between patients’ values and the public’s and may be similar to those given by health professionals (which commonly are between patient and public values). Policy decisions about resource allocation in the health services have a similar societal view as environmental decisions. Within this context, generic preference-weighted health-status classifications are widely used. These methods use patients or health professionals to describe the health states in multidimensional health classification instruments plus community-based information to aggregate this information (see overview in Hammitt\(^\text{50}\)).

4.3. How to Elicit Values and Utilities?

To combine morbidity and mortality impacts, equivalency factors are needed that allow one to combine (1) different morbidity impacts, (2) different mortality impacts, and (3) the morbidity with mortality impacts. Here we will shortly introduce methods that deal with (1) and (3). The combination of different mortality impacts is discussed in Section 4.4. These equivalency factors are sometimes called preferences, values, or utilities, but the use of these terms is not uniform. Here we use “preferences” as the most general term that does not imply certain scale characteristics or other properties. “Values” are “preferences” measured on a numerical scale and “utilities” denote “values” under risk that fulfill the requirements by von Neumann and Morgenstern.\(^\text{48}\)

The following short descriptions describe prototypical versions of common elicitation methods.\(^\text{see also 20,31,49,50}\)

- **Standard Gamble (SG).** A subject is offered a choice between two alternatives. Alternative 1 is a treatment with two possible outcomes: probability \(p\) of being restored to normal health and living another \(t\) years, and probability \((1 - p)\) of dying immediately. Alternative 2 is the certain outcome of living in a given specified state of health \(i\) for \(r\) years. The probability \(p\) is varied until the respondent is indifferent between the two alternatives. The probability \(p\) at the point of indifference is the utility weight for the specified state of health \(i\). This method provides utilities that conform with von Neumann and Morgenstern\(^\text{48}\) requirements for decisions under risk. For mild illnesses the expected probabilities \(p\) can get very close to one. Since human beings have difficulties in dealing with probabilities near zero and one, it has been suggested to transform elicited probabilities\(^\text{51,52}\) by using probability-weighting functions such as those employed in cumulative prospect theory\(^\text{53}\) to adjust for the observed overestimation of low probabilities and underestimation of high probabilities.
• **Time Tradeoff** (TTO). A subject is offered two alternatives. Alternative 1 is to live in a specified state of health $i$ for $t$ years followed by death and Alternative 2 is normal health for $x$ years. $x$ is varied until the respondent is indifferent to the choice between the two alternatives, at which point the preference weight for the specified state of health $i$ is $x/t$. This method has been widely used as it is less demanding than SG and does not suffer from the difficulties of deriving probabilities. Nevertheless, whether TTO works for minor health impairments is questioned because people have proven unwilling to trade life expectancy for minor disabilities.

• **Person Tradeoff** (PTO). A subject is offered two alternatives. Alternative 1 is to extend life for $x$ individuals in normal health and Alternative 2 is to extend life for $y$ individuals in a specified state of health $i$. $y$ is varied until the respondent is indifferent to the choice between the two alternatives, at which point the preference for the specified state of health $i$ is $x/y$.

• **Contingent Valuation** (CV) (stated preferences). Subjects can be asked in at least four different ways to estimate their willingness to pay (WTP) to avoid or willingness to accept (WTA) certain health states. One can then estimate the amount that individuals would agree to (1) pay to reach a better health state, or (2) accept to prevent a worse health state from occurring. Or, one can determine the payment they would accept in order (3) to give up the opportunity for achieving an improvement in their health, or (4) to accept a further decline in their health state. A special property of CV methods is that respondents are not only asked to weight different health states but also to relate these weights to a (health-external) monetary unit.

• **Wage-Risk Method**, **Household Production Function Method**, **Hedonic Price Method** (revealed preferences). Instead of asking people hypothetical questions, one can observe their behavior, that is, their willingness to accept increased job risks (wage-risk approach) or their willingness to pay for reducing individual risks (market approach). A major assumption in these studies is that workers and consumers actually know the change in individual risk as a consequence of their market behavior.

In the first three methods, subjects are faced with a choice between pairs of conditions. The question is: How much are you willing to sacrifice of certainty (SG), life span (TTO), and health of others (PTO), respectively, in order to improve your own quality of life (SG&TTO) or that of an imaginary patient (PTO)? Due to these different questions, it is not surprising that the derived quality weights differ for the same person and health condition if different elicitation methods are used. Although some of these differences are intended, the anchoring and unintended framing effects lead some to believe that individual preferences for health are not preexistent but constructed during the task.

It appears that especially the tradeoff between (mild) morbidity outcomes and death poses major elicitation problems. The lack of everyday experience with such questions explains the difficulty of preference formation for most people. To address this issue, Pinto Prades suggests a different version of the person tradeoff method where a number of health states are used that are close together and build up a chain from mild disabilities to severe disabilities to death. He argues that it is easier for respondents to make tradeoffs between severe illnesses and premature death. A chained approach has also been suggested for contingent valuation. First one elicits the WTP for the certainty of a complete cure from, for example, a road injury, and then a standard gamble question elicits the injury’s severity compared to death.

Many criteria have been proposed to judge the different elicitation methods, and the recommendations show a broad variety. Nord (1992) mentions that there are three reasons the different experts do not agree on the “best” method. First, they do not take into account that there are different versions of each method; second, they do not differentiate between the different applications; and third, they do not differentiate between utilitarian and preference interpretation of the outcomes of the methods. Section 5 will focus on the specific applications in environmental decision-support tools and give guidance based on application-specific criteria.

### 4.4. How to Measure and Value Premature Death?

To combine different mortality outcomes, at least two questions need to be answered: How many years of life are lost? and Is each of these years of equal value? We will discuss these questions by taking a statistical and theoretical, rather than individual, perspective. This is because we usually are unable to
identify the affected individuals in environmental decision making.

From a statistical perspective all deaths are premature because life expectancy is always positive. Life expectancy tables can be used to calculate how prematurely somebody died. Such tables need to be (1) valid for states, nations, ethnic groups, continents, or be world-averages, (2) either averages for all individuals in the chosen area, or differentiated by sex, lifestyle factors, profession, etc., and (3) either based on today’s death statistics alone (by calculating cohort life expectancies assuming that a child born today will be at each age subject to the currently observed age-specific mortality rates) or by estimating future age-specific mortality rates that will apply when the subject cohort reaches those ages. Therefore, the question: “By how many years is a death premature?” is far from trivial.

The Global Burden of Disease Study attempts to estimate years of life lost on a globally comparable level\(^\text{[31]}\) and therefore chose a worldwide standard expected years of life lost that differentiates only between age and sex.\(^\text{[20]}\) National burden of disease studies have used national rather than global life tables.\(^\text{[65,66]}\) Risk assessment and life cycle assessment often assess changes in health states due to specific interventions. In these cases, nonaffected risk factors are assumed to stay constant and no assumptions on “genetically-based” life tables are necessary. However, since many health impacts due to environmental pollution are global and may concern future generations (when sex gap and inequalities in life expectancy may be different), the approach by Murray et al.\(^\text{[31]}\) may serve as a prototype.

Is each life year of equal value? QALY- and DALY-based approaches assume that the value of one life year depends only on its health state. On the other hand, WTP estimates typically assume a constant value of fatality risk that is independent of years of life lost.\(^\text{[14,23,67]}\) Obviously, which of these assumptions is better is of great importance for the development of health metrics. Here we review the theory and evidence about this important issue.

Theory suggests that the value of a life first increases with age and then decreases and is strongly dependent on the individual’s income (see Fig. 2). These findings are based on a simple consumption model that, for example, excludes dependents.\(^\text{[68]}\) However, the shape of this curve depends strongly on the discount rate compared to the interest rate and the possibility to borrow money.\(^\text{[33,69]}\)

Empirical evidence is still weak and sometimes contradictory. Survey studies show that saving 85 70 year olds in the United States and 35 70 year olds in Sweden is equivalent to saving one 30 year old.\(^\text{[70]}\) This is strong evidence against a constant VSL, but it is also evidence against the assumptions underlying Equation (1), if typical age-specific health states and life expectancies are assumed. An inverse U-shaped

---

20 The used model modifies the present life table of Japanese women, who have the highest life expectancy worldwide.
The VSL varies only a factor of 1.5 between a 30 and 70 year old and is therefore closer to the predictions by Shepard et al. However, the authors speculate that embedding and anchoring may have affected their results.

Within the context of environmental decision-support tools, we are usually interested in social time preferences. In this setting, the first argument for time preference—risk of death—would be translated to risk of extinction, which is very small. Pure myopia would not be considered in a prescriptive tool that is concerned with intergenerational equity. This leaves the argument of decreasing marginal utility of health. Since health is generally measured per capita and not in number of individuals, the growth in health is best reflected by increasing life expectancy and its adjustment for health state (health adjusted life expectancy (HALE), see, e.g., Murray et al. (31) for disability adjusted life expectancy). While this growth in HALE can be measured, there is less known on the marginal utility of this growth. Since we are not aware that any study that deals with environmentally-caused health effects considers the growth in HALE for future effects, there is no decrease in marginal utility that would need to be accounted for by discounting.

So far we have argued within a closed nonmone tary health market and we found that no discounting is justified, at least so long as increases in HALE are neglected. However, restricting attention to a closed health market is generally unrealistic, since both individuals and societies can shift the availability of market goods through time (by savings and investment). Given this, a second school of thought claims that the opportunity costs should determine the discount rate. To illustrate, let us assume that there is a pill on the market that sells at a real cost of $100 and improves your health for the month after taking it from the state “good” to “very good.” Investing the $100 divided by one plus the market interest rate (e.g., $97) now will return $100 in a year, which can then be spent to buy the pill and experience the health benefit. Thus, a one-month improvement in health next year can be purchased by investing $97 this year.

Since the health gain stays the same in physical terms, the cost effectiveness of the pill will improve the longer you wait. Based on the same argument, a health plan may delay the inclusion of this pill in the covered part of its services. More generally, delaying...
investments in health may improve the cost effectiveness of many investments. To avoid this situation, Weinstein et al.\(^{(73)}\) suggest that the marginal internal rate of return that could be achieved by investing in alternative projects by the same actor should be used as the discount rate. Gold et al.\(^{(35)}\) suggest using the same discount rate for costs and health outcomes and applying a social discount rate.

The opportunity cost argument is only correct if the rate at which money can be transformed into health is constant (e.g., the cost and efficacy of the pill remain constant) and the relative social benefit of monetary and health increments remain constant (e.g., the monetary value of health does not change).\(^{(76)}\) Otherwise, different discount rates for costs and health may well make sense. In our example, the real cost of the pill might increase or decrease next year, altering the amount that would need to be invested now to purchase it then. Alternatively, one might prefer to enjoy the health increment now rather than next year, and be willing to spend the additional $3 (=$100 – $97) to get it now rather than next year. There is no reason to assume that the value of one HAL Y or one statistical life stays the same while real income increases.\(^{(77)}\) In short, it appears that the monetary value of health should be discounted at the market interest rate; if the value of health changes over time, the rate at which health should be discounted differs from the market rate.\(^{(78,79)}\) Therefore, we conclude that the literature has not adequately considered the question of by how much the value of a HAL Y or statistical life is changing over time. Once this value increase is considered, discounting can be applied.\(^{23}\) Available empirical evidence does not yet allow us to suggest correction functions for future values of HAL Ys or statistical life.\(^{24}\)

Therefore, we suggest the following discounting practice.

1. If health is measured as utility and one HALY stays equally valuable independent of its timing, then these HALYs are discounted at a social discount rate.
2. If the monetary value of health is measured, the following distinction is needed:
   a. If future increases in the value of HALYs or statistical life have been included in the analysis, the marginal internal rate of return that could be achieved by investing in alternative projects should be used as discount rate. For societal decision making this rate may be approximated by a social discount rate.
   b. If future increases in the value of HALYs and statistical life have been omitted in the analysis, one should discount by the difference between the (unknown) rate of value increase of HALYs and statistical life and the social discount rate. Absent other information, this net rate may be approximated by zero.

### 4.6. How to Aggregate Across Individuals?

In environmental applications, it is the social, rather than individual, welfare that must be optimized. Choices that affect groups of people are inherently more complicated than those that affect an individual because social choices can affect the distribution of consequences across people.

The neoclassical approach in economics suggests that the social welfare function should be an aggregate of individual preferences. This means that individuals are the best judges of their own welfare (consumer sovereignty), that individuals can choose rationally among options (utility maximization), that only the outcome matters (consequentialism), and that the value of any situation should be judged solely on the basis of the utility levels attained (welfarism). Neoclassical economics often assumes that it is not possible to make interpersonal utility comparisons, that is, it is not possible to say whether one individual gains more or less than another from an increase in health or wealth. Without interpersonal utility comparisons, it is possible to say that a Pareto improvement (a change that benefits some people and harms no one) improves social welfare, but one cannot say whether changes that benefit some people but harm others improve welfare.

---

23 Johannesson et al.\(^{(71)}\) find an average marginal rate of time preference for health of about 1%. Murray et al.\(^{(31)}\) and Gold et al.\(^{(35)}\) suggest both a social rate of time discounting of 3%. Others suggest using a market rate to discount the near future but a minimal discount rate for the distant future if the appropriate discount rate is uncertain.\(^{(80)}\) Since environmental decisions may have health effects in the distant future (e.g., climate change), it may be appropriate to discount such health outcomes at very low or zero rates.

24 Most empirical estimates suggest VSL varies less than proportionately with income, although a few comparisons between industrialized and developing countries suggest the variation may be greater than proportionate. Over a timespan of 16 years, the value of a statistical life (VSL) increased in Taiwan by a factor of 10 while the income per capita increased in the same period only by a factor of 2.5.\(^{(81)}\)
In benefit-cost analysis, the interpersonal utility comparison problem is “solved” by measuring all gains and losses in monetary terms—by the affected individuals’ willingness to pay for the gains and willingness to accept compensation for the losses—and assuming that a one-dollar gain contributes the same to social welfare regardless of who receives it, be he or she rich or poor, healthy or ill. Formally, a change that benefits some people but harms others is assumed to improve social welfare if it satisfies the “Kaldor-Hicks criterion.” This requires that those who benefit from the change could compensate (with money) those who are harmed, so that everyone benefits, either by the change or by the payment of compensation. If we follow the same rationale for QALYs and DALYs, then health benefits and harms to different people are evaluated by assuming a HYA contributes the same to social welfare, regardless of whether it goes to a rich person or a poor person, to a healthy or a sick one.

Positive and negative feedback loops in an increasingly interconnected world make it unlikely that there are measures that improve the health for some without harming others.\(^{(82)}\) Therefore, compensation of health gains and losses (monetary or HALYs) between individuals occur, even in the status quo situation. The next section will provide insight into when one HALY gain to one person may not compensate a loss to another individual.

4.7. Distributional and Ethical Considerations

A lot of research in medical ethics has analyzed whether people agree to maximize QALYs or minimize DALYS and WTP as a sole criterion for resource allocation. A number of additional optimization criteria have been suggested and many of them empirically tested. The following list covers the key deviations from pure utility maximization (see Nord\(^{(83)}\) for a more comprehensive review).

- Behind the veil of ignorance it was found that one group of people wants to improve first the situation for those patients that are in the worst health state.\(^{(84,85)}\) This is also known as the severity criterion.
- Next to pure “utility maximizers,” who accept the health metric as the only criterion, one also finds “diffusers,” who prefer to spend health care resources among all with disabilities, and “concentrators,” who prefer to spend the resources on fewer patients with visible improvements.\(^{(86,87)}\) Others call this the real-

**ization potential**, that is, that group with the larger improvement potential may (or may not) be treated first.

- A survey that asked for the most important allocation criterion for donor liver grafts shows that only 0.7% of the respondents followed a strict maximization of health outcome. All others also paid attention to age (prefer younger), cause for liver disease (treat innocent first), waiting time, and whether this would be the second transplant.\(^{(88)}\)
- The fair innings argument claims that everybody should enjoy the healthiest life possible until a certain age (e.g., 70 or 75 years).\(^{(89)}\)
- In the derivation and application of WTP estimates, one typically assumes that the current distribution of income among individuals is appropriate. Therefore, WTP has been criticized as violating equity principles. For global decision making, the application of local WTP for global consequences of environmental problems may lead to strong violations of the equity principle and result in giving less weight to health damages in poor countries.
- The notion of double jeopardy was introduced to spotlight disabled people. It is argued that they are disadvantaged twice: first they suffer the disability, maybe for their whole life, and second, if resource allocation follows QALY maximization, they are disadvantaged because a year of life saved counts less.\(^{(44,45,90)}\) This problem was also found when the health loss of an HIV-infected subgroup due to drinking water impurities was assessed.\(^{(25)}\)
- Due to the limited dimensionality of health metrics, it was found that the sensitivity for certain groups of health outcomes like, for example, mental or sexual ill health, might be weak and therefore set biased priorities.\(^{(91,92)}\)

This summary of arguments mostly against pure utility maximization leads to the question of whether health metrics are useful at all, whether they should be adjusted to account for the mentioned points, or whether these points should be considered in other phases of the decision-making process. Most authors,

\(^{25}\) In this case it is even a threefold disadvantage: they are already struggling with a disease, they show a higher susceptibility to drinking water infections, and their premature death would be counted less because of their lower quality weight and shorter life expectancy.\(^{(15)}\) Therefore, this subgroup was analyzed separately to allow for tailored risk management.
5. WHICH METRICS FOR WHICH TOOLS?

Realizing that both the selection of the health metric (Section 3) and the choice of elements and assumptions used within the metrics (Section 4) matter, it is prudent to ask: Which metrics and parameters are most appropriate for environmental decision-support tools?\(^{27}\) Criteria to judge on the appropriateness include, for example, congruence between assumptions in the health metric and the decision-support tool, feasibility of the analysis, possibility of communicating metrics to decisionmakers, or acceptance of the method by decisionmakers and stakeholders. In Section 4 we presented a number of elements that appear to have large practical relevance when health metrics are used. These elements include the choice of the elicitation method to derive preferences, the group used for preference elicitation, time discounting, the type of life tables to be used, and distributional aspects. The following list of eight questions may help guide the process of selecting both type of metric and these elements in the specific application.

- Do health outcomes need to be expressed in monetary units? If a monetary quantification of health benefits or damages will be needed in the course of the analysis, like, for example, in benefit-cost analysis (BCA), the use of willingness-to-pay methods is prudent. Both monetary and nonmonetary metrics have flaws for valuation of both mortality and morbidity. However, since monetary methods require not only a health/health but a health/wealth trade-off they may be cognitively more demanding than nonmonetary metrics. Therefore, we suggest using them primarily when monetary units are desirable as a measurement unit. HALY-type methods may be preferred in life cycle assessments (LCA) or comparative risk analyses (CRA) because they often do not need monetary units.

- Are collectively borne costs part of the analysis? The avoidance of collectively borne costs of illness is often a benefit component of environmental decisions. Table III helps to identify which cost components may need to be assessed in a specific application.

- Have the affected groups different age distributions? Section 4.4 discussed the dilemma that the age-dependent value of mortality risk avoidance is an area of intense research and that neither the unweighted years of life lost (QALY approach), the age-weighted years of life lost (DALYs approach), nor the assumption of age-independent value of statistical life (most WTP approaches) are supported by evidence. If the age of concerned populations varies for different decision alternatives, sensitivity analysis is recommended.

- What is the reference group for the mortality cases? If estimates for the years of life lost are needed, life tables of the appropriate reference group may be used. Studies that concern a defined geographical area (e.g., drinking water supply area), a certain subpopulation (e.g., women only) with a short time horizon may therefore select very specific life tables based on observed data. In contrast, LCA studies that assess worldwide mortality cases occurring sometime in the distant future (e.g., due to climate change) may choose constructed, future-oriented, worldwide life tables.

- Whose values are best suited for health preferences? In environmental applications, societal rather than individual decision making is usually at stake. Therefore, as concluded in Section 4.2, community values may be used in combination with information provided by patients or health professionals. Tools such as

\(^{26}\) Nord\(^{83}\) suggests that in addition the following factors are important: number of people affected, size of perceived loss in quality of life, duration of effect, responsibility of affected person for caring for others, and effect on patient’s productivity. He also suggests that sex, race, education, and income should not be used as criteria.

\(^{27}\) A description and characterization of environmental tools can be found in Hofstetter et al.\(^{82}\)
LCA that cover a large number of very different health outcomes that may occur worldwide and/or in the distant future may find such a combined approach infeasible and rely instead on health professional preferences alone.

*What elicitation method fits the purpose?* Obviously, if monetary outcomes are preferred, either stated or revealed preference methods need to be chosen. The number of such studies for environmental applications is still limited and benefit transfer often not straightforward. In addition, revealed preference methods for environmental impacts (which are often small changes) may be confounded by (too) many other factors so that this approach is not always applicable. Therefore, availability and feasibility may dictate present practice. For disability weights used in HALY-type metrics, the choice of elicitation methods can be informed by the following two subquestions: (1) Are the decisionmakers interested in maximizing the aggregate of individual preferences for their own health or interested in the average preferences on tradeoffs between different health states of others? and (2) How many health states for how many different groups need to be assessed within a budget constraint? The time tradeoff and standard gamble methods focus on the individuals’ own health; person tradeoff methods on the health of others.

*How to consider different timing of health outcomes?* Assuming that environmental decision-support tools are used in societal decision making, societal discount rates seem most appropriate. This implies that the future increase of value of HALYs and statistical life are considered as well. However, it does not imply that discount rates or value increase need to follow a simple exponential function (especially for impacts in the distant future). For life cycle assessment a zero discount rate is suggested. This is based not only on the very long assessment horizon but primarily on present practice, where information on the timing of impacts is usually unavailable.

*Which distributional and ethical choices are acceptable?* As shown in Section 4.7, health metrics make (implicit) major assumptions on distributional and ethical choices that may or may not be shared by decisionmakers. This becomes obvious when regional differences in VSL are or are not considered in international decision making. In environmental decision making intra- and intergenerational equity concerns and health shifts between subpopulations are likely. Therefore, if decisionmakers are unable to make these choices explicitly *ex ante*, total health metric scores may be presented as breakdowns considering the different affected subgroups.

Although this guidance will help exclude some options, there will always remain a number of theoretically valid approaches that cannot be excluded *ex ante*. If resources allow, uncertainty and sensitivity analysis can provide insight into the relevance of some assumptions.

### 6. CONCLUSIONS

A case study that applied three different health metrics (DALYs, QALYs, and WTP) to the example of environmental health impacts in the Netherlands revealed the empirical relevance of the choice of monetary versus nonmonetary methods and the sensitivity of the results to mild distortions that affect large shares of the population (e.g., noise impacts, allergies, effects of endocrine disruption). For the application in the environmental field we learn that the present state of practice in WTP leads in our example to the same results as a pure mortality assessment. This may be an artifact due to the lack of reliable values for age-dependent values of statistical life and due to insufficient studies that assess WTP values for morbidity outcomes. Since the valuation of premature mortality has been shown (empirically and theoretically) to be age dependent but not proportional to the years of life lost, age weighting may be a relevant characteristic to be considered. The application in the environmental arena makes this point even more important since many environmental risk factors affect old people only while some affect children only or the average population. Further, HALYs are heavily sensitive to the preference weights for mild health outcomes. Since many elicitation methods are unable to deal adequately with mild health outcomes, this needs special attention in any analysis.

Another major finding is that not only the choice of the metric is important but that it is particularly important which empirical choices (e.g., life table, time discounting, elicitation method, elicited group, and age weighting) are finally made within the metric. Eight questions, listed in Section 5, may prove helpful in making these choices.
A further implication of our analysis is that indeed—as criticized by many authors—most health metrics follow the philosophy of utility maximizing. Since decisionmakers may want to base their decisions not only on a utility metric but also on insights how different ethical and distributional preferences would affect the outcome, we suggest a semi-quantitative discussion that evaluates the influence of the following aspects: Who are the worst-off? Which group could profit most (realization-potential)? What is the age distribution? Who are the innocent? What changes if patients below the age of 70 or 75 are saved first (fair innings)? Does income matter? Are already disadvantaged subgroups concerned (double jeopardy)? and Have case-specific valuation attributes been overlooked by the generic health metrics? These considerations are usually already made in today’s decision making. Therefore, this semi-quantitative analysis will provide the hard data to support these considerations and does not replace the purpose of deriving a utility measurement.

Finally, we clarified three different situations where health discounting may matter and conclude that the discount rate depends on the way the value of future lives is counted. For HAL Ys this usually means that a social discount rate would be applied.

A major assumption we made was that information on disease type and disorder, age of onset and duration of disease, and number of affected individuals is available. However, we did not discuss how and when this information can be derived nor did we show how some of this data could be estimated. Further, for morbidity outcomes we have not studied the empirical relevance of the time-proportionality assumption made in HALYs. This assumption suggests that having the same illness for five years of one year makes the total health burden exactly five times higher. Finally, we also did not discuss the evaluation of some tricky environmentally-caused human health impacts that may become important in single case studies. Especially, we left out issues such as developmental and fertility effects due to endocrine disrupters, hereditary effects due to ionizing radiation, or development effects in fetus due to environmental causes.

For two research questions we demonstrated the relevance for environmental decision making.

1. What are empirically valid age-dependent statistical values of life for use in benefit-cost studies and which age-weighting function should be used for HALY approaches? The present evidence for the shape of the age adjustments are weak, their slopes contradict each other, and they require, due to their practical relevance, more investigations. Since these age-adjustment functions may look very different for single individuals, studies may either need to include large samples or define subgroups and contexts that allow more homogenous answers.

2. How can quality and disability weights for distortions or mild illnesses that are caused by environmental risk factors be assessed? The uncertainty in disability and quality weights for mild illnesses is high because people are not ready to trade life expectancy for eliminating disturbances and because some elicitation methods compel respondents to use probabilities, a difficult task. Further, for mild disabilities with long durations, such as noise disturbances or reduced neurocognitive development, time nonproportionality due to adaptation and adjustment may have decisive influence on the outcome.

We hope that this article contributes to a better understanding of the differences between available health metrics and a more informed choice of metric by practitioners. In addition, we hope it will stimulate additional research to help resolve some of the remaining conceptual and practical issues in measuring health for use in environmental decision-support tools.

ACKNOWLEDGMENTS

We would like to thank Jane Bare, Gordon Evans, Matt Heberling, Glenn Rice (all U.S EPA, Cincinnati), Ruedi Müller-Wenk (University St. Gall, Switzerland), and John Evans (Harvard School of Public Health, Boston) for their valuable comments on earlier drafts. Patrick Hofstetter was supported, in part, by an appointment to the Postdoctoral Research Program at the National Risk Management Research Laboratory administered by the Oak Ridge Institute for Science and Education (ORISE).

28 Major cognitive problems arise because there are one or several orders of magnitude between the severity of mild outcomes and death. Since death is mostly used as anchor, respondents are forced into tradeoffs that go beyond the elicitation of existing preferences (see also Section 4.3).
through an interagency agreement between the U.S. Department of Energy and the U.S. Environmental Protection Agency. This article may or may not reflect the views of the supporting agencies.

REFERENCES


30. Hammitt, J. K. (This issue). QALY versus WTP. *Risk Analysis*.


Selection of Human Health Metrics


