United States
Environmental Protection
Agency

Office of Research and
Development
Washington DC 20460

EPA/600/R-01/104
September 2001

EPA
US EPA Office of Research and Development

Human Health Metrics for Environmental Decision Support Tools:

Lessons from Health Economics and Decision Analysis

EPA/600/R-01/104
September, 2001

HUMAN HEALTH METRICS FOR ENVIRONMENTAL DECISION SUPPORT TOOLS:

LESSONS FROM HEALTH ECONOMICS AND DECISION ANALYSIS

BY

Patrick Hofstetter
ORISE Research Fellow at U.S. EPA
National Risk Management Research Laboratory
26W Martin Luther King Dr
Cincinnati, OH 45220

James K. Hammitt
Center for Risk Analysis
Harvard School of Public Health
718 Huntington Ave.
Boston, MA 02115

National Risk Management Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Cincinnati, Ohio, 45268

Notice

This report has been subjected to the Agency's peer and administrative review and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement of recommendation for use.

Foreword

The U.S. Environmental Protection Agency is charged by Congress with protecting the Nation's land, air, and water resources. Under a mandate of national environmental laws, the Agency strives to formulate and implement actions leading to a compatible balance between human activities and the ability of natural systems to support and nurture life. To meet this mandate, EPA's research program is providing data and technical support for solving environmental problems today and building a science knowledge base necessary to manage our ecological resources wisely, understand how pollutants affect our health, and prevent or reduce environmental risks in the future.

The National Risk Management Research Laboratory is the Agency's center for investigation of technological and management approaches for preventing and reducing risks from pollution that threaten human health and the environment. The focus of the Laboratory's research program is on methods and their cost-effectiveness for prevention and control of pollution to air, land, water, and subsurface resources; protection of water quality in public water systems; remediation of contaminated sites, sediments and ground water; prevention and control of indoor air pollution; and restoration of ecosystems. NRMRL collaborates with both public and private sector partners to foster technologies that reduce the cost of compliance and to anticipate emerging problems. NRMRL's research provides solutions to environmental problems by: developing and promoting technologies that protect and improve the environment; advancing scientific and engineering information to support regulatory and policy decisions; and providing the technical support and information transfer to ensure implementation of environmental regulations and strategies at the national, state, and community levels.

This publication has been produced as part of the Laboratory's strategic long-term research plan. It is published and made available by EPA's Office of Research and Development to assist the user community and to link researchers with their clients.

E. Timothy Oppelt, Director
National Risk Management Research Laboratory

Abstract

Environmental decision support tools often provide information that predicts a multitude of different human health effects due to environmental stressors. Medical decision making and health economics offer many metrics that allow aggregation of these different health outcomes. This paper provides a review of this literature with special attention to aspects relevant in the environmental context. Based on a characterization of medical and environmental applications, recommendations for the use of human health metrics in different environmental decision support tools have been derived. Further, three metrics (quality adjusted life years (QALYs), disability adjusted life years (DALYs) and willingness-to-pay (WTP)) have been used to compare a wide range of different environmental risk factors. In this example, WTP tends to reflect mortality outcomes only. QALYs and DALYs are sensitive to mild illnesses that affect large numbers of people, which are difficult to assess in an unbiased manner. 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 non-monetary metrics may bear the largest practical relevance out of a series of suggested research questions.

Table of Contents

Acronyms and Abbreviations

$	U.S. Dollar
15D	quality of life measurement instrument using 15 attributes (or dimensions)
BCA	Benefit Cost Analysis (same as CBA)
CA	Conjoint Analysis
CBA	Cost-Benefit Analysis (same as BCA)
CEA	Cost-Effectiveness Analysis
COI	Cost of Illness
CUA	Cost-Utility Analysis
CV	Contingent Valuation
CVA	Cost-Value Analysis
DALYs	disability adjusted life years
DC	Dichotomous choice format
EPA	Environmental Protection Agency (same as USEPA or U.S. EPA)
EUR	European currency prior to the introduction of the Euro
EuroQol	European Quality of Life measurement instrument
GDP	Gross Domestic Product
HALYs	health adjusted life years
HALYs+	Health Adjusted Life Years with age-weighting
HUI	Health Utility Index
HYE	Health-Years Equivalent
ISO	International Standard Organisation
ME	Magnitude Estimation
O3	(tropospheric) Ozone
OE	Open-ended question format
PM10	Particulate Matter smaller than 10µm
PTO	Person Trade-Off
QALYs	quality adjusted life years
Qm	chronic health state
QW	Quality Weight
QWB	Quality of Well-Being
r	risk aversion factor
SEYLL	standard expected years of life lost
SF36	short form with 36 questions/attributes
SG	Standard Gamble
t	time
TO	Tradeoff Method
TTO	Time Trade-Off
U.S. EPA	United States Environmental Protection Agency
UN	United Nations
USEPA	United States Environmental Protection Agency
UV-A/B	Ultraviolet radiation within spectrums A or B
VAS	Visual Analogue Scale
VSL	Value of a Statistical Live
WHO	World Health Organisation
WTA	Willingness to Accept
WTP	Willingness-to-Pay
YLD	Years Lived Disabled
YLL	Years of Life Lost

Acknowledgements

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. Thanks also to Jean Dye for the technical editing. 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) 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.

1. Introduction

Environmental impacts on human health are (i) relevant compared to other health impacts[1], (ii) considered as important as damages to ecosystems (Goedkoop et al. 1999, Harada et al. 2000), and (iii) often trigger a change of behavior and regulations (Morgenstern 1997). However, environmental impacts cause a myriad of different health effects for different durations (Lippmann 2000, de Hollander et al. 1999) and in environmental decisions they often have to be compared with a different set of health impacts caused by competing decision alternatives (comparative risk assessment or life cycle assessment) or with regulation costs (benefit-cost analysis) or impacts on ecosystems and resources (life cycle assessment). Therefore, a common metric for health outcomes that allows adding a wide range of different health outcomes would enable decisions that are more informed.

Applications of health metrics in environmental decision support tools have been explored in many different ways. While pure mortality statistics and years of life lost were used in early energy studies (e.g., Inhaber 1982), willingness to pay (WTP) has been used for some time now (e.g., Viscusi et al. 1991, ExternE 1995, ESEERCO 1995, USEPA 1999b). Quality adjusted life years (QALYs) have recently been used in USEPA (1998a), Hammitt et al. (1999a) and Ponce et al. (2000), and disability adjusted life years (DALYs) have been used in Hofstetter (1998), Goedkoop et al. (1999), Mara et al. (1999), de Hollander et al. (1999), Müller-Wenk (1999), Havelaar et al. (2000), Frischknecht et al. (2000).

In many environmental studies where human health metrics were used they have been selected based on the historical roots of the field[2], the needed measurement unit[3], or the authors' background[4]. Only a few general investigations concerning human health metrics for environmental decision support tools have been identified (see e.g., O'Brien et al. 1994, Hofstetter 1998, Carrothers et al. submitted). Exchange between concepts and knowledge in health economics and medical decision making on the one hand and environmental decision making on the other hand has primarily been case- or application-specific and rarely based on a broader overview and analysis.

Therefore, this article provides a summary of the concepts and findings available from the fields of health economics and medical decision making (Section 2). This summary should ease the access to these fields for researchers in environmental decision making, and also reflect the findings in the

[1] De Hollander et al. (1999) estimate that health impacts due to particles, noise, lead, ozone, radon and environmental tobacco smoke cause almost 5% of the Dutch burden of disease

[2] Most methods used in Life Cycle Impact Assessment for the assessment of human toxicity have their roots in Chemical Risk Assessment (e.g., Guinée et al. 1993, Hertwich 1999, Huijbregts 1999). 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 non-occurrence of health impacts is in this case the anchor of the health metrics scale.

[3] Externality studies used methods to assign monetary values to different health outcomes because their aim was to value environmental damages in monetary units. (e.g., Frey et al. 1985, ESEERCO 1995, ExternE 1995)

[4] A recent example may be the Comparative Risk Assessment study performed by U.S. EPA (USEPA 1998a), 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 (Havelaar et al. 2000) choose to use DALYs for a similar case study.

light of environmental applications. Some practical implications of three widely used health metrics are shown and discussed by applying them to a recent survey of environmentally caused health impacts in the Netherlands (Section 3). In order to understand the different metrics that have been suggested in the medical field and their potential transferability to the environmental arena we characterize both the medical and environmental decision support systems (Section 4). Based on this characterization, sensitive elements of health metrics can be identified and recommendations can be made for congruent health metrics for environmental decision support tools (Section 5).

Further, we raise some issues that are also relevant within medical applications (time-non-proportionality, actual age-dependency of disutility due to premature death and distributional/ethical attributes) or could be considered important aspects in environmental applications (importance of mild impairments, appropriate life tables for intergenerational and international impacts, interpretation of costs of illness).

In this article we assume that 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. This assumption is often not met because conclusive epidemiological studies are needed to supply this data. Metrics that make less restrictive assumptions have been suggested in de Rosa et al. (1985), and ILSI (1996) and TERA (1999).

2. Human health metrics: a review of the literature

The simplest form of human health metrics is to select health outcomes as reported in health statistics. Systematic health statistics were started in the 18th century in Australia and 1837 in England and Wales (WHO 1993). Many statistics have been started as mortality records only, separated by sex and age. Later, they were extended to include morbidity information. Today's standard list of diseases was initiated in 1853 and is revised at the beginning of each decade (Alderson 1988). About 100 internationally defined disease classes are adopted widely; in addition, single countries or continents use classifications that are more specific. The human health metrics we are interested in go beyond these health statistics. We are interested in a measure for the loss of health due to diseases and premature death.

Medical decision making and health economics have dealt with questions like choosing treatment methods and resource allocation for the last 30 years. This section draws on the research of these fields. Subsections 2.1 to 2.3 provide a short overview on the metrics that are of highest interest to applications in environmental decision support tools. Subsections 2.4 through 2.7 address some of the fundamental assumptions and choices behind most metrics. Subsections 2.8 and 2.9 deal with the elicitation of quality weights for morbidity outcomes, while subsection 2.10 addresses the measurement of premature death. Finally, subsections 2.11 through 2.18 review the literature with respect to aspects relevant to the application of health metrics in environmental and decision support contexts.

2.1 What to measure?

The World Health Organisation defined health in 1946 as follows: "Health is not only the absence of infirmity and disease, but also a state of physical, mental and social well-being" (WHO 1947). This broad definition captures essential elements of quality-of-life and underlies most human health metrics. Based on this definition, it is also clear that the loss of health cannot be solely measured by statistical information on mortality. It is commonly understood that mortality measures alone provide decision makers with incomplete and insensitive information about overall population health (Murray et al. 1996a, Field et al.1998). Summary measures therefore include information on mortality and morbidity and are the primary focus of this article.

Figure 1 provides a useful distinction of different assessment levels for morbidity outcomes. The further down we go on the level the more relevant becomes the information to individuals and the more relevant become local factors like health care system, family/household structure, economic development (farming *versus* service economy), cultural and religious beliefs. The metrics discussed here are mostly located on the disability or handicap level. While the former allows for applications that are more generic and international, the latter can more appropriately take into account a patient's environment.

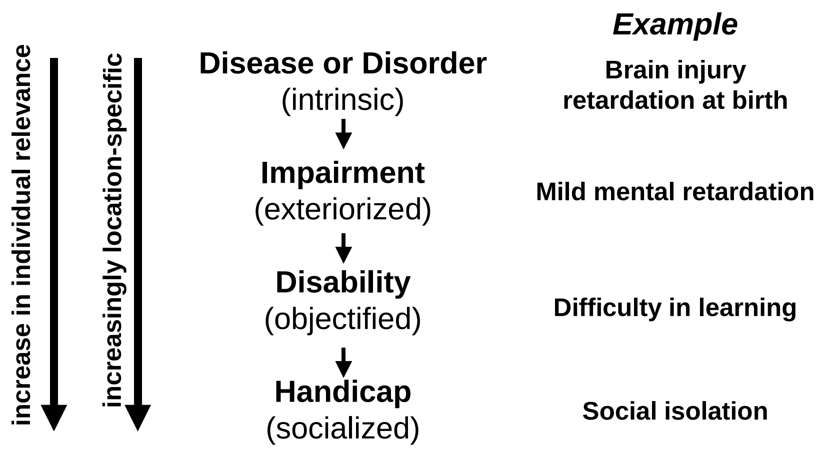

Fig. 1: Possible assessment levels for human health metrics (after Murray et al. 1996a)

While health statistics provide a snapshot in time, we are more concerned with the consequences, the disease history. This information is captured in so-called health profiles. They include information on duration and disease stages, cure, remission and co-morbidities. If the health assessment concerns a group of people[1] or *ex ante* assessments of individuals, then the relative frequency or probability for each disease stage is used as a basis to quantify the health profile.

Therefore, human health metrics summarize mortality and morbidity outcomes and attempt to measure physical, mental and social well-being on the level of disability or handicap for health profiles of individuals or the population at large. The application directs the assessment levels (see Section 5).

2.2 A classification of approaches for health metrics

Spilker et al. (1996) provide an overview of about 300 different instruments for the comparison of different health states. This wealth of instruments can be classified by few characteristics, which reduces the number of relevant approaches for this article. The following distinctions can be made:

- *Time-based versus time-point related approaches*; as introduced earlier we are seeking for instruments that assess health profiles, i.e., health state over time. Therefore, time-based approaches have been considered only.
- *Generic versus disease-specific versus problem-specific approaches*; we are interested in the disabilities due to a broad range of diseases that are caused by environmental impacts. Therefore, the more restricted instruments that focus, e.g., on asthma (Anonymous 1994) or cancer treatments (Rusthoven 1997) alone, are not sufficient for our purpose.

[1] See an example for asthma in Anonymous (1999a).

- *Indicator versus single index approaches*; not all approaches allow for an aggregation of the indicator values on different health dimensions to a single index. Here we are interested in single index approaches rather than descriptive health state instruments like SF36 (Fryback et al. 1997)[2].

- *Explicitly decomposed versus statistically inferred decomposed versus holistic approaches*; All three approaches acknowledge the multidimensional nature of health that includes aspects such as: health perceptions; social function (social relations, usual social role, intimacy/sexual function, communication/speech); psychological function (cognitive and emotional function, mood/feelings); physical function (mobility, physical activity) and impairment (sensory function/loss, symptoms/impairment)[3](Gold et al. 1996:95). In the holistic approach, judges are confronted with a full verbal description of a health outcome along with some of the above dimensions and asked for a direct utility judgment. The explicitly decomposed approach, on the other extreme, uses multi-attribute utility theory (see below) to make judgments separately on how health states influence single health dimensions and finally how to combine the different health dimensions into a single number. The statistically inferred decomposition derives the relative importance of the health dimensions by multiple regression analysis of holistic judgments on health states and the health states' scores on each dimension. While the decomposed methods reduce the cognitive load of the judges compared to the holistic approach, the explicitly decomposed approach may assume invalid properties of the aggregation structure and nature of scales. Although Frohberg et al. (1989a) recommend the statistically inferred approach because of its superior validity, all three approaches are applied in medical decision making for practical and historical reasons.

- *Composite versus whole profile*; the need for time-based metrics that are able to measure health profiles suggests that either the utility of a health profile over the time-span of interest is holistically assessed or composed by a number of time slots multiplied by their specific health state specific utility.

Based on this overview, the rest of the article will concentrate on time-based, generic, and single index approaches. Whether holistic or decomposed and composite or whole profile approaches are favorable or more feasible is less obvious.

2.3 Short Introduction to QALYs, DALYs, HYE and WTP

Figure 2 shows a hypothetical health profile[4] of an individual. The gray and black areas represent the quality adjusted life years (QALYs) and disability adjusted life years (DALYs), respectively. While

---

[2] Due to the large amount of available information on public health measured with such descriptive indicator systems many algorithms for their transformation to utility scales have been developed, see e.g., Patrick et al. 1993, Torrance et al. 1996, and Fryback et al. 1997.

[3] A selection of these dimensions is usually included in multi-dimensional quality of life instruments (EuroQol (Essink-Bot et al. 1993), 15D (Sintonen 1981), QWB (Kaplan et al. 1988), HUI I/II/III (Torrance 1986), Rosser Index (Rosser et al. 1972), but none includes all dimensions (Gold et al. 1996:108).

[4] An illustration would be yellow fever 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.

QALYs measures the actual health quality integrated over time, DALYs measure the loss compared to a hypothetical profile.

Pliskin et al. (1980) describe QALYs as utility functions under a number of different assumptions. The most general form is the risk-adjusted version:

$$QALY_m = U(Q_m,t) = [H(Q_m) * t]^r \quad [a] \quad (1)$$

where U is the utility function of the constant chronic health state Q_m during the life years t. H(Q) refers to the value function of quality (we will call it quality weight) and r is a risk-aversion factor[5]. It is common practice to discount future health outcomes if QALYs are used in cost-utility analysis (see Section 2.6).

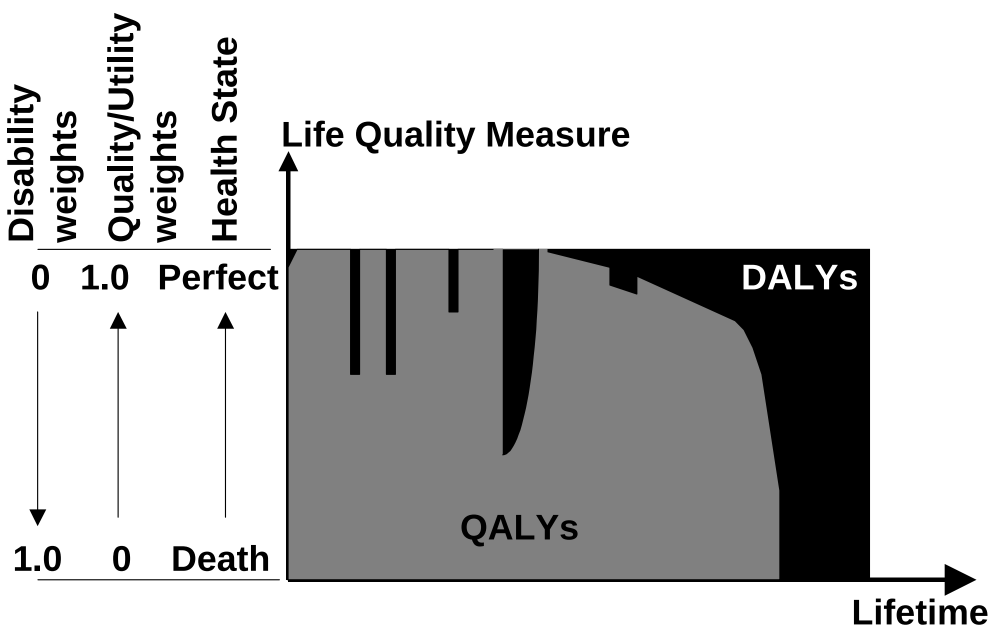

Fig.2: Graphical illustration of a health profile and its measurement by Quality Adjusted Life Years (QALY, gray area) and Disability Adjusted Life Years (DALYs, black area).

DALYs are the sum of the years life lost (YLL) and the years lived with disability (YLD) (Murray et al. 1996a):

[5] The following notion applies: r>1 risk seeking, r=1 risk neutral, r<1 risk averse. We use ‘risk aversion’ only in its strict sense, i.e., when chance matters. When chance does not matter we use the term ‘change aversion’.

$DALY_m = YLL_m + YLD_m$ [a]

$$= \text{discounting*age-weighting* (SEYLL}_m + \text{disability weight}_m\text{*disability duration}_m) \qquad (2)$$

where m is the type of disease. 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. Age-weighting is included by the expression $C \cdot a \cdot e^{-\beta \cdot a}$ where C and β are constants and set equal to 0.1658 and 0.04 respectively (see Section 2.10) and a is the age. For YLD, similar to QALYs above, the disability weight is multiplied by the disability duration. See Murray et al. (1996a:64ff) for the detailed equations for continuous and Elbasha (2000) for discrete age of onset. DALYs assume risk neutrality.

Murray et al. (1996a) allow DALYs that both use or do not use discounting and age-weighting. Therefore, there are only a few key differences between QALYs and DALYs and we introduce the term health adjusted life years (HALYs) as an umbrella term. Figure 2 and equation (2) make clear that the DALYs framework needs to define a reference life expectancy while QALYs just quantify changes from one health profile to another. This implies that the reference state used for DALYs does typically assume perfect health until death (Figure 2). Differences in the elicitation of disability and quality weights will be addressed in Sections 2.8 and 2.9.

While both QALYs and DALYs make the restrictive assumption on time-proportionality, the Health-Years Equivalent (HYE) of Mehrez et al. (1989) does not decompose the health quality and duration aspect.

$HYE_{m(t)} = U(Q_m, t)$ [a] (3)

where U is the utility function of the health state Q_m during the life years t. Although Willingness to Pay (WTP) is embedded in welfare economics and measures loss in life quality in monetary units that have an external reference, its simplest form is similar to the HYE:

$WTP_{m(t)} = V(\Delta Q_m, \Delta t)$ [$] (4)

where V is the value function of the health state change ΔQm during the time interval Δt. WTP should be understood as the rate of substitution between health and wealth. It is typically used to evaluate small changes in health states rather than to construct a total burden of disease (see Hammitt (2002) for a more detailed elaboration of the nature of WTP for premature death).

QALYs have traditionally been the most important summary measure in medical decision making. However, WTP and more recently DALYs find widespread use as well. HYE or SAVE (Nord 1992b) have not been widely used, which may be due to the increased burden of deriving standard values for all possible combinations of health states and their duration. However, one also should be aware that Azimi et al. (1998) found in 109 cost-effectiveness and cost-utility studies[6] published between 1990 and 1996, only 18% used QALYs, but 71% used no summary measures at all. However, Bell et al. (1999) collected 228 studies that used QALYs (almost all 1990-1997), which suggests a wide use of summary measures.

The underlying assumptions and problems of the chosen functions in the presented equations and the questions that arise when the variables are derived are addressed in the subsequent sections.

2.4 Social welfare function

In environmental and many medical applications, it is the *social*, rather than *individual* welfare, which must be optimized. The way social welfare is defined and assessed will influence the way preferences for health qualities are elicited (Sections 2.7ff). Therefore, principles and construction of a social welfare function need to be addressed here.

The neo-classical 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 (consequentalism), and that the value of any situation should be judged solely on the basis of the utility levels attained (welfarism). An important distinction is between individual and social choices. 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. Neoclassical economics often assumes that it is not possible to make interpersonal utility comparisons; i.e., 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.

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 one dollar gain contributes the same to social welfare regardless of who receives it, be he 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 by the change plus the payment of compensation.

[6] Nord (1999) defines cost-effectiveness analysis (CEA) by its use of natural units (mortality, number of cases) to quantify the health effects, cost-utility analysis (CUA) by its use of utility measures like QALYs to quantify the utility of health improvements, cost-benefit analysis (CBA) by its use of the Willingness-To-Pay approach to quantify the health benefits in monetary units, and cost-value analysis (CVA) by its use of a holistic assessment of the health benefits of a whole program from a societal point of view.

An alternative approach to interpersonal comparisons that is conventional in the medical cost-effectiveness literature is to measure health benefits in some form of "health-adjusted life year" (HALY, i.e., a QALY or DALY type of metric). In this case, health benefits and harms to different people are evaluated by assuming a HALY 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.

Another alternative approach captures societal or altruistic preferences. The elicitation of these preferences is very difficult. The effect of altruism on health values is somewhat subtle and uncertain, because altruism can take many forms. Altruism about another person's welfare may reflect concern for the other's total welfare, as the other person evaluates it ("pure" altruism), or it may reflect concern for only one aspect of the other person's welfare, e.g., his mortality risk ("safety-oriented" altruism, a form of "paternalistic" altruism). Bergstrom (1982) shows that a society's total willingness to pay for a publicly provided reduction in mortality risk is the same if individuals care only about their own welfare, or if they are pure altruists. Jones-Lee (1992) shows that the value is also the same in the case where individuals are paternalistic altruists. For intermediate cases where individuals care about others' welfare, but give somewhat greater weight to their physical health risks than to other aspects of their well-being, willingness to pay can be somewhat larger, on the order of 10% to 40% under reasonable assumptions.

The existence of approaches based either on individual (self-interest) or altruistic preferences may suggest that the type of welfare function depends on the decision at hand. For societal decisions in medical decision making, both approaches have been suggested. Since altruistic preferences can only be derived if self-interest can be ruled out, Nord (1999) suggests that this approach is used to support decisions on ad hoc public programs for others, while choices for private or long-term public health plans can well be based on self-interests. Environmental decision support tools may be confronted with both situations. Air pollution affects all, therefore, self-interest may be justified in a social welfare function; lead poisoning, on the other hand, will only affect families with young children who live in contaminated buildings and environments. Here, societal or altruistic preferences may come into play. The same holds for impacts that will affect people on other continents (malaria due to climate change) or future generations.

2.5 Properties of scales, attributes and the QALY-equation

The ideal metric for medical decision making and environmental decision support tools should be measured on a utility scale that would allow addition of different health episodes for the same person, add health outcomes of different persons, and allow for use in cost-utility analysis. All health metrics presented in the previous paragraph implicitly assume that such an aggregation is possible under expected utility theory. Although it is known that expected utility is not descriptive, there is some debate whether it shall be prescriptive or even normative (Raiffa 1961/1970, von Winterfeldt et al. 1986, Cohen 1996a/b, Wu 1996, Baron 1996, Douard 1996, Eeckhoudt 1996). Here we assume that, even if expected utility is not always normative, it is at least the most mature theory.

The QALYs and DALYs make additional assumptions by splitting up the time duration from the quality/disability attribute. Pliskin et al. (1980) show that the following conditions must be empirically satisfied for QALY to represent a valid utility function for health outcomes with a constant health status level over time (based on von Neumann and Morgenstern 1943, Keeney et al. 1976):

1. The two attributes duration and quality shall be *mutually independent* in their contribution to the utility (i.e., H(Q_m) for all t constant)
2. The proportion of remaining life that a person would be willing to trade off for a specific health improvement shall be independent from the expected remaining life time. This is called *constant proportional trade off*.

If it is assumed, for practical reasons, that the utility function is linear over time (r=1) then a third condition is required (Pliskin et al., 1980):

3. *Risk neutrality* regarding life years shall hold for the individual values.

In real life applications, the health status is not constant over time but follows a health path or health profile. Therefore, distinct intervals of different health states should be additive. From that request, a fourth condition has to be fulfilled (Keeney et al. 1976):

4. The value of a health state in period A shall be independent of the value of another health state in period B, i.e., *additive utility independence.*

Miyamoto et al. (1985) find r≠1 because risk neutrality is empirically not given and they confirm that the above assumptions are violated. Fryback (1998:42) states, “The most fundamental assumptions in the construction of HALY [which includes DALYs and QALYs] measures is that the part of the measure dealing with weighting health state can be obtained separately from the […] time duration part of the measure.” He acknowledges that this major assumption may well be wrong. Nord (1999) makes clear that the time-proportionality has been introduced right from the beginning, but has no empirical evidence. He also claims that time discounting is both a different issue and does not explain the full effect. Nord (1992a) also cites examples where, in one study, one day in bed performing no major activities was weighted 0.61, while another study with a non-specified duration for the health state ‘bedridden’ found a weight of 0.09![7] Multi-attribute utility theory says that simple shapes of utility functions[8] are only applicable if at least utility independence is given (Fischer 1979). However, empirical studies show that information about the expected duration of a state has an effect on the valuation of its severity (Sackett et al. 1978, Sutherland et al. 1982, Dolan 1996). McNeil et al. (1981) find that if a health state (e.g., less than perfect level of speech) is experienced for less than 5 years then individuals are unwilling to trade longevity for health improvements. Loomes et al. (1989), Bala et al. (1996/1998), and Richardson (1994) provide more evidence against the four mentioned assumptions.

[7] In both cases, the scale ranges from 0 to 1 where 0 equals death and 1 full health.
[8] like multilinear, quasiadditive and additive models

Richardson et al. (1996) and Kupperman et al. (1997) showed that composite and whole profile measurements show a poor accordance, i.e., that a known sequence of different health states over a full lifetime is judged different from the results of a calculated composition. Krabbe et al. (1998) confirms this finding by showing that additive utility independence is not fulfilled. However, MacKeigan et al. (1999) find good accordance between composite and whole profile methods for relative minor health impairments and Treadwell (1998) shows that preferential independence is satisfied in the QALY model and argues that controversial results can be explained by (negative) time discounting and lacking independence of the health states.

Gafni et al. (1993) plead against QALY for the above reasons and suggest HYE, which need not fulfill the restrictive requirements of additive independence and constant proportional trade-off as an alternative (MacKeigan et al. (1999).

However, being aware of the strong evidence against the validity of assumptions (1) through (4) many authors consider that QALY (and consequently DALYs) may still be useful because distortions are small, the composition rule is simple and the cognitive task in empirical studies is easier than, e.g., with HYE.

Whether the distortions due to the violations of all major assumptions behind QALYs (and DALYs) are indeed small enough to be accepted has not been demonstrated on a sufficiently large set of case studies.

2.6 Discounting

Discounting is generally used to account for two factors: preferences for health at different dates, and opportunities for providing health benefits at different dates. Much debate has occurred on the question whether health outcomes should be time discounted, how large the discount rate should be, and whether the rate should be the same as that used to discount costs (Weinstein et al. 1977, Gold et al.1996).

It is useful to distinguish the individual and social choice problems. For an individual, date and age are perfectly correlated and so an individual’s preferences for health at different dates and at different ages cannot be distinguished. In principle, an individual’s preferences for health at different ages are virtually unrestricted. Some individuals might consider an increment to health equally valuable at all ages, while others would consider a health increment more valuable if it occurs when they are young (positive time preference), and still others would consider the increment most valuable if it occurs when they are old (negative time preference). Moreover, preferences for health might be related to age in some non-monotonic fashion. Apparent positive time preference may be a defect of myopia. It might also arise from the latent risk of death that makes it uncertain whether one will experience future costs and benefits, or decreasing marginal utility of health (if health is expected to increase). Zero and negative time preference can be explained by dread[9] and by a preference for sequences that improve over time (Wathieu 1997).

Within the context of environmental decision support tools, we are usually interested in social time preferences and also have to deal with interpersonal and intergenerational aspects. In this setting, 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[10]. 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. (1996a) 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 non-monetary health market and we found that no discounting is justified, at least so long as increases in HALE are neglected as well. 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 (Weinstein et al. 1977, Keeler and Cretin 1983, Gold et al. 1996). 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 health plans. To avoid this situation, Weinstein et al. (1977) 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 discount rate. Gold et al. (1996) suggest in their recommendations to use the same discount rate for costs and health outcomes and to apply a social discount rate.

---

[9] Van der Pol et al. (2000) present a literature review and show that subgroups of respondents have either a zero or even negative time preferences. They also find that individuals in severe health state are more likely to have negative time preference because they want to eliminate dread (=Loewenstein hypothesis).

[10] Pigou (1932:29f) argued "there is a wide agreement that the State should protect the interests of the future *in some degree* against the effects of our irrational discounting and our preference for ourselves over our descendants. The whole movement for 'conservation' in the United States is based on this conviction. It is the clear duty of Government, which is the trustee for unborn generations as well for its present citizens, to watch over, and, if need be, by legislative enactment, to defend, the exhaustible natural resources of the country from rash and reckless spoliation."

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) (e.g., van Hout 1998). Otherwise, different discount rates for costs and health may well make sense. In our example, the 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 HALY or one statistical life stays the same while real income increases. 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 (Cropper and Sussman, 1990; Hammitt, 1993). Therefore, we conclude that the literature has not adequately considered the question by how much the value of a HALY or statistical life is changing over time. Once this value increase is considered, discounting can be applied[11]. Available empirical evidence does not yet allow us to suggest correction functions for future values of HALYs or statistical life[12].

Therefore, we recommend the following discounting practice:

1. If health is measured as utility in HALYs and one HALY stays equally valuable independent of its timing and who profits then these HALYs are discounted at a social discount rate, e.g., 3% (Murray et al. 1996, Gold et al. 1996).

2. If the value of health is measured, the following distinction is needed:

   - If future increases in the value of HALYs and 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 of 3% (Murray et al. 1996, Gold et al. 1996).

[11] Johannesson et al. (1997) find an average marginal rate of time preference for health of about 1%. Murray et al. (1996a) and Gold et al. (1996) suggest both a social rate of time discounting of 3%. Others suggest using the time preference of the market only to discount close future but to use a minimal discount rate for distant future because a damage occurring in 30 years or 40 years should not be valued much differently (Weitzmann 1998). Therefore, the discounting with a constant rate is questioned. 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.

[12] 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 time-span 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 (Hammitt et al. 2000).

- 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.

2.7 Whose values?

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, WTP or HYE we need to ask whose values should be considered in those elicitation procedures?

A recent review of 38 studies (de Wit et al. 2000) that included groups of patients and non-patients 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). 22 studies reported higher patient values, two studies showed lower patient values and three studies found contradictory results. Therefore, it matters which group or how the study population is selected. In the course of the Global Burden of Disease study (Murray et al. 1996a), it has been 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 (Üstün et al. 1999). They also find that the differences in ranking of mental versus physical conditions are larger between different groups of physicians and care givers than between countries.

Different groups that might provide preference information can be positioned in a 3-dimensional space (strength of relationship [self, family, friends, no experience], time with illness [immediate, soon, distant future, never], subjective probability of illness [certain, likely, unlikely, no chance at all]) (Dolan 1999). Patients are positioned at the origin of this system of coordinates while physicists and health professionals have usually a lot of experience but little chance of experiencing the illness soon themselves. Elicitation of preferences of people with no experience with a disability and little chance of experiencing the disability soon is a challenge. Therefore, preferences from either patients or health professionals are widely used in CEA (Bell et al. 1999).

What are the reasons for different (higher) quality weights of patients compared to health professionals or the public? It was found that

- the given description to the general public did not correspond with what patients actually suffer (Jansen et al. 2000),
- human beings are very flexible in adapting to new situations,

- human beings tend to state relative preferences that probably compare to people of similar age or fate[13] (Groot 2000),
- aversion against disability only plays in ex ante situations but patients are in 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 (Gabriel et al. 1999), and
- the whole meaning of quality of life is redefined[14].

For medical decision making, 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, i.e., 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 (Field et al. 1998). Practically speaking, the "true" weights for avoiding health cases may lay somewhere between patients' values and the public' values as the health professionals' values usually do.

It appears from this discussion that the application in environmental decision support is less dependent on patients' values, but that it may be difficult to inform the public accurately enough about the health outcomes to elicit their preferences. A two-step procedure, where patients describe in step 1) their health states in multi-dimensional quality of life instruments and the public provides in step 2) aggregated values (either with MAUT or holistically) could solve some of the problems mentioned (De Wit et al. 2000, Nord 1999). Alternatively, some of the problems with patients' preferences can be solved by eliciting preferences for *changes* in health states rather than for *absolute* health states.

We conclude that first, it is important to decide whether self-interest or altruism should be elicited. Second, it is a crucial step to make sure that the health state is well understood which can be done by choosing patients or health professionals or two-step procedures; and third - as we will discuss in the next two sections – the phrasing of the elicitation question will influence which values are activated.

Finally, one could also ask whose values for whom? Since the severity of disabilities also depends on the relevance of certain handicaps to specific groups of individuals, it has been shown that quality weights depend on the patients' occupation, gender and family status (Holmes 1997). However, environmentally induced health effects are not sensitive to these characteristics. The higher shares of environmentally affected children, elderly and already sick people can be considered by age group

[13] People tend to reduce cognitive dissonance by overstating their health state and psychological adaptations help them to shift to a new anchor (Ubel et al. 2000).
[14] Koch (2000a/b) argues that disabled people repeatedly confirm their good health because the physical disability is indeed no handicap anymore in a chronic situation. Therefore, the high quality weights of chronically ill patients make sense. Brickman et al. (1978) found that persons 1 year after winning a lottery or developing paraplegia show very little difference in happiness.

specific quality weights (Murray et al. 1996a) and co-morbidity factors respectively. Other sensitive subgroups are assumed to show no deviation from an average disability to handicap relationship. Therefore, age group and co-morbidity of affected populations should be considered in environmental decision support tools.

2.8 How to elicit values and utilities?

Here we present the elicitation methods that are used in medical decision making to derive quality weights for QALYs, disability weights for DALYs, and values for HYE and WTP. The use of the terms ‘preferences’, ‘values’ and ‘utilities’ 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 cardinal scale, and ‘utilities’ denote ‘values’ under risk that fulfill the requirements by von Neumann and Morgenstern (1943) as outlined in Section 2.5.

The following short descriptions shall describe prototypical versions of each method (see also Nord 1992a, Patrick et al. 1993:143ff, Murray et al. 1996a:71).

- *Rating Scale/Visual Analogue Scale (VAS):* A typical rating scale consists of a line with clearly defined endpoints. The most preferred health state is placed at one end of the line and the least preferred at the other. The remaining states are placed between the two endpoints so that the intervals between the placements correspond to the differences in preferences as perceived by the subject that is asked to determine the weights. This method is the easiest to administer and to understand for respondents. However, the resulting preference weights have usually only ordinal meaning.
- *Magnitude Estimation (ME):* Subjects are asked to provide the ratio of undesirability for pairs of health states. For instance, state A is felt, for example, to be two times worse than state B. A series of questions allows the subjects to locate all the health states on one scale of undesirability, where at least one health state should be perfect health or death (similar to the procedure used in the Analytical Hierarchical Process (AHP) (Saaty 1980)).
- *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 health state *i* for *t* 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 health state *i*. This method provides utilities that conform with von Neumann and Morgenstern requirements for decisions under risk. Since human beings have difficulties in dealing with (low) probabilities, it is suggested to use cumulative prospect theory (Tversky et al. 1992) to transform elicited probabilities (Stalmeier et al. 1999, Bleichrodt et al. 2000).
- *Tradeoff Method (TO)*: A subject is asked to choose a health state *i+1* so that it is indifferent between the gambles (p,r;1-p,i+1) and (p,R;1-p,i) where *p* is a constant probability, *r* and *R* are two reference health outcomes such that *R*>*r*, and *i* is first the starting health outcome and then the previously elicited health outcome. This procedure constructs an interval scale with a large

number of trade-offs between similar outcomes of equal preferential difference. (Wakker et al. 1996). It would fulfill the von Neumann and Morgenstern requirements but health outcomes are not available on a continuum, therefore, this method has so far not been applied in medical decision making (Bleichrodt et al. 2000).

- *Time Trade-Off (TTO):* A subject is offered two alternatives. Alternative 1 is health state *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 state *i* is *x/t*. Torrance et al. (1972) introduced TTO and found good accordance with SG. Therefore, this method has been widely used, as it is less demanding than standard gamble and does not suffer from the difficulties of deriving (low) probabilities. Nevertheless, whether TTO works for minor health impairments is questioned because people have proven unwilling to trade life expectancy for minor disabilities (MacKeigan et al. 1999). Therefore, others choose to use the worst health outcome rather than death (Krabbe et al. 1998). Since TTO has inherently inbuilt the consideration of time-preference, Johannesson et al. (1994) show how QALY that use TTO have to be calculated if additional time discounting is needed.

- *Person Trade-Off (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 health state *i*. *y* is varied until the respondent is indifferent to the choice between the two alternatives, at which point the preference for state *i* is *x/y*. Other forms of person trade-offs can be constructed where subjects are asked to trade-off restoring health to *x* individuals in health state *i* versus restoring health to *y* individuals in health state *j*. Patrick et al. (1973) introduced this method as "equivalence of numbers technique" and Nord (1992a) gave it the name Person Trade-Off method. The PTO most directly reflects resource allocation situations whereas SG, TTO, and VAS do not ask this question and respondents that are confronted with the implications confirm that they did not have resource allocation in mind (Nord 1995). While the methods mentioned so far are explicitly about one's own health and health-preferences, PTO is explicitly about other people's health. Pinto Prades (1997) finds that PTO is empirically superior compared to SG and VAS for societal resource allocation. He defines three versions of PTO. PTO1 has a gain/gain framing, PTO2 a gain/loss framing and PTO3 uses a number of health states that are close together and builds up a chain (similar to TO). He finds clear differences between PTO1 and PTO2 and stresses that PTO3 may work best for mild illnesses because it is both cognitively easier and easier for users to make trade-offs between severe illnesses and premature death.

- *Attribute Based Stated Choice, Conjoint Analysis (CA):* Paired comparisons of multidimensional alternatives with factorial regression analysis are the basic features of this method (Huber et al. 1993). If the comparison involves just a statement choice one speaks of a Conjoint Choice or Attribute Based Stated Choice method. If rankings or ratings are used, this is called Conjoint Ranking or Conjoint Rating Methods respectively or more generally Conjoint Analysis (Adamowicz et al. 1998). It is a very useful method when very different attributes matter in a decision and it has a high degree of realism because potentially similar alternatives are compared. For example, it was shown that the value of in-vitro fertilization can not be measured only on a health scale but the attitude of the staff, time on the waiting list or follow-up support have been considered as non-health outcomes of the medical treatment (Ryan 1999). Attribute

Based Stated Choice methods also gain popularity in determining WTP (Johnson 1998, 2000). While earlier regression analysis was restricted to linear additive models (Ryan et al. 1997) more sophisticated models are available nowadays. It has to be noted that although realistic scenarios are compared by judges the results of the regression analysis may not be acceptable to the judges. One should also be careful in the number of attributes that are presented in order to stay within the cognitive possibilities of humans (Miller 1956).

- *Contingent valuation (CV) (monetary valuation, stated preferences):* Subjects can be asked in at least four different ways to estimate their willingness-to-pay (WTP) or willingness-to-accept (WTA) certain health states. One can then measure which amount individuals would accept to pay (1) for reaching a better health state, or (2) 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 (see also Jones-Lee *et al.* (1997), and Wenstøp *et al.* (1997)). The number of studies of type (1) and (2) has rapidly increased in the 1990s for use in benefit-cost analysis (Diener et al. 1998). Next to starting point biases, anchoring biases, strategic biases, information biases and framing biases that are common pitfalls of all listed elicitation methods the monetary valuation also suffers from scope insensitivities, hypothetical biases, and payment vehicle biases (Viscusi et al. 1987, Jones-Lee et al. 1995, Baron 1997, Beattie et al. 1998, Willis et al. 1998, Blumenschein et al. 1999). Those additional problems are due to the fact that respondents are not only asked to weight different health states but also to relate these weights to a (health-external) monetary unit. An important property of CV values is their dependency on income[15]. Typical elicitation formats used for CV studies include open-ended question format (OE), (bounded) dichotomous choice format (DC), and iterative bidding. It was found that DC is most compatible with incentives and gives reasonable upper bound estimates while OE is just in a comfortable range and tends to understate the maximum WTP (strategic bias). The observation that people prefer to say yes (yea-saying effect) and the starting-point bias are potential problems. A debriefing may be important to understand potentially relevant biases (Bennett et al. 1998). Deliberative and discursive methods have been developed to deal with framing and embedding biases (Sagoff 1998); calibration factors have been suggested to adjust too-high WTP values due to the hypothetical bias[16] (Fox et al. 1998); a chained approach has been suggested that first elicits the WTP for the certainty of a complete cure from a road injury and the WTA compensation for the certainty of sustaining the same injury and then a standard gamble question elicits the injuries' severity compared to death (Carthy et al. 1999). Guidelines for good practice in the derivation of willingness-to-pay (Arrow et al. 1993) and a recent guide to CV (Carson 2000) are in place to improve the state-of-practice.
- *Wage-risk method, household production function method, hedonic price method (revealed preferences):* Instead of asking people hypothetical questions one can also observe their behavior, i.e., their willingness to accept increased job risks (wage-risk approach) or their

[15] Some critics oppose the assumption that individuals' WTP should be constrained by their ability to pay that is generally dependent on their income (Gafni 1997). However, as mentioned in Section 2.4, the applicability of this criticism solely depends on how we choose to compare utility between people.
[16] Such adjustment factors may depend on the commodity and whether it is a private or public good, i.e., is not one universal factor (Fox et al. 1998).

willingness to pay for reducing individual risks (market approach). Viscusi (1983/1993/1998) presents comprehensive overviews on studies that calculate the value of statistical life (VSL) mostly from wage-risk and few market approach and CV studies. Although Viscusi controls for many confounders that may bias the ratio between increased risk to die on the job with wage-differences between high and low-risk jobs he admits that riskier jobs may be preferred by risk-seeking individuals which means that the derived VSL may understate the true values. However, further confounders like the healthy worker effect[17] and the fact that environmental risks are perceived very different from job risks may limit the usefulness of wage-risk estimates in environmental decision support tools (Hammitt 2000b). Another basic assumption is that high-risk workers know their individual risk. Viscusi (1993) states that the valuation of morbidity is more difficult than mortality because revealed methods do not work due to lack of markets. People and society also make investments in safety features like seat belts and air bags or provide regulations to reduce risks that impose costs. These values vary widely between <0 USD and 20 trillion USD (Tengs et al. 1995) and are poor proxies for perfect risk-cost markets.

2.9 Insights in elicitation methods

The descriptions above imply that much research has been done to test the methods and that there is no consensus on which method is preferable. However, there is some consensus that methods like VAS and ME do not really ask the trade-off questions at stake, and that the VAS produces ordinal rather than cardinal scales (Nord 1992a). Nevertheless, VAS is still in use since it is the cognitively least demanding method. The lacking interval property of the scale can either be dealt with by transformation functions that compress the upper and lower tails of the scale or by its exclusive use for interpolations between health states that have been valued by trade-off methods (e.g. Murray et al. 1996a).

In the other 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) (Nord 1992a). Due to these different questions, it is not surprising that the derived quality weights differ for the same judge and health condition if different elicitation methods are used. By relying on earlier studies (Froberg et al. 1989c) and closer investigations, Nord (1992a) offers a number of reasons for the observed pattern of weights in empirical studies:

Differences in what is being valued/framing

- In SG people may show risk aversion, death aversion or reluctance of gambling with one's own health which all increase quality weights.

[17] Wage-risk studies represent only a small part of the population, the working population in risky jobs (often males at age of 20-50). The 'healthy worker effect' means that workers that feel the higher risk or that are involved in an accident drop out to find less risky jobs and that the majority of the workers that stay in such jobs have actually lower risks because of their skills. This last effect is a bias because the risk is calculated based on all events while the wage-lever may be determined by this remaining high-skill majority.

- People with positive time preference will trade life years that will be lost in the distant future for smaller health improvements right now. This effect leads in the TTO to lower weights the longer the time horizon chosen (violation of constant proportional trade-off).

- The different versions of PTO are confounded by distributional considerations. If somebody prefers not to spend all health care money on one person then the disability weights tend to be skewed to high values with little difference between severe and mild conditions. Others that prefer to invest in the persons with the worst state will produce different outcomes (inbuilt distributional criterion). Therefore, it is important whether only one or many lives will be saved in exchange for treating ill persons.

- In PTO, one sacrifices the lives of others while in TTO and SG one's own life. People with an attitude that they should not sacrifice others' lives but give priority to saving those lives will state higher quality weights. However, a test of this hypothesis could not reveal such differences between individual and altruistic values (Richardson et al. 1997).

- It depends whether one asks how good or desirable a health state is or one asks to compare different illnesses.

- Since people show a *status quo* effect, they are averse towards changes (Dolan et al. 1996).

Differences in anchors[18]

- It depends whether death or full health is used as a reference state (in those methods that do not use both).

- If *worst* versus *best imaginable health state* is used to label the 0 and 100 endpoints of a scale respectively then the scale may be understood as percentages of fitness which means that the upper state is chosen as anchor and the scale interpreted as 'percentage of fitness'. This leads to lower quality weights.

- If only 'dead' or only 'perfect health' is mentioned as endpoint, this will anchor the results.

- If the scales extend the labeled endpoints, they influence the rating as well. Dolan et al. (1996) found that a large number of health outcomes score worse than death while others do not offer such weights at all.

Labeling effects (Froberg et al. 1989c)

- It depends whether elicitation under uncertainty is presented as insurance or gamble.

- Whether one offers a cash discount or credit card surcharge matters, i.e., the presentation as a gain or loss is important (Stalmeier et al. 1999, Bleichrodt et al. 2000)[19].

[18] When preferences are partly formed during the preference elicitation process, humans tend to state preferences relative (and often close) to fixed values suggested by the elicitation procedure, i.e., are anchored by them. If other anchors would yield to different preferences for the same question, anchoring is considered to enter a bias.

[19] They do not only show the importance of this bias but also show how to debias gain/loss and probability distortions. Debiasing is a research field in decision analysis and may fertilize the development for environmental decision support tools (see, e.g., George et al. 2000 for debiasing of anchoring and adjustment biases).

While some of the effects are intended because they show that indeed different types of health outcomes are valued, the strong anchoring and unintended framing effects suggest that individual preferences for health are not pre-existent but constructed during the task (Dolan 1997). Based on reasonable good re-test-reliability of the methods shown one can conclude that preferences exist at least partly[20]. However, focus groups prior[21] to and between the elicitation procedure and post elicitation questions may help to form preferences and to detect biases (Froberg et al. 1989c, Nord 1995, Dolan 1997, Johnson et al. 1998).

Another indication of preference construction rather than elicitation is the wide spread of the weights found (Torrance 1986, Nord 1995). They consider random error as important sources of the spread. While most studies showed that the values are independent from socio-economic factors or professional level (Torrance 1986, Froberg 1989c), a more recent study found small but significant dependence on age and sex (Dolan et al. 1996). Thanks to the small effect, present evidence allows us to assume the weights' independence from socio-economic factors.

Many criteria lists have been suggested to judge the different elicitation methods (see, e.g., Froberg et al. 1989b, Richardson 1994, Gold et al. 1996, Field et al. 1998, Brazier et al. 1999) but the recommendations show a broad variety. Nord (1992a) mentions that there are three reasons that the different experts do not agree on the "best" method[22]. 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. Sections 4 and 5 will elaborate on the specific applications in environmental decision support tools and make recommendations based on application-specific criteria.

Since we use health metrics to value present and future health outcomes, it is important to know whether the derived values are temporally reliable. Research in WTP methods suggests that the temporal reliability is better than assumed (Reiling et al. 1990, Carson et al. 1997). However, Cutler et al. (1998) report different QALY weights for 1970 and 1990 (although on a ordinal scale). Since the importance of physical disabilities decreases in an information society and the amenities for physically disabled people get better, this finding is not surprising. To value health effects in the future one may want to consider such predictable trends.

2.10 How to measure premature death?

Everybody dies, but when is it premature and by how many years? From the individual's perspective, premature may mean that, e.g., one is mentally not ready to die, one wants to reach a

[20] Only one-third of the judges have changed their values during interview process (Shiell et al. 2000). However, such resistance to changing former values may also be explained by other psychological factors.
[21] This is also called warm-up process (Froberg et al. 1989c).
[22] This disagreement is not only shared by researchers but also by practitioners. Rating scale (21%), TTO (18%) and SG (12%) have been found to be the most commonly used elicitation methods in a review of 228 published CUA (Bell et al. 1999).

certain round age (e.g., 80), one wants to survive the parents (or more realistically parents do not want to survive their children), one wants (not) to survive the husband/wife, or one wants to die a "natural" cause of death. However, from a statistical perspective all deaths are premature because other individuals of the same age survive. Life expectancy tables can be used to calculate how prematurely somebody died. Such tables need to be a) valid for states, nations, ethnic groups, continents or world-averages b) either averages for all individuals in the chosen area, or differentiate by sex, lifestyle factors, profession etc. and c) 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 in the future 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: "when is a death premature and by how many years?" is far from trivial.

The global burden of disease that attempts to estimate years of life lost on a globally comparable level is the place where these questions are treated very explicitly. The following propositions were made to decide on the above question (Murray et al. 1996a:6): "I: The burden calculated for like health outcomes should be the same; and II: The non-health characteristics of the individual affected by a health outcome that should be considered in calculating the associated burden of disease should be restricted to age and sex". Based on these propositions they chose a standard expected years of life lost that differentiates only between age and sex and applied it worldwide. Although the chosen model[23] is very close to the demographics of Japanese women it was corrected for peculiarities that are not health related (like war). For Japanese men they derived a theoretical genetically caused sex-gap of 2.5 years, which is less than today's observed difference[24]. Not surprisingly, this "closing of the health gap" has been criticized for its effect of increasing men's years of life lost and the potential shift of health resources to men (Anand et al. 1997). Since the life expectancy of Japanese women is the highest worldwide and much higher than in developing countries, it was criticized that the chosen approach should not be used when single health interventions have to be evaluated because this would enter a bias to save the lives of the old (Williams 1999). Whether one agrees with these objections depends on the application in mind[25] and whether the propositions apply.

Risk assessment and life cycle assessment often assess marginal health increases or decreases due to specific interventions. In these cases, non-affected 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 smaller, i.e., the assumption of ceteris paribus does not hold anymore) the approach by Murray et al. (1996a) may serve as a prototype.

---

[23] The UN model 'Coale and Demeny West Level 26'.

[24] After this adjustment, they used the 'Coale and Demeny West Level 25' model for men – although initially developed for women.

[25] National burden of disease studies used national rather than global life-tables (Melse et al. submitted, Anonymous 1999b).

Is each life year of equal value?

Here we ask whether the implicit assumption in equation (1), p.6 – that the value of one life year depends on its health state only – empirically holds or not. On the other end of extreme assumptions, estimates of the value of a statistical life (VSL) often assumed a constant value of a VSL independent of years of life lost (Viscusi 1993, ExternE 1995). Empirical studies show that in the USA and Sweden saving 85 and 35, respectively, 70-years-old is equivalent to saving one 30-year-old (see Johannesson et al. 1997a for references). This is strong evidence against the constant VSL but does also not comply with the assumptions in equation (1), if typical age-specific health states and life expectancies are assumed.

A simple consumption model that excludes dependents shows that the VSL is strongly dependent on income and follows in the case of a perfect market a slight increase until the age of 25 and then slight decreases until age of 40 and then larger decreases (Shepard et al. 1984). Based on a similar model (see also Fig 3) it is concluded that the marginal utility of money decreases with increasing age and that the real rate of interest is crucial for knowing how much the curve deviates from a monotonically decreasing function (Ng 1992). The dependence on age in these economic models occurs because the benefit of a unit decrease in mortality risk decreases with age and opportunity costs of spending decline with age. The size of the utility discount rate compared to the interest rate, the inclusion of dependents and the possibility to borrow money alter the shape and position of the curves (Hammitt 2000b). These approaches fall short because they ignore the fact that humans are social beings where friends and family matter. This last argument may work in both directions: the end of life may be higher valued because of the social environment but may also prevent that all remaining money is spent to delay the inevitable, i.e., the dead-anyway effect (Pratt et al. 1996) may be less pertinent in a social environment. However, the inverse U-shaped curve for age-dependent VSL has also been shown by two empirical WTP studies (Johannesson et al. 1997a, Carthy et al. 1999)[26].

DALYs include an inverse U-shaped age-weighting function that was included based on a number of arguments given in Murray et al. (1996a). However, as also pointed out in a discourse in Barendregt et al. (1996), Murray et al. (1996b), and Sayers et al. (1997), this age-weighting alters the life expectancy-dependant utility of life only little since the inverse U-shaped function does not replace the life expectancy table but acts just as a multiplicative modifier with most factor values between 0.5 and 1.5. Doing so means that life years lived above the age of 50 are discounted slightly more than the life-expectancy tables already suggest. This contradicts the above mentioned empirical findings and the life cycle consumption model outcomes (see also Figure 3). If an age-weighting function should be combined in a multiplicative way with life expectancy, then this function should have a U-shape rather than an *inverse* U-shape to reflect the finding that the value per year of life lost increases with age. Therefore, we do not suggest to use the age weighting suggested in Murray et al. (1996a).

[26] 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. (1984). However, the authors speculate that embedding and anchoring may have affected their results (Johanesson et al. 1997c).

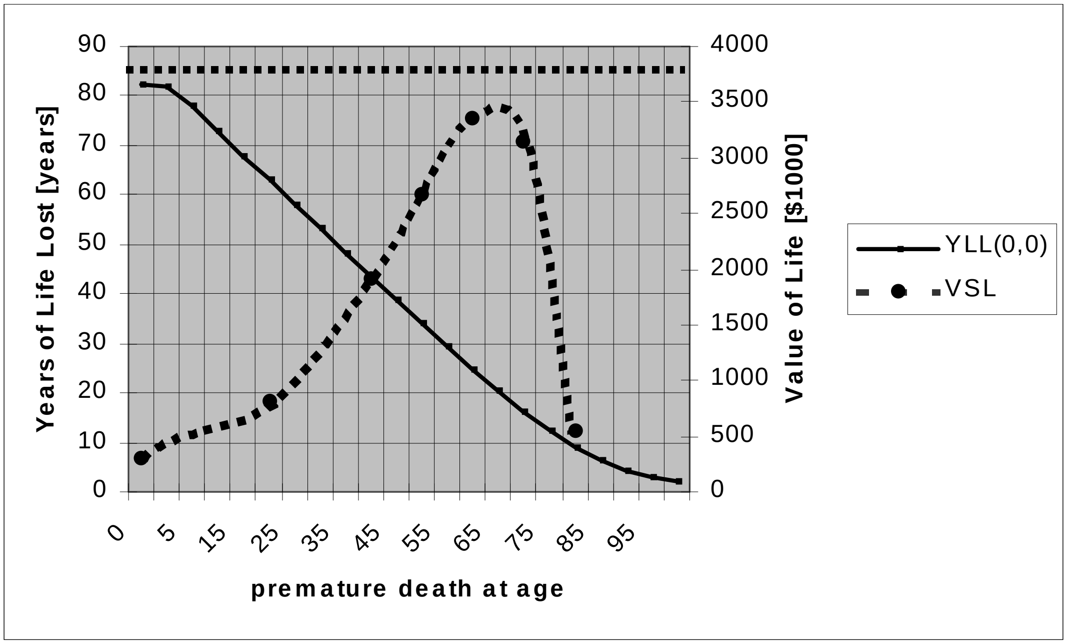

Fig 3: Examples for the different values of remaining life at different age. The solid line is measured in years life lost and represents the statistical life expectancy used in the global burden of disease (Murray et al. 1996a), the dashed line on the top is an estimate of age-independent VSL and the age-dependent VSL is taken from Ng (1992).

Since a large share of premature deaths due to environmental pollution occur at high age, it is important to know how to value these years life lost at high age. Present evidence shows that the assumption of an age-independent value of life is not supported. However, theoretical models that produce inverse U-shaped functions are over-simplistic by ignoring social interactions and their absolute values are based on a number of uncertain assumptions. The few empirical studies suffer either from potential biases (Johannesson et al. 1997c) or focus only on longevity and report much lower values than expected by common sense (Johnson et al. 1998). Therefore, an interim solution may be to rely on life expectancy alone with an additional reporting of the age-profile of the affected population or an age-weighting based on the most recent empirical findings in Carthy et al. (1999) as done in Seethaler (1999).

So far, we concentrated on the fact that the value of a life year may be a function of age and health state. However, from research in risk perception it is well known that the cause of loss and its psychometric characteristics matter when people judge risks (e.g., Fischhoff et al. 1978). Lives lost due to involuntary, unfamiliar, and catastrophic risk sources are found to be valued higher than others and lead to different WTP per life lost (Tolley et al. 1994, Ramsberg 1999, Cooksen 2000). Many environmental risks belong to involuntary, unfamiliar but chronic risks which means that the WTP is higher than average but more or less similar within this group of risks. As mentioned before, WTP is usually dependent on the individual’s ability to pay. Finally, the value of an additional year of life may also depend on the individuals' assumption whether s/he is dying prematurely or not (i.e., whether age-goal has been achieved), or on the societal assumption of a fair inning (everybody should achieve a certain age (Williams 1996), see also Section 2.14).

2.11 Time proportionality of HALYs

Section 2.5 summarized for morbidity outcomes some of the empirical evidence against the major assumption in QALYs and DALYs (=HALYs), the time proportionality. The section above provides the same evidence for mortality. Due to a lack of convincing alternatives, we concluded above that the assumption of time proportionality might be a necessary interim solution. For morbidity, the same argument was made also claiming that the deviations are small (Dolan 1996). However, for both mortality and morbidity there are examples where the deviations are major and examples could be constructed that show preference reversal if age and duration-dependency are considered respectively.

WTP studies using stated preferences (Alberini et al.1997) and conjoint analysis (Johnson et al. 1998/2000) show for short term outcomes like cough or asthma attacks strong non-proportionalities if durations are 1, 5 or 10 days. For acute and/or short-term health effects due to air pollution Johnson et al. find that $ln(d+1)$ where d are the numbers of days shows approximately a linear behavior as time factor in their attribute based stated choice analysis[27]. This is the only alternative proposal we found in the literature to incorporate the duration in a non-linear way.

A correction of time-proportionality for morbidity should be able to deal with two major effects: change aversion and adaptation. Since the environmental context implies that we can prevent health effects from occurring we are in an *ex ante* situation. The above findings can partly be explained by this effect. There is a strong aversion to get sick at all, i.e., to change the health state (status quo effect). Further, it seems to be important whether the health state is perceived to be fully reversible. Whether reversibility is assumed or not depends probably on the predicted time duration in the bad health state (Sackett et al. 1978)[28]. Therefore, aversion against both change of health state and perceived irreversibility should be accounted. When we discussed the differences between patient and non-patient values, we already mentioned that adaptation to health outcomes increases the perceived quality of life. This can also be seen as a marginal decrease in dis-utility and is not to be confused with time preference. The empirical effort to estimate the additionally needed parameters to take into account the mentioned deviations from time-proportional weights is huge. However, instead of investing in further research that confirms the lacking time-proportionality one could estimate these parameters and functions. These estimates could then be used in the screening phase of applications to get an estimate whether the assumption of time-proportionality enters a relevant bias or not. If the bias appears to be major, HYE, WTP or a program evaluation following proposals

[27] Part of the found effect may also be caused by scope insensitivity, i.e., a fixed budget for averting mild illnesses that is insensitive to the number of days.
[28] Although this early study is often cited when the time-proportionality is questioned it has not been pointed out that the study provides evidence for marginal increase rather than decrease of dis-utility.

of Nord (1999) may be most efficient and useful. In all other cases, the much simpler time proportional approaches may be acceptable.

2.12 Short-term and chronic effects

QALYs have explicitly been developed for chronic health outcomes (Pliskin et al. 1980) and DALYs concentrate usually on permanent conditions (AbouZahr et al. 2000). However, the application in medical decision making and environmental decision support tools makes it necessary that both short-term and chronic health effects can be evaluated (Alberini et al. 1997, Johnson et al. 1998, Bala et al. 2000).

Stouthard et al. (1997) distinguish diseases with an episodic pattern (e.g., asthma, migraine) and short-term conditions with full recovery (e.g., colds, gastroenteritis). The episodic diseases have been described as chronic outcomes while short-term conditions with full recovery have been presented in an annualized profile, e.g., 50 weeks of perfect health and 2 weeks of a cold. If time-proportionality applies then the latter example would lead to a quality weight of at least 0.96, even if the cold would be perceived as equally severe as death. However, as discussed above, aversion against changes in health states may justify different values. Therefore, the judges need not be forced to comply with time-proportionality for short-term conditions and the procedure suggested by Stouthard et al. (1997) may be a pragmatic solution.

2.13 Multipathology/co-morbidity

People often suffer not one health outcome but different (mild) disabilities at the same time. In Beaver Dam, Wisconsin, a township in the USA, 1356 individuals above the age of 45 rated their own health with different methods. About 20% of the individuals had no, one, two or three health conditions, respectively. The remaining 20% had as many as 4 to 10 different health conditions (Fryback et al. 1993). Epidemiological studies that are used to estimate dose-response relationships in environmental decision support tools do report all health endpoints that are considered to be caused by mechanisms triggered by the specific agent. Therefore, it does not matter whether the different health outcomes are causally related or not. However, the question arises how the quality weights can be added if different health effects affect the same individual and if this individual shows age-related deviation from perfect health? This question is rarely addressed in the literature and has been mentioned as a shortcoming of the DALYs approach (Williams 1999, Sayers 1997). Anonymous (1999) adjusts for co-morbidity by assuming a multiplicative model among morbidities. They were interested in allocating the burden of disease to different causes. Therefore, they also assume that the most severe state gets the full quality weight while the quality weights of the less severe co-morbidities are adjusted. If there are two health outcomes with QWa and QWb, and outcome (a) is the more severe outcome of (a) and (b) then

$$QWa_{comorbidity} = QWa \quad \text{and} \quad QWb_{comorbidity} = 1 - (QWa - QWa*QWb) \qquad (5)$$

Due to the high share of correlated morbidities within mental disorders and within injuries, different procedures have been suggested for these outcomes (Anonymous 1999a). Since the purpose of environmental decision support tools is not to find a just allocation to single morbidities but to estimate a decrease or increase in overall health state we only need guidance on how to calculate co-morbidities and not on how to allocate disutility to single morbidities. For this purpose, we suggest using the multiplicative model. Instead of excellent health, many CUA studies use the absence of the disease under study as the upper end for quality weight. Such quality weights have to be adjusted by the age-related quality weight (Fryback et al. 1993). We suggest that the age-related quality weight is QWa and the morbidity under study QWb and use equation (5) to adjust QWb, i.e., the age-related quality weight is kept constant. Age-related quality weights can be found in Fryback et al. (1993) and Bell et al. (1999).

2.14 Utility maximization versus distributional/ethical considerations

Although none of the discussed health metrics empirically satisfy the strong assumptions of von Neumann and Morgenstern utilities (see section 2.5) they have been developed under the assumption that health measured by these metrics should be maximized; this is called utility maximization. This policy is usually followed by consequentalists who are primarily concerned with the health outcome attained. Other policy alternatives concentrate on the process by which health is achieved or the opportunities people have to obtain health (Holmes 1995). Since the maximization of all three policy goals is usually not possible (Rawls 1971), a choice has to be made at this stage. Environmental decision support tools considered here attempt to minimize health effects. Therefore, they require the consequentalists' view, which will be discussed here in more detail.

A lot of research in medical ethics has analyzed whether people agree to maximize QALY and HYE or minimize DALYs and WTP as a sole criterion for resource allocation. A number of deviations from this sole reliance on metrics have been found:

- People want to improve the situation for the *worst-off first* (behind veil of ignorance, see e.g., Rawls 1971, Andersson et al. 1999). This is also known as the *severity criterion*, see Nord (1999) for a review[29].
- Three groups of people can be differentiated: 1) Utility maximizers that accept the health metric as the only criterion, 2) diffusers that prefer to spend health care resources among all with disabilities and not just for the patients with the largest increase in health, and 3) concentrators that prefer to spend the resources on fewer patients with visible improvements[30] (Olsen 2000, Richardson et al. 1997). Others call this the *realization potential*, i.e., that group with the larger improvement potential may (or may not) be treated first, see Nord (1999) for a review.

[29] If the quality weights show a so-called upper-end compression, i.e., that only very severe health states get quality weights below 0.65 but most health states are between 0.9 and 0.999, then this severity argument can in most cases be fulfilled by the health metric. Due to death aversion, such upper-end compression is expected from utility measures (see Nord (1999) for a review).

[30] Olson (2000) also finds that a threshold for minimal improvements may exist for the concentrators.

- While 70% of the judges of a convenience panel mentioned that the maximization criterion should be the most important allocation criterion for donor liver grafts only 0.7% finally 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 it is already the *second transplant* (Radcliffe 2000).

- *Survival* is judged by patients as much more important than perfect health. The present health metrics may underestimate the importance of survival (Nord 1999, Cohen 1996). However, Johnson et al. (1998) show that the prolongation of life at poor health gets very low or even zero WTP.

- As mentioned earlier, the *fair innings* argument claims that everybody should enjoy the healthiest life possible until a certain age (70-75 years) (Williams 1996). This is also known as equality argument, see Nord (1999) for a review.

- When values of WTP are derived one typically assumes that the current *distribution of income* among individuals is appropriate. Therefore, WTP has been criticized to violate equity principles. However, if WTP is used within a country and within the health sector alone this assumption may be unproblematic or adjustments can be made (Kenkel 1997, Donaldson 1999). The finding that socio-economic factors have no influence on health quality weights supports this claim if the population is concerned by health outcomes to the same extent or if average WTP are used for all population groups. On a global level, 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 and – if co-morbidity is calculated following equation (5) – additional health outcome may count less as well. (Singer et al. 1995, Koch 2000a/b) This problem was also found when the health loss of HIV infected subpopulation due to drinking water impurities is assessed[31].

- Due to the *limited dimensionality* of health metrics, it was found that the sensitivity for certain groups of health outcomes might be weak and therefore set biased priorities. This point was made with respect to mental health care (Chisholm et al. 1997) and sexual and reproductive health conditions (AbouZahr et al. 2000). However, several instruments consider non-physical disabilities and both Murray et al. (1996a) and Anonymous (1999a) show major shares of DALYs attributed to non-physical health outcomes.

This summary of arguments mostly against pure utility maximization leads to the question whether health metrics are useful at all, whether they should be adjusted accordingly to account for the

[31] In this case it is even a triple jeopardy: 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 (USEPA 1998a). Therefore, this subgroup was analyzed separately to allow for tailored risk management.

mentioned points or whether these points should be considered in other phases of the decision making process. Most authors, even the ones that are critical about many features of HALYs, agree that health metrics are important and useful as long as they are not seen as ultimate measures of quality of life and as long as other criteria are used as well in decision making (Dougherty 1994, Singer et al. 1995, Holmes 1995, Williams 1996). Contrary to this, Leonard et al. (1986:41) conclude "it is generally undesirable to include them [distributional considerations] in project analysis". They feel that this would distort the CBA or CUA.

Since environmental decision support tools may (risk assessment for regulation) or may not (life cycle assessment) make protective decisions that are directed towards a specific social or patient group the considerations of the mentioned points will be revisited in Section 3. The share of Norwegian politicians opting for the pure utility maximization was for the social democrats about half that of the conservatives (Nord 1999:130). Therefore, political orientations lead to different distributional judgments among politicians and let us conclude that a transparent breakdown of total HALYs or WTP or HYE has to be provided to allow for distributional judgments. Such breakdowns should be made for severity, realization of potential, groups with pre-existing disabilities, age, and timing of effect[32].

2.15 Beyond disutility: costs of illness and averting behavior

We focused so far on the individually borne disutility associated with health outcomes. However, Table I shows that there are also individually borne costs due to morbidity and collectively borne consequences. The individual WTP is supposed to include all individually borne or private costs while the social costs would include both individually and collectively borne costs. In medical decision making, the ratio between cost of a specific intervention (medical and production cost of illness (COI)[33]) and the gain in health due to that intervention is used to identify the most efficient treatments. However, in environmental decision making investments are made to avoid the cause of adverse health outcomes. The benefit of these investments is the avoidance of 'cost of illness' due to treatment and production loss, of 'cost of averting behavior', and 'intangible costs'.

External costs due to illnesses caused by environmental impacts are sometimes estimated as a multiple of COI[34] (see, e.g., ESEERCO 1995, ExternE 1995). Table II presents selected willingness to pay values to avoid health conditions that result from air pollution. The calculated WTP/COI ratios span a wide range, suggesting this rule of thumb is not very accurate.

[32] Nord (1999) 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, responsibility of affected person for caring for others, effect on patient's productivity. He also suggests that sex, race, education and income should not be used as criteria.

[33] See Gold et al. (1996) and Weinstein et al. (1997) for guidance on which cost factors are included in the nominator and denominator.

[34] Sources for COI in the USA can be found in USEPA 1998b, USDL/BLS 1999, Leigh et al. 1997, Hoffman et al. 1996, Elixhauser et al. 1999.

Tab. I: Overview on the costs of morbidity (adapted from Seethaler 1999). Dark shaded indicates 'included in health metrics', light shaded indicates 'market prices are available' and no shading indicates 'usually neglected'.

	Cost of illness (medical)	Cost of illness (production)	Cost of averting behavior	Intangible costs
Collectively borne	Treatment cost (health care, infrastructure, medication etc.)	Loss of production (GDP)	Averting expenditures (noise protection walls, water treatment plants etc.)	Disutility associated with health outcome (effects on family, friends etc.)
Individually borne	Treatment cost (health insurance, medication etc.)	Loss of production (household income)	Averting expenditures (water and air filters in private homes, no (cheap) outdoor sport during high ozone periods etc.)	Disutility associated with health outcome

If one wants to include collectively borne COI in the WTP estimates one could assume that about half of the medical costs for hospital admissions are borne collectively and add them to the individual WTP which would increase these values by 3330 and 4080 EUR for respiratory and cardiovascular hospital admissions respectively. This large increase is not found for other conditions where only minor increases can be calculated. Therefore, depending on the study's goals[35] and endpoints, each cell in Table I may be included in the calculation of health benefits for environmental decision making.

Tab. II: Values for willingness to pay (WTP) and cost of illness (COI) for five health conditions caused by air pollution (Seethaler 1999).

Health condition	WTP (1996 EUR)	COI (1996 EUR)	Ratio WTP/COI
Respiratory Hospital Admissions	7870 per admission	7910 per admission	1
Cardiovascular Hospital Admission	7870 per admission	9700 per admission	0.8
Chronic Bronchitis (adults ≥25 years)	209'000 per case	3300 per case	63
Bronchitis (children, <15 years)	131 per case	33 per case	4
Asthmatics: Asthma Attacks (person day)	31 per attack	0.55 per day	56

2.16 What is not measured by health metrics?

Following the arguments of the previous sections, we can summarize that

1. Health metrics are generally following the paradigm of utility maximization and incorporate only one out of many sets of distributional and ethical justice.

2. None of the major health metrics covers all of the cells presented in Table I. While WTP attempts to cover all individually borne costs it usually neglects collectively borne costs altogether[36]. HALYs and HYE are concerned with the individually borne intangible costs. The

[35] According to ISO (1997), Life Cycle Assessment includes effects on human health, ecosystems and natural resources. Therefore, only intangible costs would be included directly while environmental impacts due to treatment, production loss and avertable behavior would be separately considered if relevant.
[36] However, such collectively borne costs could be listed when WTP values are derived and, e.g., intangible costs may well be included when the elicitation instruments make clear that affected family members and friends shall be considered as well. Free-rider-problems may be expected with other collectively borne costs.

DALYs, based on the use of the PTO method, is the only one that may include some aspect of intangible costs that are borne by the society.

3. Quality of life has probably a broader meaning than actually reflected in health metrics.

The following comments can be made with regard the importance of these issues and how to deal with them:

ad 1: Parts of the problems occur because of the general problems with aggregating individual preferences to a social welfare function. Whether altruistic or individual preferences are more important is a question of paradigm and not a unique problem encountered only here. As suggested earlier, providing a desegregation of the damage score measured with a specific health metric will make sure that distributional considerations can be considered in decision making.

ad 2: This finding suggests that before a human health metric is chosen, it has to be known from the decision makers which cells of Table I shall be covered. In many cases this would mean that HALYs and HYE have to be complemented by information on COI and costs of averting behavior while WTP estimates may need to be complemented by collectively borne costs. Surprisingly, little research results are available on intangible costs borne by the patients' family and friends and people providing health care to the patient. This may lead to a systematic underestimation of health damages.

ad 3: A comfortable life, equality, an exciting life, happiness, health, individual freedom, mature love, pleasure, salvation, security, self-preservation, self-respect, a sense of accomplishment, a sense of community, social recognition, true friendship, wisdom, a world of beauty, a world at peace, inner harmony are all values that have been suggested to be important human values that contribute to a high quality of life (Rokeach (1973) and Kristiansen (1985)). Although some of them are not, most are related directly or indirectly to health conditions. Their inclusion or exclusion may depend on the information[37] provided in the elicitation procedure.

2.17 Practical aspects

The availability of consistently derived quality weights for a large number of health states may be considered as a practical advantage, especially if the decision support is needed within a short time or with little resources. The following are sources for such tables known to us (see also Section 3 for additional references to sources for environmental related diseases):

- *QALY weights (holistic and decomposed)*: Quality weights are published from the Beaver Dam study for 28 health conditions (Fryback et al. 1993), from the US health census for 10 health

[37] Information is understood in its broad sense including warm-up sessions, focus groups or introducing these values as explicit attributes.

states (Cutler et al. 1998), from a comprehensive review of CUA studies including almost 1000 quality weights measured by different instruments and different judge groups (Bell et al. 1999).

- *QALY recipe (explicitly decomposed):* While some of the weights above are also based on decomposed approaches Kaplan et al. (1988), Rosser et al. (1972), Patrick et al. (1993), Fryback et al. (1997) and Torrance et al. (1972/1986) provide overviews on decomposed approaches, their aggregation rules and suggest weights to be used.

- *HYE:* No compilation is known to us.

- *DALY weights:* Several hundred consistent disability weights are reported in Murray et al. (1996a) and recommended for a worldwide application. For 56 diagnostic groups separating more than 100 different disease stages disability weights for The Netherlands have been derived (Stouthard et al. 1997/2000). Environmental disease related disability weights have been provided by de Hollander et al. (1999) based on Stouthard et al. (1997). Anonymous (1999a/b) build on Murray et al. (1996a) and Stouthard et al. (1997) and add some additional disability weights (by interpolation) for the specific Australian context.

- *WTP:* An overview on morbidity costs for acute and chronic symptoms, value estimates for dysfunctions and a list of cause dependent VSL is provided by Tolley et al. (1994). Most sources are old, derived in different contexts and with different elicitation methods. Environmental disease related WTP estimates have recently been published or re-compiled by Magat et al. 1996, Alberini et al. 1997, Johnson et al. 1998/2000, Blumenschein et al. 1999, Seethaler 1999, ExternE 1999, USEPA 1999a.

This incomplete compilation suggests that there are two reasonable large and consistent data sets for world-wide and Dutch disability weights published and that the explicitly decomposed systems to calculate QALYs can be seen as another source of consistent information[38] for different regions (mostly for the North America (QWB, HUI, Rosser-Index) and Europe (EuroQoL)).

The application of health metrics implies also the knowledge on the age-distribution of affected individuals and the duration of diseases. For this purpose, information on incidence rates, prevalence and additional disease-specific knowledge has often to be combined. Methodologies and simple software tools have been developed for this matching process (Murray et al. 1996a, Anonymous 1999a and Hoogenveen et al. 2000).

[38] Consistent refers to the internal consistency of the data set. However, the scales' cardinal property can often be disputed and the health conditions to be valued need also to be consistently characterized by the quality of life scoring instrument.

2.18 Authorization of health metrics

Decision makers may prefer to rely on health metrics that have been authorized as standard or state-of-the-art approach. The global burden of disease study and its disability weights performed on behalf of the World Health Organization and the Worldbank is probably the most authorized source for a health metric[39].

On a national level Gold et al. (1996) tried to set a standard for the USA by making a number of recommendations that narrow down the number of alternatives to HALY-type of approaches. They also favor TTO as the elicitation method and recommend using a social discount rate. Since EPA performs benefit-cost analysis (BCA) rather than CUA, they use WTP. Such governmental use of an approach to support policy making can also be seen as an attempt to authorize a method. The same may hold for the Dutch burden of disease study (Melse et al. submitted).

[39] This is also reflected by the many attempts to criticize the approach (most critiques focus on points that can be criticized with almost all health metrics. More specific points have been the one of the used versions of PTO, age-weighting, and the use of one standard life table for all countries (AbouZahr et al. 2000, Anand et al. 1997, Arnesen et al. 1999, Barendregt et al. 1996, Elbasha 2000, Hanson 1999, Mansourian 1996, Sayers et al. 1997, Williams 1999/2000). However, the fact that Murray and Lopez replied on many critical articles (Murray et al. 1996b/1997/2000) may also be an indication that the approach is still within the research sphere.

3. Comparison of DALYs, QALYs and WTP based on an example

Table III presents a comparison of the three most widely used summary health metrics (DALYs, QALYs and WTP) applying them to health effects due to environmental risk factors. From this comparison, we expect further insights into the practical relevance of some of the theoretical aspects discussed in Section 2. The health effects have been assessed for The Netherlands within the Fourth National Environmental Outlook 1997-2020 and have been directly taken from de Hollander et al. (1999). For pragmatic reasons we excluded risk factors that are not strictly caused by (external) environmental pollution like accidents, environmental tobacco smoke or damp houses and also exclude a large number of carcinogens that contribute only little to the total health effects and add little insight for the comparison. The remaining five risk factors[1] are therefore neither a complete set of all environmental health effects in the Netherlands nor necessarily the most important ones. For mortality and acute morbidity incidence data with additional estimates for the duration of diseases have been used. The life years lost by premature death is estimated based on Dutch life tables that are very similar to the standard table used by Murray et al. (1996a). For chronic morbidity, prevalence data has been used (see columns 3 and 4 in Table III).

We provide here only best estimates without additional information on the uncertainty and variability. However, many of the used sources like de Hollander et al. (1999), Bell et al. (1999), Tolley et al. (1994) and USEPA (1999a) provide additional information that would allow a probabilistic analysis.

All three health metrics could be used with or without time discounting. Here, we analyze health effects in the same year and discounting should therefore not alter the presented results. However, there are two exceptions: (1) the neurocognitive effects of lead is the only chronic morbidity that has been analyzed based on an incidence basis and (2) for mortality the incidence rate was used combined with estimates of years of life lost. We assume here that the prevalence rate for these effects would roughly be the incidence cases multiplied by the assumed duration, i.e., that these incidence rates have been constant over the last decades. Accepting this assumption makes that all health effects actually happen in the same year and time discounting becomes a non-issue[2].

The disability weights for the calculation of DALYs (column 5-7) have been taken from de Hollander et al. 1999, where ‘0’ stands for perfect health and ‘1’ for death. They based their weights on Stouthard et al. (1997), Murray et al. (1996a) and an own panel of environment-oriented physicians adjusting for the health consequences typical for environmental exposure. The resulting numbers in column 6 are slightly different from the numbers in de Hollander et al. (1999) due to rounding errors. Age weighting, as suggested by Murray et al. (1996a) for one version of DALYs has not been applied in de Hollander et al. (1999).

[1] Long-term effects from particles smaller 10μm (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

[2] It needs to be reminded here that the time-tradeoff method (TTO) has an inbuilt time discounting that – in principle- would need to be corrected for (Johannesson et al. 1994).

QALYs are calculated in columns 8-11 using quality weights from different sources and sometimes using the same weight as provided by de Hollander et al. (1999) (perfect health ‘1’, death ‘0’). The quality weights are not consistent, different elicitation techniques and groups of judges have been used and in some cases rough approximations had to be made. The most relevant assumption concerns 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 only quality weights available in the literature. Since we evaluate the decrease in health due to environmental risk factors, the decrease in QALYs has been calculated (ΔQALYs). The values taken from Fryback et al. (1993) have been adjusted for co-morbidity. It is assumed either that the other values have been adjusted as well, or that the effect under study is the major health condition, or that the difference is minor. However, we did not account for the decreased utility of life years lost at higher ages due to co-morbidities. To do so one would need the information on the age-distribution of the premature death that was not provided in de Hollander et al. (1999).

The WTP values are effectively a mixture of WTP values (based on contingent valuation or/and labor market studies and hedonic price methods for noise) and COI or an estimate based on COI. This inconsistency is slightly reduced by heavily relying on one compilation of values (USEPA 1999b). All values have also been transformed to 1990 USD. Since USEPA (1999b) uses in the baseline scenario the VSL approach without adjustment for age this assumption has been adapted. More sophisticated approaches use age-adjusted VSL values (Seethaler 1999).

Tab. III: Health consequences for five environmental risk factors evaluated by three different health metrics (only best estimates are shown, number of given digits does not suggest that these are significant digits)

Risk factors	Health effects	incidence or prevalence cases per year	duration [a]	disability weight (a)	DALYs (a)	DALYs[%]	quality weight	Ref	QALY	QALY [%]	WTP or COI [1990$] per case	Ref	WTP or COI [Mio 1990$]	WTP/COI [%]
PMlt	mortality total	7114	10.9	1	77543	40.94%	0		77543	19.28%	4800000	b,c	34147	35.99%
	mortality cardiopulmonary	8041	8.2	1	65936	34.81%	0		65936	16.40%	4800000	b,c	38597	40.68%
	mortality lung cancer	439	13	1	5707	3.01%	0		5707	1.42%	4800000	b,c	2107	2.22%
	chronic respiratory symptoms, children	10138	1	0.17	1723	0.91%	0.86	d	1419	0.35%	28946	b,e	293	0.31%
	chronic bronchitis, adults	4085	1	0.31	1266	0.67%	0.86	d	572	0.14%	28946	b,e,f	118	0.12%
	Total				*152176*	*80.35%*			*151177*	*37.59%*			*75263*	*79.32%*
O3	mortality respiratory	198	0.25	r 0.7	35	0.02%	0		50	0.01%	4800000	b,c	950	1.00%
	mortality coronary heart disease	1946	0.25	r 0.7	341	0.18%	0		487	0.12%	4800000	b,c	9341	9.84%
	mortality pneumonia	751	0.25	r 0.7	131	0.07%	0		188	0.05%	4800000	b,c	3605	3.80%
	mortality other	945	0.25	r 0.7	165	0.09%	0		236	0.06%	4800000	b,c	4536	4.78%
	hospital admission, Respiratory	4490	0.038	0.64	109	0.06%	0.56	g	75	0.02%	6000	b,h,i	27	0.03%
	ERV, Respiratory	30840	0.033	0.51	519	0.27%	0.49	j	519	0.13%	194	b,k	72	0.08%
	Total				*1300*	*0.69%*			*1554*	*0.39%*			*18531*	*19.53%*
Lead (*)	Neurocognitive development (1-3 IQ-points)	1764	70	0.06	7409	3.91%	0.94	j	7409	1.84%	10005	k	18	0.02%
Noise	Psychosocial effects: severe annoyance	1767000	1	0.01	17670	9.33%	0.91	d,l	159030	39.54%	265	m,n	468	0.49%
	Psychosocial effects: sleep disturbance	1030000	1	0.01	10300	5.44%	0.92	d, o	82400	20.49%	265	m,n	273	0.29%
	Hospital admissions IHD	3830	0.038	0.35	51	0.03%	0.56	g	64	0.02%	9000	b,h,i	34	0.04%
	Mortality IHD	40	0.25	r 0.7	7	0.00%	0		10	0.00%	4800000	b,c	192	0.20%
	Total				*28028*	*14.80%*			*241504*	*60.05%*			*968*	*1.02%*
Ozone depletion	Melanoma morbidity	24	6.9	0.1	17	0.01%	0.7	p	50	0.01%	8218	q	0	0.00%
	Melanoma mortality	7	23	1	161	0.09%	0		161	0.04%	4800000	b,c	34	0.04%
	Basal	2150	0.21	0.053	24	0.01%	0.947	j	24	0.01%	4696	q	10	0.01%
	Squamous	340	1.5	0.027	14	0.01%	0.973	j	14	0.00%	8218	q	3	0.00%
	other mortality	13	20.2	1	263	0.14%	0		263	0.07%	4800000	b,c	62	0.07%
	Total				*478*	*0.25%*			*511*	*0.13%*			*109*	*0.11%*
Total					189390	1			402155	1			94888	1
mortality						79.35%				37.44%				98.61%
morbidity						20.65%				62.56%				1.39%

(*) from drinking water pipes
a) de Hollander et al. 1999
b) based on USEPA (1999) in 1990$. Most values are based on incidence cases and refer to health effects due to air pollution.
c) This central estimate is slightly higher than the values in Tolley et al. 1994 but derived based on a large body of literature reviewed in USEPA (1999a). However, Seethaler (1999) argue that the underlying studies have been biased and use values derived by a chained approach (Carthy et al. 1999) for road accident victims and adjust those values for the higher age of air pollution victims ending up with 0.9 million EUR(1996). This value is about 5 times lower than value suggested by USEPA (1999a).
d)Fryback et al. 1993 (TTO, general public >45a). Since the data allows correcting for co-morbidity, we subtract the mean for persons affected by the condition from the mean for the persons unaffected by the condition.
e) USEPA (1999b) bases their values on incidence, therefore the value of Tolley et al. (1994) is taken for a yearly value and adjusted from 1991$ to 1990$ by multiplying by 0.965.
f) Viscusi et al. 1991 derived a total value of $516000-904000 (adjusted to 1990$, based on two different elicitation methods)), considering discounting of future years this equals to an assumed duration of CB of >20a, which is confirms the order of magnitude. Krupnick et al. 1992 (see Viscusi 1993) estimate a media value of $496800-$691200 in 1990$, again the same range.

g) Sackett et al. 1978, value based on 3 months of 'Hospital confinement for an unnamed contagious disease'. Based on data one expects higher quality for shorter admission time and known rather than unknown cause. (TTO, general public)
h) Derived by dividing the total mean welfare benefits in Table H4 by the change in incidence of cases in Table D-21 which results in costs per admission (not per day) (USEPA 1999b).
i) These values are comparable to what is used in other studies (Seethaler et al. 1999) but inconsistent with findings of Johnson et al. (2000) who find much lower values below 1000$. They use conjoint analysis and different duration periods. One day alone would account for 535 1997 Can$. Multiplied by typical durations of 11 to 14 days would result in slightly lower values than reported by USEPA (1999a). However, the COI given in Seethaler et al. (1999) are alone about the same amount as the WTP reported.
j) No appropriate quality weights have been found in the literature, therefore the disability weight from de Hollander et al. 1999 has been used here.
k) Levin (1997) estimates the damage due to a decrease of one IQ point to be a loss in future earnings of 1.76% or $4600 (1988). We double this value for an average loss of two IQ points and adjust for 1990$ with a factor of 1.0785.
k) The given value is the COI for one ERV. However, Hollander et al. (1999) describe the health state as a weighted average of duration of exacerbations requiring ERV or hospital admission. Therefore, we assume we multiply the COI by the given duration.
l) Assumption that 'annoyance' can be described with 'anxiety', which is obviously a different severity level
m) Banfi et al. (2000) estimate the traffic related WTP to avoid disturbance by noise for the Netherlands using both hedonic price methods and contingent valuation and assuming a threshold of WTP at 55 dB(A). This results in 1087 million ECU(1995) per year (=740 Mio US$ 1990).
n) The total WTP to avoid disturbance from traffic noise is allocated to severe annoyance and sleeping disturbance assuming that these cases have an equal severity (as suggested by QALY and DALY). This results in (740 Mio US$/2.797 Mio cases = 265 US$ per case and year).
o) Assumption that 'sleep disturbance' matches 'sleep disorder', which is obviously a different severity level.
p) Bell et al. 1999 cite 216 (author judgments) and 249 (clinician judgment). Metastatic conditions and recurrent melanoma get both an average of 0.5, treatment causes quality weights of 0.7-0.8 and remission after surgery 0.9. An average weight of 0.7 is assumed.
q) Dickie et al. 1996 find WTP to avoid skin cancer cases in the range of $720-1200. However, they cite an EPA study that report medical treatment costs for basal and squamous cell carcinomas cost $4000 and $7000 respectively. We adjust these values for 1990$ and take the higher value as well for the melanoma. All costs are per case.
r) This disability weight applies to a period of disease before death plus the period of the premature death.

Based on these assumptions it was possible to calculate the total DALYs, QALYs and cost consequences due to the five risk factors and to compare their relative shares between the health metrics.

The following insights are important:

- The resulting DALYs and loss of QALYs can be compared to about 15 million years of life lived per year in The Netherlands. Therefore, the relative share of the burden of disease for these five environmental risk factors together compared with the total years of life lived lies between 1.3% (DALYs) and 2.7% (QALYs). The health risk costs of 95 billion USD (almost completely intangible costs) amount to about 30% of the Dutch GDP in 1990! The magnitude of this amount suggests either that major budget adjustments are warranted or that the value of a statistical life is less in this application or that the estimate of particle related health effects are too high.

- The share of (premature) mortality on the total health burden 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 the better estimate. Therefore, we can conclude that all health metrics are heavily influenced by mortality outcomes but that in this application WTP/COI seems to make a morbidity assessment unnecessary (last column Table IV).

- The assessment of the relative importance of noise is very different between the three metrics (DALYs 15%. QALYs 60%, WTP 1%). We already mentioned that the quality weight for QALYs was based on a crude assumption. A separate study to elicit such values or the use of an explicitly decomposed instrument would be needed to improve this estimate. The disability weights for sleep disturbance and severe annoyance derived by de Hollander et al. (1999) have been 0.01. Müller-Wenk (1999) derived for the same endpoints disability weight using a small convenience panel of six physicians. The mean weight was 0.048 for communication disturbances and 0.05 for sleep disturbance and a larger study that is more representative has been planned. The example of noise shows how the relative importance of a mild morbidity outcome is very sensitive on the quality weights and metric used. This special situation usually does not occur in medical decision making. The reason for the high sensitivity is, first, the large number of affected people and secondly by the large relative impact of uncertainties in small changes of the quality weights. In Sections 2.8 and 2.9 it was mentioned that most methods work worse for outcomes of low severity, since people are reluctant to trade premature death for mild disabilities at all and since the trade-off numbers get either very large (PTO) or very small (TTO, SG) or beyond the possibilities of graphical methods (VAS). This will be further discussed below.

- The increased mortality rate due to increased ozone levels is considered to affect old or already sick people. This fact is reflected in the DALYs and QALYs calculations and leads to minor health damages. However, if VSL is used without age-adjustment, increased ozone levels are (probably wrongly) judged very relevant.

- Increased UV-A and UV-B radiation is so far no problem in The Netherlands, only few cases occur and the mortality rate is very low. Uncertainty in the morbidity weights and costs hardly influence the outcome. The same holds true for most morbidity outcomes (not for noise and neurocognitive effects), where uncertainty in the morbidity weights or costs hardly matter.

- While the rank order is stable between DALYs and QALYs (only noise gets different ranking which may be an artifact), the WTP suggest that increased ozone level should get high attention while lead exposure from drinking water pipes is a very minor problem (see Table IV). This rank order reversal is due to the dominance of mortality rates in the WTP approach.

Tab. IV: Rank order of the five environmental risk factors if evaluated by different health metrics

	DALYs	ΔQALYs	WTP/COI	Mortality
Long term effects of PM_{10}	1	2	1	1
Increased tropospheric ozone concentrations	4	4	2	2
Lead from drinking water pipes	3	3	5	5
Noise	2	1	3	3
Increased UV levels due to stratospheric ozone depletion	5	5	4	4

- The ranking of risk factors and the discussion above was based on the utility maximizing paradigm. However, these health damages are not equally distributed among the population. Major health damages due to exposure to fine particles and ozone occur at higher ages or in already sick people, lead poisoning affects a small number of children with life-long consequences, noise affects those who cannot afford a living/working place free from traffic noise, and ozone depletion affects the group of people with fair skin or extensive exposure to the sun (sun-bathing, construction workers, farmers, etc.). Will this additional information on the affected population alter the ranking? Let us reconsider some of the arguments summarized in Section 2.14:

- *Improve situation for worst-off and support survival.* This would suggest that the mortality rate should be reduced and would support the ranking derived by WTP.

- *Support high realization-potential group.* The largest realization-potential can be found among health risks causing premature death with many years of life lost like the cancer cases due to ozone depletion and mortality by long-term effects of particulate exposure. This may give a higher priority to prevent ozone depletion than suggested by Table IV.

- *Improve situation for young and innocent.* Here we assume that all subjects are equally innocent since the considered environmental risk factors are only loosely attributed to lifestyle factors (maybe with the exception of sun-bathing). All risk factors affect children and young adults to some extent. However, neurocognitive effects from lead poisoning may be considered as typical risk factors affecting children and should get a higher priority than suggested by the WTP metric.

- *Allow for fair-innings.* This criterion would need a reanalysis of the data with a threshold-age of 70 to 75 years beyond which health loss would not be considered. Health damages due to particulate and ozone exposure would drop dramatically in such an analysis. Other health risks would probably be less affected.

- *Income should not matter*. Since we assumed that impacts are distributed across population uniformly, we assumed the distributional concerns “away”. However, since the WTP values for noise have mostly been derived from hedonic price methods we have an estimate of how much noise one socio-economic group (home-owners) is ready to trade for money but the same information is not available for the other income groups.

- *Correct for double jeopardy*. None of the considered environmental risk factors is supposed to affect physically handicapped individuals more than non-handicapped. However, respiratory symptoms and premature death due to particulate exposure and ozone is known to affect already sick people to a larger extent.

- *Consider overlooked dimensions.* It is not obvious that important characteristics of the included health endpoints are overlooked by the used health metrics.

These different distributional concerns point partly in different directions but may suggest that lead poisoning and ozone depletion may get slightly more importance than suggested by all health metrics. We suggest here that similar result discussions should be offered to the decision maker. A more formalized procedure would calculate the relative share of the health metric distribution among the different disadvantaged groups.

- The data need for quality weights (QALY) and WTP values could not be fully satisfied by the literature and the compiled data are inconsistent. The data basis for environmental health is presently probably best for DALYs.

In addition to the insights summarized above there are a few points worth mentioning that are potentially important but did not show up in our example:

- Time discounting was excluded by design.
- Age weighting is often applied in DALYs as a correction function of the statistically expected years of life lost (Murray et al. 1996a). However, as discussed in Section 2.10, their proposal reflects neither empirical findings nor theoretical models. Age-dependent values or utilities of life to be used in a prescriptive or even normative setting may need to be based on a societal consensus. It may well follow the ethical principles that either each year of life lost is of equal value or that each (remaining) life is of equal value.
- The remaining statistical life expectancy at the time of death is chosen to be the same for DALYs and QALYs in our example. However, DALYs as suggested by Murray et al. (1996a) have been developed for international applications with the aim to attribute all health losses to diseases. To do so they needed to state a number of equity assumptions that resulted in a life expectancy function that depends only on sex and age. This attribution mode is different from the change mode we are interested in most environmental applications (e.g., reduction of health damages thanks to clean air act or net health benefits of improved drinking water treatment). We are often interested in changes of risks. However, this is not an inherent limitation of DALYs but rather a matter of assumptions.
- The QALY framework suggests not only to control for co-morbidity when quality weights are developed for specific diseases but also to consider age-specific co-morbidity of the general population when the years of life lost due to premature death are calculated. Due to the lack of access to the age-profiles of premature deaths, we did not correct for them. However, data in Fryback et al. (1993) suggests, that a woman's year lost at the age of 65-74, 75-84 and 85+ should be counted only as 0.83, 0.79 and 0.8 respectively[3]. This is probably the appropriate way to deal with the question of marginal changes addressed in the point before and suggests that the number of QALYs due to mortality has been overestimated in our example.
- Both, the DALYs and QALYs include only the individually borne intangible costs. At least collectively and individually borne costs of illness should be added in a comprehensive assessment. WTP based on individually borne costs may be complemented by information on collectively borne costs.

[3] Measured by TTO. Men's values are 0.84, 0.84 and 0.82 respectively.

The two most stunning results directly derived from our example are the insensitivity of WTP to morbidity outcomes and the huge effect of uncertainties in the assessment of mild diseases. Both findings deserve further research:

- For the insensitivity of WTP three main problems need to be resolved: (1) age dependent VSL for environmental risks need to be further explored and developed for different cultural and economic settings; (2) the valuation of acute and chronic morbidity outcomes due to environmental risks needs to be further explored; and (3) the often observed insensitivity of WTP to magnitude of risk reduction (Hammitt et al. 1999). Promising developments that use chained approaches (Carthy et al. 1999, Viscusi et al. 1991) or attribute based stated choice analyses (Johnson et al. 1998) might ease the dollar-risk trade-offs.

- For the disability and quality weights for mild illnesses we need to address the findings that people are not ready to trade life for them and that some elicitation method compel respondents to use very low probability numbers for mild illnesses, i.e., quantify something human beings proved to fail. To ask for tradeoffs between different more or less mild illnesses may solve both problems as suggested in Pinto Prades (1997). Further, for mild disabilities with long durations like noise or reduced neurocognitive development time-non-proportionality due to adaptation and adjustment may have decisive influence on the outcome.

4. Characterization of medical applications and environmental tools

The review of the literature in Section 2 revealed a tremendous number of different metrics and within these metrics different elicitation methods, judges and assumptions are used. One obvious reason for this variety is the many different applications within medical decision making and health economics. Table V attempts to characterize some of the major applications in medical and environmental decision support[1] using the following attributes:

- *Type of diseases:* Since it makes a difference for a metric whether only chronic or mostly acute health outcomes have to be assessed, we use this attribute for characterization.
- *Need for monetary units:* If it is likely that changes in health status will be evaluated in a cost-benefit framework this favors monitization of health impacts.
- *Identifiability of victims* and *veil of ignorance*: These two attributes are correlated and should be read together. These attributes determine whether additional characteristics (disabilities, profession, family circumstances etc.) that may influence the disability weights should/can be taken into account and whether a purely individual or societal perspective is more appropriate.
- *Authoritative status:* If an assessment needs to be authoritative then the metric used needs to be acceptable not only by single decision makers but by the society at large.
- *Affected generations:* If future generations are affected then the debate on appropriate time discounting becomes very relevant. Further, the disability weighting should be done disregarding any socialized handicaps that cannot be predicted for future generations.
- *Distributional requirements:* Since the discussed health metrics follow the paradigm of utility maximization it is important to see in which applications this maximization may be sufficient and when additional distributional/ethical requirements will need to be considered.

As demonstrated in Table V these attributes differentiate well between the listed applications and tools and none of the medical applications fits exactly with one of the environmental tools. The clinical decision support for single patient does not fit at all with the environmental tools. Therefore, health metrics developed for "bedside reasoning" may not be relevant for environmental applications.

[1] Descriptions and characterizations of the environmental tools can be found in Hofstetter et al. (2002).

Tab. V: Medical and environmental decision support tools and their different attributes that may be relevant for the selection of congruent health metrics.

Applications:	Type of diseases	Need for monetary units?	Identifia-bility of victims	Veil of ignorance	Authoritative status	Affected generations	Distributional requirements
Medical decision support							
Clinical decision support for single patient	In principle all, but per application only few	no	yes	lifted	none	own	none
Technology/product assessment for pharmaceutical companies and health care providers	In principle all, but per application only few	sometimes	partly	mostly lifted	none	own	none
Tool for resource allocation of health insurance or national health planning plan	all	sometimes	no	applies	national/ binding	own plus next	important (age, race, economic status, disabled)
Global health monitoring and resource allocation (Global Burden of Disease)	all	no	no	applies	international/ not binding	own plus next	important (age, race, economic status, disabled)
Environmental decision support tools:							
Micro-tools: Life Cycle Assessment	Many chronic diseases (including episodic)	no	no	applies	usually none	>100 years	Intra- and intergenerational
Meso-tools: (Comparative) Risk Assessment for Technology Assessments	Few, mostly chronic diseases	no	no	applies	none or limited	own, 50a, >100a	may be relevant, sensitive subgroups
Macro-tools: (Comparative) Risk Assessment for regulation	Few, mostly chronic diseases	sometimes	partly	mostly lifted	national/ binding	own	relevant, sensitive subgroups
Macro-tools: Cost-Benefit Assessment for regulation	Few, acute and chronic diseases	yes	partly	mostly lifted	national/ binding	own	relevant, sensitive subgroups

5. Consequences for the choice of metrics in different applications

How do the characteristics of the application or tool determine the choice for human health metrics? Table V illustrated some of the differences between and within the medical applications and the environmental tools. How does this affect the choice of the metric, the elicitation method to derive preferences, the group for preference elicitation, time discounting, and the type of life tables to be used? Table VI summarizes our recommendations for the choices to be made according to Section 2 based on the characteristics summarized in Table V. The following arguments were used to come up with recommendations:

- *Life Tables:* The need for appropriate spatial and temporal coverage and the (im)possibility to identify subgroups with non-average mortality risks have been the guiding attributes to determine the appropriate life tables.
- *Whose values?:* Patients' preferences about their own disease are always important but may become impractical when a large number of different health outcomes need to be evaluated. In such cases, health care professionals may provide the necessary relative comparison. Depending on the degree of how socially binding the metric needs to be an additional representative panel may need to be formed (Nord 1999).
- *Time preference:* The level of individual versus societal decision making and the importance of intergenerational aspects were the guiding principles. The mentioned discount rates are illustrative for the range and do not imply that an exponential discount function needs to be chosen. It is also assumed that the future increase of value of HALYs and statistical life are considered. The zero discount rate for Life Cycle Assessment is based not only on the very long assessment horizon but also on present practice, where increase in future life expectancies are not considered.
- *Preferred elicitation method:* The main difference is here whether monetary or non-monetary values are derived. Further, the time trade-off (TTO) method with an adequate time horizon or the person trade-off method (PTO) with application compatible framing of the question have been judged to outperform other methods for the individual and societal application respectively, although the standard gamble often provides a more realistic description of the choice.
- *Level of measurement:* The better the social environment of the affected group is known the more these parameters should be included in the elicitation step (handicap level). If a large number of different social environments have to be covered or if future environments are unknown then a disability level is preferred.
- *Preferred metrics:* Both monetary and non-monetary metrics have flaws for valuation of both mortality and morbidity. However, since monetary methods require not only a health/health but a health/wealth tradeoff they are cognitively more demanding than non-monetary metrics. Therefore, we suggest using them only when monetary units are desirable[1] as a measurement unit. "HALYs+" stands for Health Adjusted Life Years with age weighting. We use this notion

[1] "Desirable" stands for decisions where trade-offs between human health and monetary expenditures are at stake.

because the column headings above specify most of the specific features that would differentiate between QALYs and DALYs and because the age weighting to be used deviates from the standard procedure in the DALYs framework. For environmental applications, we also suggest to supplement the HALYs+ with cost of illness. HYE are not considered preferable because empirical experience and data are lacking. However, this metric may well be developed for environmental applications where the number of relevant health outcomes is limited.

- *Marginal/average and distributional aspects:* If we are interested in the analysis of changes due to an intervention compared to a reference situation, e.g., present situation, then we call this a marginal analysis (where all other risk factors are kept constant). If the distributional aspects will play a major role in the decision making, we suggest to calculate the health metric scores for all relevant sub-groups and to add a semi-quantitative discussion.

We are aware that the recommendations in Table VI may be challenged in specific applications for arguments that could not be captured on this generic level. We also expect major developments in the areas of WTP that may alter our assessment within the coming years. Finally, we will list some strengths and weaknesses of the suggested metrics in the concluding Section 6.

Tab. VI: Recommendations for the choice of human health metrics and their specific assumptions.

Applications:	Life Table to calculate YLL	Whose values	Time preference (discount rate)	Preferred elicitation method	Level of measure-ment	Preferred metrics	Remarks
Medical decision support							
Clinical decision support for single patient	Clinical estimate based on diagnosis	Patient	Individual (rates vary from -x% to plus 100%)	TTO, transformed VAS, decomposed	Handicap	Non-monetary	Marginal analysis
Technology/product assessment for pharmaceutical companies and health care providers	Disease group-specific, future-oriented	Patients or health care professionals	Market (1-10%)	TTO CV, revealed preferences, attribute-based stated choice	Combined disability/ handicap	HALYs+ or WTP	Marginal analysis
Tool for resource allocation of health insurance or national health planning plan	Regional/ national life tables, present or future	Patients or combined patients/ societal values	Market/societal (1-10%)	PTO	Combined disability/ handicap	HALYs+	Distributional aspects important, mostly marginal analysis
Global health monitoring and resource allocation (Global Burden of Disease)	Universal life table for monitoring, Future-oriented regional/ national life tables for resource allocation	Health care professionals or large sample of combined patients/ societal values	Societal (1-5%)	PTO	Disability	HALYs+	Average analysis for monitoring, distributional aspects and marginal analysis important for resource allocation
Environmental decision support tools:							
Micro-tools: Life Cycle Assessment	Future-oriented regional life tables	Health care professionals or large sample of combined patients/ societal values	None (0%)	PTO	Disability	HALYs+	Marginal analysis
Meso-tools: (Comparative) Risk Assessment for Technology Assessments	Group/area-specific (all levels possible)	Depends on context	Societal (1-5% or different for longterm)	Depends on context	Combined disability/ handicap	HALYs+ plus COI, WTP plus collectively borne costs	Distributional aspects important, marginal analysis
Macro-tools: (Comparative) Risk Assessment for regulation	Present/ future national life tables	Patients or combined patients/ societal values	Societal (1-5%)	PTO, CV, revealed preferences, attribute-based stated choice	Combined disability/ handicap	HALYs+ plus COI, WTP plus collectively borne costs	Distributional aspects important
Macro-tools: Cost-Benefit Assessment for regulation	Present/ future national life tables	Patients or combined patients/ societal values	Societal (1-5%)	CV, revealed preferences, attribute-based stated choice	Combined disability/ handicap	WTP plus collectively borne costs	Distributional aspects important, marginal analysis

6. Discussion and Conclusions

This report's attempt to transfer insights from medical decision making and health economics into environmental decision support tools has proven to be fruitful. The summary and review of the respective literature made clear that not only the choice of the metric is important (whether time-proportionality is assumed (HALYs) or not (HYE, WTP) and whether the units are monetary in nature or not) but that it is particularly important which empirical choices (e.g., life table, time discounting, elicitation method, elicitation question and elicited group) are finally made within the metric. A summary of strengths and weaknesses of three of the most often applied metrics is given in Table VII.

Tab. VII: Major strengths and weaknesses of three often applied human health metrics

Selected health metrics	Strengths	Weaknesses
DALYs	consistent sets of disability weights for environmental endpoints readily available age-weighting societal perspective of disability weights and major assumptions metric unit is framed as health loss	assumption of time proportionality and risk-neutrality lacking consideration of COI and collectively borne intangible costs implementation and shape of age-weighting
QALYs	methods and limited set of quality weights available individual perspective of disability weights and major assumptions data for co-morbidity-adjustments at higher age available	assumption of time proportionality and risk-neutrality lacking consideration of COI and collectively borne intangible costs no age-weighting
WTP	metric is easier comparable to other attributes relevant in a decision process time-proportionality and risk-neutrality is not implied individually borne COI and intangible costs are considered	the methods and the set of values applicable for environmental endpoints need further research collectively borne COI and intangible costs are not included dollar/health-risk tradeoffs provoke protest bids and refusal (possible sign for non-compensatory nature of goods) and are more demanding than health/health tradeoffs

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 non-monetary 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). Further, it has been noticed that the availability of databases with consistent preference values for health outcomes differs for the three metrics where the DALYs offers presently the most comprehensible publicly available database.

The characterization of both medical and environmental decision support systems showed that their attributes vary largely within and between these groups. This may explain the large number of suggested health metrics and the different versions of the same metrics. Since the characteristics and assumptions of an application or tool should be congruent with the characteristics and assumptions of the health metrics, we indicate ranges of features of metrics that are compatible with the different applications. These recommendations (Table VI) remain preliminary, as the science is in development and the chosen categorization of applications probably too rough.

For the application in the environmental field we learn that the present state-of-the art in WTP leads in our example to a pure mortality assessment that may be an artifact due to the lack of reliable values for age-dependent statistical values of life and due to insufficient studies that assess WTP values for morbidity outcomes. 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. 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. Their 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.

A further implication of our analysis is that indeed – as criticized by many authors – most health metrics follow the philosophy of utility maximizing. Since decision makers may want to base their decisions not only on a utility metric but also on insights how different ethical and distributional modifications 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 the *income* matter; are already disadvantaged subgroups concerned (*double jeopardy*); and have *case-specific valuation attributes* been overlooked by the generic health metrics? (See Section 3 for such a discussion based on our case study). 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.

Our analysis of the application of human health metrics to environmental decision support tools was limited in detail and scope that leave open a number of potentially important questions:

- For morbidity outcomes, we have not studied the empirical relevance of the fact that the time-proportionality assumption made in HALYs does practically not hold. Potentially useful data collected within WTP studies is difficult to use because morbidity outcomes do not (in our example) matter in WTP estimates and secondly, the effects of scope insensitivities do interfere with time-non-proportionality and are difficult to separate.

- We have not investigated the practical relevance in differences in quality weights for environmental applications. These differences would be due to different elicitation methods, different question framings, or different groups of respondents. To investigate these possible differences, an empirical study would be needed that derives values for these different combinations on a reasonable number of environmentally relevant human health endpoints.

- Since environmental decision support systems sometimes capture effects that are predicted to occur in the distant future one would need to develop life tables, trend estimates for population development, quality weights in a future world with new medical treatment possibilities, future increases in the value of HALYs and statistical life, and probably most importantly, a time discounting framework that would reflect intergenerational preferences held by concerned stakeholders.

- Only a relatively small number of environmental decision support tools have been considered. However, the chosen applications are probably those that attempt to estimate health impacts on a disease or disorder level including duration and number of affected individuals.

- The availability of information on disease type and disorder, age of onset and duration of disease, and number of affected individuals has been assumed. 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.

- We also did not include all types of environmentally caused human health impacts that may become important in single case studies. Especially, we left out issues like developmental and fertility effects due to endocrine disrupters, hereditary effects due to ionizing radiation or development effects in fetus due to environmental causes.

- Further, we did not address the question whether simple exchange rates or transformation functions between different metrics exist. In the medical applications a rule of thumb says that a treatment or new drug should not cost more than 50,000 to 100,000 US$ per QALY (Hammitt 2000b). Such rules of thumb suggest that such transferability does exist. However, the case study in Section 3 and the different assumptions on time-proportionality and age-weighting make clear that such a straight forward exchange rate does *not* exist.

Next to the analysis and research into the mentioned limitations of this article we suggest to work on the following research questions due to their demonstrated relevance for environmental decision support systems:

1. Age-dependent statistical value of life or utility-adjusted years of life lost have been shown to reflect best both public values and outcomes of theoretical life-cycle models. The application of these insights was used, e.g., in Seethaler et al. (1999) to estimate age-dependent VSL, and some applications based on Murray et al. (1996a) take age weighting for DALYs into account as well. However, in both cases the underlying 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 must either include large samples or define subgroups or contexts that allow more homogenous answers.

2. Quality and disability weights for distortions or mild illnesses that are caused by environmental risk factors need to be assessed with a special emphasis on the potential biases introduced by the commonly used VAS, TTO, SG and PTO elicitation methods.

We hope that this article will contribute to 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.

References

AbouZahr C & Vaughan JP. (2000). Assessing the burden of sexual and reproductive ill-health: questions regarding the use of disability-adjusted life years. *Bulletin of the World Health Organization, 78* (5), 655-664.

Adamowicz W, Louviere J & Swait J. (1998). *Introduction to Attribute-based Stated Choice Methods*. Prepared for the National Oceanic and Atmospheric Administration, NOAA Purchase order 43AANC601388 by ADVANIS, Edmonton (Can).

Alberini A, Cropper M, Fu T-T, Krupnick A, Liu J-T, Shaw D & Harrington W. (1997). Valuing health effects of air pollution in developing countries: the case of Taiwan. *Journal of Environmental Economics and Management, 34*, 107-126.

Alderson M. (1988). *Mortality, Morbidity and Health Statistics*. New York, NY: Stockton Press.

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

Andersson F & Lyttkens CH. (1999). Preferences for equity in health behind a veil of ignorance. *Health Economics, 8*, 369-378.

Anonymous. (1994). Asthma TyPE, Health Outcome Institute.

Anonymous. (1999a). *Victorian Burden of Disease Study: Morbidity*. Melbourne, Victoria: Public Health Division. http://www.dhs.vic.gov.au/phd/9909065/index.htm (11/09/00).

Anonymous. (1999b). *Victorian Burden of Disease Study: Mortality*. Melbourne, Victoria: Public Health Division. http://www.dhs.vic.gov.au/phd/9903009/index.htm (11/09/00).

Arnesen T & Nord E. (1999). The value of DALYs life: problems with ethics and validity of disability adjusted life years. *BMJ, 319*, 1423-5.

Arrow K, Solow R, Portney PR, Leamer EE, Radner R & H Schuman (1993). Report of the NOAA Panel on Contingent valuation. *Federal Register 58*, 4601-4614.

Azimi NA & Welch HB. (1998). The effectiveness of cost-effectiveness analysis in containing costs. *J Gen Intern Med, 13*, 664-669.

Bala MV & Zarkin GA. (2000). Are QALYs and appropriate measure for valuing morbidity in acute diseases? *Health Economics, 9*, 177-180.

Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A & O'Brien B. (1996). Testing constant proportional trade-off assumption using standard gamble. *Clinical Therapeutics, 18*, 31.

Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A & O'Brien B. (1998). Valuing outcomes in health care: A comparison of willingness to pay and quality-adjusted life-years. *J Clin Epidemiol, 51*(8), 667-676.

Banfi S, Doll C, Maibach M, Rothengatter W, Schenkel P, Sieber N & Zuber J. (2000). *External Costs of Transport: Accident, Environmental and Congestion Costs of Transport in Western Europe*. INFRAS Zürich/IWW Karlsruhe.

Barendregt JJ, Bonneux L & van der Maas PJ. (1996). DALYs: the age-weights on balance. *Bulletin of the World Health Organization, 74* (94), 439-443.

Baron J. (1996). Why expected utility theory is normative but not prescriptive? *Med Decis Making, 16*, 7-9.

Baron J. (1997). Biases in the Quantitative Measurement of Values for Public Decisions. *Psychological Bulletin, 122* (1), 72-88.

Beattie J, Covey J, Donan P, Hopkins L, Jones-Lee M, Loomes G, Pidgeon N, Robonson A & Spencer A. (1998). On the contingent valuation of safety and the safety of contingent valuation: part 1- caveat investigator. *Journal of Risk and Uncertainty, 17*, 5-25.

Bell CM, Chapman RH, Stone PW, Sanberg EA & Neumann PJ. (1999). An Off-the-Shelf-Help List: A comprehensive catalogue of preference weights from published cost-utility analyses (CUAs), submitted December 9, 1999.

Bennett R & Tranter R. (1998). The dilemma concerning choice of contingent valuation willingness-to-pay elicitation format. *Journal of Environmental Planning and Management, 41* (2), 253-7.

Bergstrom TC. (1982). When is a man's life worth more than his human capital? in Jones-Lee MW (ed), The Value of Life and Safety: Proceedings of a Conference held by the Geneva Association, 3-26, Amsterdam: The Netherlands.

Bleichrodt H, Pinto JL & Wakker PP. (2000). *Making descriptive use of prospect theory to improve prescriptive applications of expected utility*. Working Paper, Erasmus University Rotterdam, NL, March 24, 2000.

Blumenschein K & Johannesson M. (1999). Use of contingent valuation to place monetary Value on pharmacy services: An overview and review of the literature. *Clinical Therapeutics, 21*(8), 1402-1417.

Brazier J & Deverill M. (1999). A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Economics, 8*, 41-51.

Brickman P, Coates D & Janoff-Bulman R. (1978). Lottery winners and accident victims: is happiness relative? *Journal of Personality and Social Psychology, 36*, 917-927.

Brock DW. (1998). Ethical issues in the development of summary measures of population health status. in Field MJ and Gold MR (eds.). *Summarizing population health: Directions for the development and application of population metrics*. Washington, D.C.: National Academy Press (73-85).

Carrothers TJ, Evans JS. & Graham JD. (2000). The lifesaving benefits of improved air quality: An uncertainty analysis. Submitted to *Risk Analysis*, May 2000.

Carson RT, Hanemann WM, Kopp RJ et al. (1997). Temporal reliability of estimates from contingent valuation, *Land Economics, 73*(2), 151-63.

Carson RT. (2000). Contingent Valuation: A users guide. *Environ.Sci.Technol, 34*, 1413-1418.

Carthy T, Chilton S, Covey J. Hopkins L, Jones-Lee M, Loomes G, Pidgeon N & Spencer A. (1999). On the contingent valuation of safety and the safety of contingent valuation: part 2-The CV/SG “Chained” Approach. *Journal of Risk and Uncertainty, 17*, 187-213.

Chisholm D, Healey A & Knapp M. (1997). QALYs and mental health care. *Soc Psychiatry Psychiatr Epidemiol, 32*, 68-75.

Cohen BJ. (1996a). Is expected utility theory normative for medical decision making? *Med Decis Making, 16*, 1-6.

Cohen BJ. (1996b). Reply: Utilitarianism, risk aversion, and expected utility theory. *Med Decis Making, 16*, 14.

Cohen J. (1996). Preferences, needs and QALYs. *Journal of Medical Ethics, 22*, 267-272.

Cookson R. (2000). Incorporating psycho-social considerations into health valuation: an experimental study. *Journal of Health Economics, 19*(3), 369-401.

Cropper ML & Sussman FG. (1990). Valuing future risks to life, Journal of Environmental, Economics and Management, 19, 160-174.

Cutler D & Richardson E. (1998). The value of health: 1970-1990. *American Economic Review, 88*, 97-100.

de Hollander AEM, Melse JM, Lebrte E & Kramers PGN. (1999). An aggregate public health indicator to represent the impact of multiple environmental exposures. *Epidemiology, 10*, 606-617.

de Rosa CT, Stara JF & Durkin PR. (1985). Ranking chemicals based on chronic toxicity data. *Toxicology and Industrial Hygiene, 1* (4), 177-191

de Wit GA, Bussbach JJW & De Charro F. (2000). Sensitivity and perspective in the valuation of health status: whose values count? *Health Economics, 9*(2), 109-126.

De Wit GA, Busschbach JJV & de Charro FTH. (2000). Sensitivity and perspective in the valuation of health status: Whose values count? *Health Economics, 9*, 109-126.

Dickie M & Gerking S. (1996). Formation of Risk Beliefs, Joint Production and Willingness to Pay to Avoid Skin Cancer. *Review of Economics & Statistics, 78* (3), 451-63.

Diener A., O’Brien B & Gafni A. (1998). Health care contingent valuation studies: a review and classification of the literature. *Health Economics, 7*, 313-326.

Dolan P, Gudex C, Kind P & Williams A. (1996). Valuing health states: A comparison of methods. *Journal of Health Economics* 15, 209-231.

Dolan P. (1996). Modelling valuations for health states: the effects of duration. *Health Policy, 38*, 189-203.

Dolan P. (1997). The nature of individual preferences: A prologue to Johannesson, Jonsson and Karlsson. *Health Economics, 6*, 91-93.

Dolan P. (1999). Whose preferences count? *Med Decis Making, 19*, 482-486.

Donaldson C. (1999). Valuing the benefits of publicly-provided health care: does ‘ability to pay’ preclude the use of ‘willingness to pay’? *Social Science & Medicine*, 49, 551-563.

Douard J. (1996). Is Risk neutrality neutral? *Med Decis Making, 16*, 10-11.

Dougherty CJ. (1994). Quality-adjusted life years and the ethical values of health care. *Am.J.Phys.Med.Rehabil, 73*, 61-65.

Eeckhoudt L. (1996). Expected utility theory – is it normative or simply “practical”? *Med Decis Making, 16*, 12-13.

Elbasha EH. (2000). Discrete time representation of the formula for calculating DALYs. *Health Economics, 9*, 353-365.

Elixhauser A & CA Steiner. (1999). *Most Common Diagnoses and Procedures in U.S. Community Hospitals, 1996*. Healthcare Cost and Utilization Project (HCUP) Research Note: Prepared by the Agency for Health Care Policy and Research (AHCPR), Rockville, MD. Publication No. 99-0046. http://www.ahcpr.gov/data/hcup/commdx. (11/13/00).

ESEERCO (Empire State Electric Energy Research Corporation). (1995). *New York State Environmental Externalities Cost Study, Volume 1: Introduction and Methods*. New York, NY: Oceana Publications Inc..

Essink-Bot ML, Stouthard ME & GJ Bonsel. (1993). Generalizability of valuation on health states collected with the EuroQol questionnaire. *Health Econ, 2*, 237-46.

ExternE. (1995). *Externalities of Energy*. European Commission EUR 16520 EN, Volume 1-6, Luxembourg: Office for Official Publications of the European Communities.

ExternE. (1999). *Externalities of Energy*. European Commission EUR 19083 EN, Volume 7-10, Luxembourg: Office for Official Publications of the European Communities.

Field MJ & Gold MR (eds.). (1998). *Summarizing population health: Directions for the development and application of population metrics.* Washington, D.C.: National Academy Press.

Fischer GW. (1979). Utility models for multiple objective decisions: do they accurately represent human preferences? *Decis Sci. 10*, 451-479.

Fischhoff B, Slovic P, Lichtenstein S, Read S & Combs B. (1978). How Safe is Safe Enough? A Psychometric Study of Attitudes towards Technological Risks and Benefits, *Policy Sciences*, 9, 127-52.

Fox JA, Shogran JF, Hayes DJ & Kliebenstein JB. (1998). CVM-X: Calibrating Contingent Values with experimental auction markets. *Amer.J.Agr.Econ, 80*, 455-465.

Frey RL, Gysin CH, Leu RE & Schmassmann N. (1985). *Energie, Umweltschäden und Umweltschutz in der Schweiz,* Basler Sozialökonomische Studien Band 27, Grüsch, CH: Verlag Rüegger.

Frischknecht R., Braunschweig A, Hofstetter P & Suter P. (2000). Human Health Damages due to Ionising Radiation in Life Cycle Impact Assessment, *Environmental Impact Assessment Review, 20*, 159-189.

Froberg DG & Kane RL. (1989a). Methodology for measuring health-state preferences – I: Measurement strategies. *J Clin Epidemiol 42*, 345-354.

Froberg DG & Kane RL. (1989b). Methodology for measuring health-state preferences – II: Scaling methods. *J Clin Epidemiol 42*, 459-471.

Froberg DG & Kane RL. (1989c). Methodology for measuring health-state preferences – III: Population and context effects. *J Clin Epidemiol*, 42, 585-592.

Fryback DG, Dasbach EJ, Klein R, Klein BEK, Dorn N., Peterson K & Martin PA. (1993). The Beaver Dam health outcomes study: Initial catalog of health state quality factors. *Med Decis Making, 13*, 89-102.

Fryback DG, Lawrence WF, Martin PA, Klein R & Klein BEK. (1997). Predicting quality of well-being scores from SF-36: results from the Beaver Dam Health Outcomes Study. *Medical Decis Making, 17* (1), 1-9.

Fryback DG. (1998). Methodological issues in measuring health status and health-related quality of life for population health measures: a brief overview of the "HALY" family of measures in Field MJ and Gold MR (eds.). *Summarizing population health: Directions for the development and application of population metrics.* Washington, D.C.: National Academy Press, 39-57.

Gabriel SE, Kneeland TS, Melton LJ, Moncur MM, Ettinger B & Tosteson ANA. (1999). Health-related quality of life in economic evaluations for Osteoporosis: Whose values should we use? *Med Decis Making, 19,* 141-148.

Gafni A & Birch S. (1993). Searching for a common currency: Critical appraisal of the scientific basis underlying European harmonization of the measurement of Health Related Quality of Life (EuroQol©). *Health Policy, 23,* 219-228.

Gafni A. (1997). Alternatives to the QALY measure for economic evaluations. *Support Care Cancer, 5,* 105-111.

George JF, Duffy K & Ahuja M. (2000). Countering the anchoring and adjustment bias with decision support systems. *Decision Support Systems, 29,* 195-206.

Goedkoop M & Spriensma R. (1999). *The Eco-indicator'99: A damage oriented method for Life Cycle Impact Assessment*. Nr. 1999/36A/B. Zoetermeer, NL: VROM, http://www.pre.nl/eco-indicator99/ei99-reports.htm (10/19/00)

Gold MR, Siegel JE, Russell LB & Weinstein MC (eds). (1996). *Cost Effectiveness in Health Medicine*. New York, NY: Oxford University Press.

Groot W. Adaptation and scale of reference bias in self-assessment of quality of life. (2000). *Journal of Health Economics*, *19* (3), 403-420.

Guinée J & Heijungs R. (1993). A proposal for the Classification of Toxic Substances within the Framework of Life Cycle Assessment of Products, *Chemosphere 26* (10): 1925-44.

Hammitt JK & Graham JD. (1999b). Willingness to pay for health protection: Inadequate sensitivity to probability? *Journal of Risk and Uncertainty, 8*, 33-62.

Hammitt JK, Belsky ES, Levy JI & Graham JD. (1999a). Residential building codes, affordability, and health protection: A risk-tradeoff approach. *Risk Analysis, 19* (6), 1037-1058

Hammitt JK, Liu J-T & Liu J-L. (2000). Survival is a Luxury Good: The Increasing Value of a Statistical Life. Manuscript at the Harvard School of Public Health, Boston, MA. August 2000.

Hammitt JK. (1993). Editorial: discounting health increments. Journal of Health Economics, 12, 117-120.

Hammitt JK. (2000a). Evaluating Contingent Valuation of Environmental Health Risks: The proportionality test. *Association of Environmental and Resource Economists Newsletter, 20* (1), 14-19.

Hammitt JK. (2000b). Valuing mortality risk: Theory and Practice. *Environ. Sci. & Techno, 34* (8), 1396-1400

Hammitt JK., (2002). QALY versus WTP. to be published in *Risk Analysis*

Hanson K. (1999). *Measuring up: Gender, Burden of Disease, and priority setting techniques in the health sector*. August 1999. Harvard Center for Population and Development Studies. www.hsph.harvard.edu/organi…lthnet/Hupapers/gender/hanson.html (4/10/00).

Harada T, Fujii Y, Nagata K, Inaba A & Mettier T. (2000). Panel test for Japanese LCA experts aiming to weight safeguard subjects. Proceedings of The Fourth International Conference on EcoBalance, 10/31-11/2/2000, Tsukaba, Japan, 201-204.

Havelaar AH, de Hollander AEM, Teunis PFM, Evers EG, Van Kranen HJ, Versteegh FM, van Koten JEM & Slob W. (2000). Balancing the risks and benefits of drinking water disinfection: disability adjusted life-years on the scale. *Environmental Health Perspectives, 108,* (4):315-321.

Hertwich EG. (1999). *Toxic Equivalency: Accounting for Human Health in Life-Cycle Impact Assessment*. Berkeley, CA: University of California.

Hoffman C; Rice D & Sung H-Y (1996). Persons with Chronic Conditions: their Prevalence and Costs, *Jour. American Medical Association 276*, (18), 1473-1479.

Hofstetter P. (1998). *Perspectives in Life Cycle Impact Assessment; A structured approach to combine models of the technosphere, ecosphere , and valuesphere*, Boston: Kluwer Academic Publishers.

Hofstetter P., Bare JC, Hammitt JK, Murphy PA, Rice GE. (2002). Tools for comparative analysis of alternatives: Competing or complementary perspectives? accepted for publication in *Risk Analysis*

Holmes AM. (1995). A quality-based societal health statistic for Canada, 1985. *Soc Sci Med, 41* (10), 1417-1427.

Holmes AM. (1997). A method to elicit utilities for interpersonal comparisons. *Med Decis Making, 17*, 10-20.

Hoogenveen RT, Gijsen R, Genugten MLL van, Kommer GJ, Schouten JSAG & Hollander AEM de. (2000). *Dutch DisMod. Constructing a set of consistent data for chronic disease modeling*. RIVM report 260751001, Bilthoven, NL: RIVM.

Huber J, Wittink DR, Fiedler JA & Miller R. (1993). The effectiveness of alternative preference elicitation procedures in predicting choices. *J Market Res, 30,* 105-114.

Huijbregts MAJ. (1999). *Priority Assessment of Toxic Substances in the Frame of LCA: Development and Application of the Multi-Media Fate, Exposure and Effect model USES-LCA*, University of Amsterdam, http://www.leidenuniv.nl/interfac/cml/lca2/index.html (11/10/00).

ILSI. (1996). *Human Health Impact Assessment in Life Cycle Assessment: Analysis by an expert panel.* Washington, DC: International Life Sciences Institute.

Inhaber H. (1982). *Energy Risk Assessment.* New York, NY: Gordon and Breach Science Publishers.

International Life Sciences Institute (ILSI). (1996). *Human health impact assessment in Life Cycle Assessment: Analysis by an expert panel.* Washington DC: International Life Science Institute

ISO. (1997). *Environmental Management - Life Cycle Assessment - Principles and Guidelines*, EN ISO 14040, Brussels, Belgium.

Jansen SJT, Stiggelbout AM, Wakker P, Nooij MA, Nordijk EM & Kievit J. (2000). Unstable preferences: a shift in valuation or an effect of the elicitation procedure? *Med Decis Making, 20,* 62-71.

Johannesson M & Johansson P-O. (1997a). Quality of life and the WTP for an increased life expectancy at an advanced age. *Journal of Public Economies, 65*, 219-228.

Johannesson M, Johansson P-O & Löfgren K-G. (1997c). On the value of changes in life expectancy: Blips versus parametric changes. *Journal of Risk and Uncertainty, 15*, 221-239.

Johannesson M, Meltzer D & O'Conor RM. (1997b). Incorporating future costs in medical cost-effectiveness analysis: Implications for the cost-effectiveness of the treatment of hypertension. *Med Decis Making, 17,* 382-389.

Johannesson M, Pliskin J & Weinstein MC. (1994). A note on QALYs, time tradeoff and discounting. *Med Decis Making, 14*, 188-193.

Johnson FR, Banzhaf MR & Desvousges WH. (2000a). Willingness to pay for improved respiratory and cardiovascular health: A multiple-format, stated-preference approach. *Health Economics, 9*, 295-317.

Johnson FR, Desvousages WH, Ruby MC, Stieb D & De Civita P. (1998). Eliciting stated health preferences: an application to willingness to pay for longevity. *Med Decis Making, 18*, 57-67.

Jones-Lee M. & Loomes G. (1997). Valuing Health and Safety: some Economic and Psychological Issues, in Nau R. *et al.* (eds.), *Economic and Environmental Risk and Uncertainty*. Dordrecht, NL: Kluwer Academic Publishers, 3-32.

Jones-Lee MW, Loomes G & Philips PR. (1995). Valuing the prevention of non-fatal road injuries: Contingent valuation vs. standard gambles. *Oxford Economic Papers, 47,* 676-695.

Jones-Lee MW. (1992). Paternalistic altruism and the value of statistical life, Economic Journal, 102, 80-90.

Kaplan RM & Anderson JP. (1988). A general health policy model: Update and applications. *Health Serv Res, 23*, 203-35.

Keeler EB. & Cretin S. (1983). Discounting of life-saving and other nonmonetary effects, *Management Science* 29, 300-306.

Keeney RI and Raiffa H. (1976). *Decision with multiple objectives: preferences and value tradeoff*. New York, NY: John Wiley & Sons.

Kenkel D. (1997). On valuing morbidity, cost-effectiveness analysis, and being rude. *Journal of Health Economics, 16,* 749-757.

Koch T. (2000). Life quality vs the 'quality of life': assumptions underlying prospective quality of life instruments in health care planning. *Social Science & Medicine, 51,* 419-427.

Koch T. (2000). The illusion of paradox. *Social Science & Medicine, 50,* 757-759.

Krabbe PFM & Bonsel GJ. (1998). Sequence effects, health profiles, and the QALY model: In search of realistic modeling. *Med Decis Making, 18*, 178-186.

Krabbe, PFM.; Essink-Bot M-L & Bonsel GJ. (1997). The Comparability and Reliability of Five Health -State Valuation Methods, *Social Science and Medicine, 45* (11), 1641-1652.

Kristiansen CM. (1985). Value correlates of preventive health behavior. *Journal of Personality and Social Psychology, 49*(3), 748-758.

Kuppermann M, Shiboski S, Feeny D, Elkin EP & Washington AE. (1997). Can preference scores for discrete states be used to derive preference scores for an entire path of events? An application to prenatal diagnosis. *Med Decis Making, 17*, 42-55.

Leigh JPl; SB. Markowitz; M. Fahs; C. Shin; & P. J. Landrigan, (1997). Occupational Injury and Illness in the United States: Estimates of Costs, Morbidity, and Mortality, *Archives of Internal Medicine, 157* (14), 1557-1568.

Leonard HB & Zeckhauser RJ. (1986). Cost-Benefit Analysis applied to risks: Its philosophy and legitimacy. In MacLean D. (Ed). *Values at Risk*. Totowa, NJ: Rowman & Allanheld Publishers, 31-48.

Levin R. (1997). Lead in drinking water. in Morgenstern R. (Ed). *Economic analysis at EPA: Assessing Regulatory Impact*. Washington, DC: Resources for the Future, 205-232.

Lippmann M (Ed). (2000). *Environmental Toxicants: Human Exposures and their Health Effects*. Second Edition. New York, NY: Wiley Interscience.

Loomes G & McKenzie L. (1989). The use of QALYs in health care decision making. *Soc Sci Med, 28*, 299-308.

MacKeigan LD, O'Brien BJ and Oh PI. (1999). Holistic versus composite preferences for lifetime treatment sequences for type 2 diabetes. *Med Decis Making, 19*, 113-121.

Magat WA, WK Viscusi & Huber J. (1996). A reference lottery metric for valuing health, *Management Science, 42,* 1118-1130.

Mansourian BG. (1996). ACHR News. *Bulletin of the World Health Organization, 74* (3), 333-337.

Mara DD & Feachem RGA. (1999). Water- and excreta-related diseases: unitary environmental classification, *Journal of Environmental Engineering, 125* (4), 334-9.

McNeil BJ, Weischselbaum R & Pauker SG. (1981). Speech and survival tradeoffs between quality and quantity of life in laryngeal cancer. *N Engl J Med, 305*, 982-987.

Mehrez A & Gafni A. (1989). Quality adjusted life years, utility theory, and health years equivalents. *Med Decis Making, 9,* 142-149.

Miller GA. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review, 93,* (2), 81-97.

Miyamoto JM & Eraker SA. (1985). Parametric estimates for a QALY utility model. *Med Decis Making, 5,* 191-213.

Moore MJ & Viscusi WK. (1988). The quantity-adjusted value of life. *Econ Inquiry, 26* (3), 369-388

Morgenstern R. (Ed). (1997). *Economic analysis at EPA: Assessing Regulatory Impact.* Washington, DC: Resources for the Future.

Müller-Wenk R. (1999). *Life-Cycle impact Assessment of Road Transport Noise*, IWÖ-Diskussionsbeitrag Nr.77, St. Gallen, CH: University of St. Gall. http://www.iwoe.unisg.ch/service/index-e.html (10/19/00).

Murray CJL & Lopez AD. (1996b). The incremental effect of age-weighting on YLLs, YLDs, and DALYs: a response. *Bulletin of the World Health Organization, 74* (4), 445-446.

Murray CJL & Lopez AD. (1997). The utility of DALYs for public health policy and research: a reply. *Bulletin of the World Health Organization, 75* (4), 377-381.

Murray CJL & Lopez AD. (2000). Progress and directions in refining the Global Burden of Disease approach; A response to Williams. *Health Economics, 9,* 69-82.

Murray CJL & Lopez AD. (Eds.). (1996a). The Global Burden of Disease, *Volume I of Global Burden of Disease and Injury Series*, WHO/ Harvard School of Public Health/ World Bank. Boston, MA: Harvard University Press.

Ng Y.-K. (1992). The older the more valuable: Divergence between utility and dollar values of life as one ages. *Journal of Economics, 55* (1), 1-16.

Nord E. (1992a). Methods for quality adjustment of life years. *Social Science and Medicine, 34,* 559-569.

Nord E. (1992b). An alternative to QALYs: the saved young life equivalent (SAVE). *BMJ, 305*, 875-7.

Nord E. (1995). The person trade-off approach to valuing health care programs. *Med Decis Making, 15*, 201-208.

Nord E. (1999). *Cost-Value Analysis in Health Care: Making Sense out of QALYs*. Cambridge, U.K.: Cambridge University Press.

O'Brien B &Viramontes JL. (1994) Willingness to Pay: A valid and reliable measure of health state preference?. *Medical Decision Making, 15*, 132-137.

Olsen JA. (2000). A note on eliciting distributive preferences for health. *Journal of Health Economics, 19*, 541-550.

Patrick DL & Erickson P. (1993). *Health status and health policy; Quality of life in health care evaluation and resource allocation.* New York, NY: Oxford University Press.

Patrick DL, Bush JW & Chen MM. (1973). Methods for measuring levels of well-being for a health status index. *Health Serv Res, 8,* 228-45.

Pigou AC. (1932). *The Economics of Welfare*, 4th edition, London, U.K.: MacMillian.

Pinto Prades J-L. (1997). Is the person trade-off a valid method for allocating health care resources? *Health Economics, 6*, 71-81.

Pliskin JS, Shepard DS & Weinstein MC. (1980). Utility functions for life years and health status. *Operational Research, 28* (1), 206-224.

Ponce RA, Bartell SM, Wong EY, LaFlamme D, Carrington C, Lee RC, Patrick DL, Faustman EM & Bolger M. (2000). Use of Quality-Adjusted Life Year weights with dose-response models for public health decisions: A case study of the risks and benefits of fish consumption. *Risk Analysis, 20* (4), 529-542.

Poulos C & Whittington D. (2000). Time preferences for life-saving programs: evidence from six less developed countries. *Environ. Sci. Technol., 34*, 1445-1455.

Pratt JW & Zeckhauser RJ. (1996). Willingness to pay and the Distribution of Risk and Wealth. *Journal of Political Economy, 104* (4), 747-763.

Raiffa H. (1961). Risk, uncertainty and the Savage axioms: comment. *Quarterly Journal of Economics, 75,* 690-694.

Raiffa H. (1970). *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Reading, MA: Addison-Wesley

Ramsberg J. (1999). Listening to the vocal citizens: how do politically active individuals choose between lifesaving programs? *Journal of Risk Research, 2* (4), 355-367.

Ratcliffe J. (2000). Public preferences for the allocation of donor liver grafts for transplantation. *Health Economic, 9*, 137-148.

Rawls J. (1971). *Theory of Justice.* Cambridge, MA: Harvard University Press.

Redelmeier DA, Heller DN & Weinstein MC. (1994). Time preference in medical economics: science or religion? *Med Decis Making 14*, 301-303.

Reiling SD, Boyle KJ, Philips ML & Anderson MW. (1990). Temporal reliability of contingent valuation. *Land Economics, 66*(2), 128-134.

Richardson J & Nord E. (1997). The importance of perspective in the measurement of quality-adjusted life years. *Med Decis Making, 17*, 33-41.

Richardson J, Hall J & Salkfeld G. (1996). The measurement of utility in multiphase health states. *Int J Technol Assess Health Care, 12*, 151-162.

Richardson J. (1994). Cost utility analysis: What should be measured? *Soc Sci Med, 39* (1), 7-21.

Rokeach M. (1973). *The nature of human values*. New York, NY: The Free Press.

Rosser RM & Watts VC. (1972). The measurement of hospital output. *Int J Epidemiol, 1*, 361-68.

Rusthoven JJ. (1997). Are quality of life, patient preferences, and costs realistic outcomes for clinical trials? *Support Care Cancer, 5*, 112-117.

Ryan M & Hughes J. (1997). Using conjoint analysis to assess women's preferences for miscarriage management. *Health Economics 6,* 261-273.

Ryan M. (1999). Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Social Science & Medicine, 48*, 535-546.

Saaty TL. (1980). *The Analytical Hierarchy Process, Planning, Priority Setting, Resource Allocation*, New York, NY: Mc Graw-Hill

Sackett DL & Torrance GW. (1978). The utility of different health states as perceived by the general public. *J Chron. Dis, 31*, 697-704.

Sagoff M. (1998). Aggregation and deliberation in valuing environmental public goods: A look beyond contingent pricing. *Ecological Economics, 24*, 213-230.

Sayers B & Fliedner TM. (1997). The critique of DALYs: a counter-reply. *Bulletin of the World Health Organization, 75* (4), 383-384.

Seethaler R. (1999). *Health costs due to road traffic-related air pollution; an impact assessment project of Austria, France and Switzerland.* Synthesis Report. Prepared for the WHO Ministerial Conference on Environment and Health, London. GVF-Bericht 1/99, Bern, Switzerland: Federal Department of Environment, Transport, Energy and Communications.

Sen AK. (1979). Personal utilities and public judgements or what's wrong with welfare economics? *Economic Journal, 89*, 537-558.

Shepard DS & Zeckhauser RJ. (1984). Survival versus consumption. *Management Science, 30*(4), 423-439.

Shiell A, Sezmour J, Hawe P & Cameron S. (2000). Are preferences over health states complete? *Health Economics, 9*, 47-55.

Singer P, McKie J, Kuhse H & Richardson J. (1995). Double jeopardy and the use of QALYs in health care allocation. *Journal of Medical Ethics, 22*, 144-50.

Sintonen H. (1981). An approach to measuring and valuing health states. *Soc Sci Med, 15C*, 55-65.

Spilker B. *et al.* (1996)., Quality of Life, Bibliography and Indexes, *Medical Care*, 28 (Suppl 12), D51-77.

Stalmeier PFM & Bezembinder TGG. (1999). The discrepancy between risky and riskless utilities: a matter of framing? *Med Decis Making, 19*, 435-447.

Stouthard MEA, Essink-Bot M-L & Bonsel GJ, on behalf of the Dutch Disability Weights Group. (2000). Disability weights for diseases: a modified protocol and results for a Western European region. *Eur J Public Health, 10*, 24-30.

Stouthard MEA, Essink-Bot M-L, Bonsel GJ, Barendregt JJ, Kramers PGN, van de Water HPA, Gunning-Schepers LJ & van der Maas PJ. (1997). *Disability weights for diseases in The Netherlands.* Rotterdam, NL: Department of Public Health/Erasmus University Rotterdam.

Suterland HU, Llewellyn-Thomas H, Boyd NF & Till JE. (1982). Attitudes toward quality of survival. The concept of “maximum endurable time”. *Med Decis Making, 2*, 299-309.

Tengs TO, Adams ME, Pliskin JS, Safran DG, Siegel JE, Weinstein MC & Graham JD. (1995). Five-hundred life-saving interventions and their cost-effectiveness, *Risk Analysis*, *15 (*3), 369-89.

Tolley G, Kenkel D & Fabian R (eds.). (1994). State-of-the-art health values. In Tolley G, Kenkel D & Fabian R.(eds.). *Valuing Health for Policy; An Economic Approach*. Chicago, IL: The University of Chicago Press, 323-344.

Tolley G, Kenkel D & Fabian R. (1994). *Valuing Health for Policy; An economic approach*. Chicago, IL: The University of Chicago Press.

Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y & Wang Q. (1996). Multiattribute utility function for a comprehensive health status classification system; Health Utility Index Mark 2. *Medical Care, 34* (7), 702-722.

Torrance GW, Thomas WH & Sackett DL. (1972). A utility maximization model for evaluation of health care programmes. *Health Serv Res, 7*, 118-33.

Torrance GW. (1986). Measurement health state utilities for economic appraisal; A review. *Journal of Health Economics, 5*, 1-30.

Toxicology Excellence for Risk Assessment (TERA). (1999). Comparative Dietary Risks: Balancing the Risks and Benefits of Fish Consumption. http://www.tera.org/news/ (11/13/00)

Treadwell JR. (1998). Tests of preferential independence in the QALY model. *Med Decis Making, 18*, 418-428.

Tversky A & Kahneman D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297-323.

Ubel PA, Richardson J & Menzel P. (2000). Societal value, the person trade-off and the dilemma of whose values to measure for cost-effectiveness analysis, *Health Economics*, *9*(2), 127-136.

Udo de Haes HA, Jolliet O, Finnveden G, Hauschild M, Krewitt W, Müller-Wenk R. (eds.) (1999). Best available practice regarding impact categories and category indicators in Life Cycle Impact Assessment, Background document for the second working group on Life Cycle Impact Assessment of SETAC-Europe (WIA-2), *Int.J.LCA*, *4*, 66-74, 167-174.

USDL/BLS. (1999). *Lost-Worktime Injuries and Illnesses: Characteristics and Resulting Time Away from Work.* Publication No. 99-102, Washington, DC: USDL/BLS. http://stats.bls.gov/oshhome.htm. (4/22/1999).

USEPA. (1998a). *Comparative Risk Framework; Methodology and Case Study*. National Center for Environmental Assessment (NCEA-C-0135), SAB External Review Draft http://www.epa.gov/ncea/frame.htm (11/9/00).

USEPA. (1998b). *Cost of Illness Handbook.* Draft prepared by Abt Associates, Cambridge, MA for the Office of Pollution Prevention and Toxics, Washington, D.C. Electronic copy provided to MSE by Dr. Nicolaas Bouwes, EPA Project Manager on January 6, 2000.

USEPA. (1999a). Valuation of human health and welfare effects of criteria pollutants. Appendix H in USEPA. *The benefits and costs of the Clean Air Act 1990-2010*, EPA Report to Congress, EPA-410-R-99-001 http://www.epa.gov/oar/oaq_caa.html (11/13/00)

USEPA. (1999b). *The benefits and costs of the Clean Air Act 1990 to 2010*, EPA Report to Congress, EPA-410-R-99-001 http://www.epa.gov/oar/oaq_caa.html (11/13/00)

Üstün TB, Rehm J, Chatterji S, Saxena S, Trotter R, Room R., Bickenbach J & WHO/NIH Joint Project CAR Study Group. (1999). Multiple-informant ranking of the disabling effects of different health conditions in 14 countries. *The Lancet*, *354*, 111-15.

Van der Pol MM & Cairns JA. (2000). Negative and zero time preference for health. *Health Economics*, *9*, 171-175.

Van Hout BA. (1998). Discounting costs and effects: A reconsideration. *Health Economics*, *7*, 581-594.

Viscusi WK. (1983). *Risk by Choice, Regulating Health and Safety in the Workplace*. Cambridge, MA: Harvard University Press. 93-113.

Viscusi WK. (1993). The value of risks to life and health. *Journal of Economic Literature, XXXI*, 1912-1946.

Viscusi WK. (1998). *Rational Risk Policy*. Oxford, U.K.: Clarendon Press, 45-68.

Viscusi, WK, WA Magat & Huber J. (1987). An Investigation of the Rationality of Consumer Valuations of Multiple Health Risks, *The Rand Journal of Economics*, *18*, 465-479.

Viscusi, WK. WA Magat & Huber J. (1991). Pricing Environmental Health Risks: Survey Assessment of Risk-Risk and Risk-Dollar Trade-Offs for Chronic Bronchitis, *J. Environmental Economics and Management, 21*, 32-51

Von Neumann J & Morgenstern O. (1943). *Theory of games and economic behavior*. Princeton: Princeton University Press, (3rd edition and second printing by Science Editions, John Wiley & Sons, New York, 1967).

von Winterfeldt D & Edwards W. (1986). *Decision Analysis and Behavioral Research*. Cambridge, U.K: Cambridge University Press.

Wakker P & Deneffe D. (1996). Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science, 42* (8), 1131-1150.

Wathieu L. (1997). Habits and the anomalies in intertemporal choice. *Management Science, 43* (11), 1552-1563.

Weinstein MC & Stason WB. (1977). Foundation of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine, 296*(31), 716-721.

Weinstein MC, Siegel JE, Garber AM, Lipscomb J, Luce BR, Manning WG & Torrance GW. (1997). Productivity costs, time costs and health-related quality of life: A response to the Erasmus group. *Health Economics, 6*, 505-510.

Weitzmann ML. (1998). Why the far-distant future should be discounted at its lowest possible rate. *Journal of Environmental Economics and Management, 36,* 201-208.

Wenstøp F, Carlsen AJ, Bergland O & Magnus P. (1997). Valuation of Environmental Goods with Expert Panels, in Climaco J. (Ed), *Multicriteria Analysis*, Berlin, D: Springer.

WHO. (1947). The Constitution of the World Health Organization, *WHO Chronical*, 1, 29.

WHO. (1993). *International Statistical Classification of Diseases and Related Health Problems* **-** *ICD-10*, Tenth Revision, Geneva, CH: WHO.

Williams A. (1996). QALYs and ethics: a health economist's perspective. *Soc Sci Med, 43* (12), 1795-1804.

Williams A. (1999). Calculating the Global Burden of Disease: Time for a strategic reappaisal? *Health Economics, 8,* 1-8.

Williams A. (2000). Comments on the response by Murray and Lopez. *Health Economics, 9*, 83-86.

Willis KG & Powe NA. (1998). Contingent valuation and real economic commitments: A private good experiment. *Journal of Environmental Planning and Management, 41*(5), 611-619.

Wu G. (1996). The strengths and limitations of expected utility theory. *Med Decis Making, 16*, 9-10.