International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 19(1): 18–33 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.300

Valuation of health states in the US study to establish disability weights: lessons from the literature

JÜRGEN REHM[1,2,3,4] & ULRICH FRICK[5]

1 Centre for Addiction and Mental Health, Toronto, Canada
2 University of Toronto, Dalla Lana School of Public Health, Toronto, Canada
3 Technische Universität Dresden, Germany
4 WHO Collaborating Centre for Substance Abuse, Zurich, Switzerland
5 Carinthia University of the Applied Sciences, Department of Healthcare Management, Feldkirchen, Austria

Key words
disability-adjusted life years (DALYs), burden of disease, disability weight, empirical assessment, psychometrics, trade-off methods

Correspondence
Jürgen Rehm, Centre for Addiction and Mental Health, 33 Russell Street, Room 2035, Toronto, Ontario, M5S 2S1 Canada.
Telephone (+1) 416-535-8501 ext. 6907
Fax: (+1) 416-260-4156
Email: jtrehm@aol.com

Received 19 September 2008; revised 5 January 2009; accepted 13 March 2009

Abstract

The metric of disability-adjusted life years (DALYs) has become the global standard of measuring burden of disease. DALYs are comprised of years of life lost due to premature mortality and years of healthy life lost due to living with disability. In order to calculate the second part of the DALY equation, disease specific disability weights have to be established, i.e. measures for the decline of health associated with these disease states, which vary between zero for perfect health and one for death. Although these disability weights are key for estimating DALYs, there have not been many comprehensive studies with empirical determinations of them. This article describes a systematic review on the state of the art with respect to empirically determining disability weights. Based on this review, a multi-method approach is outlined, which has also been implemented in a US study to measure burden of disease. This approach involves the use of psychometric methodology as well as economic trade-off methods for determining the value of health states. It is conceptualized as a disaggregated approach, where the disability weight of any health state can be calculated if the attributes of this health state are known. The US study received the collaboration of experts from more than 20 institutes of the National Institutes of Health and of the Centers for Disease Control and Prevention. First results will be available by the end of this year. *Copyright © 2010 John Wiley & Sons, Ltd.*

Background

Disability-adjusted life years as a summary measure of health

Since the first appearance of the Global Burden of Disease Study (GBD; World Bank, 1993), summary measures of population health (Murray *et al.*, 2000) have been indispensible in presenting and comparing the health status of populations, as well as informing the priority setting with respect to health care planning (van der Maas, 2003; for a historical overview see Etches *et al.*, 2006). Such summary measures combine information about mortality and non-fatal health consequences into one single value able to represent the health of a particular population (Field and Gold, 1998). In order to fulfill their main purpose, summary measures should emphasize premature mortality instead of reflecting a simple 'body count' of deaths, be based on strict epidemiological estimates, display internal consistency, enable rigorous validation of measures and estimates, and integrate health outcomes systematically into an index value (McKenna and Marks, 2002; Murray and Lopez, 2000). Best suited to these criteria, the concept of disability adjusted life years (DALYs; Murray 1994, 1996) has become the most influential summary measure for global health.

$$\mathrm{DALY} = \mathrm{YLL} + \mathrm{YLD} \quad (1)$$

where YLL is the years of life lost due to premature mortality, YLD is the years lived with disability, defined in Equation 2.

$$\mathrm{YLD} = I \times \mathrm{DW} \times \mathrm{LD} \quad (2)$$

where YLD is the years lived with disability, I is the number of incident cases, DW is the disability weight and LD is the average duration of disability (years).

As described in Equations 1 and 2, DALYs combine years of life lost due to premature mortality with years of healthy life lost due to living with disability, where the latter term is determined by the incidence of the underlying health condition, the duration of the time living with disability, and the level of disability of the underlying health state, called the disability weight (DW).

While DALYs have been quite successful in practical applications, there is still an ongoing debate regarding their theoretical framework, ethical foundations, and operationalization (for reviews see Field and Gold 1998; Murray *et al.*, 2002; see also later). Of course, many of the points raised are not limited to DALYs but also concern summary measures of health in general. We will review these areas and their criticisms only as far as they concern the design of a study deriving new disability weights (DWs) for the United States (see point on objectives later).

DWs as key element of DALYs

As seen in Equation 2, the values of DWs are key for determining DALYs. A DW is a metric for the decline of health associated with a certain heath state, varying between zero (perfect health) and one (death). In regards to the conceptual framework, it is clear that all DWs assume the existence of distinct constructs of health and disability, which, if the same DWs are applied globally, must be comparable across geographic region(s) and population(s). The construct of health should be distinguishable from both smaller concepts such as specific diseases or mortality risk (Breslow, 2006) and broader concepts such as well-being. Based on the construct of health, disability can then be defined as the decline of health in the different health states examined. All operationalizations discussed later or used in the literature to elicit DWs (overviews: Arnesen and Trommald, 2005; Doctor *et al.*, 2010; Green *et al.*, 2000; Morimoto and Fukui, 2002; Mortimer and Segal, 2008; Ryan *et al.*, 2001) rely on such an explicit construct of health. Of course, the content of the construct is partly determined by the questions asked, but most importantly by the definition of health states (Fowler, 1995; for a theoretical basis see Grice, 1975; Schwarz, 1996). For instance, if all health states are described by degrees of different attributes such as cognition, mobility, or social relationships, a respondent will infer that these dimensions should be included in their response about which state is healthier or less disabling (Wänke *et al.*, 1995). This may be relevant if the respondents, outside of the judgment situation, would not have included certain dimensions in their constructs (e.g. the attribute 'social relationships' may not be part of health for certain people, but may be included for others). Although there have been discussions on whether such a construct of health can be meaningfully assessed or if it even exists (Broome, 2004), empirical evidence has shown that most respondents are capable of answering questions about which of two people is healthier or which condition is more disabling. This indicates that an integration of relevant attributes from different domains into one category of health or decrements thereof is possible. A related question concerns the use of vectors of multiple attributes versus a single summary statement as descriptors of health ('basket presentation' versus global measure; Etches *et al.*, 2006). For the purposes of this paper it suffices

to say that for summary measures of population health, by definition, an integrated measure is needed (see review of different objectives for health measures cf. McDowell *et al.*, 2005).

The ethical discussions about DALYs have mainly concerned the potential resource allocations linked to the quantification of burden of disease. These discussions are irrelevant if one does not assume that some properties of burden of disease measures, such as DALYs, have direct implications for resource allocation. Our view certainly does not imply such implications; while the burden of disease should be weighted into such allocation decisions, other considerations for fairness or equity should also be formally integrated (Bleichrodt *et al.*, 2005; Dolan and Olsen 2001; Stolk *et al.*, 2005) into the decision-making process.

This earlier point of view does not imply that resource allocation scenarios cannot be used as one operationalization for eliciting DW (of course keeping in mind the effects of such operationalizations; see Damschroder *et al.*, 2007; Schwappach 2005; for examples). As will be laid out later, we suggest a multi-method approach for eliciting DW, including direct comparisons of disability of health states as well as indirect derivation from trade-off tasks.

Other ethical considerations, more directly linked to the DW operationalization and assessment, were made by Arnesen and Nord (1999) who argued that in the original determination of DW in the GBD 1990 study, the health value of individuals may have been unethically set equal to that of their life. Especially scenarios where the decision maker, in order to determine a DW for health states, had to choose between the lives of k healthy people and the lives of $k + x$ disabled people, there was an assumption that health and life were the same dimension. While we concede the potential risk that this argument denotes, especially if health state comparisons from a strict personal perspective are used as elicitation method (Damschroder *et al.*, 2005a), we believe that different operationalizations, such as the ones presented later, may actually invalidate this specific criticism.

On the empirical determination of DWs: objective of this paper

Given the magnitude of theoretical discussion and the importance of the topic, it is surprising that in practical terms only few efforts have been undertaken to empirically and comprehensively study DWs for the basis of estimating DALYs. More work has been done on empirically deriving weights for quality-adjusted life years (QALYs, for an overview see Mortimer and Segal, 2008) with similar methodology (Murray *et al.*, 2000; Murray *et al.*, 2002), but again there are no comprehensive efforts of including the full spectrum of morbid conditions.

All of the global World Bank and World Health Organization (WHO) statistics on the burden of disease have been based on DWs from two valuation exercises from the original 1990 GBD study (Murray and Lopez, 1996), appended by the Dutch valuation study (Stouthard *et al.*, 1997). It is an open question whether these DWs apply to the United States in the year 2005, given the empirical result of cultural specificity of key DWs, as well as the differing cultural and treatment situation (Üstün *et al.*, 1999a). With this background in mind, it was decided to assess US-specific DWs as part of a US burden of disease study.

The objective of this paper, as a part of this study, is to prepare the best possible assessment methods for establishing US-specific DWs. Solely restricted to this objective, we review theoretical and empirical results of past research on DWs and exclude arguments which have no implications for empirically establishing DW, as well as incorporate results from this study. For instance, any argument of whether or not resources should be solely based on DALYs has no direct implications on the design and measurement of DW studies, as we will also use different approaches on eliciting DW (see later). Needless to say, we also exclude any thorough discussion on cultural differences in establishing global weights (Üstün *et al.*, 1999a, Üstün *et al.*, 1999b), as we are restricted to one country. Of course, within a country cultural differences may still be relevant, but this is on a smaller scale compared to a situation where a global perspective is used.

Social psychology is the perspective of our review. Throughout this article, establishing DWs is conceptualized as a judgmental task that is solved at the very moment in which respondents are asked to compare different health states. In other words, people use the information available at the moment in which the judgment task is given and the usual laws of questionnaire construction and interview design apply (Schwarz, 1996). In doing this, we assume that there is a quantifiable construct of 'health'. Our focus is on the cognitive processes of the respondent in the judgment situation; we are concerned with how the respondent perceives and answers the judgmental tasks rather than how a philosopher or third party would. Figure 1 gives an overview of the cognitive model used here, which is an adaption from Stiggelbout and de Vogel-Voogt (2008).

In valuing health states, first the perspective of the judgmental task has to be clarified: the respondent can

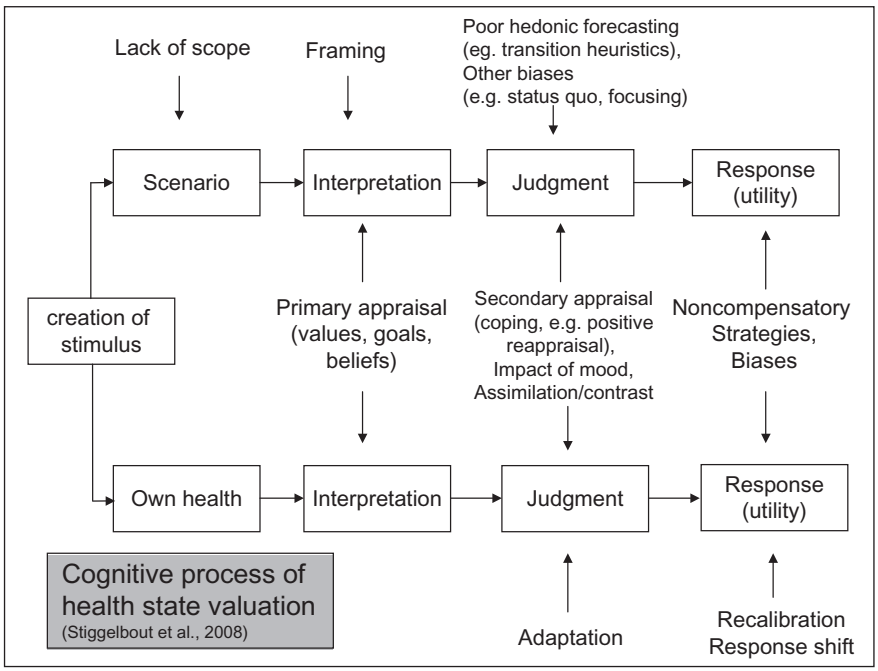

Figure 1 Theoretical model of cognitive process of health state valuation (Stiggelbout and de Vogel-Voogt, 2008).

either judge based on her/his own health or experiences or from a third person perspective. If a third person perspective is adopted, it is important that the respondent be prepared to solve a relatively difficult judgmental task. If the respondent is not prepared and just undergoes the task for other reasons, e.g. to obtain an honorarium, more or less 'random' judgments are produced because of lack of scope. The situation will then be evaluated. In a third person perspective, the scenario given by the interviewer is key, as this frames the activation of one's own relevant experiences, values, and emotions. The exact response format also contributes to the overall framing (Bless *et al.*, 1992; Schwarz and Hippler, 1987). At this point the secondary appraisal with the valuation of health states will occur. It will rely on the usual principles of human decision-making in complex circumstances (Stiggelbout and de Vogel-Voogt, 2008). These include:

- judgmental heuristics such as overweighting of rare or salient events and underweighting of the usual and expected (Gigerenzer *et al.*, 1999; Kahneman *et al.*, 1982);
- use of own mood and emotions as informative indicators for valuation of other health states (Schwarz and Clore, 2003);
- assimilation and contrast with standards stored in long-term memory and judged relevant for the current task (Sudman *et al.*, 1996);
- biases arising from the direction of change of health states (Wänke *et al.*, 1995) (equal level of improvement valued as less important compared to a change for the worse).

The judgmental processes are unconscious and are unlikely to be influenced by deliberation. Conversely, the last step of the valuation incorporates the deliberate editing of the judgment. An answer will be given, which not only reflects the subjective valuation of the health states, but is also compatible with other principles the person believes are relevant for themselves and the interview situation. For example, considerations of fairness or political correctness may influence the final judgment in addition to the subjective valuations. The cognitive model used here unlike others (e.g. prospect theory by Kahneman and Tversky, 1979) does not explicitly incorporate time perspective into the judgmental process (see Rasiel *et al.*, 2005, for an application), but known empirical consequences of (for instance) respondent's age or duration of health states are referred there in the 'secondary appraisal' step as biases and/or adaptation processes.

The measurement of DWs: methods and critical discussion

Measurement of DWs: concepts and operationalizations

DWs place differing health states onto a common continuum ranging from zero to one; perfect health has a value of zero and death has a value of one. A classical economist may sometimes separate between values of a quantitative evaluation resulting from a decision under certainty, and preferences resulting from a decision under uncertainty (Drummond *et al*., 1997). Throughout this article we refer to both concepts (i.e. values, preferences) as synonyms, because from a psychological perspective 'objective' uncertainty seems negligible when compared to characteristics as perceived by the respondent or decision-maker (e.g. subjective probability).

To arrive at the non-fatal component of a global summary measure for population health, the incidence of health states is usually recorded, weighted for the average duration of the respective state and then DWs are applied (see Equation 2). The term 'health state' is not identical with the concept of 'disease', as several diseases can coincide into the same health state (e.g. deafness resulting from different diseases) or one disease can cause various health states for the person affected during its (natural or treated) course. We will discuss various systems to classify health states later in this article.

There is no gold standard for eliciting DW for health states. Methods used stem from two distinct traditions: psychometrics (Revicki and Kaplan, 1993) and economic evaluation (Dolan, 2000). Among the most popular techniques for constructing quantitative DW from a psychometric tradition are:

- Rating scales and questionnaire-based instruments which have been shown to be transformable into a DW (McDowell, 2006; for example of an instrument see the Health Utilities Index later): the respondent uses Likert scale response formats to describe ability to perform various health related activities or emotional states.
- Ranking exercises (Klein *et al*., 2004; see also Salomon 2003, how ranks could be transformed into DWs): respondents order a number of vignettes describing health states according to their perceived level of disability.
- Magnitude estimation (ME; Beltyukova *et al*., 2008): respondents produce a direct number denoting how many times a certain health state is more disabling compared to another health state.
- Visual analogue scaling (VAS; Parkin and Devlin, 2006; van Osch and Stiggelbout, 2005): respondents give estimations of the level of disability by marking on a printed line, where zero denotes full health and one death.
- Pairwise comparison (PC), which can be transformed into a DW based on Thurstone's law of comparative judgment (Thurstone, 1927); PCs have also been used as basis for the economic 'stated preference' approach (Lancaster, 1971; Lancsar *et al*., 2007).

Economic theory based DW often use the following trade-off techniques:

- Standard gamble (SG; Gafni, 1994; Morimoto and Fukui, 2002; van Osch and Stiggelbout, 2008): respondents choose between a certain, but suboptimal health state *A* and a lottery between perfect health and death. The probabilities of health and death are varied and the point of indifference (a choice between health and death seems impossible at this point) is used to estimate the difference in utility between health state *A* and perfect health.
- Time trade-off (TTO; Buckingham and Devlin, 2006) (see later).
- Person trade-off techniques (PTO; Green, 2001; Pinto Prades, 1997) (see later).

Slightly different in logic, is the willingness to pay approach (WTP, Olsen and Smith, 2001), where respondents are asked to state how many resources they are willing to pay for achieving a certain health state.

Estimation of DWs in the GBD 1990 study (version: Murray and Lopez, 1996)

Murray and Lopez' (1996) descriptions of a health state via labels is a diagnostic term (e.g. 'active psychosis')

Active psychosis – an individual with paranoid delusions, auditory hallucinations and disorganized speech.
Paraplegia - with a rudimentary wheelchair. Associated complications may be factored in, for example, decubitus, ulcers and frequent urinary tract infections.

Figure 2 Description of a health state via disease characteristics (cited from Murray, 1996).

Int. J. Methods Psychiatr. Res. 19(1): 18–33 (2010). DOI: 10.1002/mpr

followed by some key symptoms that can either illustrate the state or give information on the severity of illness. As can be seen from the example describing 'paraplegia' in Figure 2, the potential sequelae or comorbidities are sometimes also included in the description (e.g. secondary infections). As the assessment of DWs should only concentrate on the decline of health during the time lived with the disability, and not on course, prediction of outcome or comorbidity, such descriptions should be avoided when eliciting the DW for 'paraplegia' on its own. Of course, other aspects are also important but within the burden of disease framework they deal only with the epidemiological calculations.

In the 1990 GBD, two trade-off methods, PTO and TTO, were used for eliciting DWs for health states. In the PTO method, respondents (mostly clinical experts) were presented with a scenario in which 1000 healthy individuals could be prevented from death for exactly one year through the purchase of a certain unspecified intervention 'A'. They then traded the number of individuals being in a certain (suboptimal) health state against this anchor; this meant that the decision-maker could purchase exactly the same life prolongation via an alternative intervention B at exactly the same costs, but covering more people. Thus, the number of additional patients that could outweigh the 1000 healthy life years gained under alternative A was interpreted as the metric of the distance between the optimal health and the disease state under research. In a second variant of PTO, not only could the lives of the patients in alternative B be saved for one year, but their health condition could also be completely cured.

The second trade-off method, TTO, used a personal perspective and asked the decision-makers how many years of a life expectancy, defined as 20 years, they would give up in order to be cured from the respective suboptimal health state. A second variant of the TTO used the age-adjusted conditional life expectancy of the decision-makers as the basis for trading one's life against a complete cure from the respective health state.

The PTO and TTO procedures used to elicit DWs by Murray (1996) have been the subject of considerable criticism. A summary of this criticism and the underlying empirical foundations can be found in Table 1, column 2.

Description of health states

DWs are based on valuations of health states. In practical terms, this means that in the derivation of DW, the respondent must have the health states presented with short descriptions, often no longer than a paragraph (see the health state descriptions of the 1990 GBD). There are different ways to present such health states. The most common descriptions consist of either a list of symptoms characterizing a disease, a description of functional or activity limitations associated with disease states, or a combination of both. One basic distinction concerns the question of whether health states should be presented with a genuine description specific for each health state (for an example see Figure 2 from Murray, 1996), above, or with a listing of levels of a limited number of attributes which are the same for all health states (however, if there is no functional or activity limitation, the attributes will

A person can be described as having the following problems:	
Pain or discomfort	Moderate pain or discomfort
Physical functioning	Severe limitations in physical functioning
Emotional state	Very unhappy
Fatigue	Most of the time feel tired and have little energy
Memory and thinking	Very forgetful and have great difficulty when trying to think or solve day-to-day problems
Social relationships	Severe limitations in the capacity to sustain social relationships
Anxiety	Mild levels of anxiety experienced occasionally
Speech	Unable to be understood when speaking to other people
Vision	Unable to see well enough, even with glasses or contact lenses, to read ordinary newsprint but can see well enough to recognize a friend on the other side of the street
Use of hands and fingers	Limitations in the use of hands and fingers, require the help of another person for some tasks

Figure 3 Sample description of health states using unlabeled functional and activity limitations (cited from McIntosh *et al*., 2007).

Table 1 Major criticism on the choice scenarios in the original GBD study (Murray, 1996)

Scenario	Critique	Precautionary measure
Critique on PTO-scenarios of the 1990 GBD study		
'If you purchase intervention A, you will extend the life of 1000 healthy individuals for exactly one year, at which point they will all die. If you do not purchase intervention A, they will all die today.'	Artificial 'certainty' about life expectancy (Hammerl, 2000; Hertwig, 1998) but see also QALY-HYE controversy on integrating uncertainty	Formulate realistic time frame for any public health effect mentioned in scenarios
	Logical contradiction: individuals are 'healthy', but will die today	Avoid pseudo-exact timing for mentioned health effects
PTO1: 'At the same cost: If you purchase intervention B, $n = 2000$ blind individuals' lives are extended for exactly one year. If you prefer intervention B, the number of individuals is reduced.'	No variation in efficacy of treatment (duration of life after cure) between patients; no explicit mentioning of death in alternative B (rule of rescue – McKie and Richardson, 2003 –) favours A	Allow for inter-individual variation of health effects and refer only to sums of the effects. Balance the salience of potential fatal outcomes between scenarios.
PTO2: 'If you purchase intervention B, you can cure the disability, and *n* individuals will live exactly one year. Without the intervention they all will die today.'	No variation in efficacy of treatment (duration of life after cure) between patients.	Allow for inter-individual variation of health effects and refer only to sums of the effects.
	A life-saving measure is compared to a life-saving measure that also cures the health state. => Two judgmental dimensions involved which also could display interaction effects.	Strictly construct effects involving only one dimension of potential gains (e.g. only years of life, or only health status after cure).
Critique to TTO-scenarios of the 1990 GBD study		
TTO1: 'Suppose you could expect to live 20 years in chronic pain. How many of those years would you give up to live the remaining years without pain?'	Artificial 'certainty' about life expectancy [see van Nooten and Brouwer (2004) for the impact of subjective expectations on TTO responses]	Change time frame from life expectancy with censored endpoint to proportions of fixed periods (e.g. hours per day, see Buckingham *et al*., 1996, for an example)
TTO2: 'What is the smallest number of years in perfect health that you would accept in exchange for 20 years with severe abdominal pain?'	Anchoring effect of respondent's age (Burström *et al*., 2006; Chapman and Johnson, 1999; Richards and Wierzbicki, 1990; Sherbourne *et al*., 1999)	Change personal into societal perspective
	Personal experience with scenario not controlled (Goldberg, 2006; Rehm and Strack, 1994)	Restrict to scenarios which do not require prior expertise for decision
	Adaptation to disease not anticipated (see Damschroder *et al*., 2005b)	Avoid personal perspective
	Unrealistic expectation of a continuous healthy state after cure until death: omission of competing risks. (see Spencer, 2003), for incorporation of sequence effects into elicitation method)	Avoid formulations claiming ever-lasting effects (except death)

Int. J. Methods Psychiatr. Res. 19(1): 18–33 (2010). DOI: 10.1002/mpr

sometimes not be mentioned). An example of the latter, based on functional or activity limitations, is displayed in Figure 3 [based on the Canadian Classification and Measurement System for Functional Health (CLAMES); cf. McIntosh *et al.*, 2007].

These health states can be presented either with or without the disease label (diagnosis, see examples Figures 2 and 3). Cognitively, this makes a difference as disease labels carry information not only with respect to health attributes, in part over and above the listed attributes (Sackett and Torrance, 1978), but also with respect to stigma and other forms of social evaluations (Frick *et al.*, 1988). Clearly, such a presentation of the disease labels also has an impact on which population group can and should be asked to value the health state, as the knowledge basis about certain diseases differs. Conversely, knowledge about functional limitations such as mobility limitations is universal and we expect fairly similar knowledge bases in professional and non-professional settings. In order to measure and potentially exclude effects of stigma and preconceived schemata regarding diseases, two versions of health state descriptions can be used in parallel: one with disease labels and one without.

Choice of decision-maker

The classical approach for deriving DWs was based on medical expertise (Murray and Lopez, 1996), i.e. a series of expert meetings in various countries, where the health state descriptions are short and disease oriented (see Figure 2). More recent work such as the Dutch Disability Weights Project (Stouthard *et al.*, 1997) has supplemented disease-based descriptions with descriptions based on the European Quality of Life Five-dimensions Index Plus (EQ-5D) classification, in other words based on functional and activity limitations. In principle, health state descriptions based on these limitations can be used in a general population framework (McIntosh *et al.*, 2007) as well as with medical experts or patients. Different considerations can also be brought forward in favour of one or another group of decision-makers; health professionals certainly possess the most knowledge about the health states to be compared including knowledge about functional limitations. Patients or their family members also have accumulated knowledge about their conditions, but they may not have the best knowledge about other conditions (McNamee, 2007). Another argument for this states that as the DWs will later often be used in decision-making about resource allocation, the perspective of the general population should be taken into account (Boyd *et al.*, 1990; Ubel *et al.*, 2000; Wiseman *et al.*, 2003). This argument however, is not relevant in a framework where DW should mainly reflect levels of health, independent of later resource allocation decisions.

Choices made for the US National Institutes of Health (NIH) DALY study

Methodology

We opted for a mixed methodology including:

- VAS mainly used in the warm-up exercises;
- PCs and ranking, as input to stated preference analyses;
- PTO and TTO as economic approaches.

We omitted the SG approach, as this method requires a high level of numeracy (see Woloshin *et al.*, 2001; Zikmund-Fisher *et al.*, 2007, for impact of numeracy on decision-making), and has been shown to be influenced by the effect of gain versus losses more than its alternatives (Blumenschein and Johannesson, 1998) whose results must be corrected for loss aversion and probability weighting before being comparable to other methods (van Osch and Stiggelbout, 2008).

The planned approach for statistical analyses uses factor and factor mixture analyses in order to elicit the final DWs based on input from both the psychometric and econometric methodologies (Flora and Curran, 2004; Muthen, 2006). All statistical analyses rely on the notion of a latent variable, i.e. the DW, which impacts the various operationalizations. DW is allowed to vary between zero and one only, i.e. no negative health states worse than death are considered.

Health state descriptions

We opted for the variant of describing the health states via levels of functional and activity limitations. These limitations are based on the standardized descriptions of the CLAMES system, a system developed independently by Statistics Canada for the purpose of comprehensively describing all health states in Canada with a uniform framework (see Figure 3; see also McIntosh *et al.*, 2007; http://www.statcan.ca/bsolc/english/bsolc?catno = 82-005-X20030016643). The CLAMES contains 11 health status attributes adapted from three leading generic health status instruments: the Health Utilities Index Mark III (HUI3; Feeny *et al.*, 2002; Furlong *et al.*, 2001), the Medical Outcomes Study Short-form 36 (SF-36; Ware and Sherbourne, 1992); and the European Quality of Life Five-dimensions Index Plus (EQ-5D; Brooks

Table 2 Domains of health and functioning as measured by standard measures (from Wolfson, 2003)

SF-36 (Ware and Sherbourne 1992)	HUI3 (Feeny *et al.*, 2002; Feeny, 2002)	EQ-5D (Brooks and EuroQoL Group, 1996; EuroQoL Group, 1990)	WHO WHS (WHO, 2008)	Statistics Canada CLAMES (McIntosh *et al.*, 2007)
Physical function[a]	Ambulation	Mobility	Mobility	Physical function
Role limits – physical		Self-care	Self-care	
Mental Health	Emotion[a]		Affect	Emotion
Role limits – emotional		Anxiety[a]/depression		Anxiety[b]
Pain	Pain/discomfort[a]	Pain/discomfort[a]	Pain	Pain/discomfort
Social function[a]		Usual activities	Usual activities	Social relationships
General health				
Energy[a]				Fatigue
	Memory and thinking[a]	Cognition (Dutch version)	Cognition	Memory and thinking
	Vision[a]			Vision[b]
	Hearing[a]			Hearing[b]
	Speech[a]			Speech[b]
	Use of hands and fingers[a]			Use of hands and fingers[b]

[a] Main source (at least in general terms) for domain in new Statistics Canada system.
[b] Secondary domain in new Statistics Canada system.

and EuroQoL Group, 1996; EuroQoL Group, 1990; Rabin and de Charro, 2001). CLAMES focuses on individuals' capacities (i.e. what they are able to do) with respect to the various attributes, each of which has four or five levels ranging from normal to severely limited functioning. Each health state is represented by levels of functional and activity limitation for each of 11 attributes; thus, 10,240,000 health states are possible within the system (see Figure 3 and Table 2). For the workshops, we prepared anchor descriptions for each of the levels of attributes in case respondents requested such anchors.

The choice of CLAMES also allows better flexibility with respect to integrating new health states. Once there is a description of any health state within the CLAMES system a DW can be derived, as can a valuation which specifies weights for each attribute level and, where necessary, combinations thereof.

In all tasks where health states were to be compared, great care was taken to underline the fact that all valuations should be based on the same duration for all of the valuated health states. In case a workshop insisted on specifying this duration, one month was given as this duration.

Specific operationalization of the PTO and TTO

Based on the literature, the following choices with respect to operationalization were undertaken (see Table 3).

This led to the scenarios described later (Figure 4). PTO scenarios were separately constructed for the prevention and cure of health states. Death as a potential outcome of the scenarios was never mentioned nor set as reference standard in the PTO scenarios, and was avoided as ascertained outcome for the TTO scenario. Though Buckingham *et al.* (1996) proposed some alternatives to the problematic use of a whole lifetime, arguing it as a rather abstract and unfamiliar metric when trading one's time, we kept duration of life as our measure for TTO elicitation in order to be comparable with the framework of the PTO-scenarios. But we formulated our TTO scenario from a societal perspective (deviating from the usual personal perspective, see also Burström *et al.*, 2006) in order to minimize heterogeneity and inconsistencies in the derived DWs due to personal experiences and beliefs (Bravata *et al.*, 2005). PTO elicitation was performed with the upward titration method, and TTO elicitation was performed with the ping-pong method (Lenert *et al.*, 1998).

Table 3 Considerations for constructing best operationalizations

Source of judgmental bias	Elicitation method affected	Author(s)/study	Precautionary measure
Direction of comparison (which health state is subject, and which is the referent) determines the selection of attributes relevant for the comparison => serious order effects shown in PTO-judgments	PCs, trade-off methods	(Schwarz and Sudman, 1995; Ubel *et al*., 2002; Wänke *et al*., 1995)	Balancing position of health state (l/*r*) over subjects by random assignment; providing as much as time for judgment, as subjects feel necessary for their evaluations
Perspective of judgment (personal involvement in potential gains) alters magnitude of priorities (irrespective of traded good such as people or time)	Trade-off scenarios (mostly PTO, TTO)	(Dolan *et al*., 2003; Richardson and Nord, 1997)	Strictly confining scenarios to social perspective with ex-post timeframe (utilization of treatment, not availability) => formulating TTO-scenarios without personal involvement in effects (see also Tsuchiya, 1999)
SG scores are influenced by numeracy, loss aversion, and probability weighting of respondents	SG scenarios	(van Osch and Stiggelbout, 2008; Woloshin *et al*., 2001; Zikmund-Fisher *et al*., 2007)	Omitting scenarios with probability as a measuring concept
SG assigns lower values especially to milder health states than TTO	SG scenarios	(Tsuchiya *et al*., 2006)	Omitting scenarios with probability as a measuring concept
SG is much more sensitive to framing effects (gains/losses) relative to TTO elicitation	SG and TTO scenarios	(Blumenschein and Johannesson, 1998)	Avoiding SG scenarios
Improving health (= curing 'losses') is strongly prioritized against avoiding decline (= collecting preventive 'gains')	Trade-off scenarios comparing cure to prevention:	(Schwappach, 2002)	Avoiding trade-off scenarios which mix the two types of public health effects; eliciting health state values both in a 'cure' and a 'prevention' metric separately
Health states requiring life-sustaining treatment, but judged better than death, are associated with inconsistent valuation of their duration (the longer, the worse)	TTO-scenarios involving health states with serious conditions	(Fried *et al*., 2007; Stalmeier *et al*., 2007)	Change time to a prospective period (e.g. from now until death) to time already 'consumed' (e.g. trading people of differing ages)
Individual differences determine favouring of cure over prevention. There are also subjects favouring prevention, but only a minority weighs prevention/cure – gains equally	Trade-off scenarios comparing cure to prevention	(Ubel *et al*., 1998); partly contradicts results of Schwappach (2002)	Even if evidence is contradictory, scenarios that mix cure and prevention should be avoided

Table 3 *Continued*

Source of judgmental bias	Elicitation method affected	Author(s)/study	Precautionary measure
PTO-scenarios do not result in multiplicative transitivity	PTO-scenarios	(Schwarzinger *et al.*, 2004)	Checking transitivity before using scores, respectively, using only the ordinal information of the first step of the PTO eliciting procedure, if not given
Severity of illness (initial state) increases preference for the same amount of amelioration/ recovery/improvement as a more lenient state	Only trade-off scenarios involving cure from life-threatening condition; not affected: PCs or rankings	(Nord, 1993b, 2004, 2005)	Using PCs or rankings for measurement. In trade-off comparisons, balance health state to both more serious and more lenient health states and use multivariate regression techniques to adjust net effect
A sub-optimal health state after treatment (final state) is weighted lower than would be expected from a health maximization principle, because equity of access to health care is considered more important	Only trade-off scenarios involving cure or life-saving	(Abellan-Perpinan and Pinto-Prades, 1999; Nord, 1993a)	Avoiding trade-off scenarios that could involve considerations of both health gain and equity of access

Overall and in line with the cognitive perspective on decision-making, we tried to implement tasks which would involve the decision-maker. There has to be 'experimental realism' in the task (Aronson *et al.*, 1990), with impact on the decision-maker, which does not necessarily mean that tasks have to simply copy related real tasks (as assumed by Dolan *et al.*, 2003); in fact, experimental realism often implies operationalizations that capture the underlying theoretical concepts without copying everyday situations (Aronson *et al.*, 1990; Rehm and Strack, 1994).

Choice of decision-makers

We plan to have meetings to derive DWs in all three settings discussed in the literature, i.e. with health professionals, patients, and the general population. It is hypothesized that preference judgments converge between these groups, especially if they are based on a description of functional limitations with the only exception of labelling diseases. However, if our hypothesis does not hold true, we may require a relatively large number of expert meetings to create setting-specific preference weights with relatively small confidence intervals.

Conclusions

DWs are a key ingredient for estimating DALYs or other summary measures of health. To empirically assess such weights, social psychological evidence on judgmental processes, question formulation and formatting has to be taken into consideration. We hope that this review will stimulate more empirical research on assessing DWs in both the United States and international contexts.

Acknowledgements

Financial support for this study was provided by National Institute for Alcohol Abuse and Alcoholism (NIAAA) with contract # HHSN267200700041C to conduct the study 'Alcohol- and Drug-attributable Burden of Disease and Injury in the US' to the first author. We would like to thank T.K. Li and B. Grant for their support of this study, and for organizing meetings to discuss the methodology. F. Kanteres and H. Irving helped with copy editing the manuscript and

Int. J. Methods Psychiatr. Res. 19(1): 18–33 (2010). DOI: 10.1002/mpr

Person trade-off I: rationing

Imagine that you are a decision-maker within of a health system. The health system operates with a limited budget, and you have enough money to fund only one of two treatment programmes. Although other members of your staff may offer their input, the final decision is up to you. Imagine you must choose between two treatment programmes, either of which will use up all the money that is available. The money cannot be split between the two programmes, so you will be able to fund only one of them. The other programme will not be funded, and the people who have the condition covered by that programme will go untreated. One programme will cure 1000 people with [disease state A here]; the other programme will cure 1000 people with [disease state B here]. Thus, the two programmes will treat the same number of people. Who would you cure, thereby leaving the other group without treatment?

Description of [disease states A and B] with CLAMES either with or without disease labels

Person trade-off II: preventing incidence

Program A prevents the occurrence of [disease state X here] in 1000 people in your city in two to four years time. The identity of these people is unknown. With this programme, they will live in normal health with an average life expectancy. Programme B prevents the occurrence of [disease state B here] in 1000 people in your city within the same time span. Which programme would you choose?

Description of [disease states A and B] with CLAMES either with or without disease labels

Time trade-off

Imagine the following scenario: you are in a hospital, there is a code red for fire, and while trying to leave the building, you enter a room with two 70-year-old patients, one is in health state A, and the other suffers from health state B. You do not know either of these persons. Both patients have lost consciousness. You can save only one of them by carrying him outside through the smoke. Of course you hope that the other one will be saved as well, but you cannot be sure of that.

Which person do you save?

Description of [disease states A and B] with CLAMES either with or without disease labels

Figure 4 Examples of trade-off scenarios to be used in current NIH DALY weights project.

made valuable contributions on previous versions of the text. Participation and discussions at two meetings and expert hearings within the framework of the Global Burden of Disease 2005 Study helped us to finalize this manuscript.

Declaration of interest statement

The authors have no competing interests.

References

Abellan-Perpinan J.M., Pinto-Prades J.L. (1999) Health state after treatment: A reason for discrimination? *Health Economics*, **8**, 701–707.

Arnesen T., Nord E. (1999) The value of DALY life: Problems with ethics and validity of disability adjusted life years. *BMJ*, **319**, 1423–1425.

Arnesen T., Trommald M. (2005) Are QALYs based on time trade-off comparable? – A systematic review of TTO methodologies. *Health Economics*, **14**, 39–53.

Aronson E., Ellsworth P., Carlsmith J.M., Gonzales M. (1990) *Methods of Research in Social Psychology*, 2nd edition, McGraw-Hill.

Beltyukova S.A., Stone G.E., Fox C.M. (2008) Magnitude estimation and categorical rating scaling in social sciences: A theoretical and psychometric controversy. *Journal of Applied Measures*, **9**, 151–159.

Bleichrodt H., Doctor J., Stolk E. (2005) A nonparametric elicitation of the equity-efficiency trade-off in cost-utility analysis. *Journal of Health Economics*, **24**, 655–678.

Bless H., Bohner G., Hild T., Schwarz N. (1992) Asking difficult questions: Task complexity increases the impact of response alternatives. *European Journal of Social Psychology*, **22**, 309–312.

Blumenschein K., Johannesson M. (1998) An experimental test of question framing in health state utility assessment. *Health Policy*, **45**, 187–193.

Boyd N., Sutherland H., Karen Z., Heasman D., Cummings B. (1990) Whose utilities for decision analysis? *Medical Decision Making*, **10**, 58–67.

Bravata D., Nelson L., Garber A., Goldstein M. (2005) Invariance and inconsistency in utility rankings. *Medical Decision Making*, **25**, 158–167.

Breslow L. (2006) Health measurement in the third era of health. *American Journal of Public Health*, **96**, 17–19.

Brooks R., EuroQoL Group (1996) EuroQoL: The current state of play. *Health Policy*, **37**, 53–72.

Broome J. (2004) *Weighing Lives*, Oxford University Press.

Buckingham J., Birdsall J., Douglas J. (1996) Comparing three versions of the trade-off: Time for a change? *Medical Decision Making*, **16**, 335–347.

Buckingham J., Devlin N. (2006) A theoretical framework for TTO valuations of health. *Health Economics*, **15**, 1149–1154.

Burström K., Johannesson M., Diderichsen F. (2006) A comparison of individual and social time trade-off values for health states in the general population. *Health Policy*, **76**, 359–370.

Chapman G., Johnson E. (1999) Anchoring, activation and the construction of values. *Organizational Behavior and Human Decision Processes*, **79**, 115–153.

Damschroder L.J., Roberts T.R., Goldstein C.C., Miklosovic M.E., Ubel P.A. (2005a) Trading people versus time: What is the difference? *Population Health Metrics*, **3**, 10.

Damschroder L.J., Roberts T.R., Zikmund-Fisher B.J., Ubel P.A. (2007) Why people refuse to make tradeoffs in person tradeoff elicitations: A matter of perspective? *Medical Decision Making*, **27**, 266–280.

Damschroder L.J., Zikmund-Fisher B.J., Ubel P.A. (2005b) The impact of considering adaptation in health state valuation. *Social Science & Medicine*, **61**, 267–277.

Doctor J., Bleichrodt H., Lin H.J. (2010) Health utility bias: A systematic review and meta-analytic evaluation. *Medical Decision Making*, **30**, 58–67, DOI: 10.1177/0272989X07312478.

Dolan P. (2000) The measurement of health-related quality of life for use in resource allocation decisions in health care. In: *Handbook of Health Economics* (eds Culyer A, Newhouse J), pp. 1724–1760, Elsevier.

Dolan P.A., Olsen J.A. (2001) Equity in health: The importance of different health streams. *Journal of Health Economics*, **20**, 823–834.

Dolan P., Olsen J.A., Menzel P., Richardson J. (2003) An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Economics*, **12**, 545–551.

Drummond M., O'Brien B., Stoddart G.L., Torrance G.W. (1997) *Methods for the Economic Evaluation of Health Care Programmes*, Oxford University Press.

Etches V., Frank J., Di Ruggiero E., Manuel D. (2006) Measuring population health: A review of indicators. *Annual Review of Public Health*, **27**, 29–55.

EuroQoL Group (1990) EuroQoL – a new facility for the measurement of health-related quality of life. *Health Policy*, **16**, 199–208.

Feeny D. (2002) The utility approach to assessing population health. In: *Summary Measures of Population Health. Concepts, Ethics, Measurement and Applications* (eds Murray CJL *et al.*), pp. 515–528, WHO.

Feeny D., Furlong W., Torrance G.W., Goldsmith C.H., Zhu Z., DePauw S., Denton M., Boyle M. (2002) Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care*, **40**, 113–128.

Field M.J., Gold G.M. (1998) *Summarizing Population Health: Directions for the Development and Application of Population Metrics*, National Academy Press.

Flora D.B., Curran P.J. (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, **9**, 466–491.

Fowler F. (1995) *Improving Survey Questions: Design and Evaluation*, Applied Social Research Methods Series, Sage Publications.

Frick U., Rehm J., Fichter M., Koloska R. (1988) Therapeutische Rollenorientierung und soziale Repräsentationen von stationären Alkoholismus-Patienten. *Zeitschrift für Soziologie*, **7**, 218–226.

Fried T.R., O'Leary J., Van Ness P., Fraenkel L. (2007) Inconsistency over time in the preferences of older persons with advanced illness for life-sustaining treatment. *Journal of the American Geriatrics Society*, **55**, 1007–1014.

Furlong W.J., Feeny D.H., Torrance G.W., Barr R.D. (2001) The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine*, **33**, 375–384.

Gafni A. (1994) The Standard Gamble method: What is being measured and how it is interpreted. *Health Services Research*, **29**, 207–224.

Gigerenzer G., Todd P.M., ABC Research Group (1999) *Simple Heuristics that make us Smart*, Oxford University Press.

Goldberg J. (2006) Being there is important, but getting there matters too: The role of path in the valuation process. *Medical Decision Making*, **26**, 323–337.

Green C. (2001) On the societal value of health care: What do we know about the person trade-off technique? *Health Economics*, **10**, 233–243.

Green C., Brazier J., Deverill M. (2000) Valuing health-related quality of life. A review of health state valuation techniques. *Pharmacoeconomics*, **17**, 151–165.

Grice P.H. (1975) Logic and conversation. In: *Syntax and Semantics* (eds Cole P, Morgan JL), pp. 41–58, Academic Press.

Hammerl M. (2000) Experimental realism in social psychological research [Experimentelle realitätsnähe in der sozialpsychologischen Forschung]. *Zeitschrift für Sozialpsychologie*, **31**, 143–152.

Hertwig R. (1998) Psychologie, experimentelle Ökonomie und die Frage, was gutes Experimentieren ist. *Zeitschrift für Experimentelle Psychologie*, **45**, 2–19.

Kahneman D., Tversky A. (1979) Prospect theory: An analysis of decision under risk. *Economica*, **47**, 263–261.

Kahneman D., Tversky A., Slovic P. (1982) *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press.

Klein M., Dülmer H., Ohr D., Quandt M., Rosar U. (2004) Response sets in the measurement of values: A comparison of rating and ranking procedures. *International Journal of Public Opinion Research*, **16**, 474–483.

Lancaster T. (1971) *Consumer Demand: A New Approach*, Columbia University Press.

Lancsar E., Louviere J., Flynn T. (2007) Several methods to investigate relative attribute impact in stated preference experiments. *Social Science Medicine*, **64**, 1738–1753.

Lenert L., Cher D., Goldstein M., Bergen M., Garber A. (1998) The effect of search procedures on utility elicitations. *Medical Decision Making*, **18**, 76–83.

McDowell I. (2006) *Measuring Health – a Guide to Rating Scales and Questionnaires*, Oxford University Press.

McDowell I., Spasoff R., Kristjansson B. (2005) On the classification of population health measurements. *American Journal of Public Health*, **94**, 388–393.

McIntosh C.N., Connor Gorber S., Bernier J., Berthelot J.M. (2007) Eliciting Canadian population preferences for health states using the Classification and Measurement System of Functional Health (CLAMES). *Chronic Diseases in Canada*, **28**, 29–41.

McKenna M., Marks J. (2002) Commentary on the uses of summary measures of population health. In: *Summary Measures of Population Health. Concepts, Ethics, Measurement and Applications* (eds Murray CJL *et al.*), pp. 67–74, WHO.

McKie J., Richardson J. (2003) The rule of rescue. *Social Science Medicine*, **56**, 2407–2419.

McNamee P. (2007) What difference does it make? The calculation of QALY gains from health profiles using patient and general population values. *Healthy Policy*, **84**, 321–331.

Morimoto T., Fukui T. (2002) Utilities measured by rating scale, time trade-off, and standard gamble: Review and reference for health care professionals. *Journal of Epidemiology*, **12**, 160–178.

Mortimer D., Segal L. (2008) Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QUALY-weights. *Medical Decision Making*, **28**, 66–89.

Murray C., Lopez A. (2000) Progress and directions in refining the global burden of disease approach: A response to Williams. *Health Economics*, **9**, 69–82.

Murray C., Lopez A. (1996) *The Global Burden of Disease*, Harvard School of Public Health (on behalf of the WHO and World Bank), Harvard University Press.

Murray C., Salomon J., Mathers C. (2000) A critical examination of summary measures for population health. *Bulletin of the WHO*, **78**, 981–994.

Murray C., Salomon J., Mathers C., Lopez A. (2002) *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*, WHO.

Murray C.J.L. (1994) Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bulletin of the WHO*, **72**, 429–445.

Murray C.J.L. (1996) Rethinking DALYs. In: *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020* (eds Murray CJL, Lopez A), pp. 1–98, Harvard School of Public Health, Harvard University Press.

Muthen B. (2006) Should substance use disorders be considered as categorical or dimensional? *Addiction*, **101**, 6–16.

Nord E. (1993a) The relevance of health state after treatment in prioritising between different patients. *Journal of Medical Ethics*, **19**, 37–42.

Nord E. (1993b) The trade-off between severity of illness and treatment effect in cost-value analysis of health care. *Health Policy*, **24**, 227–238.

Nord E. (2004) Concerns for the worse off: Fair innings versus severity. *Social Science Medicine*, **60**, 257–263.

Nord E. (2005) Severity of illness and priority setting: Worrisome lack of discussion of surprising finding. *Journal of Health Economics*, **25**, 170–172.

Olsen J., Smith R. (2001) Theory versus practice: A review of 'willingness-to-pay' in health and health care. *Health Economics*, **10**, 39–52.

Parkin D., Devlin N. (2006) Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics*, **15**, 653–664.

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

Rabin R., de Charro F. (2001) EQ-5D: A measure of health status from the EuroQoL group. *Annals of Medicine*, **33**, 337–343.

Rasiel E., Weinfurt K., Schulman K. (2005) Can prospect theory explain risk-seeking behavior by terminally ill patients? *Medical Decision Making*, **25**, 609–613.

Rehm J., Strack F. (1994) Kontrolltechniken. In: *Methodologische Grundlagen der Psychologie* (eds Hermann T, Tack W), pp. 508–555, Hogrefe.

Revicki D., Kaplan R. (1993) Relationship between psychometric and utility-based approaches to the measurement

of health-related quality of life. *Quality of Life Research*, **2**, 477–487.

Richards M., Wierzbicki M. (1990) Anchoring errors in clinical-like judgments. *Journal of Clinical Psychology*, **46**, 358–365.

Richardson J., Nord E. (1997) The importance of perspective in the measurement of Quality-adjusted Life Years. *Medical Decision Making*, **17**, 33–41.

Ryan M., Scott D.A., Reeves C., Bate A., van Teijlingen E.R., Russell E.M., Napper M., Robb C.M. (2001) Eliciting public preferences for healthcare: A systematic review of techniques. *Health Technology Assessment*, **5**, 1–186.

Sackett D., Torrance G. (1978) The utility of different health states as perceived by the general public. *Journal of Chronic Diseases*, **7**, 347–358.

Salomon J. (2003) Reconsidering the use of rankings in the valuation of health states: A model for estimating cardinal values from ordinal data. *Population Health Metrics*, **1**, 12.

Schwappach D. (2002) The equivalence of numbers: The social value of avoiding health decline: An experimental web-based study. *BMC Medical Informatics and Decision Making*, **2**, 3.

Schwappach D. (2005) Are preferences for equity a matter of perspective? *Medical Decision Making*, **25**, 449–459.

Schwarz N. (1996) *Cognition and Communication: Judgmental Biases, Research Methods and the Logic of Conversation*, Erlbaum.

Schwarz N., Clore G. (2003) Mood as information: 20 years later. *Psychological Inquiry*, **14**, 296–303.

Schwarz N., Hippler H.J. (1987) What response scales may tell your respondents: Informative functions of response alternatives. In: *Social Information Processing and Survey Methodology* (eds Hippler HJ, Schwarz N, Sudman S), pp. 163–178, Springer.

Schwarz N., Sudman S. (1995) *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, Jossey-Bass.

Schwarzinger M., Lanoe J., Nord E., Durand-Zaleski I. (2004) Lack of multiplicative transitivity in person trade-off responses. *Health Economics*, **13**, 171–181.

Sherbourne C.D., Keeler E., Unützer J., Lenert L., Wells K.B. (1999) Relationship between age and patients' current health state preferences. *Gerentologist*, **39**, 271–278.

Spencer A. (2003) A test of the QALY model when health varies over time. *Social Science Medicine*, **57**, 1697–1706.

Stalmeier P.F., Lamers L.M., Busschbach J.J., Krabbe P.F. (2007) On the assessment of preferences for health and duration. Maximal endurable time and better than dead preferences. *Medical Care*, **45**, 835–841.

Stiggelbout A., de Vogel-Voogt E. (2008) Health state utilities: A framework for studying the gap between the imagined and the real. *Value Health*, **11**, 76–87.

Stolk E., Pickee S., Ament A., Busschbach J. (2005) Equity in health care prioritisation: An empirical inquiry into social value. *Health Policy*, **74**, 343–355.

Stouthard M., Essink-Bot M.L., Bonsel G.J. (1997) *Disability Weights for Diseases in the Netherlands*, Erasmus University Rotterdam.

Sudman S., Bradburn N., Schwarz N. (1996) *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*, Jossey-Bass.

Thurstone L.L. (1927) A law of comparative judgement. *Psychological Review*, **34**, 273–286.

Tsuchiya A. (1999) The time trade-off method from the societal perspective. *Quality of Life Research*, **8**, 578.

Tsuchiya A.J., Brazier J., Roberts J. (2006) Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. *Journal of Health Economics*, **25**, 334–346.

Ubel P., Richardson J., Baron J. (2002) Exploring the role of order effects in person trade-off elicitations. *Health Policy*, **61**, 189–199.

Ubel P.A., 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**, 127–136.

Üstün T.B., Rehm J., Chatterji S., Saxena S., Trotter R., Room R., Bickenbach J. (1999a) Multiple-informant ranking of the disabling effects of different health conditions in 14 countries. *Lancet*, **354**, 111–115.

Üstün T.B., Saxena S., Rehm J., Bickenbach J. (1999b) Are disability weights universal? WHO/NIH Joint Project CAR Study Group. *Lancet*, **354**, 1306.

Ubel P.A., Spranca M.D., Dekay M.L., Hershey J.C., Asch D.A. (1998) Public preferences for prevention versus cure: What if an ounce of prevention is worth only an ounce of cure? *Medical Decision Making*, **18**, 141–148.

van der Maas P. (2003) How summary measures of population health are affecting health agendas. *Bulletin of the WHO*, **81**, 314.

van Nooten F., Brouwer W. (2004) The influence of subjective expectations about length and quality of life on time trade-off answers. *Health Economics*, **13**, 819–823.

van Osch S., Stiggelbout A. (2005) Understanding VAS valuations: Qualitative data on the cognitive process. *Quality of Life Research*, **14**, 2171–2175.

van Osch S., Stiggelbout A. (2008) The construction of standard gamble utilities. *Health Economics*, **17**, 31–40.

Wänke M., Schwarz N., Noelle-Neumann E. (1995) Asking comparative questions: The impact of the direction of comparison. *Public Opinion Quarterly*, **59**, 347–372.

Ware J., Sherbourne C. (1992) The MOS 36-item short form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, **30**, 473–483.

Wiseman V., Mooney G., Berry G., Tang K. (2003) Involving the general public in priority setting: Experiences from Australia. *Social Science Medicine*, **56**, 1001–1012.

Wolfson M. (2003) On the policy implications of summary measures of health status. In: *To Develop a Research Agenda and Research Resources for Health Status Assessment and Summary Health Measures* (eds National Center for Health Statistics), pp. 151–189, National Center for Health Statistics. http://www.cdc.gov/nchs/otheract/IAWG/NCHS_Report.pdf

Woloshin S., Schwartz L.M., Moncur M., Gabriel S., Tosteson A.N. (2001) Assessing values for health: Numeracy matters. *Medical Decision Making*, **21**, 382–390.

World Bank (1993) *World Development Report 1993 – Investing in Health*, World Bank/Oxford University Press.

World Health Organization (WHO) (2008) *World Health Survey*, WHO.

Zikmund-Fisher B.J., Smith D.M., Ubel P.A., Fagerlin A. (2007) Validation of the Subjective Numeracy Scale: Effects of low numeracy on comprehension of risk communication and utility elicitations. *Medical Decision Making*, **27**, 663–671.