

Distr.  
GENERAL

CES/AC.36/1998/10  
EUR/ICP/INFO 020603/10

3 August 1998

Original: ENGLISH

**STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE**

**WORLD HEALTH ORGANIZATION  
REGIONAL OFFICE FOR EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS**

Joint ECE-WHO Meeting on Health Statistics  
(Rome, Italy, 14-16 October 1998)

SESSION III : Development and use of health output indicators

**THE VALIDITY OF SUMMARY HEALTH INDEXES  
(HALE, QALYs AND DALYs)**

Invited paper submitted the National Institute of Public Health, Oslo, Norway <sup>1</sup>

**SUMMARY**

Single index valuations of health states on a scale from zero (equivalent to being dead) to unity (healthy) may be used to calculate Health Adjusted Life Expectancy (HALE), Burden of Disease in terms of DALYs and the value of health interventions in terms of Quality Adjusted Life Years (QALYs). Health measurements in terms of multidimensional profiles (e.g. the SF-36) cannot serve these purposes. The paper discusses the validity of different techniques for assigning single index values to health states, with emphasis on Multi-Attribute Utility Instruments. A case is made for seeing health state values as measures of societal value rather than individual utility in order to obtain values that are valid and meaningful over the whole range of possible states of illness.

---

<sup>1</sup> Prepared by Erik Nord, PhD.

## Introduction

There are two main ways of describing individuals' health in numerical terms. One is in terms of scores on a number of different dimensions of health, like mobility, pain, hearing and seeing. Together such unidimensional scores form multidimensional *health profiles*. Examples are the Sickness Impact Profile, the Nottingham Health Profile and the SF-36. The other way is to assign a score for overall health on a single scale from zero (equivalent to being dead) to unity (healthy). Such a single index score - hereafter referred to as a *health state value* - is usually interpreted as a measure of *health related quality of life*. Multi-attribute utility (MAU) instruments like the EuroQol (Brooks et al, 1996), the Health Utility Index (Feeny et al, 1995) and the 15-D (Sintonen and Pekurinen, 1993) provide both profiles and health state values.

Clearly, health profiles convey more information than health state values. However, profiles have the weakness that they do not always allow judgements of which of two health states is better, since one state may have higher scores on some dimensions, and the other state higher scores on other dimensions, and there is no way of judging which of the differences is more important. Health state values purport to resolve this problem by concentrating all profile information in one single number, according to which different complex health states may be ranked in terms of their overall value to the individuals concerned.

Health state values are also a way of avoiding information overload in people responsible for resource allocation decisions in health care. Rather than having to take on board complex health profiles for a number of different diagnostic groups, decision makers may be aided in their thinking by a set of single index numbers that summarize the situation of each group.

Health state values have an important additional property. Not only do they allow an ordinal ranking of health states. They also purport to express trade-offs between quality of life and quantity of life and thus allow the summation of health losses and health gains over time and across persons. This is potentially useful in three contexts. One is in the calculation of *Health Adjusted Life Expectancy* (HALE). An other is in the valuation of health interventions in terms of gained Quality Adjusted Life Years (QALYs) in cost-effectiveness analysis. For details of these two applications, see professor Harri Sintonen's paper for this conference. Thirdly, health state values may be used to calculate the *Burden of Disease* in populations in terms of Disability Adjusted Life Years (DALYs). For details, see Dr. Marie-Louise Essink-Bot's paper for this conference.

*It is important to recognize that each of these three applications - HALE, QALYs and DALYs - presuppose health measurements in terms of single index scores. Measurements in terms of profiles will not do the trick. On the other hand, the validity and meaningfulness of compressing complex health*

profiles into global health state values is debateable. The purpose of the following is to suggest how health state values need to be obtained in order for them to be meaningful.

#### Currently used techniques for eliciting health state values

Consider a complex health state, labelled A for brevity. A person's valuation of state A in terms of a single number on the 0-1 scale can be elicited in a number of ways, the most common of which are the following:

*The rating scale (RS).* The person is asked to locate state A on a straight line with end points representing the worst imaginable and the best imaginable health state respectively. State A then receives a value (on the 0-1 scale) which is directly proportional to its closeness to the upper end point of the straight line.

*The time trade-off (TTO).* The person is asked to indicate a number X of years in full health that he would consider equally desirable as for instance 10 years in state A. The value of state A (to the person in question) is then defined as the ratio  $X/10$ .

*The standard gamble (SG).* The person is asked to compare two scenarios. One is to live in state A with certainty. The other is to take a treatment that offers a probability p of living in full health and a probability 1-p of dying immediately. The person is asked to indicate the value of p that would make the two options equally desirable. The value of state A (to that person) is then set equal to p.

*The person trade-off (PTO).* The person is asked to compare two programs. One will save the life of 1 person and restore him/her to full health. The other will restore N people in state A to full health. The person is asked to indicate a value of N that would make the two programs equally worthy of funding. Say the response is  $N = 10$ . The person is then saying that the value of a cure of one person in state A is one tenth of the value of a saved life. The value of state A is then 0.9 ( $1.0 - 0.9 = 1.0 / 10$ ).

#### The validity of health state values

The above techniques for valuing health states (RS, TTO, SG and PTO) yield different results (Nord, 1992). This raises the question whether it really is possible to describe health in terms of single index values in a valid, reliable and consensual way.

To answer this question, it is useful to distinguish between two different ways in which health state values may be understood. One is in terms of *individual utility*, the other in terms of *societal value*.

Individual utility is basically an emotional category: How good is a health state or a health outcome *felt* by the individuals concerned? Total utility for a group of people, for instance for a nation as a whole, is

simply the sum of all individual utilities. Societal value, on the other hand, is a broader, ethical concept. While it is partially a function of total utility, it is also determined by concerns for fairness and equal entitlement to treatment and hence by the *distribution* of utility across individuals.

The rating scale, time trade-off and standard gamble are techniques for measuring individual utility. The person trade-off is a technique for measuring the societal value of health outcomes.

The scoring functions of multi-attribute utility instruments are based on population preference data collected by means of one or more of the rating scale, the time trade-off and the standard gamble. So are calculations of HALE and QALYs. Calculations of Burden of Disease in terms of DALYs, on the other hand, are based on preference data collected by means of the person trade-off. This is an important difference. In the following I first discuss the validity of health state values in calculating HALE. The points I make here, also apply to QALYs. In the final section I address the use of health state values in Burden of Disease measurement.

#### Health state values in calculations of HALE

The use of Health Adjusted Life Expectancy as an indicator of population health assumes the existence of three trade-offs.

First, there is a trade-off between quality of life and length of life. For instance, 75 years with considerable discomfort and disability may be regarded as equivalent to 70 years in full health. We may call this a *time-quality trade-off*.

Second, there is a trade-off between quality of life and the number of persons involved. Consider for instance a situation in which 100 people all live 75 years in a state that scores 0.8. Their average health adjusted life years is 60. Consider two possible changes in this situation. One consists in 20 people getting to live as healthy, while the other 80 people remain at level 0.8. The other possible change consists in all 100 people getting to live in a state that scores 0.84. In both cases the resulting average health adjusted life years is 63, i.e. an increase of 3 years. Effectively this means that, according to the HALE, taking 20 people from 0.8 to healthy is equivalent to taking 100 people from 0.8 to 0.84. We may call this a *person-quality trade off*.

Third, there is a trade-off between length of life and the number of persons involved. For brevity, I do not discuss this here, since it is independent of the values that are assigned to health states (see Nord (1997)).

With HALE, time-quality trade-offs and person-quality trade-offs follow mathematically from the values that are assigned to health states. One way to test the validity of these values is therefore to examine whether the trade-offs that are implied by the values, fit with the trade-offs that people in society make if they are asked directly (so-called test of reflective equilibrium). If they do not fit, the values are unsuitable in the calculation of HALE.

To perform this validity test, we need to do one more thing. That is to specify the substance, or the object, of the trade-offs. There are two possibilities: utility and societal value.

### *Utility*

Let us assume that the trade-offs that one wishes to build into HALE are in terms of utility. The proposition that 75 years of life in a scenario S is equivalent to 70 years in full health then means that the utility that individuals derive from these two scenarios is the same. A policy maker may want to know how the latter proposition can be verified.

Since utility is a matter of subjective feeling, and the strength of feelings is not directly observable, verification of the utility assignments is not straightforward. However, in theory individual utility may be assessed *indirectly* by looking at a behavioural correlate to the subjective feeling of utility. One such behavioural correlate is the quality-of-life-scores that subjects assign to themselves when asked to evaluate their own health on rating scales. However, rating scale scores have been shown not to have interval scale properties. That is, equally large intervals on different parts of such scales (for instance a movement from 0.4 to 0.6 versus a movement from 0.7 to 0.9 on a scale from zero to unity) do not carry the same significance to the individuals concerned (Morris and Durand, 1989; Nord, 1991; Richardson, 1994). Hence, rating scale measurements do not allow a meaningful summation of utility across individuals. *I therefore do not recommend the use of utilities operationalised as self ratings on rating scales in the calculation of health adjusted life expectancy.*

In economic theory and decision analysis, a more widely accepted behavioural correlate to perceived utility is the individual's willingness to sacrifice life expectancy to be cured of a state of illness or to obtain a given health improvement. People's willingness to undertake such sacrifices are elicited by means of standard gamble or time trade-off questions.

A number of MAU instruments purport to express utilities in this sense. However, the empirical support that the developers claim for this interpretation of their values is weak or non-existent ( Nord 1997, table 7). Results from studies other than those cited by the developers suggest that values assigned by MAU instruments in fact overestimate willingness to sacrifice life expectancy. Consider for instance "A person who has difficulties in moving about outdoors and has slight discomfort, but is able

to do some work and has only minor difficulties at home." MAU instruments assign values to this moderate state of illness that range from 0.60 to 0.98, with the majority lying around 0.90 ( Nord 1996a). The latter corresponds to a willingness to sacrifice 10 per cent of life expectancy to become well. However, a number of studies suggest that people who actually are at such a moderate problem level are not willing to sacrifice any life expectancy at all to be relieved of their problems. For instance, Sherbourne et al collected time trade-off and standard gamble data from close to 17.000 patients visiting primary care clinics across the USA. On average, the patients had two chronic conditions. Their average score on a rating scale from zero ("worst possible health state") to 100 ("perfect health") was 75. However, 85 per cent of the patients were not willing to sacrifice any life expectancy to be relieved of their condition. Even for patients with five different chronic diseases this percentage was as high as 65 ( Sherbourne and Sturm, 1997). Studies by O'Leary et al (1995), Fowler et al (1995), Nord (1996b) and Stavem (1996) show similar results. The implication is that *if health state valuations are to be used in measuring population health in terms of utility, and utility (or disutility rather) is measured in terms of willingness in sick and disabled people to sacrifice life expectancy to become well, then none of the existing MAU instruments seem to provide valid valuations (although the Rosser/Kind Index (1978) comes close).*

#### *Societal value*

The trade-offs that one wishes to build into HALE may alternatively be understood in terms of societal value. To say that 75 years of life with a given amount of health problems is equivalent to 70 years in full health (time-quality trade-off) then means that society values these two scenarios equally much. Similarly, the proposition that taking 20 people from 0.8 to healthy is equivalent to taking 100 people from 0.8 to 0.84 (person-quality trade-off) means that society values these two improvements equally much. Again, there is a need to make sure that these trade-offs that become built into HALE through the health state valuations that are used, fit with actual societal preferences.

There are very few data on the time-quality trade-offs that society wishes to make in valuing alternative life scenarios for other people. With respect to societal person-quality trade-offs, there are much more studies available. The bulk of these, conducted in Australia, Norway, Spain, United Kingdom and USA, are reviewed in Nord (1996a). The message from the studies is that members of the public want their health care systems to produce as much health - or utility - as possible, but within certain constraints. One is that health improvements for the severely ill are valued more highly than equally large improvements for less ill people. People also tend to feel that their right to realize their potential for health is the same, whether the potential happens to be large or small ( Nord et al, 1998). The evidence suggests the order of magnitude by which people in some industrialised countries value health improvements for people with different degrees of

severity of illness and different potentials for health improvement. To picture this order of magnitude, consider four classes of outcomes:

A: *Saving a person's life* to a life as healthy.

B: *Curing a person with a severe problem* for instance a person who sits in a wheel chair, has pain most of the time and is unable to work.

C: *Curing a person with a considerable problem* for instance a person who uses crutches for walking, has light pain intermittently and is unable to work.

D: *Curing a person with a moderate problem* for instance a person who has difficulties in moving about outdoors and has slight discomfort, but is able to do some work and has only minor difficulties at home.

In countries like Australia, England, Norway, Spain and the US the social appreciation of outcome A seems to be something like 3-6 times as high as that of class B outcomes, 10-15 times as high as that of class C outcomes and 50-200 times as high as that of class D outcomes. I emphasise that these numbers pertain to valuations of outcomes in decisions about health programs and policies (as opposed to decisions concerning identified patients in current need). A HALE statistic that purports to indicate whether one state of affairs in population health is more desirable than an other, needs to reflect this structure of concern. To do this, health states need to be assigned values in the following order of magnitude in the construction of HALE:

Severe problem ( cfr. outcome class B) : 0.65-0.85.  
Considerable problem (outcome class C): 0.90-0.94.  
Moderate problem (outcome class D) : 0.98-0.995

As shown in table 1, most MAU instruments are very far from satisfying the above requirements. In general, they assign too low values to states of moderate and slight illness, which in turn leads them to assign too high values to improvements for people in such states relative to improvements for people with severe or life threatening conditions.

#### A strategy for improving validity in calculating HALE

One response to the above state of affairs would be to adjust the valuation algorithms of MAU instruments in such a way as to produce the necessary upper end compression of values. However, here we need to distinguish between the utility perspective and the societal-value perspective.

With respect to the utility perspective, the preceding sections are based on the common assumption in health economics and decision analysis that the disutility of a state of illness may be measured as the willingness to sacrifice life expectancy to be relieved of the illness. Given this

particular definition, I am led to draw the conclusion that values from MAU instruments generally lack upper end compression, since people with illness express great reluctance to make such sacrifices when asked directly in preference studies.

If this true preference structure were built into the values assigned to health states, a serious sensitivity problem would arise: A great number of mild and moderate states would be assigned the value of 1. Real health improvements for people with such conditions would then not be captured by the MAU instruments, since there would be no differences between the values for such conditions and the value for full health. This would be a problem both in measuring population health and in cost-effectiveness analysis.

A possible strategy is to regard health state values as numbers that are supposed to express trade-offs in terms of societal value rather than utility. The rationale for this is that the preference for the preservation of life itself may be somewhat less absolute when people are asked to prioritise between different health care programs in a budgeting context than when they are asked about their willingness to sacrifice own future life years or certainty of survival. In other words, I am suggesting that more health states will be assigned values below unity if person trade-off questions are posed to the general population than if time-trade off or standard gamble questions are posed to people with illness and disability. Choosing the societal value interpretation may therefore be a way to ensure both validity and sensitivity in the practice of assigning numerical values to health states.

#### Burden of Disease in terms of DALYs

In estimating Burden of Disease in terms of DALYs, one uses *severity weights* for different illnesses. These are obtained simply by subtracting health state values from unity. The health state values are obtained by means of the person trade-off technique (Murray and Lopez, 1996). Burden of Disease is thus measured in a way which in principle accords with the recommendation above that health state values should be based on a technique that aims at capturing societal value - including concerns for fairness in health care - rather than individual utility only.

However, the particular way in which the person trade-off approach is used in current Burden of Disease assessment is quite problematic. Subjects are invited to value health states by means of two different PTO questions. One of these essentially consists in the following: Imagine that a program A will add a life year to 1000 healthy people, and that a program B will add a life year to X people with a particular disability, for instance blindness. What must X be for the two programs to be considered equally valuable? Subjects are instructed to suggest a number larger than 1000, to reflect the fact that disability, in this case blindness, represents a burden of illness. But it does not follow from this fact that the disabled are less entitled to life, or that loss of life years in disabled people should count less in

priority setting in health care than loss of life years in otherwise healthy people. Yet, this is the implication of Burden of Disease measurement as it is presently conducted. Further work with Burden of Disease statistics should therefore build on a revised version of the PTO procedure.

### Conclusion

Epidemiological and clinical information using concepts such as HALE, QALYs and DALYs purports to aid decision makers by reducing the amount of information they need to process. The concepts are more meaningful and more easily understood if they aim at expressing societal value rather than individual utility. The societal value perspective can be secured by using the person trade-off technique to elicit population preferences for health states. Work with Burden of Disease measurement shows that care needs to be exhibited with respect to how PTO questions are asked.

The fact that concepts such as HALE, QALYs and DALYs can be shown to be meaningful does not necessarily mean that they are useful. At the end of the day, decision makers must themselves judge whether they feel helped by having information in such terms added to conventional ways of describing losses and gains in health. This is an important issue for future research.

### REFERENCES

- Brooks R et al. EuroQol: The current state of play. Health Policy 1996,37,53-72.
- Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. Pharmacoeconomics 1995,7,490-502.
- Fowler FJ, Cleary PD, Massagli MP et al. The role of reluctance to give up life in the measurement of the value of health states. Medical Decision Making 1995,15,195-200.
- Kaplan RM, Anderson JP. A general health model: Update and applications. Health Services Research, 1988,23,203-235.
- Nord E, Pinto JL, Richardson J, Menzel P, Ubel P. Numerical valuations of health programs as a function of individual utility gains and societal concerns for fairness. Health Economics 1998 (forthcoming).
- Morris J, Durand A. Category rating methods: Numerical and verbal scales. Mimeo. University of York: Centre for Health Economics. 1989.
- Murray CLJ, Lopez AD. The Global Burden of Disease. Harvard University Press, 1996.
- Nord E. The validity of a visual analog scale in determining social utility weights for health states. International Journal of Health Planning and Management 1991, 6, 234-42.
- Nord E. Methods for quality adjustment of life years. Social Science & Medicine 1992, 34, 559-569.
- Nord E. The person trade-off approach to valuing health care programs. Medical Decision Making 1995,15,201-208.
- Nord E. Health status index models for use in resource allocation decisions. A critical review in the light of observed preferences for social choice. International Journal of Technology Assessment in Health Care 1996,12,31-44.

Nord E. Time trade-off scores in patients with chronic disease. Comparison with the York hypothetical TTO tariff. Paper for the EuroQol Plenary Meeting, Oslo, October 1996.

Nord E. A review of synthetic health indicators. Report for the OECD. Oslo: National Institute of Public Health, 1997.

O'Leary JF, Fairclough DL, Jankowski MK et al. Comparison of time trade-off utilities and rating scale values of cancer patients and their relatives. *Medical Decision Making* 1995,15,132-137.

Richardson J. Cost-utility analysis: What should be measured? *Social Science & Medicine* 1994,39,7-21.

Rosser R, Kind P. A scale of valuations of states of illness: Is there a social consensus? *International Journal of Epidemiology* ,1978,7,347-358.

Rosser R, Cottee M, Rabin R, Selai C. Index of health-related quality of life. In: Hopkins A (ed). *Measures of the quality of life, and the uses to which they may be put*. London: Royal College of Physicians of London, 1992.

Sherbourne C, Sturm R. Utility measures from the screener. Mimeo. Paper presented at RAND, Santa Monica, February 1997.

Sintonen H, Pekurinen M. A fifteen-dimensional measure of health-related quality of life (15 D) and its applications. In Walker SR, Rosser RM (eds). *Quality of life assessment. Key issues in the 1990s*. Dordrecht: Kluwer Academic Publishers, 1993.

Stavem K. Valuing theoretical versus own condition: A comparison using time trade-off and the EuroQol Instrument in COPD patients. Paper for the EuroQol Plenary Meeting, Oslo, October 1996.

Table 1. Health state values according to different MAUs.

<u>Model</u>	<u>Problem level</u>		
	Severe	Considerable	Moderate
15 D	.77	.86	.91-.93
HUI2	.40	.70	.90-.94
EQ-5D (rating scale)	.20	.60	.70
EQ-5D (TTO)	.20-.25	.40-.50	.80
QLHQ	.30-.40	.50-.60	.60-.70
IHQL(3D) <sup>a</sup>	.50-.70	.75-.85	.89-.93
IHQL(complex) <sup>a</sup>	.70-.75	.80-.90	.90-.94
Rosser/Kind index <sup>b</sup>	.68	.94	.97-.98
QWB <sup>c</sup>	.45-.55	.65-.70	<.80

a) Rosser et al, 1992.

b) Rosser and Kind, 1978.

c) Kaplan and Anderson, 1988.

Source for all values: Nord (1996).