Supported by sanofi-aventis

What are health utilities?

Keith Tolley MPhil
Health Economist and Director, Tolley Health Economics Ltd

- **Utilities are cardinal values** that represent the strength of an individual's preferences for specific health-related outcomes.
- Measuring health utilities involves two main steps: **defining a set of health states** of interest, and **valuing those health states.** There are **direct** or **indirect** methods of utility valuation.
- The methods that have been used to collect data on utilities include the **standard gamble approach, the time trade-off approach** and **the visual analogue approach.**
- The main indirect methods of utility measurement are: the use of **generic preference instruments** (EQ-5D, SF-6D and HUI); the use of **disease-specific preference measures;** and **mapping from a disease-specific health-related quality of life instrument to a generic instrument.**
- Generic preference-based measures are increasingly being used in **cost–utility analyses of pharmaceutical and other healthcare interventions.** In the UK, the National Institute for Health and Clinical Excellence has specified the EQ-5D as its preferred method of utility measurement.
- Utilities have been used as the **preference weights** (quality levels) within the quality-adjusted life-year model – an increasingly popular outcome measure used in pharmaceutical market access decision-making.

For further titles in the series, visit:
www.whatisseries.co.uk

Date of preparation: April 2009

NPR09/1103

What are health utilities?

In health economics, utilities are cardinal values that reflect an individual's preferences for different health outcomes. They are measured on an interval scale with zero reflecting states of health equivalent to death and one reflecting perfect health. In health economics, utilities are typically combined with survival estimates and aggregated across individuals to generate quality-adjusted life-years (QALYs) for use in cost–utility analyses of healthcare interventions (Box 1).[1,2]

Utility measurement consists of two main components. These are:

- The definition and description of a set of health states of interest
- The valuation of those health states (that is, measurement of the strength of preference for each health state).

These components can be applied in two broad ways.[3]

- **Direct measurement** of utilities can be performed for discrete condition-specific health states.
- **Indirect measurement** of utilities can be performed by applying utility algorithms to generic or disease-specific preference-based questionnaires, or by mapping from a disease-specific health-related quality of life (HRQoL) instrument onto the utility algorithm of a generic instrument such as the EQ-5D (EuroQol five dimension).

Box 1. Quality-adjusted life-years (QALYs) in practice

Let us assume that there are two treatments for an illness. Both treatments extend the life expectancy of an individual by eight years. However, treatment A results in the individual surviving the years in full health (usually represented by a utility score of 1 on a cardinal scale), while treatment B results in the individual surviving the years in a health state with a utility score of 0.5. Treatment A has led to a gain in QALYs of 8 (8 x 1), while treatment B has led to a gain in QALYs of 4 (8 x 0.5).

Direct measurement methods

There are several pertinent issues to consider in the use of direct measurement approaches, three of which are discussed below.

- What are the relevant health dimensions?
- What utility valuation method should be used?
- Whose preferences should be measured?

What are the relevant health dimensions?

It is important that the **relevant dimensions and levels of health** are assessed. If these dimensions and levels do not correspond to patient outcomes, it will not be possible to detect differences even if such differences exist.

Within the context of utility measurement, **health dimensions are referred to as attributes.** Examples of health attributes include: physical, social and cognitive function; psychological well-being; symptoms; and pain. These are incorporated into a health state description (sometimes referred to as a 'vignette') for the particular disease or condition of interest; for example, epileptic seizure frequency health state descriptions, or visual impairment severity health state descriptions.

The descriptions should ideally be based on direct patient experience, although in the interest of speed and feasibility many utility studies have used clinicians or other health professionals as proxies backed by literature evidence. The descriptions should not contain more than about five to nine attributes. Research indicates that this is the limit to the amount of information that humans can simultaneously process.[4]

Care should also be exercised to minimise context effects in health state descriptions. There may be labelling effects; for example, the use of the term anaemia/fatigue could be perceived and valued differently depending on whether or not it is labelled in the context of cancer.

What utility valuation method should be used?

There are a number of valuation methods that have been used empirically in studies. These include standard gamble (SG), time trade-off (TTO), rating scales, equivalence technique, ratio scaling and person trade-off. The first three are the most frequently used in direct measurement studies and so are the focus here.

Standard gamble

The SG approach is the classic method of measuring preferences in economics under conditions of uncertainty, and was first presented by von Neumann and Morgenstern.[5] SG utilities are fully consistent with expected utility theory and the axioms of rational individual behaviour.

The SG approach involves presenting individuals with a choice between two alternatives: a health state that is certain (for example, frequent migraine attacks) and a gamble with one better (for example, full health) and one worse (for example, death) outcome possible. Respondents are asked what probability of the better outcome would make them indifferent between remaining in the described (migraine) health state for certain or going for the risky option. Hence, if they are indifferent between the migraine state and a gamble with a 0.8 probability of the better outcome (but 0.2 probability of the worse outcome), 0.8 represents the utility of the migraine health state.

The utilities generated by the SG approach are dependent on the risk behaviour of the individuals surveyed. In general, individual behaviour is risk averse, meaning respondents will tend towards the certain outcome rather than the gamble (producing higher utilities for the health state than the risk-seeking behaviour).[6]

Time trade-off

The TTO approach involves asking subjects to consider the relative amounts of time (for example, number of life-years) they would be willing to sacrifice to avoid a certain poorer health state (for example, frequent migraines). Assuming a scenario of ten years with frequent migraines, the respondent may be indifferent between this state and a shorter lifetime of seven years, resulting in an estimated utility for the frequent migraine health state of 0.7 (seven years divided by ten years).

Utilities generated by the TTO method essentially represent cardinal 'values' based on value theory rather than expected utility theory (the basis of SG). The importance of this is that TTO assumes risk neutrality as a choice is made under conditions of certainty (that is, there are no risks); therefore, if individuals are indeed risk averse, the utility for the same health state generated by SG is expected to be higher than the value generated by TTO.

A practical issue in TTO studies is the choice of lifetime duration for TTO studies. A review found this varied between one month and 30 years.[7] The impact this has on the utility estimates generated is inconclusive.

Rating scale (visual analogue scale)

The rating scale (or visual analogue scale [VAS]) is based on psychometric theory. It consists of a single line with anchors representing best possible health and death (or some alternative). Respondents are asked to place each health state on the line such that the intervals between the placements reflect their perceived differences between the health states.

The VAS approach generate values rather than utilities, and has the least grounding in economic theory as it involves neither any element of choice nor individual decision-making under uncertainty. However, Parkin and Devlin have argued that the VAS has important advantages (versus SG and TTO) in empirical performance.[8]

Due to its simplicity the VAS is often used as a 'warm-up' to a TTO or SG exercise, as it allows the respondent to become familiar with comparing the health states.

Comparing the valuation methods

In general, health economists support the use of choice-based methods (SG or TTO) over the VAS. However, the SG approach is relatively time-consuming and people often have difficulty understanding the concept of probabilities. For many, the TTO represents a reliable and practical middle way, although the trade-off concept could still be difficult for many people to understand.

The choice of method matters as the differences in theoretical grounding and valuation approach can lead to differences in utility estimates. A review of utilities across 995 chronic and acute health states found a strong tendency for VAS to yield the lowest, TTO the middle and SG the highest utility values for the same health states.[9] Direct measurement studies may be particularly useful for valuing specific side-effects or modes of drug administration (where there are obvious and plausible differences in QoL impact).[10,11]

Whose preferences should be measured?

The valuation of preferences using SG, TTO or VAS can be performed by either the public or patients (or, in the case of young children, proxies such as caregivers/parents).

- The **primary reason for selecting patients** is that they are the people who experience the impact of the disease and treatment, so it is their preferences that should be considered of most importance. If patients are used, they can be asked to value a hypothetical health state or their own health state. While it is more challenging to recruit patients with the specific health states of interest, utilities obtained through hypothetical scenarios may not be valid predictors of preferences associated with actual experienced health states.[12]
- The **rationale for measuring the preferences of the general public** for hypothetical health states is that in a publicly funded healthcare system, it is society's resources that are being allocated and so it is the views of the general population that are most relevant.[1]

The question of whose preference to measure is important because there is some evidence that utilities differ between different population groups. In general, patient values for hypothetical health states that are likely to be worse than their current health state tend to be higher than those from the general public. This may be related to patients' coping mechanisms when addressing ill health.[13]

There are situations in which proxies are necessary; for example, when measuring young children's preferences. In a study conducted in a mixed UK population sample, parents/caregivers produced higher utilities than the general public in the context of Lennox–Gastaut syndrome – a severe form of childhood epilepsy.[11] Increasingly, the proxy values of clinicians or other healthcare professionals are recognised as not representing a valid or reliable source for utilities.[14]

A further issue is the minimum sample size for public- or patient-based valuation studies. No rules have been set for this and many utility studies have been conducted on very small sample sizes.[15] In general, a rule of thumb is that, to be considered reliable, population-based studies should have a minimum of 100 respondents representing a cross section of society. Achieving sufficiently large sample sizes becomes more difficult with patient-based valuation studies, especially if based on patients experiencing the actual health state. Also, there is a limit to the number of health states that each respondent can cope with – a rule of thumb is a maximum of ten but fewer if possible, due to interview fatigue (this also depends on the skills of the interviewer and the administration techniques used).

Indirect measurement methods

Generic utility instruments

By far the most well-known indirect method is the generic utility instrument, whereby a set of non-disease-specific health states, based on a combination of general attributes, have been valued by a general public sample.

Three instruments in common use today are the EQ-5D, the SF-6D (Short Form six dimension) and the HUI (Health Utilities Index), although there are other instruments used globally (for example, QWB [Quality of

Well-Being] and 15D [15 dimension]).[1]

For each instrument, the public values a limited number of health states and a scoring algorithm is developed using econometric modelling (EQ-5D and SF-6D) or a multiplicative multi-attribute utility function (HUI) to predict utilities for other health states not directly valued. Patients with any health condition then complete a simple questionnaire which defines the generic health state they are in, and the appropriate utility from the scoring algorithm is applied.

The instruments differ in the specific general attributes used for their descriptive systems and in the method of valuation applied – the EQ-5D has been valued using TTO, while the SF-6D is a derivative of the SF-36 (Short Form 36 item) and has been valued using SG. The HUI has two classification systems (HUI2 and HUI3) which differ in attributes but were also measured by SG. All, however, have been valued by large general public samples.

Generic utility instruments address some of the practical difficulties of conducting direct TTO or SG exercises. A large amount of research effort by separate EQ-5D, SF-6D and HUI research teams has gone into the development of standard, off-the-shelf questionnaires that can be completed in a few minutes by patients in randomised, clinical trials or observational studies. The second advantage of generic instruments from a decision-maker's perspective is that they generate general health state utilities that can be used to compare QALYs gained for interventions across patient groups and diseases to aid broad health service resource-allocation decisions.

However, there are a number of potential limitations with generic preference-based instruments.

- They may lack sensitivity in specific disease contexts.
- They are difficult to apply to acute conditions (for example, an asthma exacerbation) or when assessing smaller differences in utility (for example, those associated with mode of drug administration or specific drug-related side-effects).
- There is evidence of ceiling effects with the EQ-5D and floor effects with the SF-6D.[16]
- They generate different estimates for the same condition, which is related to their differences in valuation methods and scoring algorithms. For example, a study of the QoL of hearing-impaired people found mean utility scores were similar using the EQ-5D and SF-6D (0.77 and 0.79 respectively), but much lower using the HUI3 (0.56).[17]

Disease-specific utility instruments

In response to sensitivity issues with the generic measures, preference-based disease-specific measures are beginning to be developed. The principle here is that the utilities attached to a disease-specific HRQoL instrument are generated in the same way as with a generic instrument. Examples of preference-based scoring systems generated this way include the International Prostate Symptom Score (IPSS) for benign prostatic hyperplasia,[18] the International Index of Erectile Function (IIEE)[19] and the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR).[20] All have been valued using TTO methodology.

The construction of these instruments is promising, but they are still at a relatively early stage of development, so their value for decision-making has yet to be realised. They have the advantage of generating potentially sensitive disease-specific utilities through application of a validated HRQoL instrument in a clinical trial. However, the main practical limitation is the amount of research effort required to generate the utility scales; hence, a limited number of diseases have been covered so far.

Mapping utilities

A variant to developing a disease-specific utility measure is to map the descriptions from a disease-specific QoL instrument onto the utility algorithm of a generic instrument. This method is often used when a generic instrument is not available, but a suitable disease-specific QoL instrument has been included in clinical trials.

While it is possible to perform mapping based on clinician judgement, the rigour of this approach can be enhanced by using an empirical database containing both the HRQoL instrument and a generic instrument

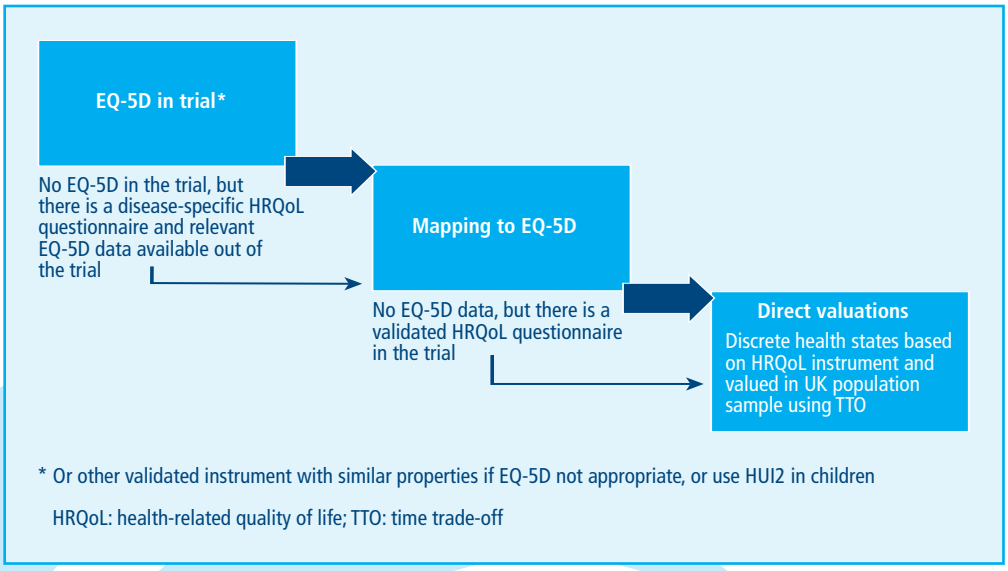

Figure 1. NICE hierarchy of utility methods[14]

such as the EQ-5D administered to the same patient to generate a mapping function. Even so, caution is needed in interpreting mapping functions derived this way, as a range of estimation methods have been applied in previous studies with varying levels of statistical robustness.[21]

Use of utilities in decision-making

As the use of economic evaluation for pharmaceutical market access decision-making increases globally, QALYs and, hence, utility assessments are being used as a key outcome measure within those evaluations.

In the UK, the calculation of QALYs is a central part of the reference case for economic evaluations of pharmaceuticals submitted to the National Institute for Health and Clinical Excellence (NICE).[14,22] NICE has a hierarchy of methods for the derivation of utilities, with the use of the EQ-5D within clinical trials being the preferred method. However, if it is not available, utilities can be derived using the TTO valuation method in community-based respondent samples (Figure 1).[14] Other bodies in the UK (for example, the Scottish Medicines Consortium [SMC]) and other countries are less prescriptive and allow a variety of methods to be used as long as they are performed robustly and the utility values generated for the health states appear plausible.

References

1. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes,* 3rd edn. Oxford: Oxford University Press, 2005.
2. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986; **5:** 1–30.
3. Szende A, Schaefer C. A taxonomy of health utility assessment methods and the role for uncertainty analysis. *Eur J Health Econ* 2006; **7:** 147–151.
4. Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 1956; **63:** 81–97.
5. von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior* 3rd edn. New York: John Wiley, 1953.
6. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions. Health Utilities Index. *Pharmacoeconomics* 1995; **7:** 503–520.
7. Arnesen T, Trommald M. Are QALYs based on time trade-off comparable?—A systematic review of TTO methodologies. *Health Econ* 2005; **14:** 39–53.
8. Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Econ* 2006; **15:** 653–664.
9. Morimoto T, Fukui T. Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals. *J Epidemiol* 2002; **12:** 160–178.
10. Osborne RH, de Abreu Lourenco R, Dalton A *et al.* Quality of life related to oral versus subcutaneous iron chelation: a time trade-off study. *Value Health* 2007; **10:** 451–456.
11. Yi Y, Verdian L, Oyee J, Tolley K, Heyes A. Eliciting utility scores for health states associated with Lennox–Gastaut syndrome. 2008; **11:** A606.
12. Jansen SJ, Stiggelbout AM, Wakker PP *et al.* Unstable preferences: a shift in valuation or an effect of the elicitation procedure? *Med Decis Making* 2000; **20:** 62–71.
13. Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Qual Life Res* 2003; **12:** 599–607.
14. National Institute for Health and Clinical Excellence. *Guide to the methods of technology appraisal (June 2008).* www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (last accessed 9 December 2008)
15. Tufts Medical Center. *Cost-Effectiveness Analysis Registry (CEA).* https://research.tufts-nemc.org/cear/default.aspx (last accessed 9 December 2008)
16. Ferreira PL, Ferreira LN, Pereira LN. How consistent are health utility values? *Qual Life Res* 2008; **17:** 1031–1042.
17. Barton GR, Bankart J, Davis AC. A comparison of the

quality of life of hearing-impaired people as estimated by three different utility measures. *Int J Audiol* 2005; **44:** 157–163.
18. Kok ET, McDonnell J, Stolk EA, Stoevelaar HJ, Busschbach JJ. The valuation of the International Prostate Symptom Score (IPSS) for use in economic evaluations. *Eur Urol* 2002; **42:** 491–497.
19. Stolk EA, Busschbach JJ. Validity and feasibility of the use of condition-specific outcome measures in economic evaluation. *Qual Life Res* 2003; **12:** 363–371.
20. McKenna SP, Ratcliffe J, Meads DM, Brazier JE. Development and validation of a preference based measure derived from the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR) for use in cost utility analyses. *Health Qual Life Outcomes* 2008; **6:** 65.
21. Brazier J, Ratcliffe J, Tsuchiya A, Salomon J. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press, 2007.
22. Brazier J. Valuing health states for use in cost-effectiveness analysis. *Pharmacoeconomics* 2008; **26:** 769–779.

What are health utilities?

First edition published 2001
Author: Stavros Petrou

Published by Hayward Medical Communications, a division of Hayward Group Ltd.

Supported by *sanofi-aventis*