What is ...?

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

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 Care 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 by health technology assessment bodies in drug access decision-making.

What is ...?

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 on to the utility algorithm of a generic instrument such as the EQ-5D (EuroQol five dimension).

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 healthcare professionals as proxies backed by literature evidence. Ideally, the descriptions should not contain more than 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)

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).

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.

An issue in TTO studies is the choice of time horizon for valuing health states in these studies. A review found this varied between one month and 30 years,[7] although ten years is commonly employed.[8] The impact this has on the utility estimates generated is inconclusive, although a ten-year time horizon may be challenging for dealing with situations such as mild, temporary or paediatric health states.[8] A further complication in TTO studies is the valuation of health states considered worse than death. Frequently, this problem is just ignored, so that only utility values between 0 and 1 are estimated in a traditional trade-off of zero to ten years in the best health state versus ten years in the poorer health state. In this classic TTO, states worse than death could, however, be valued by respondents who are willing to trade off a maximum of ten years for a health state taking part in a further exercise to measure their negative preference values for the health state in question. Alternatively there are methods such as lead (or lag time) TTO, whereby the health state to be valued is preceded (or followed) by time in full health, which enables a more natural estimation of negative values.[9,10]

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.

What are **health utilities?**

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.[11]

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 some participants 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.[12] However, this is not always the case and recent studies using both TTO and SG methods have found opposing outcomes, with significantly higher SG values for the same health states in one study in chronic myeloid leukamia[13] and significantly higher TTO values in a study in hepatic encephalopathy.[14] Direct measurement studies may be particularly useful for valuing specific side-effects, disease symptoms or modes of drug administration (where there are obvious and plausible differences in QoL impact).[15–19]

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, by 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.[20]
- 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.[21] The tendency for TTO and SG studies to be based on a general public sample is increasing, which reflects the demands of health technology assessment (HTA) bodies such as the National Institute for Health and Care Excellence (NICE) in the UK for community-based preference valuation in utility studies.[22].

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.[19] The proxy values of clinicians or other healthcare professionals do not, in most circumstances, represent a valid or reliable source for utilities.[22]

A further issue is the minimum sample size for public- or patient-based valuation

studies. 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.

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 (such as mobility or performance of usual activities), 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.

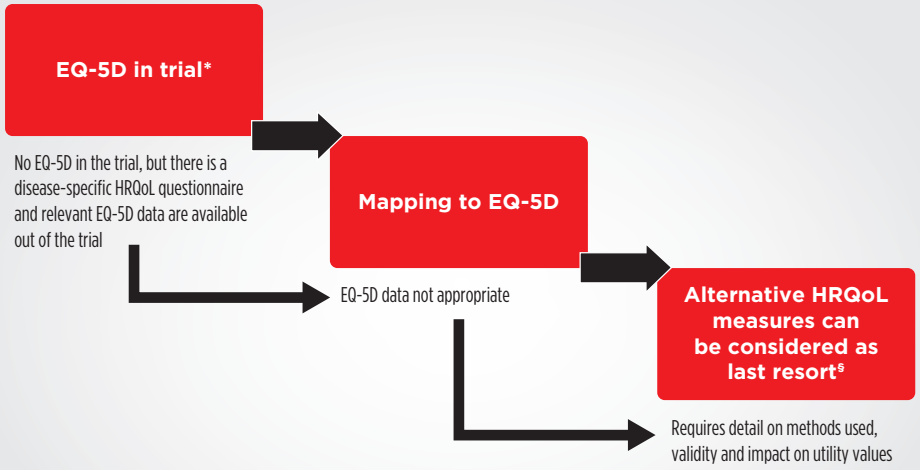
Generic utility instruments address some of the practical difficulties of conducting direct TTO or SG exercises. A large amount of research effort by separate EQ-5D, SF-6D and HUI research teams has gone into the development of standard, off-the-shelf questionnaires that can be completed in a few minutes 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. The EQ-5D has been found to perform well in studies of cancer and some skin disorders, but has been found to be inappropriate in some vision-related or hearing disorders.[23] In addition, to address criticisms of insensitivity, the EQ-5D-5L has been developed, with five instead of three levels of response for each of the five dimensions, and has been found to perform adequately in a range of chronic conditions such as chronic hepatitis,[24] multiple sclerosis[25] and others.[26]
- 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.[27]
- 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),[28] whereas a study in chronic pain found EQ-5D values were generally lower than those generated by the SF-6D.[29]
- Use of the instruments in children has limitations, although the HUI2 version and

What are **health utilities?**

Figure 1. NICE hierarchy of utility methods[22]



* Either the EQ-5D-3L or 5L can be used. For children, alternative validated preference-based measures of HRQoL designed specifically for children can be used
§ NICE guidance specifies that the valuation of HRQoL measured in patients (or carers) should be based on valuation of public preferences from a representative sample of the UK population using a choice-based approach[22]
EQ-5D: EuroQol five dimensions; HRQoL: health-related quality of life.

EQ-5D-Y (youth version) have been designed for use in children from the age of around eight years.

Disease-specific utility instruments

In response to sensitivity issues, the development of preference-based disease-specific measures has grown in recent years. 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,[30] the International Index of Erectile Function (IIEE),[31] the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR),[32] the Visual Function Questionnaire Utility Index[33] and the European Organisation for Research and Treatment of Cancer eight dimensions (EORTC-8D) classification system in cancer.[34] All have been valued using TTO methodology. Another advance is the development of disease-specific 'bolt-on' versions of the EQ-5D, such as that for use in patients with psoriasis.[35]

The construction of these instruments is moving at some pace. They have the advantage of generating potentially sensitive disease-specific utilities through application of a validated HRQoL instrument in a clinical trial. A limitation is the ability to compare utility outcomes across diseases, and the main practical limitation is the amount of research effort required to generate the utility scales. Some researchers have been cautioning that more research is needed to investigate the impact of labelling effects in preference-based disease-specific measures, and the viability of developing 'bolt-ons' to disease-specific measures to assess the utility impact of side effects and co-morbidities.[36]

Mapping utilities

A variant to developing a disease-specific utility measure is to map the descriptions from a disease-specific QoL instrument on to 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. There has also been a rapid growth in mapping (or cross-walking) studies, with most involving the development of an algorithm to map from a

disease specific HRQoL measure to the EQ-5D in diverse conditions, including overactive bladder,[37] migraine[38] and cancer.[39]

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 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.[40]

Use of utilities in decision-making

As the use of economic evaluation for pharmaceutical HTA and 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 NICE.[22] For the reference case in the economic evaluations submitted, 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 and mapping to EQ-5D being an alternative (Figure 1).[22] 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, including direct TTO and SG measurement studies, as long as they are performed robustly and the utility values generated for the health states appear plausible. In practice, HTA submissions to NICE and the SMC have included utility estimates based on a variety of methods including disease-specific and direct measurement methods ●

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. Boye KS, Matza LS, Feeny DH *et al.* Challenges to time trade-off utility assessment methods: when should you consider alternative approaches? *Expert Rev Pharmacoecon Outcomes Res.* 2014; **14:** 437–450.
9. Augustovski F, Rey-Ares, LIrazola V, Oppe M, Devlin NJ. Lead versus lag-time trade-off variants: does it make any difference? *Eur J Health Econ* 2013; **14**(Suppl 1): S25–S31.
10. Versteegh MM, Attema AE, Oppe M, Devlin NJ, Stolk EA. Time to tweak the TTO: results from a comparison of alternative specifications of the TTO. *Eur J Health Econ* 2013; **14**(Suppl 1): S43–S51.
11. Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Econ* 2006; **15:** 653–664.
12. 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.
13. Guest JF, Gray EJ, Szczudlo T, Magestro M. Utility values for specific chronic myeloid leukemia chronic phase health states from the general public in the United Kingdom. *Leuk Lymphoma* 2014; **55:** 1870–1875.
14. Guest JF, Nanuwa K, Barden R. Utility values for specific hepatic encephalopathy health states elicited from the general public in the United Kingdom. *Health Qual Life Outcomes* 2014; **12:** 89.
15. 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.
16. Doyle S, Lloyd A, Walker M. Health state utility scores in advanced non-small cell lung cancer. *Lung Cancer* 2008; **62:** 374–380.
17. Tolley K, Goad C, Yi Y *et al.* Utility elicitation study in the UK general public for late-stage chronic lymphocytic leukaemia. *Eur J Health Econ* 2013; **14:** 749–759.
18. Lauridsen JT, Lønborg J, Gundgaard J, Jensen HH. Diminishing marginal disutility of hypoglycaemic events: results from a time trade-off survey in five countries. *Qual Life Res* 2014; [Epub ahead of print].
19. Yi Y, Verdian L, Oyee J, Tolley K, Heyes A. Eliciting utility scores for health states associated with Lennox–Gastaut syndrome. *Value Health* 2008; **11:** A606.
20. 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.
21. 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.
22. National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal (April 2013).* http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9 (last accessed 27 June 2014)
23. Longworth L, Yang Y, Young T *et al.* Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol Assess* 2014; **18:** 1-224.
24. Scalone L, Ciampichini R, Fagiuoli S *et al.* Comparing the performance of the standard EQ-5D 3L with the new version EQ-5D 5L in patients with chronic hepatic diseases. *Qual Life Res* 2013; **22:** 1707–1716.
25. Fogarty E, Walsh C, Adams R *et al.* Relating health-related Quality of Life to disability progression in multiple sclerosis, using the 5-level EQ-5D. *Mult Scler* 2013. **19:** 1190–1196.
26. Agborsangaya CB, Lahtinen M, Cooke T, Johnson JA. Comparing the EQ-5D 3L and 5L: measurement properties and association with chronic conditions and multimorbidity in the general population. *Health Qual Life Outcomes* 2014; **16:** 74.
27. Ferreira PL, Ferreira LN, Pereira LN. How consistent are health utility values? *Qual Life Res* 2008; **17:** 1031–1042.
28. 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.
29. Obradovic M, Lal A, Liedgens H. Validity and responsiveness of EuroQol-5 dimension (EQ-5D) versus Short Form-6 dimension (SF-6D) questionnaire in chronic pain. *Health Qual Life Outcomes* 2013; **11:** 110.
30. 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.
31. 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.
32. McKenna SP, Ratcliffe J, Meads DM, Brazier JE. Development and

What are **health utilities?**

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.
33. Rentz AM, Kowalski JW. Development of a preference-based index from the National Eye Institute Visual Function Questionnaire-25. *JAMA Ophthalmol* 2014; **132:** 310–318.
34. Rowen D, Brazier J, Young T *et al.* Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value Health* 2011; **14:** 721–731.
35. Swinburn P, Lloyd A, Boye KS *et al.* Development of a disease-specific version of the EQ-5D-5L for use in patients suffering from psoriasis: lessons learned from a feasibility study in the UK. *Value Health* 2013; **16:** 1156–1162.
36. Brazier JE, Rowen D, Mavranezouli I *et al.* Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess.* 2012; **16:** 1–114
37. Kay S, Tolley K, Colayco D *et al.* Mapping EQ-5D utility scores from Incontinence Quality of Life Questionnaire among patients with neurogenic and idiopathic overactive bladder. *Value Health* 2013; **16:** 394-402.
38. Gillard PJ, Devine B, Varon SF, Sullivan SD. Mapping from Disease-specific measures to health-state utility values in individuals with migraine. *Value Health* 2012; **15:** 485–494.
39. Crott R, Briggs A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *Eur J Health Econ* 2010; **11:** 427–434.
40. Brazier J, Ratcliffe J, Tsuchiya A, Salomon J. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press, 2007.

First published 2001
Revised edition 2009
Revised edition 2014
This publication, along with the others in the series, is available online at
www.whatisseries.co.uk

Published by Hayward Medical Communications, a division of Hayward Group Ltd, The Pines, Fordham Road, Newmarket CB8 7LG.
Tel: 01638 723560. Fax: 01638 723561.
email: hmc@hayward.co.uk