VALUE IN HEALTH 15 (2012) 971–974

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

METHODOLOGICAL ARTICLES

Using Health State Utility Values in Models Exploring the Cost-Effectiveness of Health Technologies

Roberta Ara, MSc, Allan Wailoo, PhD*

The University of Sheffield, Sheffield, UK

ABSTRACT

Background: To improve comparability of economic data used in decision making, some agencies recommend that a particular instrument should be used to measure health state utility values (HSUVs) used in decision-analytic models. The methods used to incorporate HSUVs in models, however, are often methodologically poor and lack consistency. Inconsistencies in the methodologies used will produce discrepancies in results, undermining policy decisions informed by cost per quality-adjusted life-years. **Objective:** To provide an overview of the current evidence base relating to populating decision-analytic models with HSUVs. **Findings:** Research exploring suitable methods to accurately reflect the baseline or counterfactual HSUVs in decision-analytic models is limited, and while one study suggested that general population data may be appropriate, guidance in this area is poor. Literature describing the appropriateness of different methods used to estimate HSUVs for combined conditions is growing, but there is currently no consensus on the most appropriate methodology. While exploratory analyses suggest that a statistical regression model might improve accuracy in predicted values, the models require validation and testing in external data sets. Until additional research has been conducted in this area, the current evidence suggests that the multiplicative method is the most appropriate technique. Uncertainty in the HSUVs used in decision-analytic models is rarely fully characterized in decision-analytic models and is generally poorly reported. **Conclusions:** A substantial volume of research is required before definitive detailed evidence-based practical advice can be provided. As the methodologies used can make a substantial difference to the results generated from decision-analytic models, the differences and lack of clarity and guidance will continue to lead to inconsistencies in policy decision making.

Keywords: EQ-5D, quality of life, SF-36, utility.

Background

To facilitate comparison of results from decision-analytic models, there has been a move toward policy decision-making bodies proposing a specific preference-based measure such as the EuroQol five-dimensional (EQ-5D) questionnaire [1,2]. Inconsistencies in the way the health state utility values (HSUVs) are used will produce discrepancies in the results generated, which will undermine policy decisions informed by cost per quality-adjusted life-years (QALYs). While literature describing best practice in decision-analytic modeling is available [2–4], research exploring the practical issues arising when applying preference-based HSUVs in these models is scarce.

This article provides an overview of the current evidence base relating to issues involved in populating decision-analytic models [5]. Specifically, we look at 1) suitable HSUVs for the baseline/counterfactual health states (see definition below), 2) appropriate methods when combining or adjusting HSUVs for multiple health conditions/comorbidities (where an additional condition coexists alongside the primary condition), and 3) issues when characterizing uncertainty in HSUVs. We provide practical advice where possible and highlight where additional research is warranted. While the issues covered in this article are particularly relevant to analysts populating decision-analytic models using summary statistics reported in the literature, many are also relevant to analysts who have access to patient-level data.

Baseline/Counterfactual HSUVs

Decision-analytic models submitted to reimbursement authorities generally assess the benefits of interventions in terms of their potential to avoid or alleviate a clinical event or condition. As a consequence, in addition to the HSUVs associated with the event and condition, analysts need to know the HSUVs associated with not experiencing the event or the health condition, that is, the baseline or counterfactual values. For example, in patients with a history of cardiovascular disease (CVD), to assess the benefits of avoiding a stroke, analysts need the average HSUV for a cohort who have experienced a stroke and the average HSUV for a cohort who have not experienced

This article is based on a technical support document that was funded by the National Institute for Health and Clinical Excellence through its Decision Support Unit. The views, and any errors or omissions, expressed in this article are of the author only.

* *Address correspondence to:* Robert Ara, Health Economics and Decision Science, ScHARR, The University of Sheffield, 30 Regent Street, Sheffield S14DA, UK.

E-mail: r.m.ara@sheffield.ac.uk.

1098-3015/$36.00 – see front matter Copyright © 2012, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2012.05.003

a stroke but have a history of CVD (i.e., the baseline). Similarly, when assessing the potential benefits of a screening program for colorectal cancer, analysts need the average HSUV from a cohort who have colorectal cancer and the average HSUV from a cohort who do not have colorectal cancer (i.e., the baseline).

Evidence that can be used to represent the condition-specific baseline is often limited, and while some analysts have assumed that the alleviation of a health condition will return health-related quality of life (HRQOL) to full health (i.e., a health utility value of 1) [6], this approach is flawed. Using the previous examples, if a stroke is avoided, there will still be a detrimental effect on HRQOL due to CVD. Similarly, if bowel cancer is prevented, the average patient could still have at least one prevalent health condition that has a detrimental effect on HRQOL. It has been shown that the costs per QALY results generated when using different baseline HSUVs in the same model differ to such an extent that they could influence a policy decision based on a cost per QALY threshold [7].

Many decision models use lifetime horizons to accrue QALY gains, and the average baseline will not be constant across the full horizon modeled due to the increasing prevalence of comorbidities in older-aged cohorts and the detrimental effect on HRQOL associated with age [8]. It has been suggested that average HSUVs from the general population could be used as the baseline when condition-specific data are not available [9]. Because HSUVs obtained from the general population are informed by subgroups with many different conditions, intuitively this makes sense for less prevalent health conditions, or conditions that do not have a substantial effect on HRQOL, because removing a particular subgroup of people who have one of the conditions will not have a substantial effect on the average HSUVs.

Authors of a recent study examined the mean EQ-5D questionnaire scores for subgroups of respondents (n = 41,174) classified by self-reported health condition in the Health Survey for England [8]. The objective was to determine whether data from the general population could be used as proxy scores for the baseline (i.e., the HRQOL associated with not having the particular condition) in models. The appropriateness of the general population data was assessed by comparing the age-stratified mean EQ-5D questionnaire scores from respondents without a specific condition with matched subgroups from the general population. The study presents a number of age-stratified EQ-5D questionnaire scores categorized by broadly defined health conditions such as cardiovascular conditions, or arthritis/rheumatism or fibromyalgia. The authors reported that while data from the general population could potentially be used as proxy scores for some conditions, they may not be appropriate for all, and for some conditions, it may be more appropriate to use data from respondents who have none of the prevalent health conditions. If condition-specific data are not available, they suggest that a range of sensitivity analyses should be generated, with data from the general population used as one end of a range of plausible values.

Combining/Adjusting HSUVs

Health care decision-analytic models describe the clinical pathway followed by typical patients and can involve multiple health states representing the primary health condition, with additional health states representing comorbidities (where an additional condition coexists alongside the primary condition). An example might be when assessing the cost-effectiveness of statin treatment (which has the potential to reduce the risk of cardiovascular conditions) in patients with rheumatoid arthritis (RA) [10]. This cost-effectiveness model includes health states defined as RA but no history of CVD, RA and heart attack, or RA and stroke. Each of the individual health states in a decision-analytic model require HSUVs derived from patients whose health condition(s) mirrors the health state definitions in the model. Ideally, these would be obtained from cohorts with the conditions modeled, and it is often possible to derive the required utilities from existing catalogs informed by a comprehensive data set and appropriately classified conditions [11]. These utility values would be preferable to estimating values by using data collected from cohorts in disparate studies or subgroups with single conditions. However, because of the volume of different combinations of health states and conditions, the exact data required are not always available, and in these instances the mean HSUVs for the combined health states are frequently estimated by using the mean HSUVs obtained from patients with the single conditions [12]. There is currently no consensus on which particular method is preferred to estimate these HSUVs, and the approaches used can produce very different estimates [13,14].

The three methods typically used to estimate a mean HSUV for a combined condition when data are available only for relevant single conditions are the additive, multiplicative, or minimum methods. These assign a constant absolute decrement, a constant relative decrement, and no additional decrement over that observed for the condition with the lowest HSUV, respectively. A variation of the minimum method (the adjusted decrement estimator) has been suggested, and linear models incorporating terms to represent the three traditional methods (additive, multiplicative, and minimum) and obtained using ordinary least square regressions have been presented [12,15–17]. Specific details of the five methods are provided online.

A review of the literature in this area was conducted with articles identified by a systematic search of CINAHL, the Cochrane library, EMBASE, MEDLINE, PsycInfo, and Web of Science (1950–February 2012). The search combined terms for HRQOL (health state utility, quality of life, Euroqol, EQ5D, health utilities mark, HUI, short form six D, SF-6D, SF6D), methodologies (standard gamble, SG, time trade off, TTO, additive, multiplicative, minimum, regression, model), and terms for joint health states (joint health state, comorbid, combined health states, concurrent, multiple). This was supplemented by a forward and backward citations search in the Web of Knowledge and Google Scholar databases. The objective was to conduct a detailed critical review of existing empirical literature to gain an understanding of the reasons for differences in results and conclusions. Studies were included in the review if they estimated HSUVs for joint health conditions by using HSUVs from single conditions. Eleven studies that reported results of analyses exploring the accuracy of and/or comparing the performance of the methods used to estimate mean HSUVs were identified [13]. One article was excluded because it was an editorial informed by the results of one of the articles included in the review [18]. A second study was used to inform the discussion, but it was excluded because it reviewed the results of the early publications identified in the search but had not had access to the later publications [14].

Three of the 11 studies included used individual-level patient data (n = 50–207) directly elicited by using either standard gamble or time trade-off [16,19,20]. The remaining eight used HSUVs obtained by using generic questionnaires (EQ-5D questionnaire = 4 [15,17,21,22], six-dimensional health state short form [derived from short form 36 health survey] = 3 [12,23,24], health utilities index 3 = 1 [25]) collected during surveys (range 5,224–131,535 respondents). Two of the studies evaluated just one method, and the others compared results generated by using two, three, or more methods. The authors of the review reported that the range of actual utilities estimated influenced the accuracy of the methods and thus analysts' conclusions. For example, although the minimum outperformed the additive and multiplicative methods in one study [22], the data estimated covered a very narrow range (0.611–0.742) and two of the other studies demonstrated that the magnitude of the errors for the minimum method increased substantially when estimating lower utility values [12,17]; thus, the findings of the first study cannot be generalized beyond their data set without additional research. On a similar theme, the authors

noted that the use of mean errors when comparing methods was insufficient because these masked bias in the errors [12,17]. Finally, the accuracy of the method used was influenced by the value assigned to normal health, and the errors in estimated values increased when full health (EQ-5D questionnaire = 1) was used to determine the decrement associated with the single health conditions. The uncertainty in the estimated HSUVs has not been studied, and there is very little evidence describing results when estimating HSUVs for more than two simultaneous conditions. The authors of the review concluded that while there is currently no unequivocal evidence, the linear models obtained by using ordinary least square regressions outperformed the other methods. Because these models require validation in external data, and each quality-of-life instrument and set of preference weights would require a unique statistical model, on the basis of current evidence, however, the authors recommend the multiplicative method. We concur with this recommendation at this time but appreciate that the use of statistical models could be more appropriate once this research has been developed and validated.

Adverse events

When considering the inclusion of adverse events associated with a treatment or intervention (e.g., nausea is a side effect of treatment given for influenza) in decision-analytic models, it is essential to differentiate between acute events and chronic sequelae, and the inclusion of decrements on HRQOL associated with grades 3 to 4 (severe with marked limitation in activity—life threatening/disabling, requiring medical intervention) adverse events is particularly relevant [26]. Conversely, applying decrements for grades 1 to 2 (mild or transient discomfort—moderate limitation in activity, requiring no or minimal medical intervention) adverse events can introduce an element of double counting because the average HSUVs obtained from the main cohort are likely to include the disutility associated with these events.

A review commissioned by the National Institute for Health Research Health Technology Assessment program examined current practice when incorporating adverse events in economic models described in Health Technology Assessment reports published between 2004 and 2007 [27]. Forty-seven of the 80 studies reviewed were assessments conducted to inform National Institute for Health and Clinical Excellence appraisals. The authors recommended that a clear justification should be provided for the noninclusion of adverse effects together with an explicit report of how adverse effects are considered in the decision-analytic model. They suggested that systematic searches are required to identify the HSUVs required for adverse events and recommended research exploring the best approach to ensure that any adverse effects of interventions are captured.

Authors of a recent cross-sectional review of HRQOL data used in Health Technology Appraisals (n = 46) submitted to National Institute for Health and Clinical Excellence during the period 2004 to 2008 reported a wide range of methodological variation in the use of utility values and a lack of clarity in the reporting of detailed methods used in the submissions [28]. They found that adjustments for adverse events were made by either adding or subtracting a value (72%) from the original HSUVs, multiplying by a weight (18%), or incorporating a multivariate analysis (10%). Again, they concluded that further guidance is required to clarify the appropriateness of adjusting values and the preferred methods for undertaking these adjustments.

Capturing Uncertainty in HSUVs

Decision-analytic models in health care combine evidence from a range of sources and frequently extrapolate both costs and effects over time and between patient groups and settings. It is now standard practice to perform full probabilistic sensitivity analyses by using Monte-Carlo simulations to explore the uncertainty of input parameters [29]. The parameter uncertainty indicates the imprecision in the cost-effectiveness results and is used to inform the decision uncertainty through cost-effectiveness acceptability curves [30,31]. Ideally, each point estimate included within the model is described by a full probability distribution that reflects the uncertainty surrounding the value accurately [32,33]. The distributions used to describe the uncertainty are not selected arbitrarily but are informed by the data, the type of parameter, and the estimation process. While there is a wealth of literature describing appropriate methods for handling skewed cost data [34], little attention has been paid to the methods used to capture the uncertainty in HSUVs.

It has been reported that uncertainty around HSUVs is usually underreported and that frequently only mean values are used in decision-analytic models [35]. HRQOL data, and in particular the EQ-5D questionnaire, are not normally distributed. They are bounded by the limits of the index, often involve negative values, and are typically skewed, bimodal, or trimodal [36]. Nevertheless, the uncertainty in the mean can be adequately described by sampling from a normal distribution in the majority of cases. Exceptions include when sampling for a patient-level simulation model by using data that has a relatively low or high mean score and a wide distribution. In these cases, an alternative approach would be to describe the utility values as decrements [37] characterized by using a log normal or gamma distribution that would give a sampled utility decrement on the interval (0, positive infinity). If a lower constraint is required (e.g., −0.594 for the UK EQ-5D questionnaire index), the standard beta distribution could be scaled upward by using a height parameter (λ) producing a distribution on a (0, λ) scale.

An additional source of uncertainty, which is typically ignored, relates to the preference-based weights. Instruments such as the six-dimensional health state short form (derived from short form 36 health survey), health utilities index 3, and EQ-5D questionnaire include a number of questions relating to the respondent's health. The EQ-5D questionnaire, for example, has five questions with three possible responses to each. This gives a total of 243 (3^5) distinct health states. Because it is not practical to value all possible health states, a selection is typically valued. The statistical regression models fitted to the health states valued will consist of one or more parameter estimates (the preference weights) that are estimated with uncertainty. Although there is no reason, this source of uncertainty is typically ignored in decision-analytic models. When using patient-level data or when performing "mapping" exercises, it is simple to reflect this uncertainty by propagating the uncertainty and associated correlations in the covariance matrix in the probabilistic sensitivity analysis (see Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2012.05.003for link to EQ-5D questionnaire matrix). In addition, when the required preference-based data are not available, these data can be estimated by using mapping functions generated from HRQOL or clinical variables [1,38]. While there is a growing evidence base providing statistical regression models that can be used to estimate the required preference-based data [39], very few authors provide the statistics required to incorporate uncertainty in probabilistic analyses. These are both sources of uncertainty that are relevant to decision makers and ought to be reflected in the analysis of uncertainty in the same way as other sources of parameter uncertainty.

Decision-analytic models generally incorporate multiple health states describing changes in the health condition (e.g., disease progression, adverse events, and distinct events such as heart attacks and strokes) that may require unique HSUVs. Correlations between these HSUVs should be characterized in the probabilistic sensitivity analyses using multivariate distributions. Alternative approaches are currently being explored, and the resulting recommendations will be a useful reference for analysts.

In addition, a recent publication suggests that the standard error of measurement for a number of leading health utilities varies depending where along the health continuum the measurement is made [40]. These last two are examples of the growing volume of work in this area, and the literature should be continually reviewed to take account of emerging evidence.

Discussion

Robust research in this area is scarce, but it is clear that the methodologies employed when using HSUVs in decision-analytic models can make a substantial difference to the results generated and the differences in the methodologies will lead to inconsistent policy decision making [7]. One theme that was apparent throughout the evidence reviewed to inform this article was a lack of clarity and transparency in reports describing the methodologies used when applying HSUVs in decision-analytic models.

A substantial volume of research is required before definitive detailed evidence-based practical advice can be provided in this area including longitudinal data describing potential changes in HSUVs for subgroups of patients with specific health conditions, analyses exploring appropriate baseline data for the counterfactual health states in decision-analytic models to enable more precise calculations of the incremental heath benefits of treatment, empirical research on the most appropriate method for adjusting data to reflect comorbidities and/or adverse events, primary studies collecting data for acute events, and research to determine the class/type and duration of adverse event that should be incorporated in economic models.

Source of financial support: National Institute for Health and Clinical Excellence.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2012.05.003 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal (updated June 2008). Available from: www.nice.org.uk. [Accessed January 30, 2012].
[2] Gold MR. Panel on cost-effectiveness in health and medicine. Med Care 1996;34(12, Suppl.): DS197–9.
[3] Briggs AH. Handling uncertainty in cost-effectiveness models. Pharmacoeconomics 2000;17:479–500.
[4] Willan AR, Briggs AH. Statistical Analysis of Cost-Effectiveness Data (Ebooks Corporation ed.). Chichester, West Sussex, UK: John Wiley, 2006.
[5] Ara R, Wailoo AJ. NICE DSU Technical Support Document 12: the use of health state utility values in decision models. Available from: http://www.nicedsu.org.uk. [Accessed January 30, 2012].
[6] van Nooten F, Davies GM, Jukema JW, et al. Economic evaluation of ezetimibe combined with simvastatin for the treatment of primary hypercholesterolaemia. Netherlands Heart J 2011;19:61–7.
[7] Ara R, Brazier J. Populating an economic model with health state utility values: moving towards better practice. Value Health 2010;13:509–18.
[8] Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. Value Health 2011;14:539–45.
[9] Fryback DG, Lawrence WF. Dollars may not buy as many QALYs as we think. Med Decis Making 1997;17:276.
[10] Bansback N, Ara R, Ward S, et al. Economic evaluations in rheumatoid arthritis: a critical review of measures used to define health states. Pharmacoeconomics 2008;26:395–408.
[11] Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. Med Decis Making 2006;26:410–20.
[12] Ara R, Brazier J. Estimating health state utility values for comorbid health conditions using SF-6D data. Value Health 2011;14:740–5.
[13] Ara R, Wailoo AJ. Estimating health state utility values for comorbid health conditions: a synopsis of the current evidence base. Available from: http://eprints.whiterose.ac.uk/ [Accessed January 30, 2012].
[14] McIntosh CN. Utility scores for comorbid conditions: methodological issues and advances. In: Handbook of Disease Burdens and Quality of Life Measures. New York: Springer, 2010.
[15] Hu B, Fu AZ. Predicting utility for joint health states: a general framework and a new nonparametric estimator. Med Decis Making 2010;30:E29.
[16] Basu A, Dale W, Elstein A, et al. A linear index for predicting joint health states utilities from single health states utilities. Health Econ 2009;18:403–19.
[17] Ara R, Brazier J. Comparing EQ-5D scores for comorbid health conditions estimated using five different methods. Med Care 2012;50:452–9.
[18] Dale W. What is the best model for estimating joint health states utilities? Comparing the linear index model to the proportional decrement model. Med Decis Making 2010;30:531–3.
[19] Esnaoloa NF, Doherty DA, Johnson ML, et al. Comparison of additive and multiplicative utility predictions to predict the utilities of combined health states: abstracts from the 23rd Annual Meeting of the Society of Medical Decision Making. Med Decis Making 2001;21:517.
[20] Dale W, Basu A, Elstein A, et al. Predicting utility ratings for joint health states from single health states in prostate cancer: empirical testing of 3 alternative theories. Med Decis Making 2008;28:102.
[21] Janssen M, Bonsel G. Estimating preference weights for chronic multimorbidity: don't add, multiply. Available from: www.euroqol.org. [Accessed January 30, 2012].
[22] Fu AZ, Katan M. Utilities should not be multiplied. Med Care 2008;46:984–90.
[23] Hanmer J, Vanness D, Gangnon R, et al. Three methods tested to model SF-6D health utilities for health states involving comorbidity/co-occurring conditions. J Clin Epidemiol 2010;63:331–441.
[24] Wee H, Cheung Y, Li S, et al. The impact of diabetes mellitus and other chronic medical conditions on health related quality of life: is the whole greater than the sum of its parts? Health Qual Life Outcomes 2005;3:2.
[25] Flanagan W, McIntosh CN, Le Petit C, et al. Deriving utility scores for co-morbid conditions: a test of the multiplicative model for combining individual condition scores. Popul Health Metrics 2006;4:13.
[26] ICTDR Investigator Manual: Monitoring and reporting adverse events: appendices. Available from: http://www.icssc.org/Documents/Resources/AEManual2003AppendicesFebruary_06_2003%20final.pdf. [Accessed January 30, 2012].
[27] Craig D, McDaid C, Fonseca T, et al. Are adverse effects incorporated in economic models? A survey of current practice. Int J Technol Assess Health Care 2010;26:323–9.
[28] Tosh JC, Longworth LJ, George E. Utility values in National Institute for Health and Clinical Excellence (NICE) technology appraisals. Value Health 2011;14:102–9.
[29] Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005;14:339–47.
[30] Van Hout BA, Al MJ, Gordon GS, et al. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ 1994;3:309–19.
[31] Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Econ 2001;10:779–87.
[32] Critchfield GC, Willard KE. Probabilistic analysis of decision trees using Monte Carlo simulation. Med Decis Making 1986;6:85.
[33] Doubilet P, Begg CB, Weinstein MC, et al. Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach. Med Decis Making 1985;5:157.
[34] Briggs AH, Gray AM. Handling uncertainty in economic evaluations of healthcare interventions. BMJ 1999;319:635–8.
[35] Longworth LJ, Chandiwana D, Kandaswamy P, et al. Report to the NICE Methods Review Working Party: key issues arising from workshop on health-related utility measurement. Available from: http://www.nice.org.uk/media/4A6/59/UtilitiesBriefingPaper010607KT.pdf [Accessed January 30, 2012].
[36] Hernández Alava M, Wailoo AJ, Ara R. Tails from the Peak District: adjusted censored mixture models of EQ-5D health state utility values. Value Health 2012;14:550–61.
[37] Brazier J, Ratcliffe J. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford, UK: Oxford University Press, 2007.
[38] Longworth L, Rowen D. NICE DSU Technical Support Document 10: the use of mapping methods to estimate health state utility values. Available from: http://www.nicedsu.org.uk. [Accessed January 30, 2012].
[39] Brazier JE, Yang Y, Tsuchiya A, et al. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. Eur J Health Econ 2010;11:215–25.
[40] Palta M, Chen H, Kaplan RM, et al. Standard error of measurement of 5 health utility indexes across the range of health for use in estimating reliability and responsiveness. Med Decis Making 2011;31:260–9.