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

A B S T R A C T

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