Trends in the Measurement of Health Utilities in Published Cost-Utility Analyses

Carmen A. Brauer, MD, MSc, FRCSC,[1] Allison B. Rosen, MD, ScD,[2] Dan Greenberg, PhD,[3] Peter J. Neumann, ScD[4]

[1]Harvard Center for Risk Analysis, Harvard School of Public Health, Boston, MA, USA; [2]Division of Beneral Medicine, University of Michigan Health Systems, Department Of Health Management and Policy, University of Michigan School of Public Health, and Center for Practice Management and Outcomes Research, Ann Arbor VA Medical Center, Ann Arbor MI, USA; [3]Department of Health Systems Management, Faculty of Health Sciences & School of Management, Ben-Gurion University of Negev, Beer-Sheva, Israel; [4]Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center, Boston, MA, USA

ABSTRACT

Objective: The Panel on Cost-Effectiveness in Health and Medicine recommended the compilation of a catalog of health state utility weights for use in cost-utility analyses (CUAs), and has given methodological recommendations. This study presents an update, through 2001, to our current registry of utility weights (available at http://www.tufts-nemc.org/cearegistry; previously at http://www.hsph.harvard.edu/cearegistry), and documents recent changes in methods used for utility weight elicitation.

Methods: We searched the English-language medical literature for original CUAs reporting outcomes as cost per quality-adjusted life-year (QALY). Two trained readers independently audited each article, abstracting data on the health state descriptions, corresponding utility weights, methods of elicitation, and sources of the estimates. The utility elicitation methods from 1998 to 2001 were compared with the methods used to obtain utilities before 1998.

Results: We identified 306 CUAs published after 1998, reporting 1210 separate health-related utility estimates, bringing the total in our catalog to 2159 weights. Most fre-quently, health states pertained to the circulatory system and oncology. Methods varied substantially: 36% of authors used direct elicitation (standard gamble, time trade-off or rating scale), 23% used generic health status instruments (EQ-5D, Health Utilities Index, etc.), and 25% estimated weights based on clinical judgment. Community preferences were used in 27% of the values. Compared with pre-1998, utilities published from 1998 to 2001 were more likely to be elicited using a generic instrument, more likely elicited from community samples, and less likely derived from expert opinion, with no formally employed methodology.

Conclusions: Increasingly, analysts conducting CUAs are using generic, preference-weighted instruments, and relying on community-based preferences. Our catalog of utility weights provides a useful reference tool for producers and consumers of CUAs, but also highlights the continued need for improvement in methods and transparency.

Keywords: cost-effectiveness analysis, cost-utility analysis, health utility, quality-adjusted life-year.

Introduction

Cost-utility analysis (CUA) is a type of cost-effectiveness analysis in which health effects are measured in terms of quality-adjusted life-years (QALYs) gained or other preference-based outcome measures, allowing comparison of diverse interventions and health-care programs. Many experts and consensus groups have recommended CUA as the gold standard for conducting economic evaluations [1–5].

QALYs are calculated by weighting each time interval in a given state by its "utility"—a value between 0 and 1 that reflects the individual's preference for that health state relative to perfect health (weighted 1) and death (weighted 0). Once the utility weights are obtained for each state, they are multiplied by the time spent in the state and the products are summed to obtain the total number of QALYs (e.g., living 3 years with a utility value of 0.8 will produce 2.4 QALYs) [5]. Authors sometimes use the terms "utility," "value," "preference score," and "quality of life weight" interchangeably. For the remainder of this article, we use "utility" to encompass all of these terms.

Utility weights can be provided by direct measurement in patients, clinicians or the general population using preference-based techniques like the standard gamble and the time trade-off. Alternatively, utilities can be assigned indirectly by using a utility-weighted health status index (i.e., EuroQoL [6], Health Utility Index [7,8], Quality of Well-Being Scale [9]). In the second case, health status is assessed and then the utility score is calculated by applying preexisting weights to the health status domains. Whether preference- or

Address correspondence to: Carmen A. Brauer, BC Children's Hospital, Dept. Orthopedic Surgery, 4480 Oak St., Vancouver, BC V6H 3V4, Canada. E-mail: cbrauer@hsph.harvard.edu

nonpreference-based methods are used, consensuses groups have recommended that utility weights be derived from the community using as large a sample size as possible [1–5].

Although primary data collection of utility weights may be preferable, this may not be feasible because of the time or expense involved. One alternative is to use estimates from prior research. In 1996, the US Panel on Cost-Effectiveness in Health and Medicine (the Panel) recommended that researchers should assemble a catalog of utility weights that could provide a convenient reference tool for use by researchers and policymakers [5]. Researchers have responded by mapping preference data to health status instruments [10,11], using community responses to assess the health-related quality-of life for a wide variety of conditions [12], and developing catalogs of published utility weights [13,14]. The goal is a publicly available catalog of well-described health states and utility values to allow users to obtain relevant utilities for their analyses.

Recently, researchers at the Harvard School of Public Health and Tufts-New England Medical Center developed a comprehensive registry of CUAs, containing detailed information on studies published in the health and medical literature from 1976 to 2001 [15]. We previously examined and cataloged all utility weights for health states in CUAs published through 1997 [13]. The objectives of the current study were to update this catalog to include utilities published from 1998 to 2001, and to document recent changes in the methods used for utility elicitation. We hypothesized that after 1998, because of the recommendations of the Panel, we would begin to see improved utility elicitation methods and more transparent reporting of these methods in the literature.

Methods

This study was performed as part of a larger study reviewing all original CUAs in medicine [15–18]. Study details including information on the cost-utility ratios [17], utility values [13], and reporting practices [15,16] are described elsewhere in detail, and a comprehensive registry of these CUAs is available at (http://www.tufts-nemc.org/cearegistry). For this larger study, a Medline search of the English-language literature published between 1976 and 2001 was performed using the medical subject headings and/or text keywords "quality-adjusted," "QALY," and "cost-utility." Our search findings were validated to those of the Health Economic Evaluation Database (HEED) used by the British National Health Service [19].

As an update to this larger registry of CUAs and to our previously published catalog of utility weights reported in the CUAs published between 1976 and 1997 [13] we systematically reviewed the CUAs pub-

lished from 1998 to 2001 for data on all utilities used or reported in these studies. Two trained readers independently abstracted data on each health state description, the corresponding utility point estimate and ranges, the methods of elicitation, and the sources of the estimates. Readers met to reach consensus on the results and a third reviewer adjudicated any discrepancies. The first author (C.B.) then reviewed all articles to determine if original references needed to be consulted for clarification or supplementation of the utility data. Whenever possible, original references were consulted if utility weights were obtained from previous studies. When health state descriptions or utility estimates were not completely presented in the article or in another obtainable reference, they were excluded from the catalog and this study. When the same original source of utilities was referenced and used by more than one CUA, it was included only once. When defining the health state, we attempted, as much as possible, to record and report the terms used by the original authors. Nevertheless, we also recorded any relevant contextual information and occasionally amended the language to improve the clarity of the description.

All health state descriptions were mapped by trained physicians (C.B. and A.R.) into disease categories corresponding to several International Classification of Diseases 9th Revision (ICD-9) codes [20]. The utility catalog (http://www.tufts-nemc.org/careregistry) was then updated to include the 1998–2001 data within each appropriate disease category.

It is important to note one data collection change made in the current study (1998–2001) compared with our earlier catalog of utilities (1976–1997). Specifically, we record and report on utilities and their statistical ranges (i.e., standard deviation) in the current study, although our earlier catalog reports on any mentioned range of values around utilities regardless of statistical relationships. As a result, fewer ranges around the utility estimates are presented in this 1998–2001 update.

Analyses

The complete catalog of utility weights from 1976 to 2001 was then analyzed to determine the distribution of: 1) the utilities across disease categories; 2) the utility elicitation techniques used; 3) the different source populations (i.e., patients, general population) used; and 4) the range of estimates provided for similar health states. To explore the impact of the Panel's recommendations, we examined how preference elicitation methods used in 1998–2001 differed from those used before 1998. For comparisons between the two time periods, proportions were compared using the chi-square test. A *P*-value less than 0.05 on two-tailed tests was considered statistically significant.
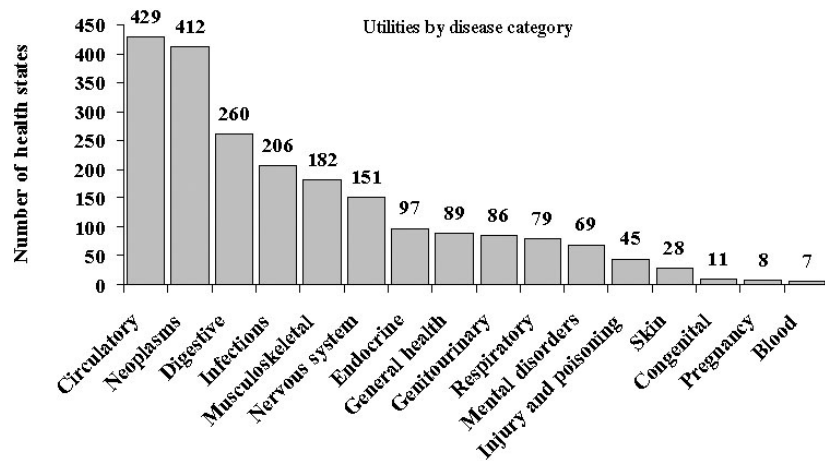
Figure 1 Number of health state descriptions and utility weights by disease category from published cost-utility analyses 1976–2001. (Total = 2159).

Results

From 1976 to 2001, 533 CUAs reported 2159 separate utilities, with 306 of the CUAs and 1210 of the utilities published after 1998. The majority of utilities for health states pertained to cardiovascular disease (19.9%) and oncology (19.1%). Many disease categories, such as mental health, blood disorders and the genitourinary system had few published utility values (Fig. 1).

In general, the health state utilities were clustered at higher values (Fig. 2). Four hundred and seventy-seven health states (22.2%) were valued above 0.9, whereas 72 (6%) were valued lower than 0.3. Six health states were valued below zero, suggesting that respondents considered these health states to be worse than death.

Health state utilities were obtained using several different methods (Table 1). Five hundred seventy-four (26.6%) health state descriptions were valued by direct time trade-off or standard gamble techniques, and 194 (9.0%) by rating scale techniques. For the 499 (23.1%) weights estimated using a generic health state classification instrument, the most commonly used instrument was the EQ-5D (EuroQol) in 161 health states (7.5%). The Quality of Well-Being (5.0%) and the Health Utilities Index (2.5%) were also commonly used.

Utilities were obtained from several different source populations. Five hundred thirty-five (24.8%) utilities were derived from clinician or author input without any formal methodology employed. A total of 590 (27.3%) health states were valued using community-based weights. This included both the cases where the authors directly elicited weights from members of the community, and those using generic instruments that relied on previously derived community weights. Five hundred thirty-six of the health state descriptions (24.8%) were valued by patients. The frequency with which authors followed the Panel recommendation, to elicit community weights using a generic instrument, is shown in Figure 3, stratified by disease category.

Different studies often reported different utility weights for the same health state. For example, since 1998, nine weights for myocardial infarction with similar health state descriptions have reported point estimates ranging from 0.58 to 0.93. In other disease
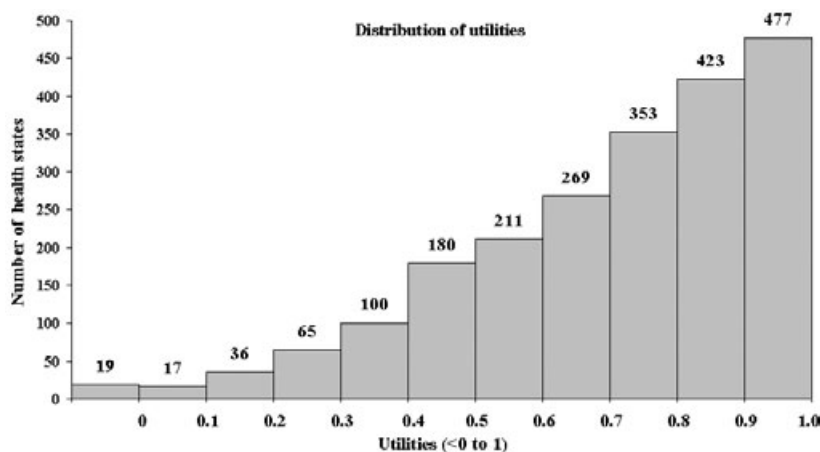

Figure 2 Distribution of utility weights for health states from published cost-utility analyses 1976–2001. (Total = 2150; nine utility weights gave only ranges and are therefore not included in this figure).

Table 1 Measurement technique and population source for utility weights in published cost-utility analyses

	% 1976–97 (N = 949)	% 1998–2001 (N = 1210)	P-value* before vs. after 1998
Measurement technique			
Generic tool	20.0	25.5	0.003
Rosser Index	6.8	0.4	<0.001
EQ-5D (EuroQol)	2.3	11.5	<0.001
Quality of Well-Being scale	5.7	4.4	
Health Utilities Index	3.4	4.3	
Time trade-off/standard gamble	25.3	27.5	
Rating scale	10.0	8.3	
Author/clinical judgment	32.5	18.8	<0.001
Source of utility weights			
Community	23.5	30.3	<0.001
Patients	26.8	23.3	
Clinician	35.8	21.0	<0.001
Author	25.1	18.7	<0.001

The total does not add up to 100% because some states did not clearly state the measurement technique used, or the source of the quality of life weights, and some states used multiple techniques or had multiple sources of estimates.
*P-value only shown if <0.05 (chi-square test).

areas, there was considerable variability in the health state descriptions (i.e., the disease severity, treatment characteristics and side-effect descriptions) resulting in wide utility ranges. For example, stroke weights ranged from −0.02 to 0.90 depending on the severity of the stroke in the health state description and the characteristics of the population surveyed.

Compared with the weights published before 1998, weights published from 1998 to 2001 were more frequently elicited from members of the community (30% up from 24%, $P < 0.001$), with fewer derived from author or expert opinion in the absence of any other

formal methodology employed (19% vs. 33%, $P < 0.001$). There was an increase in the number using a generic instrument, and some notable changes in the types of instruments used. For example, before 1997, the EQ-5D was applied in 2% of cases; after 1997 it rose to 12% (Table 1).

Discussion

We have updated a public-use catalog (http://www.tuft-nemc.org/cearegistry) of health states and corresponding utility weights, which now contains 2159 utility weights obtained from 25 years of published CUAs. In the update period (1998–2001), we were able to abstract appropriate information for more than 90% of analyses, a comparable rate to that found before 1998 [13]. In more than half of the update articles, a previously published study was used to estimate the weight. Because the majority of these estimates were new to our catalog, only 91 health states (4%) were excluded for referencing the same original source as another health state in our catalog; this may represent a trend to separately publish results of utility elicitation before using these weights in a formal CUA.

Despite the large number of utilities included, many clinical areas are still poorly represented in our utility catalog. High morbidity conditions such as musculoskeletal diseases continue to have few utility estimates available. Thus, it appears there is substantial room for further research to expand our base of catalogued utilities available to researchers and decision-makers.
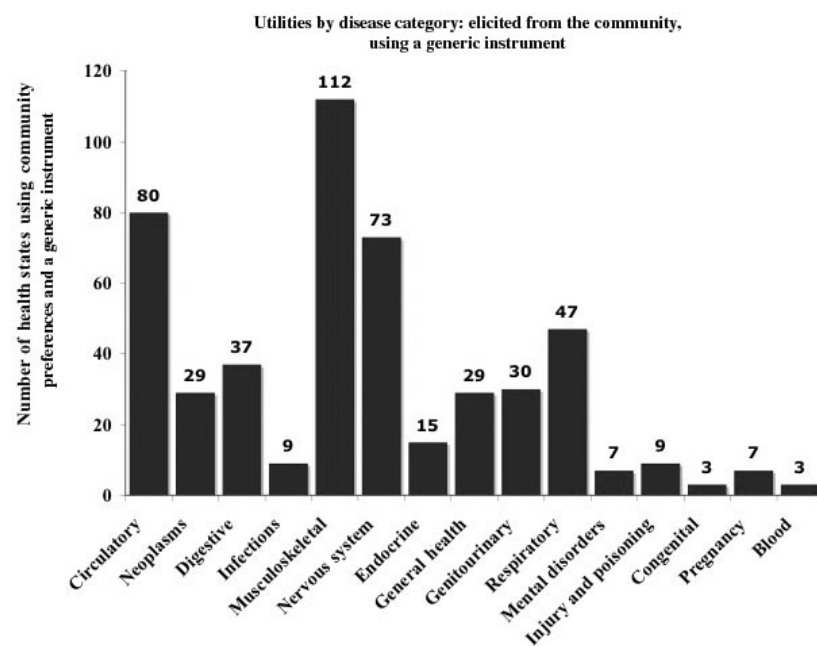

Figure 3 Number of health state descriptions and utility weights by disease category, elicited from the community, using a generic instrument, from published cost-utility analyses. (Total = 409).

Many of the health states in our catalog had wide ranges of utilities reported. Past studies suggest that the disparate utility estimates may result from differences in the health state descriptions or the populations from which the utilities were elicited [12–14,21–23]. For example, published myocardial infarction (MI) utilities are estimated at 0.58 and 0.88 in separate studies. The lower utility (0.58) was elicited using the EQ-5D in patients after peripheral vascular surgery, a subset of which had a postoperative MI [24]. In contrast, the higher utility (0.88) was elicited using time trade-off, in patients with a recent MI and a mildly depressed ejection fraction but no clinical symptoms of heart failure [25]. Subtle differences in the descriptions of seemingly similar health state may also have led to the wide variations found in the published utilities for health states.

The wide utility ranges for the same health states highlights the need for researchers to carefully consider which utility estimates they ultimately use in their analyses. Future studies would do well to use the utilities obtained from the most methodologically sound studies, in the most comparable populations, and using the most relevant health state descriptions. Despite this choice of a "best-utility estimate," for base case analyses, researchers would do well to explore a range of cataloged health state utilities in sensitivity analyses. The utilities chosen for a health state range may depend on the disease. For example in stroke with such a wide range of utilities, it may be useful to explore the range associated with the "best utility estimate" rather than the full range from the catalog. Indeed, perhaps the most important use of the catalog may be the identification of a relevant range of utility weights to include in sensitivity analyses.

There appears to be some improvement in the methods used to value health states over time. This trend after publication of the Panel's recommendations, suggests that researchers may be making an effort to adhere to the Panel's methodological recommendations. Still, many studies do not adhere to recommended practices and a large number do not make their techniques explicit. Even in the 1998–2001 time period, study authors did not clearly define the population source for 12% of utility weights, and for 18% they did not state the elicitation technique used. The persistent of these variations is troubling. Likely concerns about the comparability and credibility of analyses will persist without further improvements and standards in the field.

A number of study limitations warrant mention. Our search strategy was limited to select key words and the search was restricted to original CUAs. Therefore, some well-designed utility elicitation studies may have been omitted. Second, our data abstraction accuracy was limited by the availability of correct data in the original documents. If authors failed to offer comprehensive descriptions of health states or methods, this may be reflected in our database; although we tried to obtain the necessary information from other references or authors, we were sometimes unsuccessful.

Handling of patient preferences is one of the most challenging issues in cost-effectiveness analysis. The challenge for the future will be to find reliable and valid measurement techniques, and to use them consistently across studies and over time. In the mean time, researchers should be encouraged to present their methodology in a clear and transparent way so that analysts and decision-makers are aware of how the utilities selected affect the cost per QALY ratios found. It is our hope that posting of our catalog on the World Wide Web will help ensure broad access for researchers to identify relevant utility weights for point estimates and appropriate ranges for sensitivity analyses.

References

1 Russell LB, Gold MR, Siegel JE, et al. The role of cost-effectiveness analysis in health and medicine. Panel on cost-effectiveness in health and medicine. JAMA 1996;276:1172–7.

2 Siegel JE, Weinstein MC, Russell LB, et al. Recommendations for reporting cost-effectiveness analyses. Panel on cost-effectiveness in health and medicine. JAMA 1996;276:1339–41.

3 Weinstein MC, Siegel JE, Gold MR, et al. Recommendations of the panel on cost-effectiveness in health and medicine. JAMA 1996;276:1253–8.

4 Gold M. Panel on cost-effectiveness in health and medicine. Med Care 1996;34(Suppl.):DS197–9.

5 Gold MR, Siegel JE, Russell LB, et al. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.

6 EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199–208.

7 Feeny D, Furlong W, Boyle M, et al. Multi-attribute health status classification systems: Health Utilities Index. Parmacoeconomics 1995;7:490–502.

8 Torrance GW, Furlong W, Feeny D, et al. Multi-attribute preference functions: Health Utilities Index. Parmacoeconomics 1995;7:503–20.

9 Kaplan RM, Anderson JP, Wu AW, et al. The Quality of Well-Being scale: applications in AIDS, cystic fibrosis, and arthritis. Med Care 1989;27(Suppl.):S27–43.

10 Gold M, Franks P, Erickson P. Assessing the health of the nation. The predictive validity of a preference-based measure and self-rated health. Med Care 1996;34:163–77.

11 Gold MR, Franks P, McCoy KI, et al. Toward consistency in cost-utility analyses: using national measures to create condition-specific values. Med Care 1998;36:775–7.

12 Fryback DG, Dasback EJ, Klein R, et al. The beaver dam health outcomes study: initial catalog of health-state quality factors. Med Decis Making 1993;13:89–102.

13 Bell CM, Chapman RH, Stone PW, et al. An off-the-shelf help list: a comprehensive catalog of preference scores from published cost-utility analyses. Med Decis Making 2001;21:288–94.

14 Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. Med Care 2000;38:583–637.

15 Neumann PJ, Greenberg D, Okchanski NV, et al. Growth and quality of the cost-utility literature, 1976–2001. Value Health 2005;8:3–9.

16 Neumann PJ, Stone PW, Chapman RH, et al. The quality of reporting in published cost-utility analyses, 1976–1997. Ann Intern Med 2000;132:964–72.

17 Chapman RH, Stone PW, Sandberg EA, et al. A comprehensive league table of cost-utility ratios and a sub-table of "panel-worthy" studies. Med Decis Making 2000;20:451–67.

18 Chapman RH, Berger M, Weinstein MC, et al. When does quality-adjusting life-years matter in cost-effectiveness analysis? Health Econ 2004;13:429–36.

19 Office of Health Economics and the International Federation of Pharmaceutical Manufacturers' association's. Health Economic Evaluations Database. Available from: http://www.oheheed.com/ [Accessed June 26, 2002].

20 U.S. Public Health Service and Health Care Financing Administration. The International Classification of Diseases 9th Revision Clinical Modification ICD-9-CM. Vol. 1, 5th edn. [Online] Available from: http://cedr.lbl.gov/icd9.html [Accessed October 12, 2000].

21 Hornberger JC, Redelmeier DA, Petersen J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. J Clin Epidemiol 1992;45:505–12.

22 Read JL, Quinn RJ, Berwick DM, et al. Preferences for health outcomes: comparisons of assessment methods. Med Decis Making 1984;4:315–29.

23 Nord E. Methods for quality adjustment of life years. Soc Sci Med 1992;34:559–69.

24 Oostenbrink JB, Tangelder MJ, Busschback JJ, et al. Cost-effectiveness of oral anticoagulants versus aspirin in patients after infrainguinal bypass grafting surgery. J Vasc Surg 2001;34:254–62.

25 Sanders GD, Hlatky MA, Every NR, et al. Potential cost-effectiveness of prophylactic use of the implantable cardioverter defibrillator or amiodarone after myocardial infarction. Ann Intern Med 2001;135:870–83.