SPINE Volume 25, Number 24, pp 3161–3166

Preference-Based Health Outcome Measures in Low Back Pain

Anna N. A. Tosteson, ScD

The motivation for measurement of preference-based health outcomes in clinical practice settings and in determining health policy is reviewed. In clinical practice, where preferences for various health outcomes should guide treatment decisions, utility assessment using the visual analogue scale, time trade-off and standard gamble is described. In determining health policy, where economic evaluations motivate the need for preference-based measures, preference classification systems for valuing health are considered.

The desirability of health outcomes can be compared using preference-based health outcome measures. Such measures are distinct from health status instruments, because they characterize how health outcomes are valued rather than functional performance or capacity. It is important to recognize that two individuals with identical self-assessed health status as measured by a generic instrument, such as SF-36,[47] may value their health quite differently.

Preference-based health outcome measures are important in several contexts.[45] In this article, approaches to preference-based outcome measurement are considered in both the clinical and health policy settings. In the clinical context, it is recognized that patients' preferences should be considered in guiding clinical care.[5,37] Therefore, an understanding of how health outcomes are valued can provide important information for both the clinician and patient.[38] In the health policy context, efforts to assess the economic value of health interventions require an outcome measure that facilitates comparisons across diverse diseases. Quality-adjusted life years (QALYs), which account for morbidity by applying an explicit value to the desirability of various health outcomes, have been recommended for this purpose.[25]

Clinical Context: Shared Decision Making

The dialogue between physician and patient often touches on the benefits and harms of alternative treatment approaches. How the patient values the potential benefits and harms (*i.e.*, their preferences) may influence the treatment approach that is pursued. The process of shared decision making involves both patients and physicians in considering the probability of treatment outcomes and patient's preferences for those outcomes.[21] Current efforts to enhance shared decision making are motivated by the notion that patient preferences for health outcomes should play a central role in medical decision making.[48]

A classic example of the central role for patient preferences in determining optimal clinical choice involves treatment of benign prostatic hypertrophy.[2] Barry et al[2] used decision analysis to highlight that the "best" treatment for individual men depended on the values men assigned to life in certain health states including moderate prostatism and life with impotence. Quantitative estimates reflecting how individuals value such health states, which are sometimes referred to as utilities, are typically scaled from 0 (worst imaginable health state) to 1 (best imaginable health state). The axiomatic basis for utility assessment was first described by von Neumann and Morgenstern[46] in the 1940s. Since then, several approaches to health valuation have been described.[43]

Utility assessment requires elicitation of a value characterizing an individual's preference for a particular health state. The health state may be one that the individual is experiencing, such as current health, or it may be one that he or she is asked to imagine based on a description. The most common approaches to direct utility assessment include use of a visual analogue scale (VAS, also referred to as a rating scale), time trade-off (TTO), or standard gamble (SG).[43] Each scaling approach is briefly reviewed. For a more comprehensive treatment of measurement and scaling issues, the reader is referred to a series of articles by Froberg and Kane.[17–20]

Visual Analogue Scale

To value a health state using the VAS, subjects are asked to mark where a health state would be classified on a scale in which 0 represents the worst possible health state and 1 represents the best possible health state. This scale may be represented as a vertical line, feeling thermometer (Figure 1), or a 10-cm line. The value for the health state is then estimated as the measured distance between 0 and the respondent's mark. Category scaling, which requires subjects to rate their health using a narrower set of outcome responses, is a closely related technique.

Time Trade-Off

This approach to valuation is intended to determine the amount of time in better health the person would accept as equivalent to a fixed period of time in an impaired health state.[42] After considering a health state, it is valued using a TTO by offering subjects a choice between

From the Department of Medicine and the Center for Evaluative Clinical Sciences, Department of Community and Family Medicine, Dartmouth Medical School, Hanover, New Hampshire.
This study was supported in part by grant U01-AR45444 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases US Public Health Service.

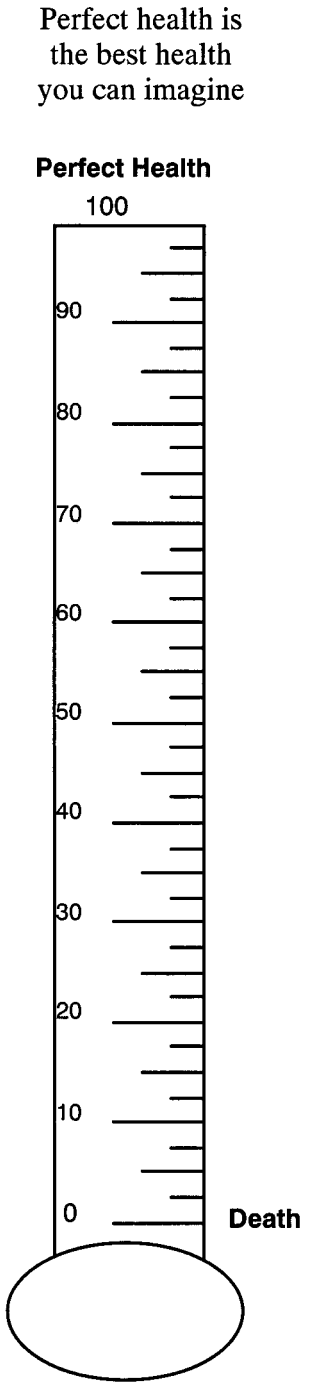

Figure 1. Feeling thermometer for assessing the value for Health State A.

two options (Figure 2). One option allows the subject to live in the health state under consideration for X years. The second option allows the subject to live in the best (or better) health state for fewer years, X − Y. The time that the subject must trade (*i.e.*, give up), Y, to live in the best health state is varied systematically until the subject is indifferent between living in the health state for X years and living in the best health state for X − Y years. At the point of indifference, the value for the health state is estimated as (X − Y)/X. For health states that are transient, sleep, rather than living for fewer years, can be used as a metaphor for trading time.[8]

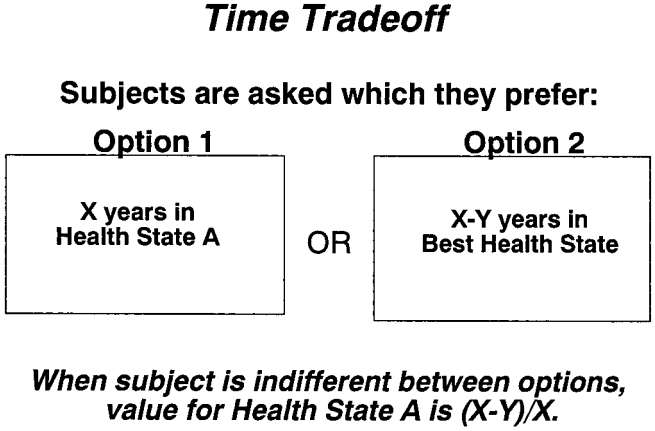

Figure 2. Time tradeoff technique for assessing utility for Health State A.

Standard Gamble

This valuation method requires subjects to choose between a risky alternative: the gamble and a sure thing. To value a health state using the SG, after considering the health state, subjects are offered a choice between two options (Figure 3). One option, the sure thing, allows the subject to remain in the health state being considered. The second option, the gamble, allows the subject a chance (p) of attaining the best health outcome and a complimentary chance, 1 − p, of attaining the worst health outcome. The chance of attaining the best outcome (p) is varied systematically until the subject is indifferent between the two options. At the point of indifference, p is defined as the utility for the health state.

In empirical studies, it generally has been found that values obtained using the VAS are lowest, with TTO values being higher, and SG values being highest. Of the methods outlined here, the VAS is considered the easiest to administer. The VAS involves fewer quantitative concepts and requires no understanding of probabilistic concepts. Findings in recent work indicate that among sub-

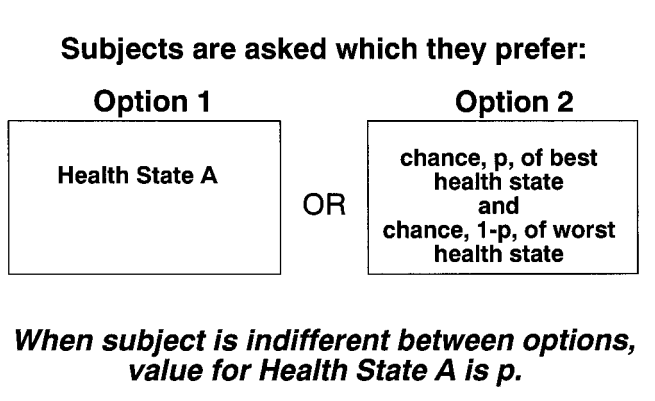

Figure 3. Standard gamble technique for assessing utility for Health State A.

jects with low numeracy (one or more errors in responding to three questions involving concepts of chance) only current health valuations using a VAS had evidence of construct validity among subjects who had low numeracy.[49] It must be noted, however, that only the SG provides a true measure of utility according to the theory described by von Neumann and Morgenstern.[46]

Implementation of the TTO or SG often involves use of either a trained interviewer and visual aids or an automated utility assessment instrument, such as U-Titer or IMPACT.[32,33,39] The latter facilitate uniform collection of data in multicenter studies.[44]

Few reports focus on utilities in patients with low back pain. Goossens et al[27] compared the VAS and SG methods for assessing utilities in clinical trial participants with fibromyalgia and chronic musculoskeletal pain and in patients with nonspecific low back pain. The VAS values were found to have an association with back pain–specific instruments and to be responsive to changes over time. In contrast, the SG had little association with back pain measures. In addition, although the VAS discriminated between patients reporting themselves as improved *versus* not improved, the SG did not. Based on these findings, the authors recommended use of the VAS for assessing patients with chronic musculoskeletal pain.

Alternative nonutility-based approaches to preference assessment also deserve consideration. Llewellyn-Thomas[34] has described a probability trade-off approach for assessing strength of preference between treatment alternatives. Although utility assessment focuses on how outcomes are valued, the probability trade-off focuses on the probabilities of the various outcomes. The probability trade-off provides a balanced description of the probability of alternative treatment outcomes and requires individuals to choose between treatments, when the probabilities of the outcomes are varied. This approach to treatment selection may represent a more natural task in the clinical setting than use of methods such as the SG, TTO, or VAS and has been used to assess patient preferences for treatment of benign prostatic hyperplasia.[35]

■ Health Policy Context: Economic Evaluation

Analysis of the cost-effectiveness of alternative health interventions requires use of a comprehensive health outcome measure to compare interventions across disease areas and to account for the impact of interventions on both morbidity and mortality. As noted earlier, the QALY has been recommended as an appropriate outcome measure for economic evaluations in health care.[25] Estimation of QALYs requires that each year of life be weighed according to its utility or value, which ranges from 0 (worst imaginable health) to 1 (best imaginable health).

In economic evaluations, it is desirable that preference weights for health outcomes reflect societal values.[25] Therefore, instead of using the utility assessment techniques described earlier to directly measure health, QALYs may be estimated using a preference classification system. Preference classification systems consist of a health status instrument, which is used to classify the population into a discrete number of health states, and a related set of societal preference weights for each discrete health state. Generally, the number of health states within a classification system is determined by both the number of health attributes considered (*e.g.*, physical function, pain) and number of levels within each attribute. For example, the EuroQol group's EQ-5D instrument has five attributes with three levels each.[14] This results in 243 health states, which are combined with a "dead" and "unconscious" state to yield 245 distinct health states.

Societal preferences are derived from a population sample for a subset of health states. Population values can be obtained for each health attribute independently (*e.g.*, values are assigned for each level of physical functioning) and/or for a subset of composite health states (*e.g.*, each level of each health attribute is specified, and the entire health state is valued). When attributes are valued independently a multiattribute utility function[31] is used to infer values for composite health states. Whether independent attributes or composite health states are valued, written outcome descriptions must be used to communicate potentially unfamiliar health states to subjects. These health states are valued using the SG, TTO, or VAS. Ultimately, statistical techniques are used to infer values for each health state in the preference classification system, because direct utility assessment for each health state is not feasible (*e.g.*, there are thousands of health states in some systems).

The most widely used generic preference classification systems include the Health Utilities Index (HUI)[3,15,40] the EQ-5D,[6,11,13,14] and the Quality of Well-being Scale (QWB).[1,30] Each of these systems is briefly reviewed and a recent system, the SF-6D,[4] is also described.

Health Utilities Index

Several versions of the HUI are available and are comprehensively reviewed by Feeny et al[16] and Torrance et al.[41] The HUI: Mark I (HUI:1) instrument was developed to evaluate outcomes in very-low-birth-weight neonates and consists of four attributes, which define 960 possible health states. The newer Mark II (HUI:2) system defines 24,000 health states using seven attributes, whereas the Mark III (HUI:3) system defines 960,000 health states using eight attributes. Both HUI:2 and HUI:3 are appropriate for assessing population health. A primary distinction between the HUI:2 and HUI:3 systems relates to the functional association between attributes. Either the Mark II or Mark III scoring system may be applied to the 15-item HUI questionnaire. Whereas the former systems are based on a multiplicative model, the latter assumes a more general multilinear relation. Feeny et al[16] note that HUI:3 may not fully capture the disutilities associated with musculoskeletal problems that do not affect the lower limbs or hands and

fingers. Permission is necessary for use of the HUI. Information on its use is available from Dr. William Furlong (Health Utilities, Inc., Dundas, Ontario, Canada).

EQ-5D

The EuroQol group's EQ-5D,[6,14] which consists of 245 distinct health states, is defined using five attributes with three levels (*e.g.*, no problem, moderate problem, severe problem) within each attribute. The attributes considered are mobility, self-care, usual activities, pain–discomfort, and anxiety–depression. The EQ-5D is generally implemented using a five-item questionnaire and a VAS that requires subjects to directly rate their current health. The VAS is not used in deriving the EQ-5D QALY, but provides a direct assessment of how each individual values her or his health state. A random sample of the United Kingdom population initially provided values for composite health states using a VAS. More recently, scoring algorithms based on TTO values in a random sample of the United Kingdom population have become available.[13] The EQ-5D is available free of charge for noncommercial uses from the EuroQol Group (http://www.euroqol.org) and has been recommended for use in low back pain studies.[10]

Quality of Well-Being Scale

The QWB scale was developed as part of a comprehensive general health policy model.[29] It combines three attributes characterizing function (mobility, physical activity, and social activity) with a fourth attribute reflecting symptoms and problems. It does not consider mental health as an attribute in valuing health. The three functional attributes are assumed to be independent (*i.e.*, an additive value function is used). Values for the QWB have been elicited using a VAS approach among a random sample of residents of San Diego, California. This instrument has been used widely in a variety of disease areas.[9] Until recently, use of this instrument required administration of lengthy in-person interviews. This version of the instrument is currently available through the Medical Outcomes Trust (http://www.qlmed.org/mot/index.html) for a nominal fee. Development of a self-administered version of the questionnaire may make its use more practical.[28] A regression approach to estimating QWB has been described using the physical function, mental health, bodily pain, and general health subscales of the SF-36 health status instrument.[23]

SF-6D

Brazier et al[4] describe a method for deriving a single preference score based on responses to the SF-36 health status instrument. This system assigns values for health states based on six attributes: physical function (six levels), role limitation (two levels), social function (five levels), bodily pain (six levels), mental health (five levels), and vitality (five levels). To date, values have been obtained from two convenience samples of the population in the United Kingdom using both SG and VAS scaling methods.

One important element to consider when choosing among instruments is the viewpoint of the instrument with regard to function. In contrast to the other preference classification systems, the HUI:2/3 systems are based on functional capacity rather than performance and have been referred to as "within the skin" measures.[41] They assess what subjects can do without the assistance of corrective devices. For example, the vision subscale is based on ability of a subject to see without corrective lenses. As such, it would be difficult to evaluate the cost-effectiveness of devices such as eyeglasses that allow individuals to overcome functional limitations. The rationale for this measurement approach is that it isolates the choices that the individuals make (*e.g.*, getting corrective lenses) from their inherent health state.

To date, preference classification systems have been used more extensively to assess health outcomes in osteoporosis[12,24,36] than in low back pain. The QWB scale was among the health outcome measures used to evaluate long-term intrathecal opioid therapy in patients with noncancer low back pain.[7] Although the EQ-5D has been recommended for use in studies of patients with low back pain,[10] few data have yet been reported about the use of this instrument. The ongoing Spine Patient Outcomes Trial (SPORT), which is examining the effectiveness and cost-effectiveness of surgical *versus* nonsurgical approaches to management of patients with chronic low back pain, should provide some comparative information on health values obtained using EQ-5D, HUI, SF-6D, and a VAS for current health.

In the health policy context, this article has been focused on preference classification systems that provide a societal value for health states. However, measurement of health outcomes using the VAS, SG, or TTO is sometimes undertaken for purposes of economic evaluation. For example, in a health economic assessment of a randomized clinical trial of behavioral rehabilitation in patients with chronic low back pain, both the VAS and SG were used to assess health outcomes.[26]

Finally, it is instructive to consider how direct utility assessment and preference classification systems compare when both are measured in the same individuals. In the population-based Beaver Dam Health study, which used several approaches to assessing health-related quality of life, average QWB and TTO scores were given for the 249 of 1356 interviewees who reported that they had been affected by severe back pain within the past year.[22] The average health values (*i.e.*, QALY) for persons with severe back pain were 0.67 (95% confidence interval [CI]: 0.66–0.68) and 0.786 (95% CI: 0.757–0.815) for QWB and TTO, respectively. Although average QWB scores were lower, both instruments provided a similar characterization of the impact of low back pain on QALYs. In particular, when mean differences among the groups affected by severe back pain and those not affected were compared, the differences in the group means were 0.07 and 0.09 for the QWB and TTO scores, respectively.

In another study addressing health state valuation, Gabriel et al[24] compared directly assessed current health values with those obtained using preference classification systems among women with osteoporosis. A close correspondence between directly measured TTO values for current health and values derived from the HUI:2 was found. In contrast, QWB scores estimated based on SF-36 subscales showed significantly lower valuations than either TTO current health values or HUI:2. The findings in this study showed that for diseases, such as osteoporosis, that have a measurable impact on physical functioning, preference classification systems may provide reasonable proxies to direct utility assessment. Because low back pain, similar to osteoporosis, has a large impact on physical function, it is likely that preference classification systems, such as HUI, will be adequate for characterizing the health impact of low back pain interventions in economic studies.

■ Summary

Preference-based health outcome measures were considered in two settings. In the clinical context, utility assessment using the VAS, TTO or SG was described. It is recognized that formal utility assessment is generally not needed in day-to-day clinical practice. When the need to explicitly value alternative health outcomes arises, the VAS is recommended because of its ease of implementation. In the health policy context, preference classification systems are recommended for valuing health when a cost-effectiveness analysis of alternative health care interventions is planned. At present, there is little evidence to base choice of one system over another for assessing the cost-effectiveness of low back pain interventions.

References

1. Anderson JP, Kaplan RM, Berry CC, et al. Interday reliability of function assessment for a health status measure: the Quality of Well-Being scale. Med Care 1989;27:1076–83.
2. Barry MJ, Mulley AG Jr, Fowler FJ, et al. Watchful waiting vs immediate transurethral resection for symptomatic prostatism. The importance of patients' preferences. JAMA 1988;259:3010–7.
3. Boyle M, Furlong W, Feeny D, et al. Reliability of the health utilities index: mark III used in the 1991 cycle 6 Canadian general social survey health questionnaire. Qual Life Res 1995;4:249–57.
4. Brazier J, Usherwood T, Harper R, et al. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol 1998;51:1115–28.
5. Brody DS. The patient's role in clinical decision-making. Ann Intern Med 1980;93:718–22.
6. Brooks R, Group E. EuroQol. The current state of play. Health Policy 1996; 37:53–72.
7. Brown J, Klapow J, Doleys D, et al. Disease-specific and generic health outcomes: a model for the evaluation of long-term intrathecal opioid therapy in noncancer low back pain patients. Clin J Pain 1999;15:122–31.
8. Buckingham J, Birdsall J, Douglas J. Comparing three versions of the time tradeoff: time for a change? Med Decis Making 1996;16:335–47.
9. Coons S, Rao S, Keininger D, et al. A comparative review of generic quality-of-life instruments. Pharmacoeconomics 2000;17:13–35.
10. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research: a proposal for standardized use [published erratum appears in Spine 1999;24:418]. Spine 1998;23:2003–13.
11. Dolan P, Gudex C, Kind P, et al. The time trade-off method: results from a general population study. Health Econ 1996;5:141–54.
12. Dolan P, Torgerson D, Kakarlapudi TK. Health-related quality of life of Colles' fracture patients. Osteoporos Int 1999;9:196–9.
13. Dolan P. Modeling Valuations for EuroQol Health Status. Med Care 1997; 35:1095–108.
14. EuroQol Group. EuroQol: A new facility for the measurement of health-related quality of life. Health Policy 1990;16:199–208.
15. Feeny D, Furlong W, Barr R, et al. A comprehensive multiattribute system for classifying the health status of survivors of childhood cancer. J Clin Oncol 1992;10:923–8.
16. Feeny D, Furlong W, Boyle M, et al. Multi-attribute health status classification systems: Health Utilities Index. Pharmacoeconomics 1995;7:490–502.
17. Froberg D, Kane R. Methodology for measuring health-state preferences-II: Scaling methods. J Clin Epidemiol 1989;42:459–71.
18. Froberg DG, Kane RL. Methodology for measuring health-state preferences: I. Measurement strategies. J Clin Epidemiol 1989;42:345–54.
19. Froberg DG, Kane RL. Methodology for measuring health-state preferences: IV. Progress and a research agenda. J Clin Epidemiol 1989;42:675–85.
20. Froberg DG, Kane RL. Methodology for measuring health-state preferences: III. Population and context effects. J Clin Epidemiol 1989;42:585–92.
21. Frosch DL, Kaplan RM. Shared decision making in clinical medicine: past research and future directions. Am J Prev Med 1999;17:285-94.
22. Fryback D, Dasbach E, Klein R, et al. The Beaver Dam health outcomes study: initial catalog of health-state quality factors. Med Decis Making 1993;13:89–102.
23. Fryback D, Lawrence W, Martin P, et al. Predicting Quality of Well-being scores from the SF-36. Med Decis Making 1997;17:1–9.
24. Gabriel SE, Kneeland TK, Melton LJ, et al. Quality of life in economic evaluations for osteoporosis: Whose utility values should we use? Med Decis Making 1999;19:141–8.
25. Gold M, Siegel J, Russell L, et al. Cost-Effectiveness in Health, Medicine. New York: Oxford University Press; 1996.
26. Goossens ME, Rutten-Van Molken MP, Kole-Snijders AM, et al. Health economic assessment of behavioural rehabilitation in chronic low back pain: a randomised clinical trial. Health Econ 1998;7:39–51.
27. Goossens ME, Vlaeyen JW, Rutten-van Molken MP, van der Linden SM. Patient utilities in chronic musculoskeletal pain: how useful is the standard gamble method? Pain 1999;80:365–75.
28. Kaplan R, Sieber W, Ganiats T. The Quality of Well-Being Scale: comparison of the interview-administered version with a self-administered questionnaire. Psychol Health 1997;12:783–91.
29. Kaplan RM, Anderson JP. A general health policy model: update and applications. Health Serv Res 1988;23:203–35.
30. Kaplan RM, Atkins CJ, Timms R. Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. J Chronic Dis 1984;37:85–95.
31. Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York: Wiley; 1976.
32. Lenert LA, Hornberger JC. Computer-assisted quality of life assessment for clinical trials. Proc American Medical Informatics Association Annual Fall Symp 1996;992–6.
33. Lenert LA, Michelson D, Flowers C, et al. IMPACT: an object-oriented graphical environment for construction of multimedia patient interviewing software. Proc Annu Symp Comput Appl Med Care 1995;319–23.
34. Llewellyn-Thomas H. Turning decision analysis around. Ann R Coll Physicians Surg Can 1994;27:177–8.
35. Llewellyn-Thomas HA, Williams JI, Levy L, et al. Using a trade-off technique to assess patients' treatment preferences for benign prostatic hyperplasia. Med Decis Making 1996;16:262–82.
36. Oleksik A, Lips P, Dawson A, et al. Health-related quality of life in postmenopausal women with low BMD with or without prevalent vertebral fractures. J Bone Miner Res 2000;15:1384–1392.
37. Owens DK. Spine Update. Patient preferences and the development of practice guidelines. Spine 1998;23:1073–9.
38. Sherbourne CD, Sturm R, Wells KB. What outcomes matter to patients? See comments. J Gen Intern Med 1999;14:357–63.
39. Sumner W, Nease R, Littenberg B. U-Titer: A utility assessment tool. In: Fifteenth Annual Symposium on Computer Applications in Medical Care; 1991: 701–705.
40. Torrance G, Zhang Y, Feeny D, et al. Multi-attribute preference functions for a comprehensive health status classification system. CHEPA working paper series No. 92-18 1992:1–61.
41. Torrance GW, Furlong W, Feeny D, et al. Multi-attribute preference functions: Health Utilities Index. Pharmacoeconomics 1995;7:503–20.
42. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Serv Res 1972;7:118–33.
43. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ 1986;5:1–30.
44. Tosteson A, Kneeland T, Nease R, et al. Feasibility of automated utility assessment in phase III clinical trails: U-Titer and a Raloxifene osteoporosis prevention trial [Abstract]. Med Decis Making 1995;15:413.

45. Tsevat J, Weeks JC, Guadagnoli E, et al. Using health-related quality-of-life information: clinical encounters, clinical trials, and health policy. J Gen Intern Med 1994;9:576–82.
46. von Neumann J, Morgenstern O. Theory of games and economic behaviour. New York: Wiley; 1944.
47. Ware J, Sherbourne C. The MOS 36-item short-form health survey (SF-36). Med Care 1992;30:473–483.
48. Weinstein JN. The missing piece: embracing shared decision making to reform health care [editorial]. Spine 2000;25:1–4.
49. Woloshin S, Schwartz L, Moncur M, Gabriel S, Tosteson A. Assessing values for current health: Numeracy matters [Abstract]. J Gen Intern Med 2000;15(suppl 2).

Address reprint requests to

Dr. Anna Tosteson
HB 7505 Clinical Research
Dartmouth–Hitchcock Medical Center
1 Medical Center Drive
Lebanon, NH 03756
E-mail: Anna.Tosteson@Dartmouth.edu