Utility Elicitation Using Single-Item Questions Compared with a Computerized Interview

LESLIE A. LENERT, MD, CATHY D. SHERBOURNE, PhD, VALERIE REYNA, PhD

Background. The use of a simpler procedure for the measurement of utilities could affect primarily the variance or both the mean and the variance of measurements. In the former case, simpler methods would be useful for population studies of preferences; however, in the latter, their use for such studies might be problematic. ***Purpose***. The purpose of this study was to compare the results of utility elicitation using single-item questions to computer elicitation using the Ping-Pong search procedure. ***Methods***. In a convenience sample of 149 primary care patients with symptoms of depression, the authors measured and compared standard gamble (SG) utilities elicited using a single-item "open question" to SG elicitations performed using a computerized interview procedure. Elicitations were performed 1 to 2 weeks apart to minimize memory effects. ***Results***. More than 90% of persons with utilities of 1.0 to the single-item standard gamble had utilities of less than 1.0 on the computer SG instrument. Consistent with this finding, the mean utilities were lower in computer interviews (0.80 vs. 0.90; $P < 0.0001$ for differences). Within subjects, utility measures had only a fair degree of correlation ($r = 0.54$). ***Conclusions***. Use of single-item questions to elicit utilities resulted in less precise estimates of utilities that were upwardly biased relative to those elicited using a more complex search procedure. ***Key words:*** computers, utility, health survey/questionnaires, data collection/questionnaires. **(Med Decis Making 2001;21:97-104)**

Utilities are numerical ratings of the desirability of health states that reflect a subject's preferences. The index method for measurement of utilities is the standard gamble.[1] In the gamble, the object is to determine when a person is indifferent between living life for certain in some health condition and a gamble with probability, P, of a better outcome (typically life in excellent health) and the probability, $1 - P$, of a worse outcome (typically death).

A long-standing matter of debate is whether utilities innately exist within individuals or are created in the process of assessment, as Slovic and others have argued.[2] If utilities innately exist within an individual, then procedures for the measurement of utilities sample an individual's true preferences. One hypothesis that might follow, if utilities exist and are "sampled" by procedures, is that simplification procedures for utility assessment should primarily influence the precision of measurements, rather than the mean value obtained. If true, although simpler methods might not be useful for measuring individuals' preferences for a health state, they might be useful preferences of populations or groups.

A recent study by Albertsen et al.[3] examined the issue of whether a simpler "paper-based" assessment method produced the same results as a more complex procedure administered by computer program. The results supported the view that simplification of assessment procedures did not affect the central tendency of utility measurements. Based on hypothesis, several research groups have developed and fielded single-item self-administered questionnaires to measure utilities.[3-7]

Received 24 June 1999, from the Veterans Affairs San Diego Healthcare System and the University of California, San Diego (LAL) and RAND Corporation, Santa Monica, California and University of Arizona, Tucson (CDS). Revision accepted for publication 2 November 2000. This study was funded by grants from the National Library of Medicine (R29-LM05562) and the Agency for Health Care Policy and Research (R01-HS08349).

Address correspondence and requests for reprints to Leslie A. Lenert, MD, Associate Professor of Medicine, Health Services Research and Development MC: 111N1, VA San Diego Healthcare System, 3350 La Jolla Village Dr., San Diego, CA 92161; phone: (619) 552-4325; fax: (619) 552-4321; e-mail: llenert@ucsd.edu.

However, a substantial body of literature supports the concept that utility elicitations are either created or strongly shaped by the procedures used to obtain numeric values. Relatively subtle changes in procedures for elicitation can have important effects on utilities.[8] For example, Lenert et al. have shown that changes in the pattern of questions used to search for a participant's indifference point result in large and persistent differences in utilities.[9] Under this view, use of a simpler procedure for assessment should change both the central tendency and the variance of population estimates of utilities.

Most research work on the measurement of utilities has been conducted in the context of having a research assistant or a computer program conduct interviews to measure utilities. In the case of measurement of utilities for current health, the primary role of the assessor is to help the subject identify his or her true indifference point. This is typically performed by providing graphical feedback and a verbal, recorded, or text description of the magnitude of the trade-off being offered the subject.

Paper-based questionnaires typically differ from most instruments performing utility elicitation in several potentially important ways: (1) the response is measured as the answer to an single-item, open-ended question rather than the result of a "forced choice" or matching task; (2) there is no graphic or verbal feedback on trade-offs; and (3) there is no training of subjects in utility elicitation procedures prior to their answering the questions of interest. Thus, if utilities are created or strongly shaped during the process of elicitation, the two processes could lead to different results. In particular, a procedure consisting of a single open-ended question does little to encourage subjects to think quantitatively. Research indicates that judgments ordinarily reflect categorical distinctions (i.e., risk vs. no risk) unless procedures encourage precision.[10,11] Thus, we would expect fewer categorical responses (e.g., utilities of 1.0) in the computerized elicitation conditions because dynamic procedures such as "Ping-Ponging" among numerical estimates encourage subjects to think with greater precision. To test this hypothesis, we compared single-item utility elicitations for current health to computerized utility elicitations performed 1 to 2 weeks later.

Methods

STUDY POPULATION

This study was conducted as an adjunct to the Partners in Care (PIC) Patient Outcomes Research Team Study.[12] The PIC study approached consecutive adult patients in managed care primary care waiting rooms and asked them to complete (anonymously) a 10- to 15-minute paper-based survey that screened for common medical conditions and symptoms of depression, and it measured general health and utilities. In each practice, consecutive patients were screened for probable depressive disorder by *DSM-IV* criteria. Patients were eligible for enrollment in PIC if they were depressed, 18 years or older, expected to use the practice as a major source of care over the next year, did not have an acute medical emergency, spoke Spanish or English, and had eligible insurance (some fee-for-service patients were ineligible in two sites).

The screener for depression consisted of the "stem items" or major mood criteria for 12-month major depressive or dysthymic disorder from the Composite International Diagnostic Interview (CIDI), 12-month version, edition 2.1.[13] The screener included items assessing presence of depressed mood or loss of interest for a week or more in the past 30 days. Probable depression was defined as a positive response on any CIDI 12-month stem items, plus any 30-day depression. This criterion had a positive predictive value of 55% against the full CIDI diagnosis of 12-month depressive disorder.

In the past several months of screening at 3 study clinic groupings (larger clinics in Los Angeles, Columbia, and Annapolis, Maryland), enrolled patients were handed a brochure by study staff that invited them to participate in a 2nd computer study of their preferences. These sites were selected because of cost constraints and because the data collection had been completed at other PIC sites prior to fielding of the computer survey. Patients indicated their willingness to participate by notifying study staff members on site or by calling an 800 number. At these sites, 7572 patients were screened, 1072 patients were eligible and had a probable disorder, 875 enrolled in the PIC study,

and 149 participated in this substudy by completing an interview either at their clinic or (in Los Angeles) at RAND. We consider this a convenience sample of depressed primary care patients.

SINGLE-ITEM INSTRUMENTS

The PIC instrument included two items designed to measure patients' preferences for their current health state. These items were preceded by a Short Form 12 Health Assessment Questionnaire[14] (SF-12), the modified CIDI depression screener described above, and items ascertaining the number of active diseases patients believed they had at the time. The 1st preference elicitation item was a simple categorical scale rating subjects' current health on a scale where 0 was death and 100 was perfect health. Patients filled in boxes with their score on a machine-readable form. The 2nd item was a standard gamble utility assessment using the following question: "Imagine that you will live the next 10 years in your current state of health, both physical and mental, and there is a treatment that would give you perfect health or kill you immediately. What must be the chance that the treatment gives you perfect health before you would accept it?" Patients filled in one box on the form if they would not accept any risk or filled other boxes with the minimum probability of successful treatment they would be willing to accept.

COMPUTER INSTRUMENT

The computerized interview was administered 1 to 2 weeks after the paper-based questionnaire. It was also a largely self-administered questionnaire but was implemented using a laptop computer and hypertext markup language (HTML). The computer questionnaire used a recorded voice to provide subjects with instructions and also trained subjects who were unfamiliar with computers on how to use them prior to starting the survey. Research assistants supervised the process. After helping subjects during the training portions of the program, research assistants withdrew, when possible, while the computer administered the survey.

In the computer questionnaire, patients first completed the SF-12 health assessment questionnaire.[14] They then practiced using both a visual analog scale and a standard gamble to rate an example health state (paralysis below the waist). After this, they rated their current health, first with the visual analog scale and then with the standard gamble.

The standard gamble question was the following: "Think about your current health. Now, suppose that your doctor tells you about a treatment that will give you perfect health. Remember that perfect health is health that is as good as you can imagine for someone your own age. Your doctor tells you this treatment has a chance of giving you perfect health, but also has a risk of death. Think about what risk you would be willing to take to have perfect health. Would you take a 1% risk of death in order to have perfect health?" Subjects clicked on a button on the computer screen to indicate a positive or negative response. Depending on subject response, different risks were presented in a Ping-Pong search grid until the subject's indifference point was identified. A negative response to the 1% risk of death item was followed by a question asking participants if they would take a risk "between 0 and 1% in order to have perfect health"; if their response to this question was also negative, participants were asked to affirm that they "would not accept any risk of death in order to have perfect health."

The computer then checked the internal consistency of subjects' responses by comparing the ordering of preferences for the training state and current health across the two methods.[15] Participants with an inconsistent ordering had this pointed out by the computer and received computer assistance with repair of their preference ratings. After measuring utilities for current health, the survey assessed utilities for three hypothetical states. The results of assessments for hypothetical states are described elsewhere.[16] The computer protocol did not include a time trade-off assessment protocol, as we did not have time to develop and test both a time trade-off and a standard gamble elicitation protocol prior to fielding of the PIC study. Readers interested in viewing the exact protocol can access it over the World Wide Web at http://preferences.ucsd.edu.

DATA ANALYSIS

To compare results from the different assessment procedures, we used the Kruskal-Wallis test. We then examined the correlation among elicitations

using the Spearman coefficient. In addition, we closely examined the rates with which each procedure resulted in utility values of 1.0. We compared procedures to determine if anchoring on 1.0 with 1 method was associated with anchoring with another method. To better understand the meaning of responses of 1.0 to single-item questions, we performed logistic regression studies examining the relationship between health status and the number of reported diseases, as well as anchoring on a utility of 1.0. We then tested for associations between utility values as measured by each method and the number of active medical problems reported by each patient. All statistical tests were performed using JMP 3.1 on a Macintosh OS computer.

Table 1 Differences between the Partners in Care (PIC) Population and Patients Participating in the Preference Elicitation Study

	PIC Study Population	Computer vs. Paper Study
Number	795	149
Los Angeles clinic site (%)	65.4	7.8**
Major depression (%)	56	57
Black race (%)	29	2**
Other minorities (%)	15	17
Female gender (%)	71	68
High school education or less (%)	42	32**
Married (%)	57	52
Age (years)	43.7	44.7**
PCS	45.1	47.6*
MCS	35.1	33.7

Note: PCS = physical composite score; MCS = mental summary score.
*$P < 0.05$. **$P < 0.001$ vs. PIC population.

Results

Table 1 shows a comparison of the demographics of the volunteers participating in the computer study to the demographics of all PIC study enrollees at the test sites. Factors statistically associated with participation as a volunteer were older age, white race, and higher educational level. Most patients (95%) participating in the study had symptoms of acute depression, and 34% had a history of acute depressive symptoms consistent with dysthymia. The mean SF-12 physical score in patients participating in the study was 43.8 (*SD* = 11.0), and the mean mental health score was 37.6 (*SD* = 11.0.). SF-12 scores were consistent with the patient population having moderately impaired mental health and mildly impaired physical health. (The population norms for SF-12 scores are 50 with a standard deviation of 10.[17]) Patients reported a wide variety of other medical disorders regarding their current health, including musculoskeletal disease (67%), cardiovascular disease (35%), neurologic disease (31%), pulmonary disease (28%), endocrine disease (24%), mental health disorders other than depression (22%), gastrointestinal disease (19%), and cancer (4%).

Of the 149 enrollees, 23 patients (15%) had left standard gamble answers blank. Two patients (1.3%) could not complete the utility ratings for current health in the computer survey. One failure was due to the subject having difficulty understanding the materials, and the other was due to inability to access the studies database over the Internet to record results. In the computer survey, 19% of subjects initially had an internal inconsistency in the rank order of their preferences for their current health versus paralysis below the waist (the practice condition rating condition). Preference ratings were repaired successfully in all subjects with this error.

Both analog scale ratings and standard gamble utilities differed between the 2 survey methods. The paper-based survey that produced visual analog scale ratings was 0.08 units lower on average than the computer visual analog scale ratings ($P < 0.0001$, paired *t*-test). The mean utility value for the computer standard gamble was lower (0.78; 0.019 standard error) than the mean utility value for the single-item questionnaire (0.90; 0.020 standard error). Differences between standard gamble procedures were highly statistically significant ($P < 0.0001$, Kruskal-Wallis test).

All measurement methods had statistically significant correlation with other methods; however, the degree of correlation was only fair. Correlation between the computer and paper analog scale procedures was 0.47 (Spearman ρ). Correlation between the computer standard gamble and the single-item procedure was lower (ρ = 0.330).

An important difference in the psychometric properties of the computer and the paper instruments was the rate of persons assigning a utility of 1.0 to their health state. In the paper-based questionnaires, 53% of subjects assigned a utility of 1.0 in the standard gamble question. Only 15% of participants assigned a utility of 1.0 in the computer

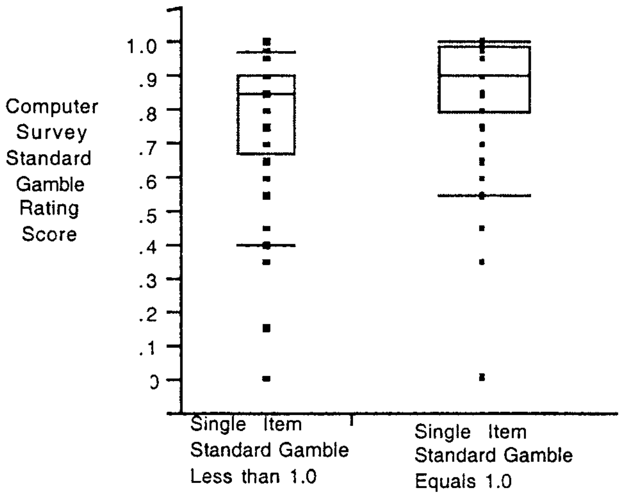

FIGURE 1. Distribution of responses of participants with utilities of 1.0 by either paper-based standard gamble or paper-based time trade-off. Boxes display the 25th to 75th percentile range with the bar at the median value. The outside bars mark the 10th and 90th percentiles.

questionnaire. Assigning a utility of 1.0 was associated with subjects' education level for the paper-based method (P = 0.002 for the paper standard gamble) but not for the computer method (P = 0.76). Subjects with a high school education level or less were more likely to report assign utilities of 1.0 in the paper-based approach. No other associations were observed between demographic factors (age, ethnicity, and gender) and anchoring on a utility of 1.0 for any of the methods.

The range of computer standard-gamble utilities for subjects who did and did not assign a utility value of 1.0 with paper methods is shown in Figure 1. As shown in the figure, there are statistically significant differences in computer utilities between subjects who did and who did not assign a utility of 1.0. This suggests that the utilities measured with single items do have some general correspondence with the computer elicitations. However, assigning a paper-based standard gamble or the time trade-off utility of 1.0 was associated with a wide range of computer standard gamble utilities (albeit somewhat narrower than not assigning a utility of 1.0). The median rating of persons on the computer instrument who had a utility of 1.0 on the paper-based instrument was 0.9. More than 90% had values that were less than 1.0.

We then examined the association between reported medical conditions and utilities. To simplify analyses, we grouped patients into rough quartiles based on the number of reported medical conditions: 0 conditions, 1 condition, 2 conditions, and 3 or more conditions. We then tested for associations between the reported disease burden and utilities as measured by each procedure. The results are shown in Figure 2. Utilities were not

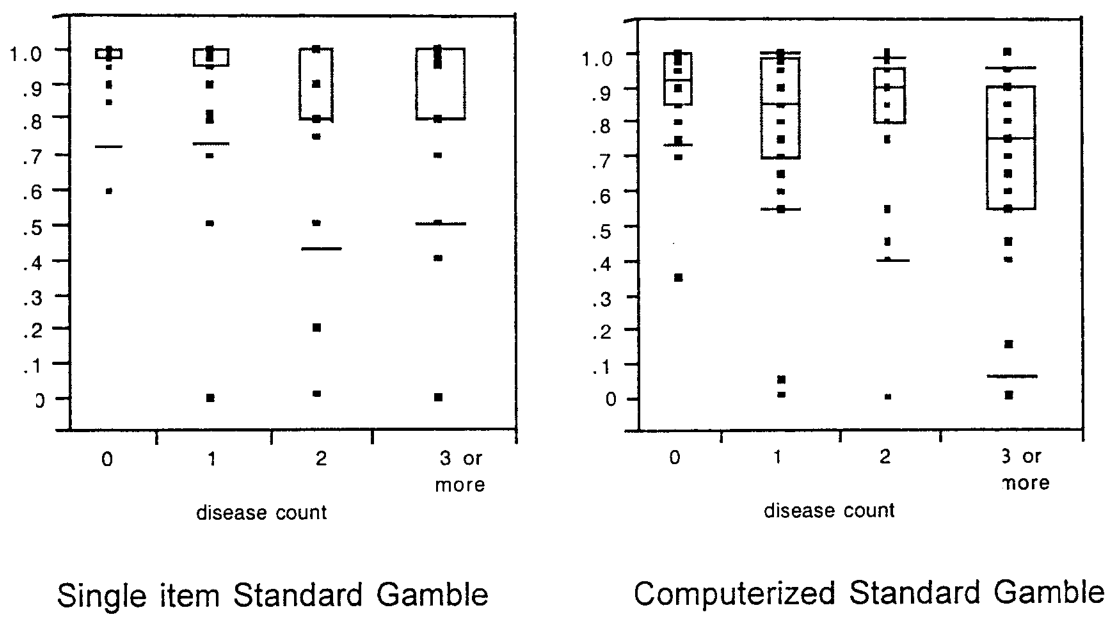

FIGURE 2. Distributions of utilities for patients with different numbers of self-reported diseases using the single-item gamble and the Ping-Pong (computerized) gamble. Boxes display the 25th to 75th percentile range with the bar at the median value (visible when less than 1.0). The outside bars mark the 10th and 90th percentile (visible when < 1.0). Distributions were statistically different (P < 0.001).

statistically different across different numbers of reported medical conditions for the paper-based procedure ($P = 0.780$, Kruskal-Wallis test). However, there was a highly statistically significant difference in utilities across numbers of reported diseases with the computer-based standard gamble method ($P < 0.001$).

Discussion

In this study, we compare results of standard gamble utility elicitation for participants' current health state performed using a single-item question to elicitation performed using a computer questionnaire implementing a Ping-Pong search procedure. The findings illustrate an important principal for investigators seeking to apply these methods: The results of elicitation using a simpler method for measurement cannot be assumed to differ from a complex method only in the amount of random error in the measurement. Use of different procedures for measurement may lead to an entirely different distribution of responses.

The sensitivity of results to the method of administration, to some extent, lessens the usefulness of utilities as "benchmark" measures of the significance of health impairments across diseases. Unless experiments use exactly the same procedures, specific numeric values may not be directly comparable. However, the results may be useful for relative comparisons. For example, a consistent finding across a number of studies using disparate methods is a relatively low utility for health states associated with depression compared with other illnesses.[12,17-19]

Both standard gamble procedures used in this study are reliable measures of "utility." Work by Sherbourne et al.[6] has shown that the single-item questions had a 1-day test-retest reliability of 0.67 for the standard gamble (Spearman correlation coefficient). The computer instrument used in this study had a 3-week test reliability (intraclass correlation coefficient of 0.83[20]). Both instruments appear to be sensitive to reported levels of disease burden; however, there is some difference in degree of responsiveness. Although in this study, the paper-based instrument was not sensitive to the number of comorbid conditions in this study, it was responsive in the larger (17,000 patients) PIC sample.[6,7]

The primary difference between results obtained using the different procedures was the rate of reporting utilities of 1.0. More than 50% of participants assigned ratings of 1.0 in the paper-based standard gamble and time trade-off questionnaires. Only about 16% of patients assigned utility values of 1.0 using the computerized standard gamble. Thus, the prediction that computer elicitation supported greater precision was confirmed. The differences in results of utility measurements may be attributable to any number of factors: use of visual displays, forced-choice questions versus open-ended questions, or error correction methods used in the computer elicitation survey. All of these elements probably encourage subjects to reason about their preferences with greater precision.

The utility elicitation procedures used different language for presenting trade-offs or "frames." The single-item standard gamble instrument asked patients to specify the minimum chance of survival they would accept for a cure of their medical problems. The computer instrument asked patients to specify the maximum risk of death they would accept. Previous work in this area by McNeil et al.[20] has shown that the frame for presentation of risks (chance of cure or risk of death) can influence subjects' willingness to accept risky treatments. Use of a positive frame (chance of survival) generally led subjects to accept greater risks in McNeil et al.'s experiments. The use of a positive frame for the 1-item standard gamble should have promoted increased willingness to take risks, and the negative frame of the computer gamble should have inhibited willingness to assume risk. Therefore, the finding of differences between the 2 methods cannot be explained on this basis.

The finding of a high rate of subjects responding with utilities of 1.0 with single-item questions is not unique to this study. Other studies using single-item paper-based methods have found similar rates of reporting of utilities of 1.0.[7] If subjects do have a greater tendency to anchor on 1.0 with single-item elicitations, measurements performed using single-item procedures would have different (smaller) scaling than measurements performed using more complex elicitation procedures. For example, as shown in Figure 2, the difference in utilities for patients with 1 reported disease and 3 diseases was much smaller in the paper-based survey than the computer survey. The smaller scaling of single-item instruments has implications

for the use of utilities derived from these measures in cost-utility analysis. Analysts might need to adjust the threshold value for the cost-utility ratios for results to be comparable with other studies (e.g., the "cutoff" threshold value for a "cost-effective treatment" might need to be increased from $50,000 per quality-adjusted life year [QALY] gained to perhaps $75,000 per QALY or $100,000). Obviously, this would complicate the interpretation of findings from such an analysis.

The present study examines utilities for participants' current health. Previously published studies comparing utilities for hypothetical states that used the simple and complex procedures show conflicting results. In Albertsen et al.,[3] researchers used a within-subjects design to compare utilities for hypothetical states with prostate cancer that were elicited using a self-administered questionnaire and a computer program. They found no effect on the population mean. In Dolan et al.[21] and Dolan and Sutton,[22] researchers used a randomized design to compare preferences for states based on the Eq-5-D health status model. Participants completed a self-administered booklet or a face-to-face interview in which a research assistant measured utilities assisted by visual props. Results showed large differences median and distribution of utilities between the 2 methods of administration.[21] Subsequent analyses found differences in the relationship between analog scale and standard gamble ratings[22] between the 2 groups, suggesting fundamental differences in the structure of preferences.

Limitations

This study does not explore what specific elements of a utility elicitation produce changes in the distribution of responses. Its purpose was primarily to compare procedures representative of a growing body of work using single-item paper-based instruments with the results of a computer interview designed to foster quantitative reasoning. The study used a sequential design in which all subjects completed a paper-based procedure assessment prior to the use of the computer instrument. This design assumes that participation in the initial paper-based survey did not systematically change patients' values in some way and thus cause apparent differences in utilities. The delay of 1 to 2 weeks or more between use of the paper-based instrument and the computer instrument should have minimized this effect.

A further limitation is that the study was conducted in a group of patients with symptoms of depression. Depression can affect cognition.[23] This might influence patients' reasoning so that they more frequently think categorically, resulting in a greater frequency of anchoring on 1.0.

Conclusions

Utility elicitations performed using different procedures may each yield "valid" representations of preferences by all standard criteria but may not be directly comparable. Although there is no gold standard for preference measurement, in this survey, many of the persons who responded with answers suggesting utilities of 1.0 in single-item standard gamble questionnaires had other lower utilities when utilities were assessed by computer interview. This is not to say that single-item standard gamble measures do not capture the aspects of preferences of a population. However, a substantial portion of participants may not participate in gambles. This would result in smaller differences in utility between healthy and ill states and less sensitivity to detect changes in the health of populations.

Preparation of this manuscript was supported in part by grants to the last author from the National Science Foundation (SBR973-0143), the National Institutes of Health (NIH 316201 and P50AT00008), the Academic Medicine and Managed Care Forum (Integrated Therapeutics Group/Schering Plough, Inc.), and the United States Department of Commerce (SUB005).

References

1. Bennett KJ, Torrance GW. Measuring health preferences and utilities: rating scale, time trade-off and standard gamble methods. In: Spliker B (ed). Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia: Lippincott-Raven, 1996;235-265.
2. Slovic P. The construction of preference. American Psychologist. 1995;50(5):364-371.
3. Albertsen PC, Nease RF Jr, Potosky AL. Assessment of patient preferences among men with prostate cancer [see comments]. J Urol. 1998;159:158-163.
4. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ. 1996;5:141-154.

5. Lundberg L, Johannesson M, Isacson DG, Borgquist L. The relationship between health-state utilities and the SF-12 in a general population. Med Decis Making. 1999;19:128-140.
6. Sherbourne CD, Keeler E, Unutzer J, Lenert L, Wells KB. Relationship between age and patients' current health state preferences. Gerontologist. 1999;39:271-278.
7. Wells KB, Sherbourne CD. Functioning and utility for current health of patients with depression or chronic medical conditions in managed, primary care practices. Arch Gen Psychiatry. 1999;56:897-904.
8. Tversky A, Sattath S, Slovic P. Contingent weighting in judgement and choice. Psychological Review. 1988;95:371-384.
9. Lenert LA, Cher DJ, Goldstein MK, Bergen MR, Garber A. The effect of search procedures on utility elicitations. Med Decis Making. 1998;18:76-83.
10. Reyna VF, Brainerd CJ. Fuzzy trace theory: an interim synthesis. Learning and Individual Differences. 1995;7:1-75.
11. Reyna VF, Brainerd CJ. Fuzzy-trace theory and framing effects in choice: gist extraction, truncation, and conversion. Journal of Behavioral Decision Making. 1991;4:249-262.
12. Wells KB. The design of Partners in Care: evaluating the cost-effectiveness of improving care for depression in primary care [in process citation]. Soc Psychiatry Psychiatr Epidemiol. 1999;34:20-29.
13. World Health Organization. Composite International Diagnostic Interview (CIDI): Interviewer's Manual. Geneva: World Health Organization, 1993;103.
14. Ware JJ, Kosinski M, Keller S. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Medical Care. 1996;34:220-233.
15. Lenert LA, Treadwell J. Effect of failure to maintain procedural invariance on utility elicitations. Medical Decision Making. 1999;19:473-481.
16. Lenert LA, Treadwell JR, Schwartz CE. Associations between health status and utilities implications for policy. Med Care. 1999;37:479-489.
17. Ware J, Kosinski M, Keller S. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. Boston: The Health Institute, New England Medical Center, 1995.
18. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. J Chronic Dis. 1978;31:697-704.
19. Fryback DG, Dasbach 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.
20. McNeil BJ, Pauker SG, Sox HC Jr, Tversky A. On the elicitation of preferences for alternative therapies. N Engl J Med. 1982;306:1259-1262.
21. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. J Health Econ. 1996;15:209-231.
22. Dolan P, Sutton M. Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values. Soc Sci Med. 1997;44:1519-1530.
23. Silberman EK, Weingartner H, Post RM. Thinking disorder in depression. Arch Gen Psychiatry. 1983;40:775-480.