What Should Be Reported in a Methods Section on Utility Assessment?

PEEP F. M. STALMEIER, PhD, M. K. GOLDSTEIN, MD, A. M. HOLMES, PhD, L. LENERT, MD, MS, J. MIYAMOTO, PhD, A. M. STIGGELBOUT, PhD, G. W. TORRANCE, PhD, J. TSEVAT, MD, MPH

Background. The measurement of utilities, or preferences, for health states may be affected by the technique used. Unfortunately, in papers reporting utilities, it is often difficult to infer how the utility measurement was carried out. ***Purpose***. To present a list of components that, when described, provide sufficient detail of the utility assessment. ***Methods***. An initial list was prepared by one of the authors. A panel of 8 experts was formed to add additional components. The components were drawn from 6 clusters that focus on the design of the study, the administration procedure, the health state descriptions, the description of the utility assessment method, the description of the indifference procedure, and the use of visual aids or software programs. The list was updated and redistributed among a total of 14 experts, and the components were judged for their importance of being mentioned in a Methods section. ***Results***. More than 40 components were generated. Ten components were identified as necessary to include even in an article not focusing on utility measurement: how utility questions were administered, how health states were described, which utility assessment method(s) was used, the response and completion rates, specification of the duration of the health states, which software program (if any) was used, the description of the worst health state (lower anchor of the scale), whether a matching or choice indifference search procedure was used, when the assessment was conducted relative to treatment, and which (if any) visual aids were used. The interjudge reliability was satisfactory (Cronbach's alpha = 0.85). ***Discussion***. The list of components important for utility papers may be used in various ways, for instance, as a checklist while writing, reviewing, or reading a Methods section or while designing experiments. Guidelines are provided for a few components. ***Key words:*** utility assessment; cost-utility assessment. (**Med Decis Making 2001;21:200–207)**

Health utility measures, such as the visual analogue scale, time trade-off, standard gamble, person trade-off, Health Utilities Index, EuroQol EQ-5D, and quality of well-being are widely used in health care studies.[1–4] There are many different ways to implement those methods. After reading through the Methods section of a paper, it is often hard to determine what exactly was done. Lack of clarity in the methods is problematic because the implementation of the instrument is thought to affect the feasibility, reliability, and validity of the measures.[1,3,5–8]

Our aim was to specify those components of utility assessment that, when described in the Methods section, provide a sufficiently detailed account of the assessment procedure to enable readers to assess the quality of the study and

Received 22 May 2000 from the Medical Psychology, Academic Medical Center, Amsterdam, and the Joint Center for Radiation Oncology Arnhem-Nijmegen (RADIAN), University Medical Center Nijmegen, and the Nijmegen Institute for Cognition and Information (NICI), Nijmegen, the Netherlands (PFMS); VA Palo Alto Health Care System and Department of Medicine, Stanford University School of Medicine, Stanford, California (MKG); School of Public and Environmental Affairs, Indiana University–Purdue University, Indianapolis, Indiana (AMH); Department of Medicine, Veterans Affairs Healthcare System and the University of California, San Diego (LL); Department of Psychology, University of Washington, Seattle (JM); Department of Medical Decision Making, Leiden University Medical Center, the Netherlands (AMS); Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada, Innovus Research, Inc., Burlington, Ontario, Canada, and Health Utilities, Inc., Dundas, Ontario, Canada (GWT); Section of Outcomes Research, Division of General Internal Medicine, University of Cincinnati Medical Center, Cincinnati, Ohio (JT). Revision accepted for publication 10 October 2000. Financial support of this study was provided partly by a grant from the Dutch Cancer Society, Amsterdam, the Netherlands (project #NUKC 97-1585). The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. The following author is employed by the sponsor: PFMS. The opinions expressed are those of the authors and do not necessarily reflect the opinions of other affiliated institutions.

Address correspondence and reprint requests to Dr. Stalmeier: NICI, P.O. Box 9104, 6500 HE Nijmegen, the Netherlands; e-mail: stalmeier@nici.kun.nl.

evaluate its results in the context of the methods used. We did not attempt to prescribe how to conduct utility measurement in health care settings. Our main aim was simply to present a consensus on issues that are important to describe when utilities are assessed in health care.

One way to specify important components is through reviewing the literature. A review should yield components that exert a sizeable effect on the measured utilities. A problem with a systematic review is that quite often Methods sections offer insufficient detail; this is precisely what triggered this study. Therefore, we chose the approach of querying a panel of experts.

In a broad sense, utility measurement can only be sound if the whole research protocol is sound. For instance, sampling and recruitment methods influence the results of utility assessment.[9] Because such issues are germane to all types of clinical research, however, we do not address them here. Testing of the axioms of expected utility was also considered to be outside the scope of this report because it is not essential for papers in the general health care literature. Thus, the panel decided to focus on the description of issues specific to utility assessment.

The survey of experts was undertaken with direct utility assessment methods, such as the time trade-off and the standard gamble, in mind. However, most of the components identified below are also relevant to other methods, for instance, the visual analogue scale and person trade-off.

Our resulting list of components is intended to be relevant for utility assessment in clinical settings. The recommendations reported here are also relevant to papers reporting on utility assessments to estimate standard weights (scoring functions) for multiattribute health indices such as the Health Utilities Index or EuroQol EQ-5D. However, in clinical applications, such indices are often used for routine assessment and monitoring of patients or for assessing and describing patients' health-related quality of life.† Our recommendations are largely irrelevant for such clinical applications because they never employ a direct utility assessment but instead estimate societal preference weights indirectly through applying standard multiattribute scoring functions.

† Of course, standard weights are not useful if the clinical setting is a bedside decision making for individual patients.

Method

PROCEDURE

The project was initiated through an announcement call on the e-mail network of the Utilities Interest Group of the Society for Medical Decision Making. This network contained about 23 entries. Six experts agreed to join the project and 2 additional experts were enlisted at the Utilities Interest Group meeting of the 1998 Medical Decision Making conference. Thus, 8 experts, the authors, joined the project. These experts from economics, psychology, health sciences, information sciences, decision sciences, and clinical medicine investigate fundamental and applied issues in utility measurement.

An initial list of 21 potentially important components was prepared by one of the authors (PFMS). The components were organized into 6 clusters that focus on the design of the study, the administration procedure, the health state descriptions, the description of the utility assessment method, the description of the indifference search procedure, and the use of visual aids or software programs. The list was distributed among the group along with a solicitation for additional components. All communications were done via e-mail. Almost all experts provided additional components, and 18 items were added for a total of 39.

To reach consensus on which issues were most important, the list was redistributed and the 8 experts rated all items for importance. A 100-point rating scale with anchors ranging from *not at all important* to *very important* was used to rate each component. It was emphasized that the rating should reflect the importance of mentioning a particular component in a Methods section and not, for instance, the desirability of including that component in utility assessment or the correctness of a particular implementation.

The experts' ratings were to be given keeping in mind a detailed Methods section in a paper focusing on utility assessment. The experts were also asked to indicate 10 components that should always be included in a 3-sentence summary of the utility assessment method for a general readership.

After the ratings were collected, we felt that additional experts should be given the opportunity to inspect our list to add and judge items. Seven experts were invited, of whom 6 agreed to participate. These experts found the list adequate, and the existing items were rated. During the preparation of the final report, 3 new items emerged and were rated independently. Thus, the final list contained 42 items.

ANALYSES

Descriptive statistics (means and standard deviations) were calculated for ratings of all 42 items. The remaining analyses involved only the first 39 items, which were judged simultaneously. To assess interjudge reliability, we calculated Cronbach's alpha across the judges. To account for disparity in the ranges of ratings, the averaged ratings were compared with the averaged z scores of the ratings; these 2 statistics were strongly related (Pearson's $r = 0.995$), so only the averaged ratings are presented.

Because the additional panel of 6 experts joined the project later on, the agreement of their ratings with ratings from the first 8 experts was assessed by correlating the averaged ratings in each group.

Results and Discussion

Ratings of 1 expert were omitted because more than 50% of the rating data were missing. Out of the remaining ratings from 13 experts, 12 ratings from 6 experts were missing and imputed from the mean values of the remaining experts. All 14 experts provided their statement of components that they felt should always be included in a 3-sentence summary.

WHAT SHOULD BE REPORTED IN A METHODS SECTION FOCUSING ON UTILITY ASSESSMENT?

Below, we present the list of components with their associated ratings. The components are preceded by an identification code listed in boldface and followed by the associated mean importance rating (range = 0 to 100) and standard deviation. When necessary, the components are explained in more detail. This explanation is usually purely descriptive; in some cases (C1, C5, C7, C9, C10, D4, and F1) guidelines have been offered. Three components (B1e, C13, and D4b) were added as amendments. Each of these 3 components was rated in isolation by 14, 14, and 13 experts, respectively. The interjudge reliability (Cronbach's alpha) of the importance ratings was 0.85. The agreement between the averaged ratings of the first 8 experts and the additional 6 experts was good ($r = 0.63$).

A. Design

A1. When exactly were utility measurements done, for example, before or after a medical treatment? 76 (26)

A2. The timing of utility assessment relative to other questionnaires. 49 (25)

Often, additional questionnaires for quality of life, costs, or psychological assessment are employed. Component A2 describes the timing of the utility assessment relative to these other questionnaires.

B. Administration

B1. How were the utility questions administered (e.g., by interviewer, mailed questionnaires, computer, the Internet, or self-administered under general supervision)? 92 (17)

If by interview,

- **a**. What was the interview setting (e.g., face-to-face, by telephone, or in the hospital)? 78 (20)
- **b**. Where were the utility measurements done (city)? 43 (23)
- **c**. How were interviewers trained? 58 (23)
- **d**. How was between-interviewer reliability assessed? 59 (21)
- **e**. What was the interview duration?‡ 38 (24)

B2. Response and completion rates. 85 (22)

B3. Efforts (if any) to increase response or completion rates. 50 (20)

C. Health State Descriptions

C1. Description of health states, if any. 90 (11)

‡ This item was rated independently.

In a decision tree, health states correspond to outcomes or Markov states. It is desirable to show 1 or 2 representative health state descriptions in a table or appendix. Issues to consider are the following: first, second, or third person; which dimensions of health were described; if applicable, characteristics of actors such as age, race, and gender used in any multimedia presentations; and whether a narrative or point-form format was used in the written descriptions.[10]

C2. How is "perfect health" described? 78 (14)

Items C2 and C3 refer to the anchors used in the valuation task. The concept of perfect health for the upper anchor may not necessarily be absolute. For instance, when utilities are obtained from the elderly, the description of perfect health could pertain to a young person or to a similarly aged person.

C3. How is "worst health" described? 87 (15)

Usually, "dead" is used as the lower anchor, but often "worst imaginable health" is used. Sometimes, the subject is asked to select the worst state from a list of states that may include dead. Subsequently, this state is used as the lower anchor. It should be clear for which states the utilities are set to 0 and 1.

- **a**. If "dead" was used, how immediate was death? 60 (31)

Usually, death is described as instantaneous, but in some implementations death is postponed for 1 week or even longer (see also C9, below).

C4. If utility for "own health" was assessed, was own health specified further? 70 (22)

This specification could, for instance, include physical, psychological, and social well-being. In some implementations, respondents first have to describe their own health by selecting levels in a multiattribute system like the Health Utilities Index or the EuroQol EQ-5D.

C5. Was the presentation order of the health states randomized? If not, what was the order? 67 (22)

Randomization or balancing are standard experimental precautions used to guard against order effects. In general, it is desirable to use randomization or balancing unless other reasons (feasibility, for instance) go against it.

C6. Were fixed survival durations used for the health states? 90 (11)

- **a**. If so, what duration(s) was used? 91 (8)

C7. How were the survival durations characterized (e.g., a survival outcome might be described as *x* years of survival followed by death or as a life expectancy of *x* years of survival)? 68 (20)

The panel felt that presenting survival durations as life expectancies, without mentioning "followed by death," is undesirable.

C8. How was the fixed survival duration chosen (e.g., from life tables of the general population or data from studies on a particular disease)? 62 (21)

C9. Was it made explicit that each duration was followed by death? 72 (20)

This concept is related to C7, above. In a decision-making context, one measures the utility of being in a particular disease (health state) in the steady state. Therefore, death can have the following qualities: quick, painless, costless (like dying in one's sleep), unpredicted, and unrelated to the health state described. The interviewer should keep these issues in mind and correct possible misunderstandings.

C10. Was the subject instructed to assume that survival does not occur with knowledge of the date of death? 58 (23)

Respondents may believe that even 3 years of survival would be worth a lot because one can pack many valuable experiences into 3 years of life. The problem here is that the value of 3 years has been inflated by the incorrect assumption that "3 years of survival" means 3 years with the knowledge that one has only 3 years. The panel feels that researchers should at least be aware of this possible misunderstanding and correct it when it crops up.

C11. Description of treatments, if any (a treatment corresponds to a decision option in a decision tree). 58 (22)

C12. What was the subject instructed to assume, if anything, regarding costs to him or her or family about the possible outcomes? 48 (29)

C13. Was the health state labeled or unlabeled?§ 85 (13)

To clarify this issue, suppose one prepares 2 health descriptions for "lung cancer." The first description simply states the physical, psychological, and social issues without reference to lung cancer per se. The second description is identical to the first, but the label "lung cancer" is added on top to identify the health state. Such a label may affect the valuation.[11]

D. Description of the Utility Assessment Method

D1. Which method was chosen (e.g., visual analogue scale, time trade-off, standard gamble, willingness to pay)? 95 (14)

D2. If more than 1 utility measure was used, was the presentation order randomized? If not, what was the order? 77 (22)

D3. How was the choice introduced? 77 (17)

Several formats are available: "you look into a crystal ball," "you make choices for a good friend," "suppose you may choose between . . . ." Sometimes, researchers instruct respondents to imagine they are in the state to be evaluated, whereas others present a choice between options without saying anything on this issue. Every implementation induces particular framing effects:[12] for instance, when the time trade-off method is used, some implementations explicitly state that "the choice is between 10 years in health state *Q* and 8 years in good health, that is, giving up 2 years of good health." Through adding "giving up 2 years of good health," attention may be shifted from a benefit, 8 years of good health, to a loss, giving up 2 years of good health.

D4. If more than 1 health state was assessed, were they rank ordered first?¶ 77 (13)

Rank ordering of health states including death is generally recommended as a standard procedure at the beginning of the interview. It familiarizes the respondents with the health states, builds a foundation for later valuation tasks, and alerts the interviewer that some health states could be considered to be worse than dead.

- **a**. If so, was death included in the ordering procedure? 67 (30)
- **b**. If there were states worse than dead, how were they handled? 79 (25)

Some researchers set the utilities of such states to 0. Others pursue the assessment of negative utilities. Because many utility studies do not ask about states' being worse than dead, it is not an issue in most utility reports. Where it is an issue, the handling of states worse than dead should be described.

D5. Were subjects confronted with inconsistencies in their scores, such as a change in the health state ordering as inferred from the different utility assessment methods? 68 (30)

E. Indifference Procedures

E1. Was a matching or choice indifference search procedure used? 88 (9)

Matching consists of asking for a 1-shot (single) indifference response, for example, "How many years in good health do you consider to be

§ This item was rated independently.

¶ This item was rated independently.

equivalent to 10 years in your actual health?" The choice procedure employs several choices (iterations) to narrow in on the indifference point. Matching and choice can yield different attribute weights.[7,13,14]

In the case of choice,
a. What were the first 2 choices? 65 (19)

In some applications, the first 2 choices involve the upper and lower anchors of the scale. The lower end of that range may be a 100% probability of death, or 0 years in good health; the upper end may be a 100% probability of perfect health, or the maximum duration in perfect health. Misunderstandings on the part of the respondent can be detected by starting at these extremes, for instance, when the subject prefers his own health to a 100% probability of perfect health. Nevertheless, in some applications the lower end is avoided so as not to frighten the patient with the prospect of death.

b. Which particular indifference search procedure was used? 79 (16)

A variety of indifference search procedures are available: the Ping-Pong method, the bisection method, a method using a random starting point followed by bisection, starting from the lower anchor and going up (titrating up), starting from the upper anchor and going down (titrating down), or using the best utility estimate as the starting point. Each procedure may induce its own biases.[15]

c. What were the criteria for terminating the indifference search procedure (e.g., terminating when the range is narrowed to 10%)? 75 (18)
d. Did the subject give a final guess within the narrowed-down indifference range? 62 (22)

F. Visual Aids and Software Programs

F1. Which software program (if any) was used? 83 (21)

Currently, a number of computerized utility assessment programs are available, for instance, U-Titer,[16] U-Maker,[17] iMPACT,[18,19] and Gambler.[20] If such a program is used, the paper describing the program should be cited whenever available.

If a software program was used,
a. Was it used by the subject alone, or was someone present in the start-up phase to answer questions or detect misconceptions? 78 (18)
b. If someone was present, how was he or she trained? 54 (23)

F2. What visual aids, if any, were used (e.g., rulers, pies, probability wheels, or other means of visualizing probabilities or trade-offs)? 80 (21)

F3. Were any aspects of the interview controlled by computer? 66 (23)

Sometimes, an aspect of the utility assessment is preprogrammed into the computer. For instance, the computer may control the next choice to be presented in the indifference procedure or the presentation order of the health states.

WHAT SHOULD GO INTO A SHORT DESCRIPTION?

The experts were also asked to indicate those 10 components that should never be omitted, even in a brief (e.g., 3-sentence) summary. A total of 145 selections were made by the 14 experts, with the top 10 components selected a total of 100 times (see Table 1). The agreement between the rankings implied by the percentages and ratings was excellent. Eight of the 10 most highly rated components described in the complete component list also appear in Table 1. Only item A1 concerning the timing of utility assessment received a relatively low rating, perhaps because it was the first component to be rated. Other exceptions included item C13 on the labeling of the health state and item D4b on the handling of states worse than dead, which were not considered in this analysis because they were added later as amendments. The correlation between the percentages and averaged ratings in Table 1 is high (0.73, $P < 0.01$), especially if one considers that the correlation is calculated over a restricted range of ratings and percentages.

Table 1 • Components to Be Included in a 3-Sentence Summary of a Utility Assessment Procedure

Component (item number)	Endorsing (%)	Mean Rating
How were utility questions administered? (B1)	93	92
Health state description? (C1)	86	90
Which utility assessment method? (D1)	86	95
Response and completion rates? (B2)	79	85
Use and specification of fixed duration? (C6 and C6a)	79	91
Which software program (if any) was used? (F1)	71	83
Description of worst health state? (C3)	57	87
Matching versus choice indifference procedure? (E1)	57	88
When was assessment done relative to treatment? (A1)	57	76
Which visual aids (if any) were used? (F2)	50	80

General Discussion

By querying a panel of 14 experts, we obtained a list of components considered important to mention in a Methods section on utility assessment. Although it was not our aim to present guidelines, we have given a few recommendations. These recommendations were unanimously agreed on by the panel.

The list of components may be used by researchers in different ways. The primary purpose of the list is for use as a checklist by authors writing a Methods section on utility assessment. Readers and referees may use it to determine whether potentially important components were left out. Thus, we hope that future studies will present a more detailed picture of the utility assessment procedure. Second, describing these components in the Methods section will facilitate the conduct of systematic reviews on utility measurement by providing information on a number of important elements of utility assessment. Third, the list may be used by researchers when designing their experiments because all these components should be dealt with when planning the utility assessment procedure. Fourth, the list may be used to identify topics for further research; for instance, little is known about how the description of the worst health state affects utility measurement.

The ratings and rankings reflect the current state of the science of utility analysis in a health care setting. The relative importance of these items might shift depending on the specific application or as the science of utility assessment progresses. For instance, when utilities are compared across studies, it is important to be able to compare how the utilities were assessed in each of the studies. If several utility assessment techniques are compared within a single study, however, it is less important to describe all of the components as long as they are held constant across techniques.

WHAT SHOULD GO INTO A SHORT DESCRIPTION?

The components that should always be included in even a short summary are listed in Table 1. Some components, for instance, “how were the utility questions administered?” and “response and completion rates,” are clearly relevant for judging the validity of the underlying study. The other components can affect the utilities themselves. An elaborate review of the underlying evidence of these effects is beyond the scope of this article.

Mentioning the components in Table 1 should conform to the space limitations for papers that do not primarily focus on utility measurement. These components can often be covered in 2 sentences: one describing the respondents, the sample size, and the completion rate; and the other describing the utility assessment method (standard gamble, time trade-off, willingness to pay, etc.), the indifference search procedure (matching-based or choice-based), how utilities were collected (interview, telephone, etc.), the visual aids or software, and the durations used for the health states. A 3rd sentence might be needed to list the health states measured, if that is not clear from the rest of the paper. (Alternatively, the completion rates can be placed in the 1st paragraph of the Results section or in a table, and the health state description can be put in an appendix.)

The panel discussed whether it is really mandatory to report all items in Table 1, as the final 6 items will interest only specialists in utility measurement and not a general medical readership. From our data, several inferences are possible. Some experts believed that the top 4 (or 5) could be considered “necessary” and that the rest are “highly recommended.” All experts agreed, however, that all 10 items are very valuable for understanding the utility methods used and assessing the validity of the utilities. Therefore, information regarding all 10 items should be published somewhere, either in a separate methods paper or in an appendix.

LIMITATIONS

We have focused on the Methods section from the point of view of utility assessment. As delineated in the introduction, other study aspects, for example, sample characteristics, remain important. We also considered whether tests of expected utility assumptions (e.g., substitutability axiom), quality adjusted life years assumptions (e.g., constant proportional trade-off), or psychometric properties (e.g., test-retest reliability) should be mentioned in a Methods section. Again, the panel decided that this would be outside the scope of our report, which focuses on the description of the utility assessment method. Of course, in a methodological paper, such tests should be mentioned whenever relevant.

With respect to our analyses, one could argue that the ratings should be treated as non-independent because the experts interacted with each other when new items were added to the list. Still, each expert rated all items independently, blinded to the ratings of the others. Therefore, treating the ratings as independent is reasonable.

Another possible limitation is the small sample of 14 experts and the convenience nature of the sample. We know of no enumerated list of investigators with expertise in utility assessment in health care from which a random sample could be drawn. The sample size may affect the precision of the ratings in the complete component list and the rankings in Table 1. We showed, however, that the agreement between the ratings and rankings is excellent. We do not believe that adding more experts would affect the recommendations as the correlation between the averaged ratings of the initial group of 8 experts and the additional group of 6 experts was good ($r = 0.63$).

In summary, we believe that these ratings can serve as guidelines for Methods sections for papers reporting results of direct utility assessment.

The authors thank the additional panel of 6 experts who participated in this study: P Dolan, PhD, Sheffield Health Economics Group and Department of Economics, Sheffield, U.K.; D Feeny, PhD, Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, Alberta, Canada, and Institute of Health Economics, Alberta, Canada; DG Fryback, PhD, Department of Preventive Medicine, University of Wisconsin–Madison; P Kind, PhD, Center for Health Economics, University of York, Heslington, U.K.; DL Patrick, MSPH, PhD, University of Washington–Seattle; and JE Siegel, ScD, Arlington Health Foundation, Virginia.

References

1. Froberg DG, Kane RL. Methodology for measuring health-state preferences-II: scaling methods. J Clin Epidemiol. 1989;42:459–71.
2. Nord E. Methods for quality adjustment of life years. Soc Sci Med. 1992;34:559–69.
3. MR Gold, JE Siegel, LB Russell, MC Weinstein, eds. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
4. Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 2nd ed. Oxford: Oxford University Press, 1997.
5. Birnbaum MH. Issues in utility measurement. Org Behav Hum Dec Proc. 1992;52:319-30.
6. Camerer C. Individual decision making. In: Kagel JH, Roth AE, eds. Handbook of Experimental Economics. Princeton, NJ: Princeton University Press, 1995.
7. Borcherding K, Schmeer S, Weber M. Biases in multiattribute weight elicitation. In: Caverni J-P, Bar-Hillel M, Hutton Barron F, Jungermann H, eds. Contributions to Decision Making–I. Amsterdam: Elsevier, 1995, p 3–28.
8. Baron J. Biases in the quantitative measurement of values for public decisions. Psych Bull. 1997;122:72–88.
9. Boyd NF, Sutherland HJ, Heasman KZ, Tritchler DL, Cummings BJ. Whose utilities for decision analysis? Med Decis Making. 1990;10:58–67.
10. Llewellyn-Thomas HA, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. The measurement of patients' values in medicine. Med Decis Making. 1982;2:449–62.
11. Gerard K, Dobson M, Hall J. Framing and labelling effects in health descriptions: quality adjusted life years for treatment of breast cancer. J Clin Epidemiol. 1993;46:77–84.
12. Kahneman D, Tversky A. Prospect theory: an analysis of decision making under risk. Econometrica. 1979;47:263–91.
13. Tversky A, Sattath S, Slovic P. Contingent weighting in judgement and choice. Psych Rev. 1988;95:371–84.
14. Hawkins SA. Information processing strategies in riskless preference reversals: the prominence effect. Org Behav Hum Dec Proc. 1994;59:1–26.
15. 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.
16. Sumner W, Nease RF, Littenberg B. Status report on automated utility assessment with U-Titer. Med Decis Making. 1993;13:390.
17. Frank Sonnenberg, New Brunswick (NJ).
18. Lenert L. The reliability and internal consistency of an Internet program for measuring utilities. Qual Life Res. Forthcoming.
19. Lenert LA, Michelson D, Flowers C, Bergen MR. Impact: an object-oriented graphical environment for construction of multimedia patient interviewing software. Proc Annu Symp Comput Appl Med Care [online]. 1995:319–23. Available from: http://prefdev.ucsd.edu.
20. Gonzalez E, Eckman MH, Pauker SG. "Gambler": a computer workstation for patient utility assessment. Med Decis Making. 1992;12:350.