International Journal of Technology Assessment in Health Care, **21:4** (2005), 511–516.

A test of prospect theory

David Feeny
University of Alberta and *Institute of Health Economics*

Ken Eng
Institute of Health Economics

Objectives: Prospect theory (PT) hypothesizes that people judge states relative to a reference point, usually assumed to be their current health. States better than the reference point are valued on a concave portion of the utility function; worse states are valued on a convex portion. Using prospectively collected utility scores, the objective is to test empirically implications of PT.
Methods: Osteoarthritis (OA) patients undergoing total hip arthroplasty periodically provided standard gamble scores for three OA hypothetical states describing mild, moderate, and severe OA as well as their subjectively defined current state (SDCS). Our hypothesis was that most patients improved between the pre- and postsurgery assessments. According to PT, scores for hypothetical states previously $>$ SDCS but now $<$ SDCS should be lower at the postsurgery assessment.
Results: Fourteen patients met the criteria for testing the hypothesis. Predictions were confirmed for 0 patients; there was no change or mixed results for 6 patients (42.9 percent); and scores moved in the direction opposite to that predicted by PT for 8 patients (57.1 percent).
Conclusions: In general, the direction and magnitude of the changes in hypothetical-state scores do not conform to the predictions of PT.

Keywords: Prospect theory, Standard gamble, Utility, Total hip arthroplasty

Several investigators have noted that patients with experience with a health state often provide higher utility scores for that state than do members of the general population for whom the state is hypothetical (1;9;23). Prominent (and not mutually exclusive) explanations for this phenomenon include adaptation and the implications of prospect theory.

The rationale for the adaptation explanation is that patients who have experienced the health state, especially if they have experienced the state for some time, learn to cope, make adjustments, and learn to use aids and devices that ameliorate the impact and burden. The adaptation rationale can also include the process by which patients reduce their expectations or redefine what they consider to be normal or acceptable to accommodate their diminished capacity. The latter rationale is consistent with recent literature on response shift (see for instance references 18;19).

According to prospect theory from psychology and economics (10; see also 8;14;25), the systematic difference in

Conflict of Interest. It should be noted that David Feeny has a proprietary interest in Health Utilities Incorporated, Dundas, Ontario, Canada. HUInc. owns the copyright to and distributes HUI materials.
The authors acknowledge the contributions of the interviewer in the study on health-related quality while waiting for and undergoing elective total hip arthroplasty, Susan Webster-Bogaert, whose insights and observations led to the analyses reported in this paper. The authors also acknowledge the contributions of other investigators on that study which generated the data used for the analyses reported here, including Dr. Jeffrey L. Mahon (principal investigator), Dr. Robert Bourne, Dr. Cecil Rorabeck, and Larry Stitt. Furthermore, the authors gratefully acknowledge the help of the patients and surgeons who participated in the study (London, Ontario, orthopedic surgeons Drs. Harvey Bailey, David Chess, Wayne Grainger, Paul Kim, and Richard McCalden). The authors also acknowledge helpful comments made by George. W. Torrance.
The study "The Effect of Waiting for Elective Hip Arthroplasty on Health-Related Quality of Life" was supported by a grant from Physician Services Incorporated (PSI) of Ontario to Dr. Jeffrey Mahon (grant no. 94-30). The analyses reported in this paper were supported by grants from the Alberta Heritage Foundation for Medical Research (AHFMR no. 199909) and the Institute of Health Economics to David Feeny. PSI, IHE, and AHFMR played no role in the design, interpretation, or analysis of the project and have not reviewed or approved of this manuscript. The paper was presented as a poster at the 10th Annual Conference of the International Society for Quality of Life Research, November 12–15, 2003, Prague. It should be noted that David Feeny has a proprietary interest in Health Utilities Incorporated (HUI), Dundas, Ontario, Canada. HUInc. owns the copyright to and distributes HUI materials.

point of reference between patients and members of the general population could account for the higher scores provided by patients. In the context of the utility approach to assessing health-related quality of life (HRQL), prospect theory (see for instance references 4;12;13;16;20; related results are reported in van Osch et al. [24] and Wittenberg et al. [26]) postulates that people evaluate health states relative to their personal reference point, usually their current health state. The reference point serves as a point of inflection in their utility (preference) function. The utility function is concave above the reference point (states preferred to the current state) and convex below the reference point (states dis-preferred to the current state). Thus, given that members of the general population are in general healthier than patients with chronic conditions, members of the general population evaluate the chronic health state being experienced by patients as worse than their own current health and, thus, on the convex portion of their utility function. In contrast, the patient evaluates the chronic state as their point of reference.

Several investigators have provided empirical evidence from studies that have included both prospective and retrospective assessments of HRQL that is consistent with response shift and/or prospect theory (Norman [17] provides a useful summary). In several studies, patients were asked serially to evaluate their HRQL. Patients who experienced changes in their health status were also asked to rate their previous HRQL retrospectively. Frequently, patients who had improved provided retrospective scores that were much lower than scores obtained prospectively. The explanation according to prospect theory is as follows. When the patients provided the evaluations in "real" time prospectively, their current health state before they experienced an improvement in health status was their reference point. In contrast, when asked to provide a retrospective evaluation from the vantage of their new reference point, a much preferred state of health experienced after their health status had improved, the "baseline" health state was then evaluated on the convex portion of their utility function, generating a much lower score. Norman [17] also notes that, according to the implicit theory of change approach, there are systematic problems with recall that can also account for the results observed. Patients do not really recall their baseline health state but do believe that they have improved. Patients then infer what the baseline health state must have been given their current health state and their impression of the magnitude of change experienced. This phenomenon could also account for the observation that scores obtained retrospectively are lower than scores obtained prospectively.

Much of the existing evidence in support of prospect theory in the evaluation of utility scores for health states has relied on cross-sectional comparisons or prospective studies that have included retrospective assessments. Retrospective assessments, however, are potentially problematic; they may not be reliable and may be biased. The study reported in this paper provides an opportunity to test the implications of prospect theory among patients whose health status changed substantially over time without the use of retrospective assessments.

METHODS

Patients and Procedures

The paper is based on a study investigating the HRQL burden of waiting for, undergoing, and recovering from elective total hip arthroplasty (THA) for osteoarthritis (OA). The study has been described previously (2;3;5;6;15) and will be described here briefly. Before commencing the study, approval was obtained from the local Human Ethics Committee.

All patients who were referred for "hip disease" between November 1993 and 1996 to any of seven surgeons performing THA in London, Ontario, were potentially eligible to participate in the study. Eligible patients who provided consent were invited to attend an outpatient department for a baseline assessment. Upon arrival to the clinic, the following data were collected from the patients: (i) age, gender, home address, employment status, duration of hip disease symptoms, and presence of comorbid conditions; (ii) health status and HRQL, using several generic, preference-based, and disease-specific measures; and (iii) visual analogue scale (VAS) and standard gamble (SG) scores for three hypothetical states (mild, moderate, and severe OA) and for the patient's subjectively defined current health state (SDCS). Interviewer administration was used for the VAS and SG; self-completion questionnaires were used for the HRQL instruments. Patients who were put onto a waiting list for THA continued to participate in a longitudinal study examining HRQL after THA (15).

Hypothetical States

To provide a context for patients to provide preference-based scores (VAS and SG) for their SDCS and assist in the interpretation of SG scores, health-state descriptions for three hypothetical health states, corresponding to mild, moderate, and severe OA such that a patient was eligible for elective THA, were developed (see Table 1). The hypothetical-state descriptions were based on an earlier study of THA (11). Each OA hypothetical-state description included six items: pain and stiffness on exertion, use of walking aids, use of analgesics, pain at night, ability to do housework, and socializing. The content of the hypothetical-states descriptions was designed to capture the most salient domains of health status in the context of OA and THA.

At the baseline assessment (assessment no. 1) and at subsequent complete assessments, patients completed the full battery of health status and HRQL measures. They also evaluated the three hypothetical states and their SDCS.

Assessment of Preferences: VAS and SG

In the first step in the preference interview, patients ranked health states on a VAS scale known as the Feeling

Table 1. Hypothetical-State Descriptions of Mild, Moderate, and Severe Osteoarthritis in Patients Who Are Candidates for Elective Total Hip Arthroplasty

Mild
1. You have slight pain and stiffness.
2. You don't need a walker or cane but sometimes use a piece of furniture to steady yourself.
3. You only need to use aspirin or plain Tylenol occasionally.
4. You have no night pain and you sleep well.
5. You are able to do all housework and chores if you take your time.
6. Your social activity with family and friends is only slightly decreased.
Moderate
1. You have moderate pain and stiffness upon exertion.
2. You need to use a cane to walk more than one city block.
3. You occasionally need to take pain pills such as Tylenol and Codeine.
4. You sometimes experience night pain, which is relieved by a position change and/or pain pills.
5. You can only do light housework or chores.
6. You socialize with family and friends but more than 1 hour is painful and tiring.
Severe
1. You have constant pain and stiffness.
2. You must use a walker at all times.
3. You regularly use pain pills such as Tylenol with Codeine and/or anti-inflammatory medication.
4. You sleep poorly at night.
5. You are unable to do any housework and/or chores.
6. It is very difficult for you to socialize with family and friends for even a few minutes.

Note: Patients were not shown the mild, moderate, or severe labels.
Source: Laupacis et al. [11].

Thermometer (FT) (7). The top of the scale was labeled as "Most Desirable"; the bottom of the scale was labeled as "Least Desirable." Patients evaluated the three hypothetical states, perfect health (absence of OA) and dead, on the FT. Patients were then asked to place their SDCS on the FT.

Next, patients were asked to evaluate the same states in the same order as on the VAS using the SG. In the SG, respondents are offered a choice between a lottery and an intermediately ranked health state as a sure thing (21). The lottery consists of probability p of a highly desirable outcome (perfect health) and probability 1-p of a highly undesirable outcome (dead). The probability p is varied until the patient is indifferent between the lottery and the sure thing. The more desirable the sure thing (for instance, the mild OA state), the higher the probability of perfect health the patient will require to be indifferent. To assist patients, a Chance Board (CB) was used to ask the SG questions (7). The CB presents the probability information for the lottery in pie charts along with health state descriptions for perfect health, dead, and the intermediate state. (Furlong et al. [7] provide details on the conduct of preference elicitation interviews.) Interviews were conducted by a trained professional interviewer.

Starting at 6 months after assessment no. 1, and every 6 months thereafter, patients were asked to return to the outpatient clinic for a complete assessment (including direct assessment of utilities). (Patients who were not recommended for THA were dropped from the study.) Follow-up continued for those patients for whom THA was recommended and who decided to proceed to surgery. Final and complete assessments (including direct assessment of utilities) were done at least 3 months after THA.

A Priori Hypotheses Based on Prospect Theory

In general, the health status of patients declined while they waited for THA and then improved dramatically after THA (15). Thus, it would be expected that the reference point of many patients would change substantially over the course of the study. Furthermore, some patients are likely to have "jumped" a hypothetical state. For instance, if at the presurgery assessment, a patient viewed herself (Self 1 in Figure 1) as worse than a particular hypothetical state, say the mild (Mild 1) state, and then after surgery viewed herself as better than her presurgery score for the mild state (Self 2 in Figure 1), the presurgery evaluation of the mild state would be from a reference point, Self 1, which is below Mild 1 (and, thus, Mild 1 would be on the concave portion of the utility function), and the evaluation of Mild 1 after surgery would be from a reference point, Self 2, above the mild state (and, thus, would be on the convex portion of the utility function). This finding would suggest that the presurgery score for the mild state would exceed the postsurgery score for the mild state:Mild 1 > Mild 2.

Thus, it is possible to test a prediction of prospect theory by identifying patients whose scores for their SDCS increased over time if the change in the score for their SDCS was large enough so that they "jumped" one (or more) of the hypothetical-state scores. In addition, patients eligible for testing the hypotheses need to have ranked the

https://doi.org/10.1017/S0266462305050713 Published online by Cambridge University Press

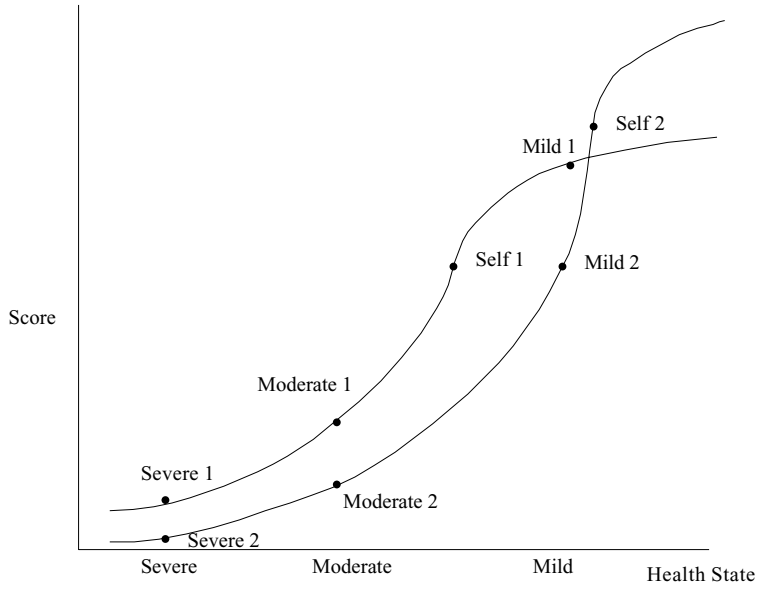

Figure 1. Illustration of prospect theory. Source: based on Lenert et al. (12;13).

hypothetical states in the expected order: mild preferred to moderate, moderate preferred to severe.

Analysis

The testing of the predictions of prospect theory will include a count of the number of patients for whom changes in hypothetical-state scores are in the direction predicted. (Because some patients jumped more than one hypothetical state, the number of cases exceeds the number of patients.) Changes in the direction predicted will be interpreted as consistent with the hypothesis; changes in opposite direction will be interpreted as inconsistent with the hypothesis. In the primary analysis, zero change in the relevant hypothetical-state scores will be interpreted as neither confirming nor refuting the hypothesis. In a secondary analysis, a patient for whom there was a mix of no change in scores in one or more cases along with change in the direction predicted by PT in one or more cases will be classified as supporting the hypothesis. A patient for whom there was a mix of no change in score and change in the direction opposite to that predicted by PT will be classified as inconsistent with the hypothesis. Patients who had changes both in the direction predicted and opposite the direction predicted or no change in scores will be interpreted as neither supporting nor refuting the hypothesis. Although test–retest reliability of the hypothetical-state scores in the study was 0.83, there is measurement noise in SG scores. Further in interpreting the results of the test of the hypothesis, it is important to consider the magnitude of the difference in scores that will be considered to be important. We will define a difference of 0.05 or more as a meaningful difference. In general choice-based scores, such as SG scores, are regarded as more valid than scores obtained from the VAS (22). Therefore, SG scores will be used to test the implications of PT.

Table 2. Demographic and Clinical Characteristics of Patients at Baseline ($n = 102$)

Sex, % male	50
Age, mean in years	68
Comorbidities, % with problem or disease	
Cerebrovascular disease	7
Cancer	13
Hypertension	39
Coronary artery disease	15
Diabetes mellitus	5
Mean duration of symptoms, years	6.15
Employment status at time of enrollment in study, %	
Working full-time	10
Retired	81
Other	9

RESULTS

The demographic and clinical characteristics of the patients enrolled in the waiting-list study are described in Table 2. As one would expect, most patients were retired and had experienced OA symptoms for several years.

The hypothesis is about patients whose HRQL improved over time. Of the 109 patients, 7 are eliminated because we have no surgery dates for them. Another three are eliminated because of missing hypothetical-state scores. Of the remaining ninety-nine patients with pre- and postsurgery SDCS scores, ninety provided scores for the hypothetical states that conformed to the expected rank ordering; nine did not. Eighteen patients had $SDCS_{postsurgery} < SDCS_{presurgery}$ scores and, thus, are not eligible for testing the hypothesis.

https://doi.org/10.1017/S0266462305050713 Published online by Cambridge University Press

Table 3. Tests of the Predictions of Prospect Theory ($n = 14$ Patients; 29 Cases)

Patients	Cases: change in standard gamble score: postsurgery minus presurgery score
Change in direction predicted by prospect theory	Not applicable
0 patients	
Mixed: no change in utility score or change in direction predicted by prospect theory	0.0, −0.5, −0.5 0.0, −0.1
2 patients	
Mixed: changes both in and opposite to the direction predicted by prospect theory	−0.4, 0.3, 0.4
1 patient	
No change in score	0.0
1 patient	
Mixed: no change or change opposite to direction predicted by prospect theory	0.0, 0.2, 0.2 0.0, 0.0, 0.1, 0.1, 0.1
2 patients	
Change in opposite direction as predicted by prospect theory	0.4 0.3, 0.3, 0.1 0.4, 0.1
8 patients	0.4, 0.6 0.1 0.1 0.1 0.2

Twenty-six patients had $SDCS_{postsurgery}$ scores equal to their $SDCS_{presurgery}$ scores (ineligible). Forty-six patients met the criteria that the rank order of the hypothetical states was as expected and their $SDCS_{postsurgery}$ scores were greater than their $SDCS_{presurgery}$ scores. The score for the SDCS jumped one or more the hypothetical-state scores for 14 of the 46 patients.

Results are presented in Table 3. Patients meeting specific criteria, such as consistent with the predictions of PT, are listed in the left-hand panel of Table 3. The magnitude of the change in utility scores for the hypothetical health states for each patient is listed in the right-hand panel of Table 3. There were fourteen patients eligible for testing the hypothesis. These fourteen patients generated twenty-nine cases for which the score for the SDCS jumped the score for a hypothetical health state.

The changes in hypothetical-state SG scores were completely (in all cases for that patient) in the direction predicted by prospect theory for no patients. For six patients (42.9 percent), there were no changes in the SG scores for the relevant hypothetical states or the results for those patients were mixed. For eight patients (57.1 percent of patients), the changes in SG scores were in the direction opposite to the predictions of prospect theory.

Using the criteria for the secondary analysis, there were two patients (14.3 percent) for whom there were no changes or the changes were in the direction predicted by PT. Results were mixed (no change or changes in both directions) for two patients (14.3 percent). There were ten patients (71.4 percent) for whom there were no changes or the changes were in the direction opposite to that predicted.

DISCUSSION

In general, the results do not support the predictions of prospect theory. When hypothetical-state SG scores changed, they changed by 0.1 or more; the changes in SG scores observed were clearly important (7) and greater than the threshold for clinically important differences of 0.05. Measurement imprecision is an unlikely explanation for the poor predictive performance of prospect theory.

The decline in health status and HRQL while waiting for THA and improvements observed when comparing post- and presurgery scores is well documented and quantitatively important (3). For instance, the mean change in SG score between the pre- and postsurgery assessments was 0.17; the mean change in the overall Health Utilities Index Mark 2 (HUI2) score was 0.18; the mean change in overall HUI Mark 3 score was 0.25; the mean change in the SF-36 physical functioning domain score was 25.4; and the mean change in the SF-36 bodily pain domain score was 27.5. Results from disease-specific measures agree with respect to the direction and order of magnitude of the change experienced. Clearly, the HRQL of these patients changed substantially. Thus, one would have expected a substantial change in reference point.

A major advantage of the study is the lack of reliance on retrospective assessments in testing the predictions of prospect theory. Thus, results are unlikely to be confounded by imprecise or biased recall (17).

A limitation of the study, however, is the assumption, common in the literature on prospect theory in health care, that patients use their current health as the reference point. While the study used rigorous direct utility measurement procedures and did ask patients to evaluate their current health, the study did not include any questions that might have helped to identify any reference point they might have used in such an evaluation. Another limitation of the study is the small sample size.

To improve empirical work on prospect theory in the context of the utility assessments of health states, future studies might include extensive debriefing about what respondents were thinking about when they evaluated the health states. Future studies might experiment with approaches to identifying what reference point, if any, respondents are considering.

Clearly, the results of this study need to be replicated and confirmed in other contexts and with much larger sample sizes before the prospect theory explanation for response shift is discarded. Nonetheless, one can speculate that perhaps learning to live with chronic conditions rather than changes in the reference point are perhaps more important in explaining

https://doi.org/10.1017/S0266462305050713 Published online by Cambridge University Press

the differences in scores provided by patients and members of the general population.

CONTACT INFORMATION

David Feeny, PhD (david.feeny@ualberta.ca), Professor of Economics and Public Health Sciences, University of Alberta, Edmonton, Alberta T6G 2H4; Fellow, Institute of Health Economics,10405 Jasper Avenue, Suite 1200, Edmonton, Alberta T5J 3N4, Canada;
Ken Eng, MA (keng@ihe.ca), Research Associate, Institute of Health Economics, 10405 Jasper Avenue, Suite 1200, Edmonton, Alberta T5J 3N4, Canada

REFERENCES

1. Baron J, Asch DA, Fagerlin A, et al. Effect of assessment method on the discrepancy between judgements of health disorders people have and do not have: A web study. *Med Decis Making*. 2003;23:422-434.
2. Blanchard C, Feeny D, Mahon JL, et al. Is the health utilities index valid in total hip arthroplasty patients? *Qual Life Res*. 2004;13:339-348.
3. Blanchard C, Feeny D, Mahon JL, et al. Is the health utilities index responsive in total hip arthroplasty patients? *J Clin Epidemiol*. 2003;56:1046-1054.
4. Bleichrodt H, Pinto JL, Abellan-Perpinan JM. A consistency test of the time trade-off. *J Health Econ*. 2003;22:1037-1052.
5. Feeny D, Blanchard C, Mahon JL, et al. Comparing community-preference based and direct standard gamble utility scores: Evidence from elective total hip arthroplasty. *Int J Technol Assess Health Care*. 2003;19:362-372.
6. Feeny D, Blanchard CM, Mahon JL, et al. The stability of utility scores: Test-retest reliability and the interpretation of utility scores in elective total hip arthroplasty. *Qual Life Res*. 2004;13:15-22.
7. Furlong W, Feeny D, Torrance GW, Barr R, Horsman J. Guide to design and development of health-state utility instrumentation. McMaster University Centre for Health Economics and Policy Analysis Working Paper No 90-9, June 1990.
8. Glimcher PW, Rustichini A. Neuroeconomics: The consilence of brain and decision. *Science*. 2004;306:447-452.
9. Gold MR, Patrick DL, Torrance GW, et al. Identifying and valuing outcomes. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996:82-134.
10. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. *Econometrica*. 1979;47:263-291.
11. Laupacis A, Bourne R, Rorabeck C, et al., Effect of elective total hip replacement upon health-related quality of life. *J Bone Joint Surg Am*. 1993;75:1619-1626.
12. Lenert LA, Treadwell JR, Schwartz CE. Associations between health status and utilities: Implications for policy. *Med Care*. 1999;37:479-489.
13. Lenert LA. Treadwell JR, Schwartz CE. Associations between health status and utilities: Indirect evidence for response shift. In: Schwartz CE, Sprangers MAG, eds. *Adaptation to changing health status: Response shift in quality-of-life research*. Washington, DC: American Psychological Association; 2000: 201-209.
14. Levy M, Levy H, Prospect theory: Much ado about nothing?" *Manag Sci*. 2002;48:1334-1349.
15. Mahon JL, Bourne R, Rorabeck C, et al. The effect of waiting for elective total hip arthroplasty on health-related quality of life. *Can Med Assoc J*. 2002;167:1115-1121.
16. Oliver A. The internal consistency of the standard gamble: Tests after adjusting for prospect theory. *J Health Econ*. 2003;22:659-674.
17. Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. *Qual Life Res*. 2003;12:239-249.
18. Postulart D, Eddy MM. Adang, Response shift and adaptation in chronically ill patients. *Med Decis Making*. 2000;20:186-193.
19. Schwartz CE, Sprangers MAG, eds. *Adaptation to changing health status: Response shift in quality-of-life research*. Washington, DC: American Psychological Association; 2000.
20. Stratmann-Schoene D, Klose T. Health values and prospect theory: A comment. *Med Decis Making*. 2001;21:57-59.
21. Torrance GW. Measurement of health state utilities for economic appraisal—a review. *J Health Econ*. 1986;15:1-30.
22. Torrance GW, Feeny D, Furlong W. Visual analogue scales: Do they have a role in the measurement of preferences for health states? *Med Decis Making*. 2001;21:329-334.
23. Ubel PA, Lowenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Qual Life Res*. 2003;12:599-607.
24. van Osch SMC, Wakker PP, Van den Hout WB, Stiggelbout AM. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Making*. 2004;24:511-517.
25. Wakker PP. The data of Levy and Levy (2002) 'Prospect theory: Much ado about nothing?' actually support prospect theory. *Manag Sci*. 2003;49:979-981.
26. Wittenberg E, Winer EP, Weeks JC. Patient utilities for advanced cancer: Effect of current health on values. *Med Care*. 2005;43:173-181.