Empirical Validation of Patient versus Population Preferences in Calculating QALYs

Eva-Julia Weyler and Afschin Gandjour

A fundamental assumption of the quality-adjusted life year model is mutual utility independence between life years and health status. However, this assumption may not hold for severe health states: living in a severe health state may cause disutility beyond a so-called maximal endurable time (MET). It is unknown, however, whether persons without experience of a disease, who are often used in health state valuation exercises, account for MET. Using data from 159 respondents from two convenience samples in Germany who were presented a health state description of depression, this study shows that persons without experience of depression had a lower rate of MET than persons with a history of depression. Furthermore, they had more preference reversals in case of MET, thus violating a fundamental principle of rational choice theory. While these findings suggest that severe health states should be assessed by patients rather than the community, confirmation in additional studies outside Germany and based on other health-state valuation techniques and diseases is recommended.

Key Words. QALYs, patient preferences, maximal endurable time, preference reversals

Cost-effectiveness analyses in health care often use quality-adjusted life years (QALYs) as a measure of health outcome. This is in line with recommendations by authorities such as the U.S. Panel on Cost-effectiveness in Health and Medicine and the U.K. National Institute for Health and Clinical Excellence. QALYs are the product of life years and a representation of preference for different health states (preference weight or score). Preference weights are anchored on a scale from 0 to 1, where 0 and 1 represent death and

full health, respectively. Techniques to obtain preferences include direct preference measures such as the time trade-off (TTO) method (Torrance 1972), the standard gamble (SG) method (Torrance 1976), and visual analogue scales. They have in common that they use full health and death as reference states. Yet only the SG method is based directly on the axioms of von Neumann–Morgenstern expected utility theory (von Neumann and Morgenstern 1953) and therefore often considered theoretically superior.

An open question is whose preference assessment should count when calculating QALYs, and this question continues to be intensely debated in the health economics literature (Gold et al. 1996; Dolan 1999; Nord 1999; Menzel et al. 2002; Ubel, Loewenstein, and Jepson 2003; Brazier et al. 2005; Gandjour 2010). In recent years, the debate has centered on preferences of patients versus community members (general population/public). As patients generally provide higher valuations of the same health states than community members (Peeters and Stiggelbout 2010), it is necessary to resolve this debate.

The validity of QALYs depends on a set of restrictive assumptions. One fundamental assumption is that survival duration and health state are mutually utility independent, that is, each is utility independent of the other (Pliskin, Shepard, and Weinstein 1980). This assumption implies that an increase in survival duration leads to higher utility. However, this assumption may not hold for severe health states: living in a severe health state may cause disutility beyond a so-called maximal endurable time (MET). MET was first shown by Sutherland et al. (1982) and subsequently confirmed in a number of studies (Bala et al. 1999; Stalmeier et al. 2001; Duru et al. 2002; Dolan and Stalmeier 2003; Franic et al. 2003; Robinson and Spencer 2006; Stalmeier et al. 2007). Still, the QALY model could account for MET if the expected duration of a health state were explicitly considered in the TTO or SG exercise. The TTO method would then yield an overproportional trade-off.¹ To the best of our knowledge there has been only one study (Stalmeier et al. 2001) analyzing this issue. This study was conducted among 176 students and compared direct preferences for health profiles with QALYs calculated based on the TTO method and the expected duration of health states. It showed that 103

Address correspondence to Dr. Afschin Gandjour, Ph.D., Pennington Biomedical Research Center, Louisiana State University, 6400 Perkins Road, Suite G3047, Baton Rouge, LA 70808; e-mail: afschin.gandjour@pbrc.edu. Dr. Eva-Julia Weyler, Ph.D., M.D., is with the Faculty of Management, Economics, and Social Sciences, University of Cologne, Cologne, Germany. Afschin Gandjour, was at the time of the study with the The James A. Baker III Institute for Public Policy, Rice University, Houston, TX, and also with the Institute of Health Economics and Clinical Epidemiology, University of Cologne, Cologne, Germany.

respondents had MET preferences when directly choosing between health profiles with different durations of a single health state. Out of these 79 (77 percent) did not show MET preferences when health states assessed by the TTO method, however. Hence, these respondents showed a preference reversal (PR), that is, their preference was dependent on the method of elicitation.

PRs present a notable deviation from economic theory of rational choice and violate the principle of procedure invariance. PRs were first described with regard to monetary outcomes (Lindman 1965; Slovic and Lichtenstein 1968). As noted in a recent review on PRs (Oliver and Sorenson 2008), there has been very little research on PRs in the health economics literature so far.

An important but neglected question is how patients and community members compare in terms of procedure invariance and PRs. As stated above, the discussion of whose preferences should count in economic evaluations has not been resolved. So far the debate has primarily used theoretical arguments (Gold et al. 1996; Dolan 1999; Nord 1999; Menzel et al. 2002; Ubel, Loewenstein, and Jepson 2003; Brazier et al. 2005; Gandjour 2010). For example, Gandjour (2010) recently argued that use of patient preferences is founded in economic theory while there is no compelling theoretical basis for community preferences. Yet we would also like to know whether this foundation of patient preferences holds empirically. That is, we would like to know the empirical validity of patient versus community preferences in relation to economic theory.

The purpose of this study was therefore to determine the empirical validity of using patient versus nonpatient (population) preferences when calculating QALYs. To this end, the study compared the rate of MET and PRs between the two respondent groups. Thus, this study attempted to analyze whether nonpatients account for MET and deviate no more from economic theory than patients do. This study was also the first to analyze PRs among respondents to the SG questionnaire, which is often considered as a gold standard in health-state valuation. All health states included depression of varying length and intensity.

METHODS

Participants were recruited from the health economics diploma program at the University of Cologne, Germany, and in a primary care physician's office in Hamburg, Germany. They independently filled out questionnaires while an interviewer was available for questions.

In order to assess depressive health states all participants were provided with a description of severe depression based on the ICD-10 criteria. That is, all participants—whether patients and nonpatients—assessed the same health states. We classified participants as patients if they had either an acute depression or a history of depression. In order to screen for acute or past depression we used the PHQ-2 (Kroenke, Spitzer, and Williams 2003), a validated two-item depression screening questionnaire. In addition to asking for symptoms of depression the questionnaire also elicited information on the level of school education, age, gender as well as daily time spent on work, family/partner, and leisure.

The SG questionnaire we used asked for the risk of death the respondent was willing to accept in order to be cured (see Appendix SA2 for questionnaire). Note that this question is slightly different from the question implied by the axioms of expected utility theory, which is the probability of death rendering a person indifferent between the gamble and the certain outcome. One may speculate that if one asks for the risk of death a respondent is willing to accept in order to be cured, the gamble (with the accepted risk of death) could be valued slightly higher than the certain outcome. The reason we used this approach was to reduce complexity. In any case, this approach is well accepted in the literature (see, e.g., Williams and Kind 1992; Bala and Zarkin 2000; Weinstein 2005). Furthermore, as our study is based on a comparison (between patients and nonpatients), any potential bias affects both groups and thus may cancel out.

In order to detect MET and PRs among patients and nonpatients, we followed the approach by Stalmeier et al. (2001), except for using the SG and not the TTO method. The following description follows closely their paper. Let (Y, h) denote living Y years in health state h , followed by immediate death. For example, the health state “living 5 days per week in severe depression over 10 years” is denoted by $(10, 5)$. A typical PR is depicted in Figure 1. The first line indicates that, when asked directly, the 10-year duration is preferred over the longer 20-year duration with 5 days of severe depression per week. This simple preference establishes MET preference. The $\sim$ symbol indicates that

Figure 1: Hypothetical Example of Maximal Endurable Time and Preference Reversal (Adapted from Stalmeier et al. 2001)

$$\begin{array}{ccc} (10, 5) & > & (20, 5) \\ & \sim & \\ (4, \text{healthy}) & < & (8, \text{healthy}) \end{array}$$

the respondent is indifferent between (10, 5) and (4, healthy); likewise, he is indifferent between (20, 5) and (8, healthy). The number of life years in full health (4 and 8, respectively) is calculated by multiplying SG weights (0.4) by the duration in years. Because it can be safely assumed that (8, healthy) is preferred to (4, healthy), the last line indicates that the preference order derived from the SG task is reversed.

While MET is a necessary condition for PR, not every person with MET has PR: if a short duration of depression leads to a SG weight that is much larger than for a long duration, more QALYs result (despite a short duration) and a PR does not occur. In our experiment this situation is given when a short duration (10 years) is preferred to a long duration (20 years) in the direct comparison and the SG weight for the short duration is more than twice that for a long duration.

In order to detect any potential threshold for MET, we varied the number of depressive days per week between 1 and 7 (see Appendix SA2).

In order to detect relationships between the occurrence of MET and PR, respectively, on the one hand and location of survey, patient status, level of school education, age, gender, and daily activities on the other hand we conducted a logistic regression analysis using *SPSS* version 12.0 for Windows (*SPSS* Inc., Chicago, IL, USA). Location of survey, patient status, level of school education, and gender were coded as binary variables. We considered $p < .05$ to be statistically significant.

RESULTS

A total of 174 persons filled out a questionnaire. Fifteen questionnaires had to be excluded, mainly due to incomplete (6) and implausible answers (5) such as estimates outside the defined range. Therefore, a total of 159 questionnaires were considered in the analysis, 91 and 68 of which were obtained at the university and physician practice, respectively. Characteristics of study participants are provided in Table 1. The average age was 33 years (range: 19–76; standard deviation: 14.1). The average time spent on work, leisure, and family/partner was 6.2, 3.9, and 3.7 hours, respectively. Noteworthy, the proportion of participants with prior or acute depression was higher than in the normal population (36 percent versus 17 percent [Jacobi et al. 2004]). The proportion was higher for women than men (40 percent versus 30 percent) which is in line with epidemiological data (Jacobi et al. 2004). Forty percent of participants with prior or acute depression were recruited from the university setting.

Table 1: Characteristics of Study Participants ($n = 159$)

	<i>n</i>	%
Location of survey		
University	91	57
Physician practice	68	43
Depression status		
Patients	57	36
Nonpatients	102	64
Age, years		
< 26	57	36
26–50	79	50
51+	23	14
Gender		
Female	92	58
Male	67	42
Education		
Abitur (highest level of school education)	112	70
Realschule (2nd highest)	32	20
Hauptschule (lowest)	15	10
Without school diploma	—	—

Utility scores of patients were consistently as high as or higher than those of nonpatients, both for the assessment over 10 and 20 years (Table 2). When comparing assessments over 10 and 20 years, nonpatients consistently provided higher utility scores for a 20-year horizon. Patients, on the other hand, showed lower scores for a 20-year horizon when assessing depression for 5 or more days per week.

Table 2: Preferences (Standard Gamble Utilities) of Patients and Nonpatients for Living 10 or 20 Years with Different Number of Days of Depressive Symptoms Per Week

<i>Number of Depressive Days per Week</i>	<i>10 Years</i>		<i>20 Years</i>	
	<i>Nonpatients</i>	<i>Patients</i>	<i>Nonpatients</i>	<i>Patients</i>
1	0.88 (0.16)	0.89 (0.23)	0.89 (0.18)	0.89 (0.19)
2	0.82 (0.17)	0.85 (0.23)	0.84 (0.19)	0.85 (0.20)
3	0.71 (0.21)	0.78 (0.23)	0.76 (0.22)	0.79 (0.21)
4	0.59 (0.23)	0.69 (0.25)	0.63 (0.24)	0.69 (0.23)
5	0.47 (0.24)	0.60 (0.28)	0.49 (0.30)	0.58 (0.28)
6	0.36 (0.23)	0.51 (0.31)	0.40 (0.30)	0.49 (0.31)
7	0.28 (0.26)	0.41 (0.32)	0.29 (0.29)	0.30 (0.92)

Note. Standard deviations are given in parentheses.

MET and PR

MET was found in 110 persons (69 percent). The probability of MET increased with the number of depressive days. MET occurred at an average of 5.7 days depression per week (standard deviation: 1.2 days). The corresponding utility score was 0.47. That is, at this utility score preference changed from more survival to less. The probability of PR also increased with the number of depressive days and was found in 65 respondents, all of whom had MET preference. Figure 2 shows that MET occurred more often in patients than nonpatients (95 percent versus 55 percent, $p < .01$), while PRs conditional on MET were more common among nonpatients (70 percent versus 30 percent, $p < .01$).

In the following, we illustrate the logistic regression model that describes the relationship between the occurrence of MET on the one hand and patient

Figure 2: Maximal Endurable Time (MET) and Preference Reversal (PR) in Patients and Nonpatients

status, level of school education, age, gender, and daily activities on the other hand. MET occurred significantly more often in participants classified as patients (Table 3). Furthermore, those spending little time with their family also showed more MET preferences (Table 3). In order to assess the goodness of the fit, we used two well-known tests: the likelihood-ratio test (Agresti 2002) and the Hosmer–Lemeshow test (Hosmer and Lemeshow 2000). For the first test, the likelihood-ratio statistic, which followed a χ^2 distribution with $9 - 1 = 8$ degrees of freedom, was estimated. It was 29.00, larger than the tabulated value of $\chi^2_{8;0.001} = 26.12$. The fit was therefore considered to be acceptable. The Hosmer–Lemeshow goodness-of-fit statistic (with 10 groups) was estimated to be 10.29 with a p -value of .25, again implying an acceptable fit (p -values larger than .05 are considered acceptable). In order to detect multicollinearity, we examined bivariate correlations among independent variables. No pair of variables was highly correlated except for location of survey and education ($r = 0.55$). However, the variance inflation factor ($= 1/(1 - R^2)$) was 1.70 and therefore clearly below 10, which is sometimes used as a cutoff (Wooldridge 2009, p. 99). We did not include interactions between independent variables as we did not find interactions to make sense conceptually and did not want to “fish” for statistically significant results.

In an ancillary analysis, we also analyzed whether the severity of depression that patients had experienced would affect these results. To this end, we introduced two dummy variables, representing a low ($= 3$) and high (> 3)

Table 3: Results of the Logistic Regression Analyzing the Relationship between Maximal Endurable Time and Individual Variables

<i>Variable</i>	<i>Coefficient (β)</i>	<i>e^{β} (95% Confidence Interval)</i>	<i>Wald Statistic</i>
Constant	2.834	17.011	8.975**
LOC	− 0.723	0.486 (0.158–1.493)	1.590
GEN	0.511	1.668 (0.712–3.907)	1.387
AGE	− 0.091	0.913 (0.465–1.791)	0.071
EDU	− 0.228	0.796 (0.260–2.436)	0.159
WORK	− 0.241	0.786 (0.377–1.640)	0.412
LEISURE	− 0.213	0.808 (0.408–1.600)	0.373
FAM	− 1.376	0.253 (0.120–0.531)	13.179**
DEP	2.618	13.709 (3.525–53.308)	14.276**

**Significant at $\alpha = 0.05$.

AGE, age in years; DEP, present or past depression (0 = no; 1 = yes); EDU, education (0 = no Abitur; 1 = Abitur = highest level of school education); FAM, daily time spend with family/partner (hours); GEN, gender (0 = female; 1 = male); LEISURE, daily leisure time (hours); LOC, location of survey (0 = university; 1 = physician practice); WORK, daily work time (hours).

PHQ-2 score, respectively. The logistic regression model showed that both a low and high score were associated with a significantly higher probability of MET (results not shown).

We also analyzed the relationship between the occurrence of PRs in patients with MET on the one hand and patient status, level of school education, age, gender, and daily activities on the other hand. PRs occurred significantly more often in females and those participants who spend more time at work (Table 4). Higher educated participants also had a higher rate of PRs. Furthermore, PRs occurred more commonly among nonpatients. According to the likelihood-ratio statistic (32.61) and the Hosmer–Lemeshow goodness-of-fit statistic (6.38 with a *p*-value of .61), the fit of the model was acceptable. The variance inflation factor, which quantifies the severity of multicollinearity, was 1.22 and thus below the cutoff of 10.

DISCUSSION

A fundamental assumption of the QALY model is mutual utility independence between life years and health status (Pliskin, Shepard, and Weinstein 1980). That is, the time horizon used for utility elicitation does not influence utility weights. This paper demonstrates, based on two convenience samples and a variant of the SG method, a violation of the utility independence

Table 4: Results of the Logistic Regression Analyzing the Relationship between Preference Reversals and Individual Variables for Patients with MET

<i>Variable</i>	<i>Coefficient (β)</i>	<i>e^{β} (95% Confidence Interval)</i>	<i>Wald Statistic</i>
Constant	− 1.081	0.339	1.287
LOC	0.242	0.785 (0.226–2.729)	0.785
GEN	− 1.902	0.149 (0.049–0.454)	11.249**
AGE	− 0.004	0.996 (0.948–1.047)	0.020
EDU	1.748	5.743 (1.722–19.154)	8.088**
WORK	1.287	3.621 (1.602–8.148)	9.567**
LEISURE	− 0.339	1.403 (0.668–2.948)	0.800
FAM	0.554	0.575 (0.238–1.390)	1.512
DEP	− 1.955	0.142 (0.047–0.428)	11.991**

**Significant at $\alpha = 0.05$.

AGE, age in years; DEP, present or past depression (0 = no; 1 = yes); EDU, education (0 = no Abitur; 1 = Abitur = highest level of school education); FAM, daily time spend with family/partner (hours); GEN, gender (0 = female; 1 = male); LEISURE, daily leisure time (hours); LOC, location of survey (0 = university; 1 = physician practice); MET, maximal endurable time; WORK, daily work time (hours).

assumption particularly for severe depression and thus confirms the results of other studies also using the SG method (Bala et al. 1999; Duru et al. 2002; Franic et al. 2003).

More important, this paper shows that people without experience of depression did appreciate the burden of depression in the short term, by providing utility scores similar to or lower than persons with a history of depression. On the other hand, they failed to express a preference for death with long-term disease, as indicated by the contradictory finding of fewer MET preferences despite lower utility scores. Moreover, they had more PRs, thus violating a fundamental principle of rational choice theory.

In contrast to our study, a recent U.S. study by Pyne et al. (2009) showed that depressed patients report lower SG scores for depression health states than the general population. Reasons for this difference may include cultural backgrounds (Germany versus U.S.), demographic variables, and description of depressive symptoms. Still, the results by Pyne and colleagues rather confirm our conclusion that people without experience of depression may underestimate the impact of depression on people's lives.

Furthermore, we would like to address the issue of generalizability of our results. In favor of generalizability we would like to point out that depression is an important outcome of many chronic diseases (Wells, Golding, and Burnam 1988; Bisschop et al. 2004) and chronic diseases are a frequent target of cost-effectiveness analysis due to their large financial burden. Accordingly, many cost-effectiveness analyses may be affected by the described phenomenon. On the other hand, our sample included rather young and well-educated individuals and while we adjusted for age and education in our analysis, we certainly cannot exclude that another sample with more elderly and less-educated individuals would provide a different result. Also, as noted above, results may not be transferable from country to country. Based on these limitations it seems premature to conclude based on the findings of this study that severe health states should be assessed by patients rather than the community. Given the importance of this question for the conduct of cost-effectiveness analysis in health care, confirmation in additional studies that are conducted outside Germany and consideration of other health-state valuation techniques and diseases is recommended.

The occurrence of MET with long-term horizons suggests that health-state valuations should be based on the true time horizon, which may be longer than the typical 10-year timeframe. This suggestion particularly refers to the assessment of severe health states, for which depression for several days a week is just one example. Still, our results indicate that even for people with

experience of depression the SG method does not fully capture MET and thus leads to PRs in a considerable percentage of cases (30 percent). This finding supports the conclusion by Stalmeier et al. (2001) that the standard QALY model is incorrect for describing preferences among poor health states. Therefore, the search for a health-state valuation method that adequately captures MET needs to continue.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors would like to thank two anonymous reviewers for very helpful comments on an earlier draft.

Disclosures: None.

Disclaimers: None.

NOTE

1. The TTO technique determines the proportion of remaining life years in poor health one is willing to give up or trade in exchange for perfect health.

REFERENCES

- Agresti, A. 2002. "An Introduction to Categorical Data Analysis." 2nd edition. New York: John Wiley and Sons.
- Bala, M. V., L. L. Wood, G. A. Zarkin, E. C. Norton, A. Gafni, and B. J. O'Brien. 1999. "Are Health States 'Timeless'? The Case of the Standard Gamble Method." *Journal of Clinical Epidemiology* 52 (11): 1047–53.
- Bala, M. V., and G. A. Zarkin. 2000. "Are QALYs an Appropriate Measure for Valuing Morbidity in Acute Diseases?" *Health Economics* 9 (2): 177–80.
- Bisschop, M. I., D. M. Kriegsman, D. J. Deeg, A. T. Beekman, and W. van Tilburg. 2004. "The Longitudinal Relation between Chronic Diseases and Depression in Older Persons in the Community: The Longitudinal Aging Study, Amsterdam." *Journal of Clinical Epidemiology* 57 (2): 187–94.
- Brazier, J., R. Akehurst, A. Brennan, P. Dolan, K. Claxton, C. McCabe, M. Sculpher, and A. Tsuchiya. 2005. "Should Patients Have a Greater Role in Valuing Health States?" *Applied Health Economics and Health Policy* 4 (4): 201–8.
- Dolan, P. 1999. "Whose Preferences Count?" *Medical Decision Making* 19: 482–6.
- Dolan, P., and P. Stalmeier. 2003. "The Validity of Time Trade-Off Values in Calculating QALYs: Constant Proportional Time Trade-Off versus the Proportional Heuristic." *Journal of Health Economics* 22 (3): 445–58.

- Duru, G., J. P. Auray, A. Beresniak, M. Lamure, A. Paine, and N. Nicoloyannis. 2002. "Limitations of the Methods Used for Calculating Quality-Adjusted Life-Year Values." *Pharmacoeconomics* 20 (7): 463–73.
- Franic, D. M., D. S. Pathak, and A. Gafni. 2003. "Are Health States 'Timeless'? A Case Study of an Acute Condition: Post-Chemotherapy Nausea and Vomiting." *Journal of Evaluation in Clinical Practice* 9 (1): 69–82.
- Gandjour, A. 2010. "Validity of Patient vs Population Preferences in Calculating QALYs." *Medical Decision Making* 30 (4): E57–63.
- Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein. 1996. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.
- Hosmer, W., and S. Lemeshow. 2000. *Applied Logistic Regression*. 2nd edition. New York: John Wiley and Sons.
- Jacobi, F., H. U. Wittchen, C. Holting, M. Höfler, H. Pfister, N. Müller, and R. Lieb. 2004. "Prevalence, Comorbidity and Correlates of Mental Disorders in the General Population: Results from the German Health Interview and Examination Survey (GHS)." *Psychological Medicine* 34 (4): 597–611.
- Kroenke, K., R. L. Spitzer, and J. B. Williams. 2003. "The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener." *Medical Care* 41 (11): 1284–92.
- Lindman, H. R. 1965. *The Measurement of Utilities and Subjective Probabilities [dissertation]*. Ann Arbor: University of Michigan.
- Menzel, P., P. Dolan, J. Richardson, and J. A. Olsen. 2002. "The Role of Adaptation to Disability and Disease in Health State Valuation: A Preliminary Normative Analysis." *Social Science and Medicine* 55: 2149–58.
- Nord, E. 1999. *Cost-Value Analysis in Health Care: Making Sense out of QALYs*, pp. 82–90. Cambridge, England: Cambridge University Press.
- Oliver, A., and C. Sorenson. 2008. "Importance of Preference Reversals in the Valuation of Health and Healthcare." *Expert Review Pharmacoeconomics and Outcomes Research* 8 (2): 95–9.
- Peeters, Y., and A. M. Stiggelbout. 2010. "Health State Valuations of Patients and the General Public Analytically Compared: A Meta-Analytical Comparison of Patient and Population Health State Utilities." *Value Health* 13 (6): 306–9.
- Pliskin, J. S., D. S. Shepard, and M. C. Weinstein. 1980. "Utility Functions for Life Years and Health Status." *Operations Research* 28: 206–24.
- Pyne, J. M., J. C. Fortney, S. Tripathi, D. Feeny, P. Ubel, and J. Brazier. 2009. "How Bad Is Depression? Preference Score Estimates from Depressed Patients and the General Population." *Health Services Research* 44 (4): 1406–23.
- Robinson, A., and A. Spencer. 2006. "Exploring Challenges to TTO Utilities: Valuing States Worse Than Dead." *Health Economics* 15 (4): 393–402.
- Slovic, P., and S. Lichtenstein. 1968. "Relative Importance of Probabilities and Payoffs in Risk Taking." *Journal of Experimental Psychology* 78: 1–18.
- Stalmeier, P. F., G. B. Chapman, A. G. de Boer, and J. J. van Lanschot. 2001. "A Fallacy of the Multiplicative QALY Model for Low-Quality Weights in Students and Patients Judging Hypothetical Health States." *International Journal of Technology Assessment in Health Care*. 17 (4): 488–96.

- Stalmeier, P. F., L. M. Lamers, J. J. Busschbach, and P. F. Krabbe. 2007. "On the Assessment of Preferences for Health and Duration: Maximal Endurable Time and Better Than Dead Preferences." *Medical Care* 45 (9): 835–41.
- Sutherland, H. J., H. Llewellyn-Thomas, N. F. Boyd, and J. E. Till. 1982. "Attitudes Toward Quality of Survival: The Concept of "Maximal Endurable Time."" *Medical Decision Making* 2 (3): 299–309.
- Torrance, G. W. 1972. "A Utility Maximization Model for Evaluation of Health Care Programs." *Health Services Research* 7: 118–33.
- . 1976. "Social Preferences for Health States: An Empirical Evaluation of Three Measurement Techniques." *Socioeconomic Planning Science* 10: 129–36.
- Ubel, P. A., G. Loewenstein, and C. Jepson. 2003. "Whose Quality of Life? A Commentary Exploring Discrepancies between Health State Evaluations of Patients and the General Public." *Quality of Life Research* 12: 599–607.
- von Neumann, J., and O. Morgenstern. 1953. *Theory of Games and Behavior*. Princeton, NJ: Princeton University Press.
- Weinstein, M. C. 2005. "Quality-Adjusted Life Years: Application to Food Safety Priority Setting." In *Toward Safer Food: Perspectives on Risk and Priority Setting*, edited by S. A. Hoffman and M. R. Taylor, pp. 227–40. Washington: Resources for the Future.
- Wells, K. B., J. M. Golding, and M. A. Burnam. 1988. "Psychiatric Disorder in a Sample of the General Population with and without Chronic Medical Conditions." *The American Journal of Psychiatry* 145 (8): 976–81.
- Williams, A., and P. Kind. 1992. "The Present State of Play about QALYs." In *Measures of the Quality of Life*, edited by A. Hopkins, pp. 21–34. London: Royal College of Physicians.
- Wooldridge, J. M. 2009. *Introductory Econometrics: A Modern Approach*, 4th edition. Mason, OH: South-Western.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

- Appendix SA1: Author Matrix.
- Appendix SA2. Questionnaire.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.