International Journal of Technology Assessment in Health Care, **17:4** (2001), 488–496.

A FALLACY OF THE MULTIPLICATIVE QALY MODEL FOR LOW-QUALITY WEIGHTS IN STUDENTS AND PATIENTS JUDGING HYPOTHETICAL HEALTH STATES

Peep F. M. Stalmeier
Academic Medical Centre, Amsterdam, University Medical Center Nijmegen, and University of Nijmegen

Gretchen B. Chapman
Rutgers University

Angela G. E. M. de Boer
Jan J. B. van Lanschot
Academic Medical Centre, Amsterdam

Abstract

Objectives: In quality-adjusted life-years (QALY) models, it is customary to weigh life-years with quality of life via multiplication. As a consequence, for positive health states a longer duration has more QALYs than a shorter duration (i.e., longer is better). However, we have found that for poor health states, many prefer to live only a limited amount of time (i.e., longer is worse). Such preferences are said to be maximum endurable time (MET). In the present contribution, the following questions are asked: a) How low does the utility have to be in order for a MET to arise? and b) Do MET preferences occur when patients judge hypothetical health states?
Methods and Results: We reanalyzed data from 176 students for the hypothetical health states of "living with migraines" and "living with metastasized cancer." For utilities smaller than 0.7 (ranging from 0 to 1), the MET preference rate was larger than 50%. High MET preference rates were also found in two new studies on migraine and esophageal cancer patients, who evaluated hypothetical health states related to their disease.
Conclusions: We discuss the interpretation of the MET preferences and the preference reversal phenomenon. Standard QALY models imply that longer is better. However, we find that more often, longer is worse for poorly evaluated health states. Consider the following question: are 3 years with a weight of 0.3 equally as valuable as 1 year with a weight of 0.9? Our results suggest that the 3-year period may be less valuable because for poor health, many will prefer a 1-year over a 3-year period.

Keywords: Technology assessment, Health state valuation, QALYs, Maximum endurable time preferences, Preference reversals

Part of this research was made possible through grant OG-96-041 from the Ziekenfondsraad Amsterdam, and grants NUKC 93-136 and NUKC 97-1585 of the Dutch Cancer Society. The authors thank Janneke Willems for her help with interviewing the esophageal cancer patients in close cooperation with A.G.E.M. de Boer, and Jennifer Winquist for interviewing the migraine patients.

In quality-adjusted life-years (QALY) analyses, it is customary to weigh the life-years gained with quality of life. This weight recognizes that an episode in perfect health is preferred to an episode of equal duration in impaired health. Specifically, in the multiplicative QALY model, the duration Y of a health state is multiplied by the weight W for that health state, to yield a number $Y \times W$ that is equivalent to a number of years in full health (10). As a consequence, a longer duration in a health state with a low weight of 0.2, for example, on a scale from zero for death to one for perfect health, has more QALYs than a shorter duration with the same weight, i.e., "longer is better." In the same vein, it is assumed that 3 years with a weight of 0.3 are as valuable as 1 year with a weight of 0.9.

The number of QALYs are assumed to represent underlying preferences. Specifically, if health episode A has more QALYs than health episode B, then it is assumed that A is preferred to B. We discovered a preference reversal that strongly suggests that the multiplicative QALY model is incorrect (5;6). The preference reversal suggests that low weights or utilities correspond to health states for which many prefer a shorter duration to a longer duration in that health state.

Previous findings were obtained in a student sample evaluating hypothetical health states (5;6). This paper has three contributions: a) it presents a reanalysis of the student data; b) it extends previous findings to patient populations who were familiar with the hypothetical health states to be evaluated; and c) the interpretation of the maximum endurable time (MET) preferences and preference reversals is discussed.

THE PREFERENCE REVERSAL

Preference reversals with monetary gambles were first found by Lindman (3) and by Lichtenstein and Slovic (2). They were confirmed by Grether and Plott (1). For example, let ($x, yY) denote a delayed payment in which one receives $x in y years from now. When asked to choose between the delayed payments ($1,600, 1.5Y and $3,550, 10Y), many prefer the first payment (choice task). However, when asked to state the instantaneous monetary value of these delayed payments, the majority assigns a higher value to the second payment (matching task). Originally, preference reversals were explained as violations of transitivity. Now, the prevailing explanation is by a violation of *procedure invariance* (9). The two procedures yield different results because the money dimension receives less weight in the choice task, as compared to the matching task.

Our preference reversal in the domain of health valuation was discovered when quality weights were obtained for poor health states in which one prefers to live only a limited amount of time. Such preferences are labeled MET preferences (7). Corresponding QALYs are depicted in Figure 1. QALYs increase proportionally with duration for the positive health states H1 and H2; in contrast, for health state H3, QALYs decrease beyond the MET, located at S on the duration axis.

In our experiments, healthy college students assessed health states with the time tradeoff (TTO) method. A detailed account of the experiments can be found in Stalmeier et al. (5;6). Let (Y, H) denote living Y years in health state H, followed by immediate death. If in the TTO task, one is indifferent between (20, H) and (15, healthy), then the TTO weight for health state H is 15/20 (8).

One of the health states we used is "living 4.5 days per week with continuous migraines," denoted by (4.5 M). A typical preference reversal is depicted in Figure 2. The first line in Figure 2 indicates that, when asked directly, the 10-year duration is preferred over the longer 20-year duration with 4.5 days of migraines per week. This simple preference establishes a MET preference. The $\sim$ symbol indicates that, in the TTO task, the respondent is indifferent between (10, 4.5 Migr) and (4, Healthy); likewise, he is indifferent between (20, 4.5 Migr) and (8, Healthy). Because it can be safely assumed that (8, Healthy)

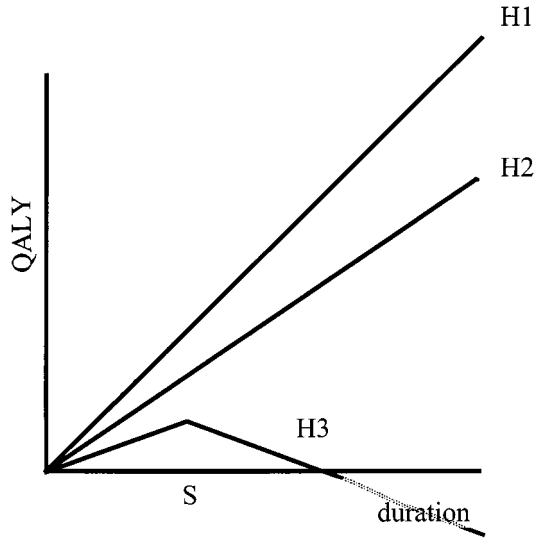

Figure 1. QALYs versus duration for different health states. For a health state with a maximum endurable time at S, the preference curve is single-peaked.

is preferred to (4, Healthy), the last line indicates that the preference order between the migraine episodes, derived from the TTO task, is reversed.

This preference reversal is not at all uncommon. For 176 students who evaluated poor health states, MET preferences were obtained in 103 students. For 79 of these 103 students, a preference reversal occurred ($p < .0001$). For the remaining 24 students, no preference reversal occurred because either fewer healthy years were equated to the longer migraine episode or equal healthy durations were equated to the longer and shorter migraine episodes.

$$\begin{array}{ccc} (10,\ 4.5\ \text{Migr}) & \succ & (20,\ 4.5\ \text{Migr}) \\ \sim & & \sim \\ (4,\ \text{Healthy}) & \prec & (8,\ \text{Healthy}) \end{array}$$

Figure 2. The preference reversal. The preference in the first line establishes the MET preference; the preference sign reversal in the first and last lines establishes the preference reversal.

The reversal is robust: even when it was explained that the above response pattern entails an inconsistent choice pattern, and given the opportunity to undo the reversal, most students chose not to do so (6). Only 4 of 27 students undid their preference reversal.

In the present contribution, we address the following research questions: a) how poor do these health states have to be in order for MET preferences to arise? and b) do MET preferences occur in patient populations judging hypothetical health states related to their disease? The first question is answered by a reanalysis of previously obtained data. The second question is answered by an analysis of new data obtained in migraine and esophageal cancer patients.

METHODS

Reanalysis of Student Data

Figure 3 presents a reanalysis of the student data (6). The percentage of MET preferences is plotted against the TTO quality weights. The TTO weight is based upon the trade-off for

https://doi.org/10.1017/S026646230110704X Published online by Cambridge University Press

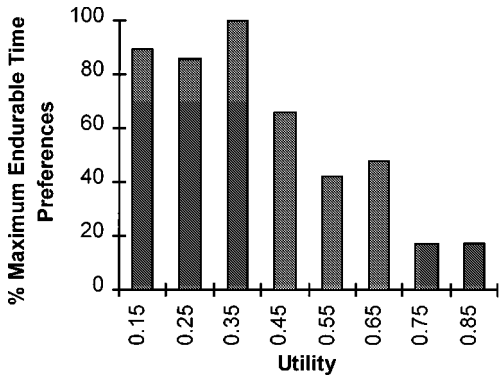

Figure 3. Percentage of students with MET preferences against time trade-off utility.

the shorter duration, for instance, on the trade-off for the 10 years' duration in Figure 2. Figure 3 aggregates data from 176 students and 3 health states: living with metastasized breast cancer, living with 4.5 days of migraines per week, and living with 5 days of migraines per week. The result is surprising: for weights smaller than 0.7, MET preferences are the rule rather than the exception. For weights above 0.7, as one would expect, MET preferences are rare.

Experiment 1: Migraine Patients

In the first experiment, the issue is whether MET preferences occur in a patient sample evaluating a highly familiar but still hypothetical health state. We also checked for preference reversals.

Patients. Patients were from the Neurology Department outpatient clinic at the University of Illinois Hospital in Chicago. All patients had a history of migraines and were thus highly familiar with the health state of living with migraines.

Procedure. Migraine patients were interviewed from June 1996 until May 1997, using a program on a notebook computer. The interview lasted about 10 minutes. Patients were asked the following preference question:

Imagine that when you have a migraine, it lasts a whole day (24 hours). For each of the questions below, tell us which option is more attractive.

Health State 1: You have a migraine x days per week and you feel fine the other $(7 - x)$ days per week. You live for 10 more years and then you die.

Health State 2: You have a migraine x days per week and you feel fine the other $(7 - x)$ days per week. You live for 20 more years and then you die. Which health state do you think is more attractive: 1 or 2?

Now x, the number of days with migraine, was varied via bisection to arrive at the MET preference threshold, i.e., until the patient just preferred the 10 years' duration with (x Migr) to the 20 years' duration. Next, TTO questions were asked, first for the 10 years' duration, then for the 20 years' duration. The 10 years' TTO question was:

Now imagine that you have a migraine x days per week and you feel fine the other $(7 - x)$ days per week. You live for 10 more years and then you die. Also imagine that all of your migraines could be completely eliminated, but you would live somewhat less than 10 years. For each of the questions below, tell us which option you find more attractive:

https://doi.org/10.1017/S026646230110704X Published online by Cambridge University Press

- Option 1: Migraine x days per week, live 10 years; or
- Option 2: Migraine 0 days per week, live y years.

Which option is more attractive to you: 1 or 2?

The duration y was varied to arrive at indifference. The starting duration for y was a random number within the range of 0 and 10 years. Next, y was varied via bisection. The search terminated when the final indifference duration for y was pinpointed to a 1-year difference.

For the 20 years' TTO question, the starting duration for y equaled half of the final indifference duration obtained in the 10 years' TTO question for the following reason. Once a MET preference is established, a preference reversal occurs when the TTO duration for the 20 years' question is *longer* than the TTO for the 10 years' question (see bottom line of Figure 2). If we had used a starting duration y in the 20 years' TTO question longer than the final indifference point from the 10' years TTO question, a longer indifference duration might have been induced in the 20 years' question through anchoring on the starting duration y, resulting in a preference reversal. By using a starting duration shorter than the indifference duration from the 10 years' question, we make preference reversals through anchoring less likely, improving the reliability of any remaining preference reversal.

At the end of the interview, the answers on the MET preference question and the two TTO questions were summarized on one screen, and patients were given the opportunity to correct any answer.

Results. Fifteen patients participated, but the answers of one patient were deleted due to incorrect data collection. In the remaining 14 patients, the averaged 10 years' TTO value for the various migraine health states was 0.63 with a standard deviation of 0.3 years; the averaged 20 years' TTO value for the various migraine health states was 0.54 with a standard deviation of 0.3 years. Ten patients exhibited MET preferences; thus, the MET preference rate was 10/14 or 71% at an averaged utility of 0.63. This rate appears to be even higher than the interpolated rate of about 40% at a utility of 0.63 from Figure 3. However, these rates cannot be readily compared as in the migraine sample, where the health states were worsened to obtain MET preferences. In the student sample, fixed health states were judged.

For the 10 patients with MET preferences, the averaged migraine duration x for which MET preferences occurred was 5.9 days per week with a standard deviation of 1.2 days. Migraine patients experience a much lower frequency of migraines; thus, we consider the patient preferences to be obtained for a hypothetical, albeit familiar health state.

Of the 10 patients with MET preferences, seven patients exhibited a preference reversal. In these 10 patients, the averaged 10 years TTO value for the various migraine health states was 0.57 with a standard deviation of 0.3; the averaged 20 years' TTO value for the various migraine health states was 0.47 with a standard deviation of 0.3. After seeing the summary of their answers, 2 of these 7 patients undid their preference reversal by changing their TTO responses.

Figure 4 presents MET and preference reversal rates for the migraine and student population; for comparison's sake, the numbers represent frequencies recorded before students or patients were given the opportunity to correct their answers. Considered on their own, the migraine data do not rule out that the preference reversals in patients are caused by random responses, since preference reversals were found in only 7 of 10 patients (of whom 2 undid their preference reversal). However, the number of 7 preference reversals of 10 MET preferences is close to the proportion of 77% that we observed in the student population. Thus, in combination with the very reliable preference reversals in the much larger student sample, we interpret the data in Figure 4 as suggesting that once a MET is established, preference reversals occur in the student sample as well as in the patient sample.

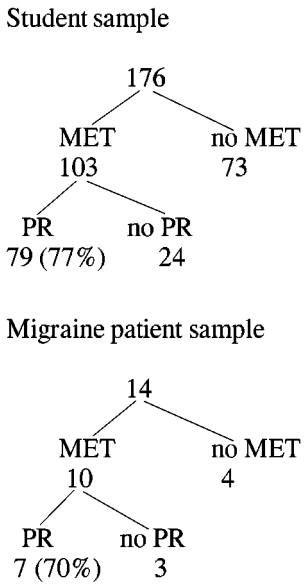

Figure 4. Comparison of MET and preference reversal frequencies in a student and a migraine patient sample. (no) MET = (no) maximum endurable time preferences; (no) PR = (no) preference reversal.

Experiment 2: Esophagectomy Patients

In the second experiment, we study the percentage of MET preferences as a function of utilities in esophageal cancer patients. We did not check for preference reversals.

Setting. The sample consisted of esophageal cancer patients participating in a randomized clinical trial. The trial compared transhiatal and transthoracic resection procedures for the treatment of esophageal cancer. The transhiatal procedure involves a resection through the abdomen and the neck, while the transthoracic procedure involves extended lymph node resection in the abdomen and the chest. The second procedure is more taxing, with the possibility of a better long-term survival.

Interview Procedure. Standard gamble utilities were collected in an interview 3 to 12 months after the resection. The interviews took place in the outpatient department of the university hospitals in Amsterdam and Rotterdam in the Netherlands when the patients were scheduled for a control follow-up. The interview was canceled if cancer had recurred. Data collection took place from February 1997 until November 1999.

Health States. Patients read unlabeled health state descriptions of eight health states related to esophageal cancer presented in random order. One of the health states involved skeletal metastases, and is described in Appendix 1. This health state is hypothetical since selected patients had not experienced recurrent cancer. In these patients, metastases are quite likely to occur since the 5-year survival is about 25%.

Standard Gamble. The probability equivalent gamble was used to elicit utilities. Patients were confronted with the following choice:

Suppose you have to choose between two options: (option A) taking a gamble with a probability p of perfect health and a probability $(1 - p)$ of dying within 1 week, or (option B) living with health state skeletal metastases. Which option would you choose?

https://doi.org/10.1017/S026646230110704X Published online by Cambridge University Press

Cards with the health state description were placed on a probability wheel that was used to visualize the probabilities and options.

Probability equivalents were obtained via a bracketing procedure that involved forced choices. The first two choices were with p set equal to 0 or 100, in random order. Next, we varied p until the patient expressed indifference between the gamble and the sure option. The starting number p was chosen randomly to minimize anchoring effects. From that starting number, we approached the indifference point via a bisection procedure. Once the indifference point was pinpointed within a final range of 5% to 10%, the patients were asked to state the indifference point.

MET Preferences. At the end of the interview, patients had to answer the following preference question, where it was made clear that death occurred at the end of the 2- and 4-year durations.

Which do you prefer:

- Option 1: living 2 years with skeletal metastases; or
- Option 2: living 4 years with skeletal metastases

Results. Responses were obtained from 29 patients. Two patients were discarded due to erratic data collection. Of the remaining 27 patients, the average age was 62 years (range, 23–79 years), while 23 were male, and 4 were female.

In these 27 patients, the averaged standard gamble utility for skeletal metastases was 0.39 with a standard deviation of 0.32. Twelve patients exhibited MET preferences; thus, the MET preference rate was 44% at a utility of 0.39. This rate is lower than the interpolated rate of about 65% at a utility of 0.4 from Figure 3 ($p = .042$ for a binomial test against a test proportion of 65%).

In Figure 5, the percentage of MET preferences is plotted as a function of the standard gamble utilities. The histogram from the patient sample is more irregular than that of the student population (Figure 3), probably due to the small sample size; the percentage of MET preferences was calculated for 10, 12, and 5 patients in the utility ranges of 0–0.33,

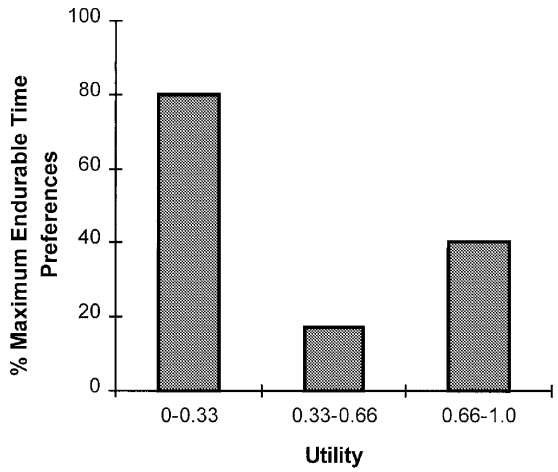

Figure 5. Percentage of esophageal cancer patients with MET preferences against standard gamble utilities for the skeletal metastases health state.

https://doi.org/10.1017/S026646230110704X Published online by Cambridge University Press

0.33–0.66, and 0.66–1.0, respectively. More MET preferences occur for utilities below 0.33 as compared to the range 0.33–0.66 ($\chi^2 = -8.9,\ p = .01$).

DISCUSSION

The data from the student sample suggest that MET preferences occur frequently for utilities smaller than 0.7. MET preferences are also found in the patient samples judging hypothetical health states related to their disease. The MET preferences are not due to random responses: the absence of MET preferences for high utilities and the abundance of MET preferences for low utilities rule out that MET preferences are a consequence of random responses, because with random responses one would expect a 50% rate of MET preferences for low and high utilities alike. In addition, preference reversals were tested and obtained in the migraine patient and student samples. However, the occurrence of preference reversals in the migraine sample should be interpreted with care in light of the small patient samples.

That MET preferences and preference reversals also occur in patient populations suggests that previously obtained MET preferences and preference reversals in the student population do not arise from students' unfamiliarity with the hypothetical health states of migraine and metastasis, and are thus not a consequence of confusion. It supports the notion that the preference reversal is caused by a cognitive reasoning heuristic (6), a rule of thumb used by many respondents, be it patients or students. It suggests that for hypothetical and poor health states, MET preferences and preference reversals will be the rule rather than the exception.

In fact, high and reliable MET preference rates in the general population have been reported by Robinson and colleagues (4). They found that 83.7% of respondents (n = 3,395) rated at least one health state worse than death in the TTO, while rating it better than dead in the Visual Analogue Scale (VAS). Their explanation is that respondents ignore the duration of the health state when completing their VAS, and that the 10-year duration of the health state is more salient in the TTO than in the VAS. Their explanation points to a MET preference: health state H with a shorter duration (VAS) is preferred to death, and death is preferred to health state H with a longer duration (TTO); hence, the shorter duration of health state H is preferred over the longer duration in H, establishing a MET preference. Robinson et al. (4) obtained somewhat lower rates of MET preferences in experiments using in-depth interviews: MET preferences were obtained in 29 of 43 respondents (67.4%, $p = .03$ for a binomial test against a test proportion of 50%).

MET Preferences and Preference Reversal

When low utilities are given to hypothetical health states, we found that MET preferences are rather frequent. However, standard QALY calculations predict that "longer is better." Thus, these standard calculations may be misleading because they misrepresent the actual preferences for the shorter duration.

One might hope to detect the presence of a MET through assessing the time trade-off weights for different durations. However, for the majority (about 75%) of the respondents with a MET, this attempt will fail as a preference reversal occurs for the time trade-off task. This preference reversal effectively hides the MET preferences when values are assessed with the time trade-off task. It is still an open issue whether other commonly used utility assessment methods, such as the gamble or the rating tasks, can pick up the existence of a MET preference.

However, the presence of a MET preference may be checked via simple preference questions as shown in the first line of the preference reversal scheme of Figure 2. By varying the durations in this simple preference question, the location S of the MET duration

https://doi.org/10.1017/S026646230110704X Published online by Cambridge University Press

may be established. These additional preference questions in the data collection phase will result in a more complete picture of the actual preferences of the respondent at the expense of a more time-consuming preference assessment.

Further research is needed to establish these phenomena and address their consequences. Whether MET preferences and preference reversals occur in patients judging their own actual poor health is a subject currently under investigation.

Finally, these results shed light on the following question: are 3 years with a weight of 0.3 equally valuable as 1 year with a weight of 0.9? Our results suggest that the 3-year period may be less valuable simply because for poor health, many may prefer a duration of 1 year to a duration of 3 years.

POLICY IMPLICATIONS

The standard QALY model is incorrect for describing preferences among poor hypothetical health states. This conclusion holds for samples from the general population and for patients. For decision alternatives involving poor health states, the standard QALY calculations overestimate the strength of preference.

REFERENCES

1. Grether D, Plott CR. Economic theory of choice and the preference reversal phenomenon. *Am Econ Rev.* 1979;69:623-638.
2. Lichtenstein S, Slovic P. Response-induced reversals of preferences in gambling: An extended replication in Las Vegas. *J Exper Psychol.* 1973;101:16-20.
3. Lindman HR. Inconsistent preferences among gambles. *J Exper Psychol.* 1971;89:390-397.
4. Robinson A, Dolan P, Williams A. Valuing health status using VAS and TTO: What lies behind the numbers? *Soc Sci Med.* 1997;45:1289-1297.
5. Stalmeier PFM, Bezembinder ThGG, Unic IJ. Proportional heuristics in time trade off and conjoint measurement. *Med Decis Making.* 1996;16:36-44.
6. Stalmeier PFM, Wakker PP, Bezembinder TGG. Preference reversals: Violations of unidimensional procedure invariance. *J Exper Psychol.* 1997;23:1196-1205.
7. Sutherland HJ, Llewelyn-Thomas H, Boyd NF, Till JE. Attitudes toward quality of survival: The concept of "maximum endurable time." *Med Decis Making.* 1982;2:299-309.
8. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res.* 1972;7:118-133.
9. Tversky A, Slovic P, Kahneman D. The causes of preference reversal. *Am Econ Rev.* 1990;80:204-217.
10. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med.* 1977;296:716-721.

APPENDIX A

EXAMPLE FOR THE HEALTH STATE "LIVING WITH RECURRENT DISEASE INVOLVING SKELETAL METASTASES," MALE VERSION

This patient had surgical resection for an esophageal tumor. The illness returned in his bones. He has pain and will receive radiotherapy. He has no appetite and is tired. He is shocked by the return of the illness and he is gloomy about the future. He has social support from those nearby.