Medical Decision Making

http://mdm.sagepub.com/

The QALY Model and Individual Preferences for Health States and Health Profiles over Time: A Systematic Review of the Literature

Aki Tsuchiya and Paul Dolan

Med Decis Making 2005 25: 460
DOI: 10.1177/0272989X05276854

The QALY Model and Individual Preferences for Health States and Health Profiles over Time: A Systematic Review of the Literature

Aki Tsuchiya, PhD, Paul Dolan, D. Phil

The numbers of quality-adjusted life years (QALYs) gained are increasingly being used to represent the gains in individual utility from treatment. This requires that the value of a health improvement to an individual is a simple product of gains in quality of life and length of life. The article reports on a systematic review of the literature on 2 issues: whether the value of a state is affected by how long the state lasts, and by states that come before or after it. It was found that individual preferences over health are influenced by the duration of health states and their sequence. However, although there is much variation across individual respondents, the assumptions tend to hold much better when valuations are aggregated across respondents, which is encouraging for economic evaluations that rely on using average (mean or median) values. ***Key words:*** *QALYs; individual preferences; utility measurement.* ***(Med Decis Making 2005;25:460–467)***

Because people experience health benefits as improvements in their quality of life and/or as increases in their length of life, the quality-adjusted life year (QALY) attempts to combine the value of these attributes into a single index number. At a broad conceptual level, the value of a QALY is the value of 1 year spent in full health. This is then taken as a benchmark value against which all other health profiles (of whatever duration, in whatever combination through time) are valued. However, because there is an infinite number of such combinations of health states, establishing the benchmark value of each in QALY terms would be quite impractical, and some simplifying assumptions are introduced.

In this article, we present the results from a systematic review of the literature that was designed to examine the extent to which people's preferences satisfy some of the key assumptions of the QALY model explained below. Our aim has not been to be prescriptive about which elements of the QALY approach *should be* adhered to, and we leave it for others to make their judgments about the normative significance of some of our findings. We also consider this to be a review of empirical tests of QALY assumptions and, although we present a summary of study design (such as the sample size and composition and the country of origin), we have made no attempt to assess the quality of empirical studies. Because people's preferences are so heavily influenced by the ways in which questions are put to them, it has not really been possible to systematically assess the quality of the empirical evidence. For instance, there are no obvious criteria that allow us to rank between a marginally poorly designed postal survey with a large and representative sample and a marginally better designed interview with a small and nonrepresentative sample. This is in contrast to trial evidence, for example, where the criteria for assessing the quality of studies are well established.

In what follows, section 2 sets out the QALY model and the assumptions that are tested here. Section 3 describes how the systematic review was undertaken and presents some summary data about the studies included in the review. Section 4 discusses the empirical evidence, and section 5 provides a summary of the findings and some conclusions.

DOI: 10.1177/0272989X05276854

Received 17 December 2003 from Centre for Well-being in Public Policy, School of Health and Related Research, University of Sheffield (AT), and Centre for Well-being in Public Policy, University of Sheffield (PD). Revision accepted for publication 4 October 2004.

Address correspondence to Aki Tsuchiya, CWiPP, School of Health and Related Research, 30 Regent Street, Sheffield, S1 4DA; phone: 0114.222.0710; fax: 0114.272.4095; e-mail: a.tsuchiya@sheffield.ac.uk.

THE QALY MODEL AND ITS ASSUMPTIONS

In the simplest case, with no uncertainty, no temporal discounting, and no changes in health over time, the value of a health gain from treatment for an individual, $QALY_G$, can be represented as

$$QALY_G = T_1 Q_1 - T_0 Q_0, \qquad (1a)$$

where T is the number of years of survival, Q represents health state values, and the subscripts 1 and 0 represent health with and without treatment, respectively.[1]

Alternatively, introducing uncertainty and temporal discounting, and assuming discrete time so that changes in health occur only when moving from one period to the next, the expected net gain of a treatment to any one individual can be expressed as

$$QALY_G = \Sigma_h \Sigma_t p_{1ht} Q_{ht} - \Sigma_h \Sigma_t p_{0ht} Q_{ht}, \qquad (1b)$$

where p_{1ht} and p_{0ht} represent the probabilities of an individual finding himself in health state h in time period t with and without treatment, respectively. Q_{ht} is the value of health state h at time t (the subscript t here allows for constant rate temporal discounting so that $Q_{ht} = \frac{Q_h}{(1+r)_t}$, where r is the discount rate).

This algorithm—the QALY model—is an expression of the value to an individual associated with a given intervention. If the quality of life associated with "full health" were to be assigned a value of 1, then the algorithm could be considered to express health gains measured in "objective physical units," that is, life years. Any state of health less than this is adjusted for its quality and hence assigned a lower value. Against a background based on expected utility theory, Pliskin and others first set out a set of sufficient assumptions for this simple model to represent individual utility over health states and duration.[2] For health profiles of constant quality (i.e., "chronic" states), these are mutual utility independence between quality of life and duration, constant proportional tradeoff, and risk neutrality over life years. Bleichrodt and others presented a smaller set of sufficient assumptions consisting of risk neutrality and the "zero condition" (which implies that for a duration of zero life years, all health state values are equivalent).[3] Alternatively, Miyamoto and others further demonstrated that, with nonlinear utility functions, this set becomes the zero condition and "standard gamble invariance" a special case of the utility independence of duration of survival from quality of life.[4]

On the other hand, besides risk attitude and time preference (which are issues not specific to health and QALYs), the empirical literature on whether the QALY model holds has typically addressed one or both of the following 2 questions:

1. Is the value of a state affected by how long the state lasts?
2. Is the value of a state affected by the states that come before or after it?

The first question is related to 3 concepts: utility independence of quality of life from the duration of survival, constant proportional time tradeoff, and maximum endurable time. The 1st 2 terms both mean that the value of a health state is independent of its duration.[2,5,6] *Utility independence* here means that values elicited using Standard Gamble (SG) with some given fixed duration are unaffected by this specific choice of duration. *Constant proportional time tradeoff* means that values elicited using time tradeoff (TTO) are not affected by duration, that is, the same proportional amount of time is traded off independently of the absolute duration presented in the scenario. When these concepts are applied to VAS (Visual Analogue Scale) values, they mean that VAS scores should not be affected by how long the state lasts. *Maximal endurable time* means that, for some severe states (independent of which method is used to value them), the value of those states becomes negative after some threshold duration. When maximal endurable time takes effect, utility independence and constant proportional time tradeoff are violated.

The 2nd question can be broken down into 2 issues. The 1st is whether or not *additive separability* holds; that is, the value of a health state should be independent of what precedes or follows it.[7] Under zero discounting, additive separability means that the value of a complete health profile would be equal to the sum of the value of individual health states that make up that profile, irrespective of the order of the states. Obviously, the present value of the 2 profiles will not coincide under nonzero discounting, but then, the difference should be a function of a positive discount rate alone. The 2nd is whether or not *preference independence* holds. This requires that "given two profiles that have the same health state during interval i, preference between them does not depend on the level of health during interval i."[8] The testing of this concept does not rely on any assumptions concerning time preference. However, there is a large literature in experimental psychology that addresses the issue of how people's perceptions are affected by "troughs and peaks," or sequence effects (see ref no. 9, which includes a brief review). This suggests that additive separability and preference independence in the context of QALYs may not be satisfied.

Thus, this article reports on a systematic literature review on these 2 questions. As can be seen, these 2 questions are taken from the set of sufficient assumptions set out in the literature. As they are individual assumptions within sets of sufficient assumptions, demonstrating that any one of these is satisfied individually will not validate the QALY model (although it may count as an additional piece of evidence in favor of the validity of the model). On the other hand, because each of these assumptions is also a necessary assumption, demonstrating that any of these is not satisfied has the potential to invalidate the QALY model.

THE LITERATURE SEARCH

The aim of the search strategy was to identify systematically all issues relating to the 2 topics of the review through the retrieval of published and unpublished papers. A method called "citation pearl growing"[10] was employed, using the citation search facility of the Institute of Scientific Information (ISI) citation indexes and through reference list searching (as also used in ref no. 1). These databases cover the science (including biomedical science), social sciences (including economics), and arts and humanities literature. In addition, the Web sites, publication lists, and research registers of relevant organizations were searched and relevant experts were consulted. The process of citation searching begins from an initial list of relevant references, which were put together from the authors' own collections. Finally, a keyword search strategy was developed, based on the indexing terms of included studies, to check the completeness of the primary search method. The search was restricted to papers in the English language, dated 2002 or earlier.

After 3 rounds of searching, no additional unique references were retrieved. These searches provided 601 references. Using the titles and abstracts of retrieved references, the 1st author undertook the 1st stage of assessment for inclusion, and then the 2nd author checked a sample of the references excluded at this stage. Full papers were assessed for inclusion independently by both authors. Through this process, 71 papers were identified as relevant, including 20 with empirical data. Table 1 provides information on the empirical studies, in terms of study design, sample population, and sample size. It can be seen from this table that most of the empirical studies have used structured interviews with students or patients, and have often had sample sizes less than 100. The table also shows the country in which the study was conducted. Most of the studies have been carried out in North America, followed by the UK and Europe.

EMPIRICAL EVIDENCE ON THE QALY MODEL

Is the Value of a State Unaffected by How Long the State Lasts?

Utility Independence

McNeil and others interviewed 37 volunteers (25 executives and 12 firefighters) and asked them to value speech loss for various lengths of time.[11] They found that although respondents, on average, accepted a 14% risk of death to avoid speech loss, none accepted a positive risk of death when survival was shorter than 5 years. Bleichrodt and Johannesson asked 172 students to fill in a questionnaire with SG questions of 10- and 30-year durations, followed by death.[6] The authors concluded that utility independence is violated at the aggregate level, with 10-year SG values higher than 30-year SG ones. Bala and others interviewed 114 elderly people using 20-year SG and a 1-year SG, both followed by death.[12] About 25% satisfy utility independence, but there is no systematic pattern in the responses of those who do not.

Constant Proportional Time Tradeoff

Several studies have shown constant proportional time tradeoff to be a pretty good approximation of preferences at the aggregate level. In a questionnaire survey, Pliskin and others asked 10 respondents (physicians, economists, and statisticians) the number of years they will sacrifice to avoid severe or mild angina pain.[2] The same question was asked with 5-year and 15-year survival baselines. At the individual level, most respondents violated constant proportional time tradeoff. However, at the aggregate level, there is little difference between the tradeoffs from the 5-year TTO and the 15-year TTO. Cook and others interviewed more than 500 patients with gallstone disease, and at the aggregate level, tradeoffs for states lasting 12 months and 12 years followed by death were not significantly different from one another.[13] Bleichrodt and Johannesson found that 10-year TTO and 30-year TTO values (followed by death) did not differ from one another at the aggregate level.[6]

However, other studies have shown constant proportional time tradeoff to be violated. All of these except the one by Unic and others have found that shorter periods of time are associated with less tradeoffs (i.e.,

Table 1 Empirical References

Author(s)	Year	Reference No.	Design	Sample	Sample Size	Country of Study	Assumptions Tested	Technique Used
Bala et al.	1999	12	SI	GP (c)	114	USA	UI	SG
Bleichrodt, Johannesson	1997	6	SQ	S	172	Sweden, the Netherlands	UI, CPT	SG, TTO
Buckingham et al.	1996	18	PQ	GP (r)	1500+	UK	CPT	TTO
Cook et al.	1994	13	SI	P	500+	Australia	CPT	TTO
Dolan	1996	19	SI	GP (r)	236	UK	CPT	TTO, VAS
Krabbe, Bonsel	1998	25	E	S	104	the Netherlands	AS	TTO
Kuppermannn et al.	1997	24	SI	P	121	USA	AS	SG, VAS
MacKeigan et al.	1999	26	SI	P	89	Canada	AS	VAS, TTO
McNeil et al.	1981	11	SI	GP (c)	37	USA	CPT	SG
Miyamoto, Eraker	1988	16	SI	P	64	USA	CPT	TTO
Pliskin et al.	1980	2	SI	HP	10	USA	CPT	TTO
Richardson et al.	1996	23	SI	GP (c)	63	Australia	AS	VAS, TTO, SG
Sackett, Torrance	1978	15	SI	GP (r) P	246 29	Canada	CPT	TTO
Spencer	2000	27	SI	GP (c)	29	UK	AS	SG
Stalmeier et al.	1997	17	SI	S	176	the Netherlands	CPT	RP, TTO
Stalmeier et al.	1996	22	SI	S	86	the Netherlands	MET	TTO
Sutherland et al.	1982	21	SI	HPA	20	Canada	MET	SG
Treadwell	1998	8	SQ	S	163	USA	PI	RP
Treadwell et al.	2000	29	SQ	P	67	USA	PI	RP
Unic et al.	1998	14	SI	GP (c)	54	the Netherlands	CPT	TTO

Note: PQ = postal questionnaire; SQ = self-completion questionnaire; SI = structured interview; E = experiment. GP (r) = general public (random/quota); GP (c) = general public (convenience); S = students; P = patients; HPA = health professionals or academic staff. UI = utility independence; CPT = constant proportional time tradeoff; MET = maximal endurable time; PI = preference independence; AS = additive separability. SG = standard gamble; TTO = time tradeoff; VAS = visual analogue scale; RP = ranking or pairwise choice.

higher implied health state values).[14] Sackett and Torrance interviewed 246 members of the public and 29 patients on home dialysis.[15] They asked respondents to value 15 scenarios covering various health conditions from tuberculosis to kidney transplant, with durations of 3 months, 8 years, and the life expectancy of a respondent, all followed by death. They found that values declined with duration. In a study on utility independence of duration on quality of life where 64 hospital inpatients with a range of conditions were interviewed, Miyamoto and Eraker also explored constant proportional time tradeoff, and they reported that about 25% of respondents did not trade off any time to improve their current health when the duration was under 1 year, whereas time was traded off when the duration was more than a year.[16]

Stalmeier and others asked 4 groups of university and high school students (total respondents 176) to rank 2 scenarios, one living for a longer time with a severe health condition and dying, and another living for a shorter time with the same health condition and dying.[17] The proportion of those who ranked the shorter scenario over the longer one varied from 44% to 71%. The vast majority of these (73% to 94%) displayed a *preference reversal*, where their TTO value for the shorter scenario was lower than that for the longer scenario. Furthermore, regarding those respondents whose preferences were not reversed, the authors go on to discuss the possibility of a "proportional heuristic" in the TTO. When respondents are asked to give the number of healthy years that is equivalent to living in a given state for 10 years, and then the same for 20 years, respondents may give proportional answers not because they satisfy constant proportional time tradeoff but because they see that the nummeraire of the exercise has been doubled. Because this indicates that certain tests of constant proportional time tradeoff may be too easy to pass, this has important implications for earlier studies that demonstrated satisfaction of this requirement.

Buckingham and others conducted a postal survey of more than 4000 members of the public, with more than 1500 usable replies.[18] They reported aggregate results from 3 different TTO formats for a condition that lasts for the rest of one's life. These were a daily TTO, which was about trading off the number of hours awake per day; a yearly TTO, which was about trading off the number of active days per year; and a lifetime TTO, which was about trading off years of life expectancy. Assuming that time spent sleeping, "lost" days, and lost years are all valued at zero, constant proportional time tradeoff will require that the proportion of a day that is traded off is equal to the proportions of a year and of a lifetime that are traded off in exchange for full health. The study found that the yearly values are the highest and the daily values the lowest. This suggests that the relationship between the length of the period and the size of the tradeoff may not be linear. There has been one study that has looked at the effect of duration on VAS responses. From interviews with 236 members of the general public, Dolan reports values for health states lasting for 1 month, 1 year, and 10 years, "and what happens thereafter is not known and should not be taken into account."[19] In general, the shorter the duration, the higher the value. Olsen has presented a method whereby positive implicit time preference rates can be derived for such responses.[20]

Maximal Endurable Time

Sutherland and others interviewed 20 health professionals (physicians, biophysicists, biologists) and asked them to value 7 states, each lasting for 3 months, 8 years, and the respondent's life expectancy, each followed by death, using the SG.[21] They were also asked for the preference between each scenario and death. A maximal endurable time was observed for up to 75% of respondents, depending on the health state. The worse a health state was considered to be, the more respondents indicated maximal endurable time. Stalmeier and others asked 3 groups of female university and high school students (totaling 86 respondents) to value breast cancer–related health states.[22] Fifty-eight percent indicated maximal endurable time such that 25 years with metastasized breast cancer (implicitly followed by death) was preferred to 50 years in the same state (again implicitly followed by death). However, 74% of these also indicated preference reversals in TTO such that the number of healthy years equivalent to 25 years with metastasized breast cancer was proportionally smaller than that for 50 years in the same state.

Is the Value of a State Unaffected by the States That Come before or after It?

Additive Separability

Richardson and others interviewed 63 women who did not have breast cancer to value 4 breast cancer–related health scenarios using VAS, TTO, and SG.[23] Three scenarios consisted of a single health state, whereas the last one was a profile combining these 3 states in deteriorating order followed by death. Using a 3% and a 9% discount rate, they found that the number of QALYs calculated indirectly from the individual health states was 30% to 50% higher than number of QALYs calculated from the direct value of the profile. The authors argue that "the knowledge of future death casts a shadow over, or devalues, the enjoyment of earlier life years." Thus, there is the possibility that the results are driven by the dread of suffering and death at the end of the scenario in addition to a systematic violation of the additivity assumption.

Kuppermann and others interviewed 121 pregnant women and asked them to value (using VAS and SG) 8 "paths," involving 2 prenatal diagnostic tests for chromosomal abnormalities of the fetus at different stages of the pregnancy, different test results, and outcomes including spontaneous abortion of the fetus possibly related to the test and the effect on the woman's fertility afterward.[24] The paths were then broken down into discrete states, and the direct valuation of the paths was compared to the indirect values calculated from the values of the discrete states, assuming no temporal discounting. At the individual level, preferences were not additive, and there does not seem to be any obvious pattern. At the aggregate level, the mean direct value could be predicted from the mean values of the discrete states, but this was not by means of an additive model weighted by duration, as suggested by the additivity assumption. The results were not affected by the introduction of a 5% discount rate. In general terms, the indirect values of the paths tended to be higher than the direct values, including the case where the path was not a deteriorating one.

Krabbe and Bonsel asked 104 (mostly medical) students to value 13 hypothetical health states on 2 separate occasions using the TTO.[25] The health states lasted for 10 years. On the 1st occasion, the respondents were given 2 alternatives, one of living in a fixed state (EQ5D state 21232) and the other of living for x years in the "best imaginable" state followed by (10-x) years in the "worst imaginable" state. On the 2nd occasion, the 2nd alternative was changed to live for z years in the "worst imaginable" state followed by (10-z) years in the best imaginable state. Under both formats, after the 10-year period, health was to return to the current level. If additive separability holds, then, with appropriate discounting, the number of years spent in the best health state in the 2 scenarios should coincide. This held for two thirds of respondents when a discount rate of 5% was used for everybody. Thus, on the one hand, by allowing for individual discount rates, a higher proportion of respondents may have achieved convergence of the numbers of years. On the other hand, there is also the possibility that the discount rate that makes the numbers of years converge may not reflect the genuine temporal preference of the individual, in which case two thirds could be an overestimate. A small proportion of the remaining wanted "best things first," whereas the majority wanted a "happy ending."

Mackeigan and others interviewed 89 patients with type 2 diabetes.[26] Nine scenarios, covering 30 years and followed by death, consisting of diet therapy, insulin use, 3 "mono" therapies, 3 "dual" therapies, and 1 "triple" therapy, were valued using VAS and TTO. The study found that the indirect and direct values of the combination therapies were not statistically significantly different from one another. However, the agreement between the 2 approaches was poor, suggesting that the differences between the health states may have been too small to invoke the sequence effect. Spencer conducted interviews with 29 members of the public that tested for additive separability in 2 ways while controlling for risk attitude and time preference.[27] In the 1st test, using the SG method, the *difference* between profiles *x-y* and *x-z* was compared to the *difference* between profiles *w-y* and *w-z*, where all profiles lasted 10 years and were followed by death. The differences were statistically significant, thus violating additive separability. The 2nd test was first proposed by Bleichrodt[28] and consists of a choice between 2 gambles: one offers a 50-50 chance of the best and worst health states, and the other involves a 50-50 chance of the best-then-worst profile and the worst-then-best profile. The respondents were split roughly in half, 13 preferring the former gamble and 15 preferring the latter, whereas 1 was indifferent. This suggests a violation of additive separability, but it is not systematic and so could, in the extreme, simply represent noise in the valuation process.

Preference Independence

Treadwell presented 163 psychology students with pairwise choices of health profiles.[8] Each combination consisted of 2 scenario pairs: *A* with *B* and *A′* with *B′*, all with a 30-year duration followed by death, constructed such that independence is satisfied when a respondent who prefers *A* (*B*) in the 1st pair also prefers *A′* (*B′*) in the 2nd pair. The author concludes "independence was more commonly satisfied than it was violated." Of 42 combinations tested, the requirement was satisfied in 36. Treadwell and others asked 67 outpatients with type-C hepatitis to fill out a questionnaire that asked them 6 pairwise choices of health profiles.[29] The profiles were either both followed by "normal" health or both ended in death. About two thirds of respondents satisfied independence. However, when respondents were asked to give reasons for their choices, explanations implying sequence effects were observed, for example, to "get [bad states] out of the way" or to have a relatively good state before death.

CONCLUSIONS

Let us summarize the empirical evidence relating to the 2 questions posed at the beginning:

1. Preferences over different health states when they are valued using different fixed durations. There have been 2 empirical studies addressing utility independence of SG responses from duration. The respondents in these studies did not satisfy this, although there is no clear pattern in the violations. There have been 8 studies that have looked at whether constant proportional time tradeoff holds for TTO responses. In general, the results suggest that the assumption holds at the aggregate level but is violated (albeit in a largely nonsystematic way) at the individual level. Shorter durations typically have higher values, and longer durations are sometimes associated with a maximal endurable time, after which time death is preferred to additional survival in the state.
2. Preferences over profiles of different health states. The 5 studies that have addressed additive separability suggest that this requirement does not hold, but we cannot really point to any clear systematic violations. Two studies have addressed preference independence, and both found that the majority of respondents satisfy the requirement.

Thus, contrary to the assumptions of the QALY model, it would seem that an individual's preferences over health are influenced by the duration of health states and their sequence. Given that each of these are necessary conditions for the QALY model to hold, they cast serious doubt to the validity of the QALY model as a representation of individual utility with respect to their own health. Unfortunately, none of these factors appears to impact the QALY model in a straightforward way and so it is not possible at this stage to provide a simple algorithm to adjust the QALY model to better represent individual preferences over own health. However, there have been 2 developments to generalize the QALY model to overcome known and systematic violations. The 1st is the HYE (health years equivalents), introduced by Mehrez and Gafni.[30] Mehrez and Gafni argue that the standard QALY concept is flawed because, although the quality adjustment component of the QALY is preference based, the life year component is not. To reflect this, they proposed the HYE, which is based on measuring the value of whole profiles directly, as opposed to constructing this through values of individual states. Therefore, it does not require the additive separability assumption or preference independence.[31–35] However, its major practical disadvantage is that it is virtually impossible to estimate a value set for all possible profiles, given the infinite number of profiles there would be.

The 2nd development concerns generalizations of expected utility theory. The theory has offered the main theoretical background to the QALY model, and yet the extent to which individual choice behavior violates its axioms is well documented. The new developments base the QALY model on, for instance, rank-dependent expected utility theory.[36–38] This line of research consists of identifying theoretical models that satisfy both some notion of what is rational and real choice behavior, to better explain the way the human mind behaves when faced with choices regarding health. However, it should also be noted that expected utility theory could remain as the theoretical basis on which to make policy choices, even if actual individual choices violate their axioms. Or, in other words, the particular notion of rationality that best fits real individual behavior does not have to be the one that forms the basis for policy choices.

It should also be noted that, once we turn to putting the numbers to policy use (as opposed to positive uses), it is usually not the individual preferences but the aggregate (mean or median) preferences that are applied. Although not all studies report whether or not aggregate preferences satisfy the assumptions of the QALY model, when they are reported, they appear to perform much better than individual preferences. Moreover, many of the violations at the individual level do not follow a systematic pattern, that is, some people violate an axiom in one direction and others violate it in another direction, which might simply represent noise in the valuation process. Ultimately, it is a matter of judgment about whether the inability of the QALY model to accurately represent all individual preferences is compensated for by the fact that it more accurately represents aggregate preferences.

ACKNOWLEDGMENTS

This review was funded by the NHS Executive R&D Programme, Health Technology Assessment Grant, on The Validity of Aggregation Methods in Cost-Utility Analysis. We would like to thank Suzy Paisley and Rebecca Shaw for their assistance with the systematic review, and Gretchen Chapman, John Miyamoto, Jan Abel Olsen, Alan Williams, and an anonymous referee for their helpful comments. The usual disclaimer applies.

REFERENCES

1. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. Health Econ. 2004; forthcoming.

2. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. Operations Res. 1980;28(1):206–24.

3. Bleichrodt H, Wakker P, Johannesson M. Characterizing QALYs by risk neutrality. J Risk Uncertainty. 1997;15(2):107–14.

4. Miyamoto JM, Wakker PP, Bleichrodt H, Peters HJM. The zero-condition: a simplifying assumption in QALY measurement and multiattribute utility. Manage Sci. 1998;44(6):839–49.

5. Loomes G, McKenzie L. The use of Qalys in health-care decision-making. Soc Sci Med. 1989;28(4):299–308.

6. Bleichrodt H, Johannesson M. The validity of QALYs: an experimental test of constant proportional tradeoff and utility independence. Med Decis Making. 1997;17(1):21–32.

7. Broome J. Qalys. J Public Econ. 1993;50(2):149–67.

8. Treadwell JR. Tests of preferential independence in the QALY model. Med Decis Making. 1998;18(4):418–28.

9. Ariely D, Loewenstein G. When does duration matter in judgment and decision making? J Exp Psychol Gen. 2000;129(4):508–23.

10. Hartley RJ. Online Searching: Principles and Practice. New Providence, NJ: Bowker-Saur; 1990.

11. Mcneil BJ, Weichselbaum R, Pauker SG. Speech and survival: trade offs between quality and quantity of life in laryngeal cancer. N Engl J Med. 1981;305:982–7.

12. Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A, O'Brien BJ. Are health states "timeless"? The case of the standard gamble method. J Clin Epidemiol. 1999;52(11):1047–53.

13. Cook J, Richardson J, Street A. A cost-utility analysis of treatment options for gallstone disease—methodological issues and results. Health Econ. 1994;3(3):157–68.

14. Unic I, Stalmeier P, Vefhoef LvDW. Assessment of the time-trade-off values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. Med Decis Making. 1998;18(3):268–77.

15. Sackett D, Torrance G. The utility of different health states as perceived by the general public. J Chronic Dis. 1978;31:697–704.

16. Miyamoto JM, Eraker SA. A multiplicative model of the utility of survival duration and health quality. J Exp Psychol Gen. 1988;117(1):3–20.

17. Stalmeier P, Wakker P, Bezembinder T. Preference reversals: violations of unidimensional procedure invariance. J Exp Psychol Hum Percept Perform. 1997;23(94):1196–205.

18. Buckingham JK, Birdsall J, Douglas JG. Comparing three versions of the time tradeoff: time for a change? Med Decis Making. 1996;16(4):335–47.

19. Dolan P. Modelling valuations for health states: the effect of duration. Health Policy. 1996;38(3):189–203.

20. Olsen JA. Persons vs years—2 ways of eliciting implicit weights. Health Econ. 1994;3(1):39–46.

21. Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. Attitude toward quality of survival: the concept of maximal endurable time. Med Decis Making. 1982;2:299–309.

22. Stalmeier PFM, Bezembinder TGG, Unic IJ. Proportional heuristics in time tradeoff and conjoint measurement. Med Decis Making. 1996;16(1):36–44.

23. Richardson J, Hall J, Salkeld G. The measurement of utility in multiphase health states. Int J Technol Assess Health Care. 1996;12(1):151–62.

24. Kuppermann M, Shiboski S, Feeny D, Elkin EP, Washington AE. Can preference scores for discrete states be used to derive preference scores for an entire path of events? An application to prenatal diagnosis. Med Decis Making. 1997;17(1):42–55.

25. Krabbe PFM, Bonsel GJ. Sequence effects, health profiles, and the QALY model: in search of realistic modeling. Med Decis Making. 1998;18(2):178–86.

26. MacKeigan LD, O'Brien BJ, Oh PI. Holistic versus composite preferences for lifetime treatment sequences for type 2 diabetes. Med Decis Making. 1999;19(2):113–21.

27. Spencer A. Testing the Additive Independence Assumption in the QALY Model. 427. 2000. London, Queen Mary, University of London, Department of Economics Discussion Paper.

28. Bleichrodt H. Qalys and Hyes—under what conditions are they equivalent. J Health Econ. 1995;14(1):17–37.

29. Treadwell JR, Kearney D, Davila M. Health profile preferences of hepatitis C patients. Digest Dis Sci. 2000;45(2):345–50.

30. Mehrez A, Gafni A. Quality-adjusted life years, utility-theory, and healthy-years equivalents. Med Decis Making. 1989;9(2):142–9.

31. Mehrez A, Gafni A. The healthy-years equivalents—how to measure them using the Standard Gamble approach. Med Decis Making. 1991;11(2):140–6.

32. Buckingham K. A note on hye (healthy years equivalent). J Health Econ. 1993;12(3):301–9.

33. Gafni A, Birch S, Mehrez A. Economics, health and health economics - Hyes versus Qalys. J Health Econ. 1993;12(3):325–39.

34. Johannesson M. Quality-adjusted life-years versus healthy-years equivalents—a comment. J Health Econ. 1995;14(1):9–16.

35. Dolan P. A note on QALYs versus HYEs—health states versus health profiles. Int J Technol Assess Health Care. 2000;16(4):1220–4.

36. Wakker P, Stiggelbout A. Explaining distortions in utility elicitation through the rank-dependent model for risky choices. Med Decis Making. 1995;15(2):80–6.

37. Bleichrodt H, Quiggin J. Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis? J Health Econ. 2003;18(6):681–708.

38. Miyamoto J. Quality-adjusted life years (QALY) utility models under expected utility and rank dependent utility assumptions. J Math Psychol. 1999;43(2):201–37.