A Head-to-Head Comparison of the Standard Quality-Adjusted Life Year Model With the Annual Profile Model

Janssen, Mathieu F.; Birnie, Erwin; Bonsel, Gouke J.

Published in:
Value in Health

DOI:
10.1016/j.jval.2020.11.021

Publication date:
2021

ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Methodology

A Head-to-Head Comparison of the Standard Quality-Adjusted Life Year Model With the Annual Profile Model

Mathieu F. Janssen, PhD, Erwin Birnie, PhD, Gouke J. Bonsel, MD, PhD

ABSTRACT

Objectives: The standard quality-adjusted life year (QALY) model (SQM) assumes time–utility independence within constant health states and additive independence when health varies over time. The validity of SQM has been challenged through reported violations of these assumptions. An alternative approach that relaxes these assumptions is to assign a single valuation to an entire health profile: an integral assessment of disease severity over time. Here, we compare SQM with the annual profile model (APM) and test SQM for additive independence.

Methods: Eighty-two respondents valued 6 episodic conditions, including 4 of short duration, with SQM and APM, using the time trade-off method. Inter-rater reliability was assessed using intraclass correlation coefficients. Face validity was tested by asking respondents how well they were able to imagine the health states under SQM and APM. We calculated SQM QALY values for a 1-year time period, allowing for a direct comparison with APM values. For the short-term conditions we expected higher QALY values for SQM, violating additive independence.

Results: APM showed higher interrater reliability (intraclass correlation coefficient of 0.53 vs 0.18, respectively) and better face validity than SQM, with 6% (APM) vs 21% (SQM) of all respondents reporting difficulties. Additive independence of SQM was violated in 5 of the 6 conditions (including the 4 short duration health states), with higher QALY values under SQM (mean difference 0.04).

Conclusion: The impact of short-term conditions is systematically underestimated under SQM when compared to a health profile model. APM is a less restrictive model and demonstrates better validity.

Keywords: methodology, quality-adjusted life-years, short-term diseases, utility measurement.

VALUE HEALTH. 2021; 24(5):707–713

Introduction

The quality-adjusted life year (QALY) has become the standard outcome measure in the economic evaluation of healthcare programs and interventions. The standard QALY model (SQM) operates under at least 2 strong assumptions. First, when morbidity (disease severity) is constant over a time interval, time–utility independence is assumed, that is, disease duration and the utility of disease severity are assumed to be independent of each other. SQM is multiplicative for constant health over time (T) and the utility value (U) of the health state (Q) equals: $U(Q, T) = T \times U(Q)$, hence $U(T) = T$. This assumption is usually referred to as mutual utility independence when QALYs are estimated with the standard gamble method, and as constant proportional trade-off when the time trade-off method is used.[1-3] Second, when health varies over time, additive independence is assumed, which means that the utility of a health state in period T_1 is independent of the utility of another health state in period T_2.[3-6] Additive independence, a special form of time-utility independence, implies that QALYs are calculated by adding the utility values of the different health states over time: $U(Q_A+Q_B) = U(Q_A) + U(Q_B)$. For a health profile or sequence of health states over time period T, $U(Q_1; Q_2; \ldots; Q_T)$, the number of QALYs according to these assumptions is denoted by:

$$U(Q,T) = \sum_{t=1}^{T} U(Q_t)$$

Figure 1 shows how QALYs are calculated under SQM, for two 5-day periods of severely poor health (eg, extreme pain, severe depression or anxiety), in an otherwise healthy year. Assume the severe health state is assigned a utility value of 0.1. The total number of QALYs under SQM can be represented by the area under the curve. Within each health state, the utility values are multiplied by the time duration (ie, assuming time–utility independence). Subsequently, the values for the separate health

A previous version of this article has been published as part of the thesis: M.F. Janssen, "Sense and Sensitivity of Summary Measures of Health" (2008), PhD Thesis, Faculty of Medicine, University of Amsterdam, Amsterdam, The Netherlands.

Figure 1. Standard quality-adjusted life year model calculations for 2 short, severe episodes, each lasting 5 days.

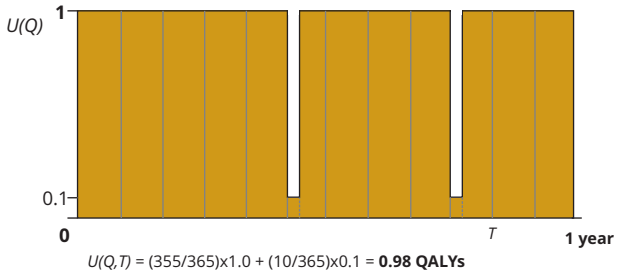


Table 1. The 6 conditions with EQ-5D-5L+C description and duration.

Condition	Annual profile (days)	EQ-5D-5L+C
Acute tonsillitis, 1 week	7	123411
Concussion, 2 weeks	14	224313
Migraine, monthly recurrent	12×1	124413
Severe depression, 6 months	183	135353
Severe gastrointestinal disorder, 10 days	10	145412
Wrist fracture, 4 months	122	144211

states are added (ie, assuming additive independence), amounting to 0.98 QALYs, which is nearly perfect health. It is questionable whether a patient (or a proxy or clinical expert) will view this year as nearly perfect health. Examples are epilepsy and recurrent psychotic episodes, which are regarded as serious conditions that one would assume to be reflected in its corresponding utility values. As the 2 assumptions preclude any interaction between disease severity and disease duration, SQM might lead to an overestimation of QALYs in such temporary severe conditions; this in turn questions the validity of SQM for these conditions. Although there is some evidence that time-utility independence and additive independence hold at the aggregate level,[1-2,7] numerous studies conclude that the assumptions are violated at both the individual and aggregate level.[2,8-19]

A health (time) profile approach has been suggested as an alternative to SQM. This approach obtains direct, integral valuations as a way of relaxing these restrictive assumptions.[14,20-23] Although the theoretical superiority of profile models to SQM is widely acknowledged,[5,19-21,23-25] SQM is still the dominant practice to estimate QALYs, probably because SQM values are easier to measure and arguably easier to implement into cost-effectiveness models. After the introduction of the healthy years equivalent (HYE) method, perhaps the first time profile method published,[26] it is remarkable that there has been so little empirical work to assess the value of profile models in comparison to SQM. To empirically demonstrate that a profile approach is a valid and tenable alternative to SQM, 3 criteria need to be met: (1) the alternative to the QALY model should be at least as robust as SQM in terms of feasibility and reliability; (2) the new model should produce systematic differences in outcome values with SQM, related to expected violations of the SQM assumptions; (3) the direction of these differences in QALY values can be predicted, that is, QALY values obtained by the new model are systematically lower or higher than values obtained by SQM. Earlier we demonstrated good to excellent feasibility and reliability (criterion 1) of the annual profile model (APM) for a set of 46 disease stages, of which 36 were episodic.[27] In this study, we compare SQM with APM for 6 episodic diseases, including 4 of short duration. Comparisons are made in terms of inter-rater reliability, face validity, and additive independence of SQM (criterion 2). We hypothesize that for the 4 short conditions, additive independence does not hold and SQM values will be higher than the APM values (criterion 3). Based on available evidence, SQM values are also expected to be higher than APM values for the 2 conditions of longer duration, but maybe not significantly so.[14,16-17,28]

Methods

Annual Profile Vignettes

From the available health profile formats, we selected the annual profile model as reference, taking a 1-year perspective as time horizon.[14,18,26,29,30] We used 6 conditions for the SQM-APM comparison (see Table 1) of which 4 conditions of short duration (7-14 days range) and 2 conditions of a longer duration: 4 months (wrist fracture) and 6 months (severe depression). To capture substantial and realistic variations in terms of disease severity and disease duration, the 6 conditions varied in both duration and severity to represent typical disease courses over time, such as middle-term (wrist fracture and severe depression), short-term (acute tonsillitis, concussion, and severe gastrointestinal disorder), and chronic recurrent/episodic (migraine) conditions. All disease and condition information, including information on duration, was based on empirical evidence.[27] The full health state descriptions are shown in Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.021.

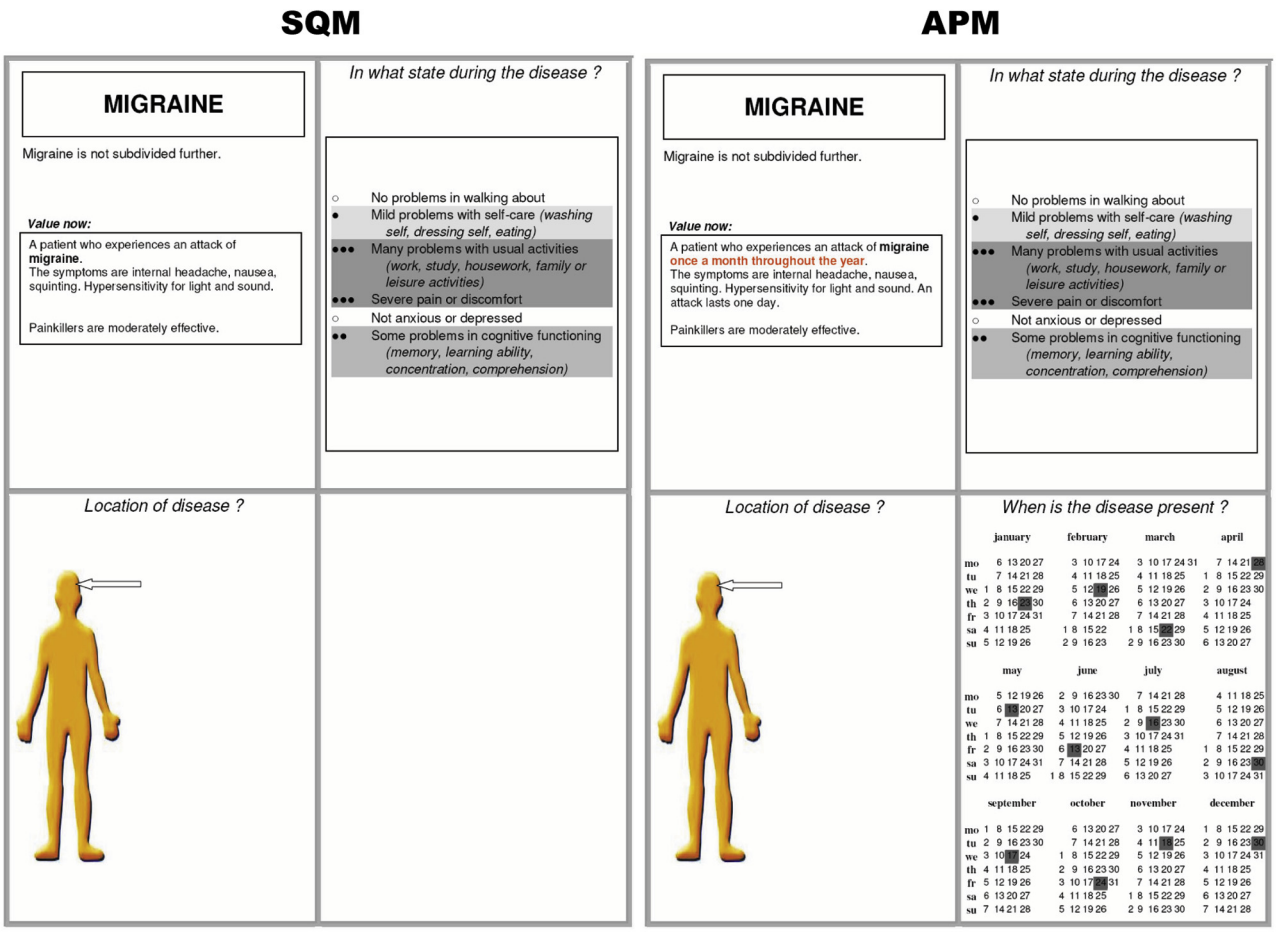
For APM, the conditions were represented in the form of standardized preformatted A4 sized vignettes, which consist of 4 quadrants of information (see Fig. 2): (1) a condition label with a naturalistic description; (2) a functional description using the EQ-5D-5L+C, an extended version of the EQ-5D descriptive system; (3) the course of the condition over a period of 1 year (annual profile); and (4) a visual representation of the condition. For SQM the vignettes were similar, except for quadrant III, which by definition was empty, and for quadrant I, from which information on duration was removed (see Fig. 2).

Regarding quadrant II, the extended EQ-5D+C consisted of an extra dimension of cognitive functioning[31] and an increase of the number of levels from 3 to 5.[32] Grayscales are used in the EQ-5D-5L+C to depict the severity of each dimension, with increasing grayscale (darker) when the severity of the level description of a dimension increases. Regarding quadrant III, the annual profile is represented as a calendar, with grayscales depicting the condition severity corresponding to the EQ-5D-5L+C profile (see Table 1). The EQ-5D-5L+C description was based primarily on detailed patient data from 1 large representative GP registration and validated by a large database (N = 400,912) of GP visits recorded as part of the second Dutch National Survey of General Practice,[33] and additional information from literature and disease specific handbooks.

Participants and Study Design

Eighty-two participants took part in 1 of 2 panel sessions and a follow-up postal survey 2 weeks later. All participants were recruited from the general population and were familiar with

Figure 2. Typical vignettes for SQM and APM.



APM indicates annual profile model; QALY, quality-adjusted life year; SQM, standard QALY model.

valuing health states with the time trade-off method since they participated in a previous study using the same valuation protocol.[27] In the face-to-face panel sessions, guided by a strict protocol, respondents valued the 6 vignettes with SQM, and in the survey the same 6 vignettes were valued with APM, in the same order. Because respondents were familiar with APM from the previous study, it was decided to have respondents perform the SQM valuations in the panel sessions so additional clarification could be given if necessary. Instructors stimulated deliberation among respondents so that each respondent could arrive at a well-considered individual valuation. The protocol is described in more detail in Janssen et al (2008). Participants received a fee of €60 for full participation (March 2005).

Valuation Procedure

The 6 vignettes were valued with a 1-year version of the time trade-off (TTO) method. In the TTO, respondents were asked to trade off a number of healthy days, weeks, or months as of now, with the maximum of 1 year, to avoid spending time in the health state being valued. A refined response form was developed that allowed respondents to form their valuation in 2 steps, the first approximate and the second more precise (see Fig. 3). Respondents were instructed to consider the 1 year described by the vignette only and to disregard what could happen after that year. In SQM, respondents were instructed to imagine that the condition described by the vignette lasted for the entire time horizon of the TTO version used (1 year in our case).[14,17,34]

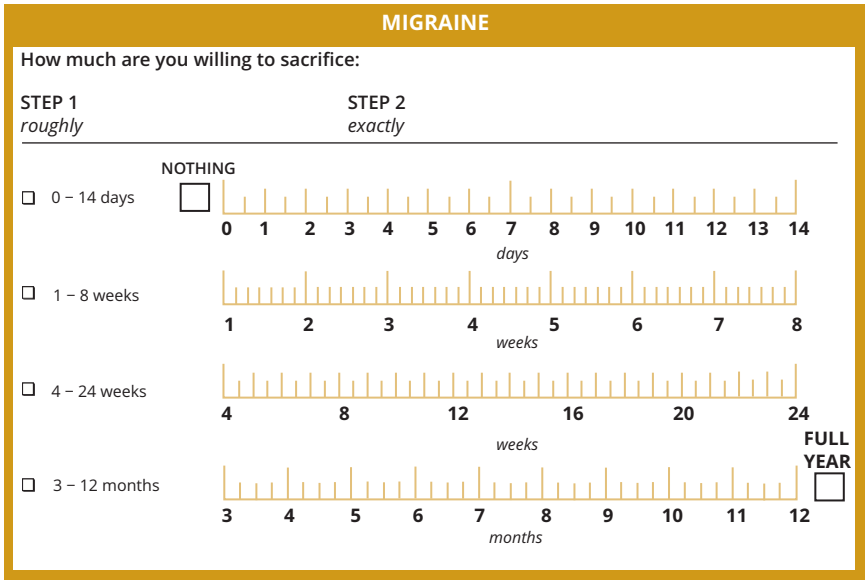
Analysis

The TTO valuation scores were converted by a linear transformation into a 0-to-1 score. Interrater reliability for both SQM and APM was assessed by the intraclass correlation coefficient (ICC, 2-way random effects model, absolute agreement test), which explicitly takes systematic variability between respondents (raters) into account. We followed available guidelines for interpreting interrater agreement statistics by Cicchetti et al (1981): <0.40 poor; 0.40–0.59 fair; 0.60–0.79 good; ≥ 0.80 excellent.[35] QALY values for SQM were estimated by calculating the area under the curve for each of the 6 health states: $QALY_{SQM} = 1 - ((1 - TTO_{SQM}) \times$ duration of the health state). TTO values under APM by definition are integral QALY values in itself: $QALY_{APM} = TTO_{APM}$. Because of the short duration of the conditions, no discounting for QALYs under SQM was applied because this would result in negligible differences.

Face validity was tested for both SQM and APM by asking respondents how well they were able to imagine and value the health states represented by both models (no trouble, some difficulties, hard to imagine).

Additive independence was tested by a paired *t* test on the individual $QALY_{SQM}$ and $QALY_{APM}$ values. If these values differ

Figure 3. Time trade-off response form.



significantly, APM produces systematically different QALY values, and the assumption of additive independence of SQM is rejected. Subgroup analysis of additive independence was carried out excluding any respondents who reported difficulties with the tasks.

Results

The mean age of the participants was 53.6 years, with 42.7% being males. Of the 82 respondents who attended the panel sessions, 81 returned the survey.

The interrater reliability intraclass correlation coefficients was 0.18 for SQM and 0.53 for APM, which indicates low reliability for SQM and acceptable reliability for APM. Face validity was judged differently: 49% of the respondents had no trouble imagining the health states under SQM, whereas 63% had no trouble under APM; 21% of the respondents found the health states under SQM hard to imagine, and 6% under APM.

Table 2 shows the mean TTO values with 95% confidence intervals (CI) for SQM and APM. SQM values range from 0.47 (severe depression) to 0.84 (wrist fracture) and APM from 0.63 (severe depression) to 0.98 (acute tonsillitis). Table 3 shows the resulting $QALY_{SQM}$ and $QALY_{APM}$ values, with APM showing lower QALY values for all 6 conditions. An APM/SQM ratio of disability values, calculated as $([1.0 - QALY_{APM}]/[1.0 - QALY_{SQM}])$, indicates that 52% of respondents reported 1.5 times more disutility under APM, 44% report 2 times more disutility, and 14% report 10 times more disutility.

The test for additive independence showed significant differences for the 4 short-term conditions in the expected direction (higher QALY values for SQM), as well as for severe depression, suggesting interaction (dependence) between time and utility (see Table 3). For wrist fracture, the assumption of additive independence of SQM could not be rejected. Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.11.021 shows scatter plots for SQM versus APM QALY values for all conditions. The subgroup analysis excluding the respondents who indicated it was hard to imagine the health states under SQM or APM resulted in similar findings, with similar levels of significance and in all but 1 health state with slightly larger mean differences.

Discussion

A head-to-head comparison of the standard QALY model with an annual profile model showed APM to be more reliable and to demonstrate better face validity. Additive independence of SQM was rejected for 5 out of 6 selected conditions. As hypothesized, SQM produced significantly higher QALY values for conditions of short duration. The 3 empirical criteria required to establish that APM is a valid and tenable alternative to SQM were met in these cases.

Inter-rater reliability for APM was sufficient and comparable with results reported elsewhere.[36] SQM demonstrated very low inter-rater reliability. One other study using SQM for temporary health states also reported low reliability.[37] Face validity was high for APM and worse for SQM. The latter might be due to difficulty in imagining constantly continuing health problems for a year for conditions such as tonsillitis or migraine. We acknowledge that different SQM results on face validity might have been obtained if only generic disease information had been included in the vignettes. However, since the current generic systems are limited in covering essential aspects of many conditions, our intention was to assess health states as realistically as possible, and for this purpose a condition label and specific information on the condition adds relevant information in our view.

As hypothesized, SQM produced significantly higher QALY values than APM in 4 prespecified conditions, which is in accordance with the available evidence.[14,16-17,19,28,38] The fact that these conditions should be considered, according to SQM values, as almost perfect health (QALY values of 0.987–0.996) seems to disagree with clinical practice: neither are these conditions considered clinically as "nearly perfect health", nor are they left untreated (revealed preference argument).

Our choice of a 1-year time horizon in particular was based on practice rather than theory, as it covers most nonchronic conditions and is convenient for being the common standard unit of

Table 2. Mean SQM and APM TTO values (N = 82).

Condition	TTO_{SQM}	
	Mean	95% CI
Acute tonsillitis	0.780	0.725–0.835
Concussion	0.710	0.650–0.770
Migraine	0.610	0.537–0.682
Severe depression	0.467	0.394–0.540
Severe gastrointestinal disorder	0.669	0.605–0.732
Wrist fracture	0.841	0.796–0.886

	TTO_{APM}	
	Mean	95% CI
Acute tonsillitis, 1 week	0.979	0.965–0.992
Concussion, 2 weeks	0.972	0.962–0.982
Migraine, monthly recurrent, 12 days	0.956	0.943–0.970
Severe depression, 6 months	0.629	0.566–0.693
Severe gastrointestinal disorder, 10 days	0.958	0.943–0.973
Wrist fracture, 4 months	0.944	0.926–0.962

APM indicates annual profile model; CI, confidence interval; QALY, quality-adjusted life year; SQM, standard QALY model; TTO, time trade-off.

time in QALYs and in most epidemiologic sources and clinical and health registries. Moreover, the seminal Global Burden of Disease initiative uses the disability-adjusted life-year (DALY), which, like the QALY, is based on a 1-year metric.[39] An annual perspective seems a natural way of looking forward or backward when evaluating personal life affairs such as health. One year is a reasonable evaluation timespan, as can be illustrated by numerous examples in medicine and beyond where the 1-year perspective is universally used in prognostic or past performance statements. The case of this article can similarly made for longer or shorter profiles, although it remains to be empirically established whether similar results would be found, especially for longer time profiles where adaption and life expectancy might impact valuation.

The first formal time profile model published to our knowledge, the HYE method,[26] was severely criticized, mainly because of its 2-stage valuation technique; first a standard gamble, then a time trade-off on the outcome of the standard gamble.[40-43] Criticism did not focus on the groundbreaking concept of the valuation of a health profile over time (until death). Cognitive overload can be expected when (complex) health profiles over time are valued. The HYE method, which takes lifetime profiles as its base, is likely to suffer from this problem.[21] In contrast to HYEs, APM is confined to 1 year, and we adopted a carefully operationalized protocol (empirical vignette construction, protocolized panel sessions, refined TTO form), which showed that the cognitive burden of our method is limited, as we demonstrated high face validity, acceptable reliability, and good-to-excellent feasibility and reliability in an earlier study.[27]

There are other methodological approaches specifically developed for the valuation of temporary and short-term conditions, although these methods do not relax the assumption of additive independence, including the TTO with specified duration, the waiting trade-off, and the sleep trade-off.[44] The most commonly used alternative is the experimental chained TTO procedure.[45] Chained TTO uses an anchor state other than death (which requires a rescaling of the obtained values) and a time horizon equal to the duration of the episode of the condition being valued.[46] This solves the problem of the unrealistic nature of the time horizon for short-term conditions when using a 1-year (or longer) version of TTO. The alternative anchor state aims to increase the sensitivity of the scale for temporary conditions, but it could introduce bias or error through the rescaling procedure. More importantly, the chained TTO procedure still does not address the dependence between time and utility.[8,37]

Some methodological issues and limitations need to be addressed. First, APM was tested for only a set of 6 conditions. A generalization of APM for a larger set of conditions, showing more variation in both condition duration and severity, is required. Second, because the APM vignettes were valued in the panel sessions but the SQM vignettes were valued in the postal survey, there could have been an effect of mode of administration. However, all respondents were already familiar with APM, and because face validity was high for APM, we are confident any bias arising from using the postal survey would be minimal. There also could have been ordering effects, both for the ordering of the 6 vignettes and for the ordering of the SQM and APM methods. Finally, a practical disadvantage of a time profile model is that it may require a relatively large number of different time-dependent health profiles to be valued to cover the course of a single disease or condition with many different disease courses over time. This criticism was raised by Tsuchiya and Dolan (2005) with regard to lifetime profiles.[20] However, for APM the number of profiles is likely to be limited, since the time horizon is confined to a single year. Moreover, when performing a pharmacoeconomic evaluation, the number of time profiles will be limited since there usually is only 1 disease to be valued with a limited number of

Table 3. Additive independence: SQM versus APM QALYs* (N = 82).

Condition	$QALY_{SQM}$	$QALY_{APM}$	mean difference	95% CI mean difference	*P*
Acute tonsillitis, 1 week	0.996	0.979	0.017	0.004–0.030	.011
Concussion, 2 weeks	0.989	0.972	0.017	0.007–0.026	<.001
Migraine, monthly recurrent, 12 days	0.987	0.956	0.031	0.018–0.044	<.001
Severe depression, 6 months	0.734	0.629	0.104	0.046–0.162	<.001
Severe gastrointestinal disorder, 10 days	0.991	0.958	0.033	0.018–0.048	<.001
Wrist fracture, 4 months	0.947	0.944	0.003	−0.020 to 0.026	.799

APM indicates annual profile model; CI, confidence interval; QALY, quality-adjusted life year; SQM, standard QALY model; TTO, time trade-off.
*According to APM, $QALY_{APM} = TTO_{APM}$.

(average) disease courses over time. Note that essentially SQM faces a similar problem. To capture a single disease course over time under SQM, many measurement points in time would be needed to realistically assess the variation of utility over time. If SQM would aim for the same "realistic" assessment that APM uses, SQM, too, needs many (more) profiles and valuations.

Because APM does not require the strict and easily violated SQM assumptions of time-utility independence and additive independence, construct validity of APM is likely to be better.[2,8-18] When shorter time cycles than 1 year are used in economic models, additive independence has to be assumed for APM. When using APM values to calculate QALYs or DALYs for periods longer than 1 year, time-utility independence is assumed. We argue that APM is a stronger model than SQM for theoretical reasons, because although SQM always requires (at least) 2 assumptions (time-utility independence and additive independence), APM requires no assumption for the duration of the profile and only 1 assumption for periods shorter than or exceeding the time horizon of the profile. Most available evidence (as described above), including the current study, show that both these assumptions do not hold for SQM. Whether these assumptions hold for APM under aforementioned conditions (either shorter or longer time periods than 1 year) remains to be investigated. When health shows variation beyond 1 year, the only way to relax both time-utility independence and additive independence is to tailor the profile to the duration of the disease or condition, so that all variations in health status are captured (eg, for chronic diseases with deteriorating health status a life expectancy model could be used). Most nonchronic diseases have their entire disease course over time within 1 year, which makes APM a superior model for practical reasons for many applications (eg, economic evaluations of treatments) of most nonchronic diseases. Finally, APM captures time preference and sequence effects, and it is also capable of capturing risk awareness in a realistic manner, by including risk of, for example, disease occurrence or mortality in the vignette description.

We conclude that APM is a more valid model than SQM, and SQM underestimates loss of health instead of APM overestimating the loss of health. To our knowledge, this is the first study that shows that the impact of short-term diseases is underestimated with SQM when compared to an elaborate alternative: the annual profile model. Although most differences in QALY values are not that large quantitatively, taking, for example, known minimally important differences into account,[47,48] the resulting differences can be large for public health applications, particularly for high-prevalence diseases. In burden of disease studies and economic evaluations involving temporary and episodic conditions, we therefore recommend the use of health profile methods such as APM because they bridge the current gap between SQM and revealed preferences.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2020.11.021.

Article and Author Information

Accepted for Publication: November 30, 2020

Published Online: March 12, 2021

doi: https://doi.org/10.1016/j.jval.2020.11.021

Author Affiliations: Section Medical Psychology and Psychotherapy, Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands (Janssen); Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands (Birnie); Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands (Birnie); Department of Public Health, Erasmus MC, Rotterdam, The Netherlands (Bonsel).

Correspondence: Mathieu F. Janssen, Section Medical Psychology and Psychotherapy, Department of Psychiatry, Erasmus MC, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands. Email: mf.bas.janssen@gmail.com

Author Contributions: *Concept and design:* Janssen, Birnie, Bonsel
Acquisition of data: Janssen, Bonsel
Analysis and interpretation of data: Janssen, Birnie, Bonsel
Drafting of the manuscript: Janssen, Birnie, Bonsel
Critical revision of the paper for important intellectual content: Janssen, Birnie, Bonsel
Statistical analysis: Janssen
Provision of study materials or patients: Janssen, Bonsel
Supervision: Bonsel

Conflict of Interest Disclosures: The authors reported no conflicts of interest.

Funding/Support: The authors received no financial support for this research.

Acknowledgment: The authors would like to thank 2 anonymous reviewers for valuable comments.

REFERENCES

1. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Operations Res*. 1980;28:206–224.
2. Bleichrodt H, Johannesson M. The validity of QALYs: an experimental test of constant proportional tradeoff and utility independence. *Med Decis Making*. 1997;17:21–32.
3. Johannesson M, Jönsson B, Karlsson G. Outcome measurement in economic evaluation. *Health Econ*. 1996;5:279–296.
4. Keeney RL, Raiffa H. *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge: Cambridge University Press; 1993.
5. Bleichrodt H. Qalys and Hyes - under what conditions are they equivalent. *J Health Econ*. 1995;14:17–37.
6. Fishburn PC. Independence in utility theory with whole product sets. *Operations Res*. 1965;13:28–45.
7. Cook J, Richardson J, Street A. A cost utility analysis of treatment options for gallstone disease: methodological issues and results. *Health Econ*. 1994;3:157–168.
8. Bala MV, Wood LL, Zarkin GA, et al. Are health states "timeless"? The case of the standard gamble method. *J Clin Epidemiol*. 1999;52:1047–1053.
9. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. *J Chronic Dis*. 1978;31:697–704.
10. Miyamoto JM, Eraker SA. Parameter estimates for a QALY utility model. *Med Decis Making*. 1985;5:191–213.
11. Stalmeier PFM, Wakker PP, Bezembinder TGG. Preference reversals: violations of unidimensional procedure invariance. *J Exp Psychol Human*. 1997;23:1196–1205.
12. Buckingham JK, Birdsall J, Douglas JG. Comparing three versions of the time tradeoff: time for a change? *Med Decis Making*. 1996;16:335–347.
13. Kirsch J, McGuire A. Establishing health state valuations for disease specific states: an example from heart disease. *Health Econ*. 2000;9:149–158.
14. Richardson J, Hall J, Salkeld G. The measurement of utility in multiphase health states. *Int J Technol Assess Health Care*. 1996;12:151–162.
15. Dolan P, Gudex C. Time preference, duration and health state valuations. *Health Econ*. 1995;4:289–299.
16. Kuppermann M, Shiboski S, Feeny D, et al. 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;17:42–55.
17. Lee GM, Salomon JA, LeBaron CW, Lieu TA. Health-state valuations for pertussis: methods for valuing short-term health states. *Health Qual Life Outcomes*. 2005;3:17.
18. Krabbe PFM, Bonsel GJ. Sequence effects, health profiles, and the QALY model: in search of realistic modeling. *Med Decis Making*. 1998;18:178–186.
19. Brazier J, Dolan P, Karampela K, Towers I. Does the whole equal the sum of the parts? Patient-assigned utility scores for IBS-related health states and profiles. *Health Econ*. 2006;15:543–551.
20. Tsuchiya A, Dolan P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. *Med Decis Making*. 2005;25:460–467.
21. Towers I, Spencer A, Brazier J. Healty year equivalents versus quality-adjusted life years: the debate continues. *Expert Rev Pharmacoeconomics Outcomes Res*. 2005;5:245–254.

22. Holmes AM. Measurement of short term health effects in economic evaluations. *Pharmacoecon*. 1998;13:171–174.
23. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press; 2005.
24. Brazier J, Deverill M, Green C, et al. A review of the use of health status measures in economic evaluation. *Health Technol Assess*. 1999;3:1–164.
25. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.
26. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making*. 1989;9:142–149.
27. Janssen MF, Birnie E, Bonsel GJ. Feasibility and reliability of the annual profile model for deriving QALYs when health varies over time. *Med Decis Making*. 2006;28:500–510.
28. Abellan-Perpinan JM, Pinto-Prades JL, Mendez-Martinez I, Badia-Llach X. Towards a better QALY model. *Health Econ*. 2006;15:665–676.
29. Stouthard ME, Essink-Bot ML, Bonsel GJ. Disability weights for diseases. a modified protocol and results for a Western European region. *Eur J Public Health*. 2000;10:24–30.
30. Spencer A. A test of the QALY model when health varies over time. *Soc Sci Med*. 2003;57:1697–1706.
31. Krabbe PF, Stouthard ME, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. *J Clin Epidemiol*. 1999;52:293–301.
32. Mahapatra P, Nanda L, Rajshree KT. The 6D5L description system for health state valuation. In: Murray CJ, Salomon JA, Mathers CD, Lopez AD, eds. *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*. Geneva: World Health Organization; 2002.
33. Linden van der MW, Westert GP, Bakker DH, Schellevis FG. *The second Dutch survey of general practice*. Utrecht/Bilthoven: NIVEL/RIVM; 2004.
34. Bala MV, Zarkin GA. Are QALYs an appropriate measure for valuing morbidity in acute diseases? *Health Econ*. 2000;9:177–180.
35. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic*. 1981;86:127–137.
36. Schwarzinger M, Stouthard ME, Burstrom K, Nord E. Cross-national agreement on disability weights: the European Disability Weights Project. *Popul Health Metr*. 2003;1:9.
37. Locadia M, Stalmeier PF, Oort FJ, et al. A comparison of 3 valuation methods for temporary health states in patients treated with oral anticoagulants. *Med Decis Making*. 2004;24:625–633.
38. Verschuuren M. *Quality Adjusted Life Years and Trade Off Exercises*. Utrecht: University Medical Center Utrecht; 2006.
39. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet*. 2018;392:1789–1858.
40. Buckingham K. A note on HYE (healthy years equivalent). *J Health Econ*. 1993;12:301–309.
41. Culyer AJ, Wagstaff A. QALYs (quality-adjusted life-years) versus HYEs (healthy years equivalents). *J Health Econ*. 1993;12:311–323.
42. Loomes G. The myth of the HYE (healthy year equivalent). *J Health Econ*. 1995;14:1–7.
43. Morrison BW. Antihypertensive therapy and quality of life. *N Engl J Med*. 1987;316:52.
44. Wright DR, Wittenber E, Swan JS, et al. Methods for measuring temporary health states for cost-utility analyses. *Pharmacoecon*. 2009;27:713–723.
45. Stoniute J, Mott DJ, Shen J. Challenges in valuing temporary health states for economic evaluation: a review of empirical applications of the chained time trade-off method. *Value Health*. 2018;21:605–611.
46. Torrance GW. Measurement of health utilities for economic appraisal. a review. *J Health Econ*. 1986;5:1–30.
47. McClure NS, Sayah FA, Ohinmaa A, Johnson JA. Minimally important difference of the EQ-5D-5L Index Score in adults with type 2 diabetes. *Value Health*. 2018;21:1090–1097.
48. McClure NS, Sayah FA, Xie F, Luo N, Johnson JA. Instrument-defined estimates of the minimally important difference for EQ-5D-5L Index Scores. *Value Health*. 2017;20:644–650.