Received: 18 October 2019 | Revised: 13 August 2021 | Accepted: 22 October 2021

DOI: 10.1002/hec.4457

RESEARCH ARTICLE

Health Economics WILEY

Not all respondents use a multiplicative utility function in choice experiments for health state valuations, which should be reflected in the elicitation format (or statistical analysis)

Marcel F. Jonker[1,2,3] | **Richard Norman**[4]

[1]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, the Netherlands

[2]Erasmus Choice Modelling Centre, Erasmus University Rotterdam, Rotterdam, the Netherlands

[3]Duke Clinical Research Institute, Duke University, Durham, North Carolina, USA

[4]School of Public Health, Curtin University, Perth, Western Australia, Australia

Correspondence
Marcel F. Jonker, Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA, Rotterdam, the Netherlands.
Email: marcel@mfjonker.com

Funding information
EuroQol Research Foundation

Abstract

Discrete choice experiments (DCEs) that include health states and duration are becoming a common method for estimating quality-adjusted life year (QALY) tariffs. These DCEs need to be analyzed under the assumption that respondents treat health and duration multiplicatively. However, in the most commonly used DCE duration format there is no guarantee that respondents actually do so; in fact, respondents can easily simplify the choice tasks by considering health and duration separately. This would result in valid DCE responses but preclude subsequent QALY tariff calculations. Using a Bayesian latent class model and data from two existing valuation studies, our analyses confirm that in both datasets the majority of respondents do not appear to have used a multiplicative utility function. Moreover, a statistical correction for respondents who used an incorrect function changes the range of the QALY weights. Hence our results imply that one can neither assume that respondents use the theoretically required multiplicative utility function nor assume that the type of utility function that respondents use does not affect the estimated QALY weights. As a solution, we advise researchers to use an alternative, more constrained DCE elicitation format that avoids these behavioral problems.

KEYWORDS
discrete choice experiment, health state valuations

1 | BACKGROUND

Conventional economic evaluations in health typically employ the quality-adjusted life year (QALY) framework, in which the value of a chronic health state is defined as the product of life expectancy and health-related quality of life. Health states are valued on a scale where 1 represents full health, and 0 represent death, or health states considered equivalent to death. This framework is a useful approach for facilitating comparison between interventions that have benefits manifesting in different ways. The assumptions that are required for this framework are widely known, and are testable

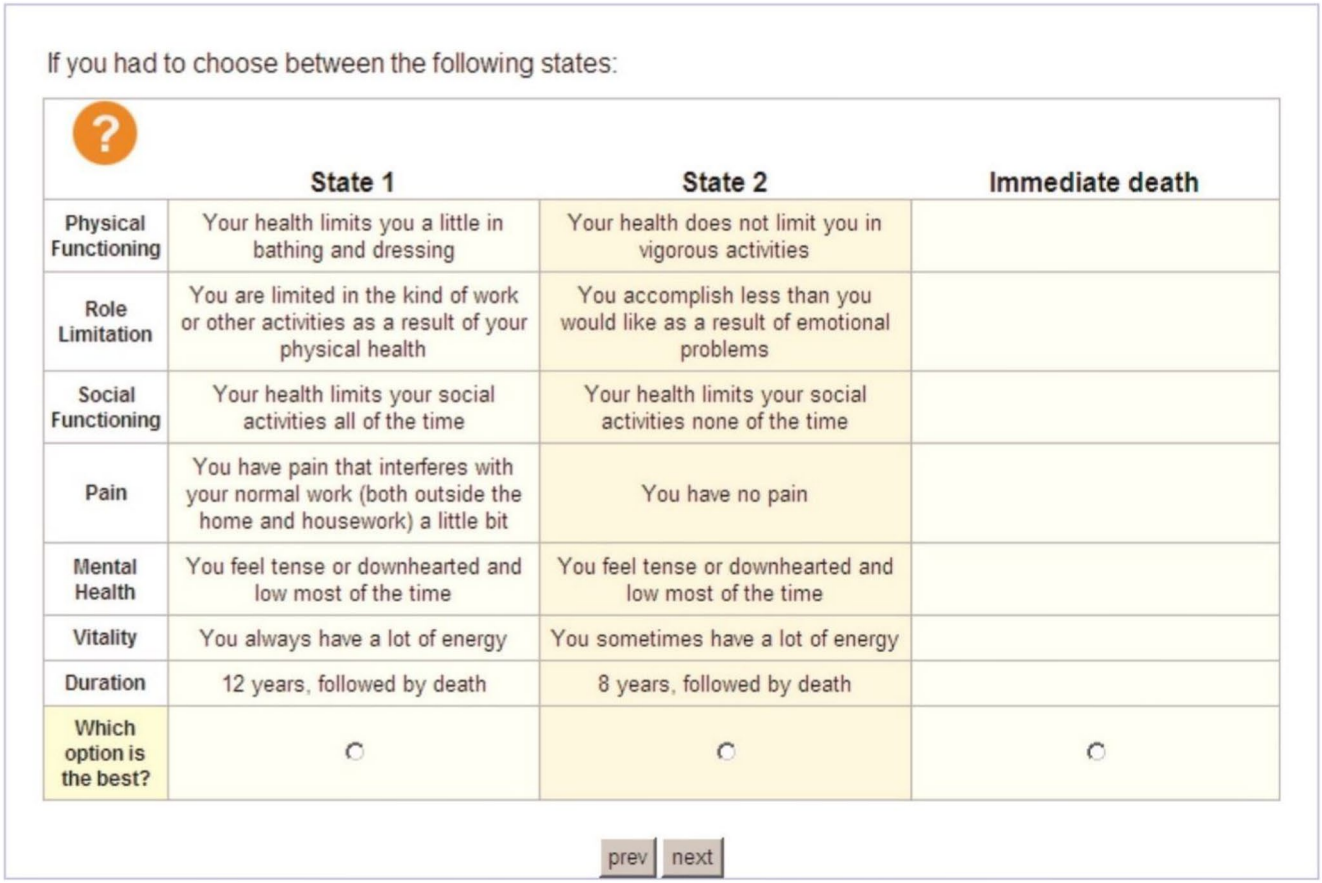
(Bleichrodt et al., 1997). For example, the constant proportional trade-off assumption has been considered widely, with recent studies exploring methods of adjusting for non-linearity of utility with respect to time (Craig et al., 2018; Jonker et al., 2018a).

There are a variety of methods that have been used to estimate tariffs for health states, the key sacrifice-based methods being Time Trade-Off (TTO), Standard Gamble, and Discrete Choice Experiments (DCEs) (Brazier et al., 2007). The use of DCEs has grown significantly in health generally (Soekhai et al., 2019), and in the valuation of health states specifically (Mulhern et al., 2019). A key reason for this growth is that they produce similar results irrespective of mode of administration, a result which does not hold for TTO (Mulhern et al., 2013; Norman et al., 2010). Thus, they can be administered online, usually without an interviewer, reducing cost and increasing potential geographical spread of valuation surveys.

Discrete choice experiment studies can be subdivided into those that present health states with different life expectancies versus those that either have a fixed life expectancy, or do not state a life expectancy. An example of the latter would be the DCE that is included as part of the EuroQol Group's standard protocol for the valuation of health states in the EQ-5D-5L (Devlin et al., 2018; Ramos-Goni et al., 2017). An issue with the use of an approach without durations or with fixed durations, is that, while it provides values on a latent scale, it does not easily anchor on the required 0-1 scale. The common solution to this, as initially advocated by Flynn et al. and subsequently introduced into the health economics literature by Bansback et al., is to include a duration attribute (Bansback et al., 2012; Flynn, 2010). The principle of this is that it allows a quantification of the trade-offs that respondents make both between dimensions of quality of life, and between quality of life and length of life; trade-offs that are essential for the estimation of QALY tariffs.

In the most commonly used DCE duration (DCEd) elicitation format (also known as DCE with duration, or DCE_{TTO}), two different impaired health states are presented with different levels of duration (c.f. Figure 1). Respondents are required to base their choice between the two health states on the overall utility of each choice option, which is defined in the QALY framework as the product of the respondents' utility for each health state multiplied by the corresponding duration. Without such a multiplicative utility function, it is not possible to calculate theoretically appropriate QALY tariffs. However, a multiplicative utility function requires respondents to be able to and willing to perform a series of complicated evaluations. That is, in each choice task, respondents have to evaluate the relative attractiveness of the health states, multiply each by different duration levels, and subsequently choose the option that gives them the highest overall utility. From a theoretical perspective, respondents could easily simplify the choice tasks by avoiding the required multiplication with duration and instead treat duration as a standard, additive attribute. Such a linear additive utility function would correctly take into account that longer duration of life has positive utility to the respondent, and that health problems have a negative utility. Moreover, there is nothing in the choice format that prevents respondents from adopting a linear additive utility function. Hence the use of a linear additive utility function does not violate the question that is asked to respondents; however, it does contradict the assumption built into the QALY framework and analyzing data assuming a multiplicative utility function for respondents that used a linear additive utility function could induce substantial bias in the resultant QALY tariffs.

Thus far, there is no evidence that substantiates that respondents actually use the theoretically imposed multiplicative utility function. At the same time, there is also no evidence that respondents are *not* using the theoretically required utility function, yet there is ample evidence that at least a subset of respondents in DCE research tends to simplify complicated choice tasks, for example by only focusing on a few instead of all of the included attributes (i.e., so-called attribute non-attendance, see e.g., Hole et al., 2016, Jonker et al., 2018b). Accordingly, this paper aims to establish the type of utility function that respondents most likely use in DCE-duration datasets, and aims to provide a quantitative assessment of the impact that respondents who simplify the choice tasks can have on the resulting QALY tariffs. Hence the two key purposes of this work are to (1) establish whether or not people actually make choices that match the theoretically required multiplicative utility function (as opposed to resorting to a simpler yet equally feasible linear additive utility function) and (2) estimate the impact on (i.e., bias of) the estimated QALY tariffs when not adequately taking into account that some respondents may not have used the theoretically required multiplicative utility function.


If you had to choose between the following states:

	State 1	State 2	Immediate death
Physical Functioning	Your health limits you a little in bathing and dressing	Your health does not limit you in vigorous activities	
Role Limitation	You are limited in the kind of work or other activities as a result of your physical health	You accomplish less than you would like as a result of emotional problems	
Social Functioning	Your health limits your social activities all of the time	Your health limits your social activities none of the time	
Pain	You have pain that interferes with your normal work (both outside the home and housework) a little bit	You have no pain	
Mental Health	You feel tense or downhearted and low most of the time	You feel tense or downhearted and low most of the time	
Vitality	You always have a lot of energy	You sometimes have a lot of energy	
Duration	12 years, followed by death	8 years, followed by death	
Which option is the best?	○	○	○

prev next

FIGURE 1 Example Discrete choice experiment (DCE) duration choice task. Note that immediate death is included as an alternative-specific, third choice option in both datasets that are analyzed in this paper. Including an immediate death state is optional and not universally recommended for DCE duration valuation studies. The estimates of the immediate death parameter are not used in the main text; please see the online supplemental for an assessment of the impact of anchoring the quality-adjusted life year (QALY) tariffs based on the immediate death parameters [Colour figure can be viewed at wileyonlinelibrary.com]

2 | METHODS

2.1 | Modeling approach

The general idea of the modeling approach used in this paper is that the overall utility (U_{ijt}) that respondent i obtains from alternative j in choice task t can be derived from one of two a-priori equally reasonable and theoretically sound utility functions, $U_1$and U_2, that is,

$$U_{ijt} = \varphi_i U_{1ijt} + (1-\varphi_i) U_{2ijt} + \varepsilon_{ijt}, \quad i = 1,..,I; j = 1,..,J; t = 1,..,T. \tag{1}$$

In Equation (1), the selection parameter ($\varphi_i \in [0,1]$) reflects the respondents' probabilities of having used the first utility function as opposed to the second, and the error term (ε_{ijt}) is assumed to be independently and identically Gumbel distributed.

The first utility function is the standard multiplicative utility function and defined as the quality of the health state (H_{ijt}) multiplied with the duration of life (D_{ijt}) in years, that is,

$$U_{1ijt} = H_{ijt} * D_{ijt}. \tag{2}$$

The quality of the health states (H_{ijt}) is defined as the dot product of the K dummy coded health state characteristics ($X_{ijt1},..,X_{ijtK}$) and preference coefficients ($\beta_1,..,\beta_K$), that is,

$$H_{ijt} = \sum_{k=1}^{K} \beta_k * X_{ijtk}\,, \tag{3}$$

with the first element of X equal to 1 and the last equal to 0 if $j = 1,2$ and the opposite if $j = 3$. Accordingly, β_1 is defined as the perfect health intercept and β_K as the immediate death intercept, although the latter only needs to be included if the DCE design includes immediate death as an alternative-specific, third choice option (cf. Figure 1).

The second utility function is the additive utility function and defined as the sum of the health state quality (H^*_{ijt}) and the utility attributed to the dummy-coded duration of life (D^*_{ijt}) levels, that is,

$$U_{2\,ijt} = H^*_{ijt} + D^*_{ijt}\,. \tag{4}$$

As before, the quality of the health states (H^*_{ijt}) is defined as the dot product of the K dummy-coded health state characteristics excluding the perfect health intercept that is only relevant in the multiplicative utility function (*i.e.* $X_{ijt2},..,X_{ijtK+1}$), and the associated preference coefficients ($\gamma_1,..,\gamma_K$):

$$H^*_{ijt} = \sum_{k=1}^{K} \gamma_k * X_{ijt,k+1} \tag{5}$$

whereas D^*_{ijt} is defined as the dot product of the M dummy-coded duration levels (Q_{ijtm}) and associated preference coefficients ($\gamma_{K+1},..,\gamma_{K+M}$), that is,

$$D^*_{ijt} = \sum_{m=1}^{M} \gamma_{K+m} * Q_{ijtm}\,. \tag{6}$$

In the default model specification, all φ_i are fixed at 1. This implies that a standard multiplicative utility function is used for all respondents, which is the default modeling option for QALY tariff estimations using DCE duration data. To test the hypothesis that there is at least a subgroup of respondents that is more likely to use a linear additive utility function than the theoretically required multiplicative utility function, the alternative model specification includes φ_i as model parameters to be estimated. The second specification is thus a latent class model with two classes, in which the first class captures the theoretically required multiplicative utility function and the second class the linear additive utility function.

In latent class models, each respondent is assigned to the different classes with its own probability. To establish the fraction of respondents who were more likely to have used the simpler linear additive utility function rather than the multiplicative utility function, respondents' mean φ_i estimates are used. Respondents with $\varphi_i < 0.5$ are considered to have used a linear additive whereas respondents with $\varphi_i \geq 0.5$ are considered to have used a multiplicative utility function. Additionally, to obtain an estimate of the sensitivity of the presented results, all respondents were also classified into one of three groups. These groups were (1) respondents that almost certainly used a multiplicative utility function, (2) respondents that almost certainly used a linear additive utility function, and (3) respondents for which the observed choice tasks provide insufficient information for a sufficiently reliable classification. Four different cut-off values were used to assign respondents into each of these groups, that is $\varphi_i \leq 0.25$ versus $\varphi_i \geq 0.75$, $\varphi_i \leq 0.20$ versus $\varphi_i \geq 0.80$, $\varphi_i \leq 0.10$ versus $\varphi_i \geq 0.90$, and $\varphi_i \leq 0.05$ versus $\varphi_i \geq 0.95$, each being increasingly more conservative than the default $\varphi_i < 0.5$ versus $\varphi_i \geq 0.5$ cut-off rule.

All model specifications were programmed in the BUGS language, which means that Bayesian Markov Chain Monte Carlo (MCMC) methods were used to fit the model parameters. The following prior distributions were used:

Specification 1: The multiplicative logit model
$\varphi_i \equiv 1$, $\beta \sim$ Normal(0,0.01)

Specification 2: The latent class logit model
$\varphi_i \sim Bernoulli(0.5)$, $\beta \sim$ Normal(0,0.01), $\gamma \sim$ Normal(0,0.01)

Both models were fitted using OpenBUGS using two MCMC chains of 20,000 draws for the multiplicative and 40,000 draws for the latent class logit models, respectively. Half of the draws were discarded as burn-in iterations and convergence was evaluated based on a visual inspection of the MCMC chains and the diagnostics as implemented in the OpenBUGS software.

Because the parameter estimates of specifications 1 and 2 are on different (latent) utility scales they are not directly comparable. The parameter estimates on the utility scale are included in the online supplemental, whereas parameter estimates transformed onto the QALY scale are included in the main text. These QALY tariffs were calculated by dividing all elements of β by the first element of β, which represents the full-health intercept, and are directly comparable. Based on the position of the immediate death parameter it was possible to further re-scale the QALY tariffs, the estimates of which are reported in Appendix B as part of the sensitivity analysis. Note that it is not possible to calculate QALY tariffs from the linear additive specifications.

2.2 | Datasets used

The two datasets both come from Australia, and consider preferences for health states in two different instruments, specifically the EQ-5D-5L (Norman et al., 2013) and the SF-6D (Norman et al., 2014). The details of the data collection, survey design, and base case analysis are given elsewhere. Briefly, both studies conducted general population valuation studies using an online panel of respondents who had previously stated a willingness to participate in such research. Both studies asked respondents to state a preference between combinations of health states and duration and used an efficient DCE design in which different impaired health states were combined with different durations of life. Both studies had a relatively large sample size (973 for the EQ-5D-5L and 1017 for the SF-6D), and both assumed that respondents considered a multiplicative function (i.e., that all $\varphi_i = 1$). The studies differed in number of choice tasks per respondents (10 for the EQ-5D-5L and 15 for the SF-6D).

3 | RESULTS

Table 1 presents the aggregated class membership percentages calculated using the various cut-off values based on the mean φ_i estimates. When all respondents are either assigned to the additive or multiplicative utility function, 76 and 71 percent of the respondents in the latent class conditional logit models are considered to have used an additive utility function in the EQ-5D and SF-6D data, respectively. Conversely, only 24 and 29 percent used the required multiplicative utility function for the QALY tariff calculations.

When more conservative cut-off values are used and, consequently, an "uncertain" category of respondents that could have used either utility function is introduced, the percentage of respondents who are thought to have definitely used a linear additive utility function decreases. As shown in Table 1, the share of the linear additive utility function reduces from 76 to 64 and from 71 to 48 percent if the most stringent cut-off rule is applied. But with more stringent cut-off rules the percentage of respondents who are designated as having used a multiplicative utility function also decreases. Hence, irrespective of the cut-off values used, the percentage of respondents assigned to the linear additive utility function is larger than the percentage of respondents assigned to the multiplicative utility function.

Tables 2 and 3 present the calculated QALY tariffs from the EQ-5D and SF-6D datasets, respectively. Whereas the full sample results are based on all respondents, irrespective of whether they used a multiplicative or linear additive utility

TABLE 1 Aggregated class membership (in percentages), by cut-off value

Dataset	Cut-off values	Additive utility	Unclear	Multiplicative utility
EQ-5D	$\varphi_i < 0.50$ versus $\varphi_i \geq 0.50$	76	0	24
	$\varphi_i \leq 0.25$ versus $0.25 < \varphi_i < 0.75$ versus $\varphi_i \geq 0.75$	74	5	21
	$\varphi_i \leq 0.20$ versus $0.20 < \varphi_i < 0.80$ versus $\varphi_i \geq 0.80$	73	7	20
	$\varphi_i \leq 0.10$ versus $0.10 < \varphi_i < 0.90$ versus $\varphi_i \geq 0.90$	70	11	19
	$\varphi_i \leq 0.05$ versus $0.05 < \varphi_i < 0.95$ versus $\varphi_i \geq 0.95$	64	18	18
SF-6D	$\varphi_i < 0.50$ versus $\varphi_i \geq 0.50$	71	0	29
	$\varphi_i \leq 0.25$ versus $0.25 < \varphi_i < 0.75$ versus $\varphi_i \geq 0.75$	66	9	25
	$\varphi_i \leq 0.20$ versus $0.20 < \varphi_i < 0.80$ versus $\varphi_i \geq 0.80$	63	14	23
	$\varphi_i \leq 0.10$ versus $0.10 < \varphi_i < 0.90$ versus $\varphi_i \geq 0.90$	56	24	20
	$\varphi_i \leq 0.05$ versus $0.05 < \varphi_i < 0.95$ versus $\varphi_i \geq 0.95$	48	33	19

TABLE 2 EQ-5D-5L quality-adjusted life year weights*

Attributes/levels	Entire sample	Multiplicative class only
Full health	1.00 (n/a)	1.00 (n/a)
Mobility 2	−0.08 (−0.13,−0.04)	−0.15 (−0.27,−0.03)
Mobility 3	−0.10 (−0.14,−0.05)	−0.15 (−0.26,−0.03)
Mobility 4	−0.28 (−0.33,−0.24)	−0.36 (−0.50,−0.23)
Mobility 5	−0.37 (−0.42,−0.32)	−0.40 (−0.54,−0.28)
Self-care 2	−0.07 (−0.11,−0.02)	−0.03 (−0.15, 0.11)
Self-care 3	−0.10 (−0.14,−0.05)	−0.13 (−0.25,−0.00)
Self-care 4	−0.24 (−0.28,−0.20)	−0.30 (−0.42,−0.18)
Self-care 5	−0.34 (−0.39,−0.30)	−0.44 (−0.58,−0.31)
Usual activities 2	−0.11 (−0.15,−0.06)	−0.07 (−0.20, 0.07)
Usual activities 3	−0.13 (−0.17,−0.08)	−0.17 (−0.29,−0.05)
Usual activities 4	−0.29 (−0.34,−0.25)	−0.25 (−0.39,−0.12)
Usual activities 5	−0.31 (−0.35,−0.26)	−0.24 (−0.37,−0.11)
Pain/discomfort 2	−0.07 (−0.12,−0.03)	−0.13 (−0.24,−0.01)
Pain/discomfort 3	−0.08 (−0.12,−0.03)	−0.15 (−0.25,−0.03)
Pain/discomfort 4	−0.27 (−0.31,−0.22)	−0.39 (−0.52,−0.27)
Pain/discomfort 5	−0.35 (−0.40,−0.31)	−0.61 (−0.78,−0.46)
Anxiety/depression 2	−0.16 (−0.20,−0.11)	−0.30 (−0.42,−0.19)
Anxiety/depression 3	−0.24 (−0.28,−0.20)	−0.34 (−0.46,−0.23)
Anxiety/depression 4	−0.44 (−0.49,−0.39)	−0.66 (−0.84,−0.52)
Anxiety/depression 5	−0.42 (−0.47,−0.37)	−0.72 (−0.89,−0.57)
“Pits” (5-5-5-5-5)	−0.79 (−0.90,−0.69)	−1.40 (−1.80,−1.08)
Sample/class size (%)	100%	24%

* 95% Bayesian credible intervals in parentheses.

function, the multiplicative latent class results reflect the QALY tariffs solely derived from those who used the theoretically required multiplicative utility function. As shown, most Bayesian 95% credible intervals do not include zero and almost all parameters have the expected sign. For both datasets, the latent class QALY weight estimates are very different from those derived from the entire sample, which confirms that a model-based correction of the QALY tariff estimates for the influence of respondents who used a non-multiplicative utility function has a strong effect on the calculated QALY tariffs.

4 | DISCUSSION

4.1 | Overview of results

The analyses reported here give a strong indication that, in two different datasets using different instruments, the majority of participating respondents did not use the multiplicative utility function that is required for the construction of QALY tariffs. These respondents did not necessarily simplify the choice tasks because they were paying insufficient attention or were using a heuristic. In contrast, the additive utility function correctly takes into account that more health problems are a bad thing and that more duration is a good thing, with the overall attractiveness of the profile being determined by the sum of the two (cf. Equations 4–6). Still, the presented results indicate that these respondents violated a key assumption of the QALY framework. This is a significant finding because tariffs that are based on the entire sample differ significantly from tariffs that are derived from the subset of respondents who use the theoretically required multiplicative utility function. More specifically, including respondents who used a linear additive utility function in the QALY tariff calculations by assuming they used a multiplicative approach results in a sizable upwards bias in the tariff with smaller (i.e., less negative) decrements.

TABLE 3 SF-6D quality-adjusted life year weights*

Attributes/levels	Entire sample	Multiplicative class only
Full health	1.00 (n/a)	1.00 (n/a)
Physical functioning 2	−0.04 (−0.07,−0.01)	−0.14 (−0.24,−0.04)
Physical functioning 3	−0.08 (−0.10,−0.05)	−0.13 (−0.23,−0.03)
Physical functioning 4	−0.14 (−0.16,−0.11)	−0.27 (−0.38,−0.17)
Physical functioning 5	−0.15 (−0.17,−0.12)	−0.36 (−0.47,−0.25)
Physical functioning 6	−0.31 (−0.34,−0.28)	−0.48 (−0.61,−0.37)
Role limitations 2	−0.09 (−0.12,−0.07)	−0.09 (−0.19, 0.01)
Role limitations 3	−0.06 (−0.09,−0.04)	−0.07 (−0.16, 0.03)
Role limitations 4	−0.13 (−0.15,−0.10)	−0.14 (−0.23,−0.05)
Social functioning 2	−0.02 (−0.05, 0.01)	0.05 (−0.06, 0.17)
Social functioning 3	−0.03 (−0.05,−0.00)	−0.01 (−0.10, 0.10)
Social functioning 4	−0.11 (−0.14,−0.09)	−0.05 (−0.14, 0.04)
Social functioning 5	−0.12 (−0.14,−0.09)	−0.17 (−0.26,−0.08)
Pain 2	−0.08 (−0.11,−0.05)	−0.09 (−0.19, 0.02)
Pain 3	−0.18 (−0.21,−0.16)	−0.21 (−0.31,−0.11)
Pain 4	−0.21 (−0.24,−0.18)	−0.23 (−0.33,−0.12)
Pain 5	−0.30 (−0.32,−0.27)	−0.39 (−0.51,−0.28)
Pain 6	−0.29 (−0.32,−0.26)	−0.39 (−0.52,−0.27)
Mental health 2	−0.07 (−0.09,−0.04)	−0.12 (−0.22,−0.03)
Mental health 3	−0.08 (−0.11,−0.05)	−0.15 (−0.24,−0.06)
Mental health 4	−0.19 (−0.22,−0.16)	−0.33 (−0.44,−0.23)
Mental health 5	−0.29 (−0.31,−0.26)	−0.36 (−0.47,−0.27)
Vitality 2	−0.01 (−0.03, 0.02)	−0.02 (−0.11, 0.08)
Vitality 3	−0.04 (−0.07,−0.01)	−0.07 (−0.17, 0.04)
Vitality 4	−0.21 (−0.24,−0.19)	−0.22 (−0.32,−0.13)
Vitality 5	−0.26 (−0.28,−0.23)	−0.29 (−0.38,−0.19)
"Pits" (6-4-5-6-5-5)	−0.38 (−0.44,−0.32)	−0.83 (−1.07,−0.62)
Sample/class size (%)	100%	29%

* 95% Bayesian credible intervals in parentheses.

The study has a number of strengths. The findings translate across multiple datasets, suggesting it is not a problem unique to a single instrument. Moreover, as shown in the online supplemental, the presented results are robust to the accommodation of preference heterogeneity in the modeling approach, remain robust when the estimates of alternative specific immediate death health states were used to anchor the calculated QALY tariffs at zero, and, using Monte Carlo simulations, confirmed to be based on identified latent class logit models with sufficient statistical power.

The study also has several potential limitations. First, fitting latent class logit models requires adequate information to be able to distinguish, at the individual level, between the additive and multiplicative utility functions. Unlike models with a single, fixed utility function, there is limited opportunity to borrow strength from the population-level estimates, meaning that the φ_i parameters crucially rely on the information obtained from the individual-level data. In this respect, being able to fit the models as included in this paper was possible because of the efficiently optimized DCE designs in both the EQ-5D and SF-6D datasets. However, neither of the datasets that were used was specifically optimized to be able to distinguish between different utility functions. With more appropriately optimized DCE designs and/or with a larger number of choice tasks per respondent it seems reasonable to assume that fewer respondents would be classified in the intermediate "uncertain" category.

Second, respondents who neither used a multiplicative nor additive utility function are always assigned to one of the two latent classes, which potentially overestimates the true size of the latent classes. Although an important limitation, it should be noted that the main conclusions of the paper are unaffected by and thus robust to this effect and thus robust to this effect. For instance, there is no third utility function with similar theoretical validity that obviously needs to be

included in the modeling approach. Moreover, the Monte Carlo simulations in Online Supplemental A confirm that the latent class models that were used can correctly differentiate between respondents who use an additive or multiplicative utility function and our results indicated that a large majority of respondents was more likely to have used an additive rather than multiplicative utility function. Given that these two utility functions are non-nested and behaviorally very different, the reported class shares for the multiplicative utility function can thus be considered as an upper bound. If additional latent classes were to be added, this could either keep the class share of the multiplicative model unaffected or would further reduce it, and thereby strengthen the conclusion that not all respondents use the theoretically required multiplicative utility function.

Third, while the two DCEs differed in a number of key respects, there were areas of commonality, which might explain the low proportion of respondents using a multiplicative approach. For instance, both studies were conducted in a single country (Australia) using an online panel of respondents. It may be that either of these two characteristics is associated with a higher probability of using an additive approach when making choices. However, it is unlikely that conducting data collection elsewhere, or face-to-face would move the results back to the special case where all $\varphi_i = 1$. This is something that can be tested in datasets collected in different ways, and in other locations.

4.2 | Future research

Most importantly, it is unclear to which the observed percentage of respondents who use a linear additive utility function is an inherent consequence of the unrestricted DCE duration format used, or whether adequate instruction and warm-up tasks would be able to significantly alleviate the problem.

4.3 | Recommendations

The primary conclusion of this paper is that one cannot simply assume that respondents use the theoretically required multiplicative utility function in an unconstrained DCE duration format. On the one hand, respondents could be instructed to use a multiplicative utility function when evaluating the choice tasks, including several training tasks to help them do so. However, as it is unclear as to whether this is sufficient to ensure unbiased QALY tariff estimates, it should be mentioned that there are alternative DCE duration formats that avoid the described problem altogether. An example is the matched pairs format that was introduced by Jonker et al. (2017), which comprises (a) comparisons between different impaired health states presented at identical durations of life and (b) comparisons between impaired health states combined with a longer duration and health states without health problems presented at a shorter duration of life. Alternative design constraints could also be used, as long as they adequately restrict the occurrence of different impaired health states presented at different duration levels. After all, an unconstrained and therefore statistically more efficient DCE duration design is not an aim in itself. Instead, the goal is to obtain preference estimates using DCE designs that are sufficiently efficient and allow for the derivation of unbiased QALY weights.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the EuroQol Research Foundation. The authors are members of the EuroQol Group but the views expressed in this article do not necessarily reflect those of the EuroQol Group.

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Marcel F. Jonker https://orcid.org/0000-0001-8433-1402
Richard Norman https://orcid.org/0000-0002-3112-3893

REFERENCES

Bansback, N., Brazier, J., Tsuchiya, A., & Anis, A. (2012). Using a discrete choice experiment to estimate societal health state utility values. *Journal of Health Economics*, *31*, 306–318.

Bleichrodt, N., Wakker, P., & Johannesson, M. (1997). Characterizing QALYs by risk neutrality. *Journal of Risk and Uncertainty*, *15*, 107–114.

Brazier, J., Ratcliffe, J., Salomon, J. A., & Tsuchiya, A. (2007). *Measuring and valuing health benefits for economic evaluations*. Oxford University Press.

Craig, B., Rand, K., Bailey, H., & Stalmeijer, P. F. (2018). Quality-adjusted life-years without constant proportionality. *Value in Health*, *21*(9), 1124–1131.

Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & van Hout, B. (2018). Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Economics*, *27*, 7–22.

Flynn, T. (2010). Using conjoint analysis to estimate health state values for cost-utility analysis: Issues to consider. *PharmacoEconomics*, *28*, 711–722.

Hole, A. R., Richard, N., & Rosalie, V. (2016). Response patterns in health state valuation using endogenous attribute attendance and latent class analysis. *Health Economics*, *25*(2), 212–224.

Jonker, M. F., Attema, A. E., Donkers, B., Stolk, E. A., & Versteegh, M. M. (2017). Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. *Health Economics*, *26*, 1534–1547.

Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018a). Advocating a paradigm shift in health-state valuations: The estimation of time-preference corrected QALY tariffs. *Value in Health*, *21*, 993–1001.

Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018b). Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. *Value in Health*, *21*(7), 767–771.

Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N., & Tsuchiya, A. (2013). Binary choice health state valuation and mode of administration: Head-to-head comparison of online and CAPI. *Value in Health*, *16*, 104–113.

Mulhern, B., Norman, R., Street, D. J., & Viney, R. (2019). One method, many methodological choices: A structured review of discrete-choice experiments for health state valuation. *PharmacoEconomics*, *37*, 29–43.

Norman, R., Cronin, P., & Viney, R. (2013). A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. *Applied Health Economics and Health Policy*, *11*(3), 287–298.

Norman, R., King, M., Clarke, D., Viney, R., Cronin, P., & Street, D. (2010). Does mode of administration matter? Comparison of on line and face-to-face administration of a time trade-off task. *Quality of Life Research*, *19*, 499–508.

Norman, R., Viney, R., Brazier, J., Burgess, L., Cronin, P., King, M., Ratcliffe, J., & Street, D. (2014). Valuing SF-6D health states using a discrete choice experiment. *Medical Decision Making*, *34*(6), 773–786.

Ramos-Goni, J. M., Oppe, M., Slaap, B., Busschbach, J. J., & Stolk, E. (2017). Quality control process for EQ-5D-5L valuation studies. *Value in Health*, *20*, 466–473.

Soekhai, V., de Bekker-Grob, E. W., Ellis, A. R., & Vass, C. M. (2019). Discrete choice experiments in health economics: Past, present and future. *PharmacoEconomics*, *37*(2), 201–226.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Jonker, M. F., & Norman, R. (2022). Not all respondents use a multiplicative utility function in choice experiments for health state valuations, which should be reflected in the elicitation format (or statistical analysis). *Health Economics*, *31*(2), 431–439. https://doi.org/10.1002/hec.4457