ScienceDirect

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

Preference-Based Assessments

A Feasible Estimation of a "Corrected" EQ-5D Social Tariff

Jose-Maria Abellan-Perpiñan, PhD, Jorge-Eduardo Martinez-Perez, PhD, Fernando-Ignacio Sanchez-Martinez, PhD, Jose-Luis Pinto-Prades, PhD

ABSTRACT

Objectives: To demonstrate the feasibility of estimating a social tariff free of utility curvature and probability weighting biases and to test transferability between riskless and risky contexts.

Methods: Valuations for a selection of EQ-5D-3L health states were collected from a large and representative sample (N = 1676) of the Spanish general population through computer-assisted personal interviewing. Two elicitation methods were used: the traditional time trade-off (TTO) and a novel risky-TTO procedure. Both methods are equivalent for better than death states, which allowed us to test transferability of utilities across riskless and risky contexts. Corrective procedures applied are based on rank-dependent utility theory, identifying parameter estimates at the individual level. All corrections are health-state specific, which is a unique feature of our corrective approach.

Results: Two corrected value sets for the EQ-5D-3L system are estimated, highlighting the feasibility of developing national tariffs under nonexpected utility theories, such as rank-dependent utility. Furthermore, transferability was not supported for at least half of the health states valued by our sample.

Conclusions: It is feasible to estimate a social tariff by using interviewing techniques, sample sizes, and sample representativeness equivalent to prior studies designed to generate national value sets for the EQ-5D. Utilities obtained in distinct contexts may not be interchangeable. Our findings caution against routinely taking transferability of utility for granted.

Keywords: corrective approach, debiasing, EQ-5D, health-state utilities, social tariffs.

VALUE HEALTH. 2024; 27(9):1243–1250

Highlights

- To date there is no national social tariff or value set fully corrected under nonexpected utility assumptions.
- This article presents 2 value sets for the EQ-5D-3L system estimated by applying corrections based on rank-dependent utility theory, identifying parameter estimates at individual level. All corrections done are health-state specific, which is a unique feature of our corrective approach. These findings highlight the feasibility of developing national tariffs under nonexpected utility theories.
- Our results suggest that utilities obtained in distinct contexts may not be interchangeable. Transferability across riskless and risky domains was not supported for at least half of the health states valued by our sample. Consequently, it seems that ensuring tariffs are not context dependent is as relevant and challenging as correcting value sets.

Introduction

Quality-adjusted life year calculations combine health-utilities with life duration.[1] Although utilities can be elicited directly using the time trade-off (TTO) or the standard gamble (SG), health-related multiattribute utility instruments are often used instead. These instruments are based on preferences elicited from a representative sample of the general population, giving rise to different value sets, also called social tariffs. Take, for example, the case of EQ-5D-3L and SF-6D (SF-36)v1, whose algorithms, however, predict significantly different value sets.[2,3]

One reason why EQ-5D-3L and SF-6D (SF-36)v1 tariffs diverge is because they are based on different health-state utility measurement methods.[4] Whereas the EQ-5D-3L tariff has usually been estimated on the basis of TTO measurements,[5,6] the SF-6D (SF-36) v1 tariff is frequently based on SG measurements.[7] Empirical studies have shown that the SG (a risky method) and the TTO (a riskless one) yield different utilities.[8-10]

One explanation for the discrepancy posits the existence of 2 utility functions, 1 for riskless and another for risky decision contexts.[11-13] Alternatively, a unifying concept of utility is proposed, with discrepancies attributed to biases or systematic deviations from expected utility (EU).[14,15] In this way, both SG and TTO measurements can be distorted because people process probabilities and life years in a nonlinear manner.[16]

Probability weighting is a bias explained by both rank-dependent utility (RDU) theory[17] and prospect theory (PT).[18,19] As Wakker and Stiggelbout[20] (under RDU) and Bleichrodt[21] (under PT) showed, the SG may lead to utilities that are biased upward if people underweight large probabilities, which is the common pattern observed.[22,23] In turn, the TTO is affected by utility curvature for life duration (ie, discounting). As shown by Bleichrodt,[21] if the utility of life years is concave instead of linear, TTO utilities are biased downward. Empirical evidence supports concavity of utility for life duration.[24-26]

If SG and TTO utilities are biased, then resultant social tariffs will also be biased, which will undermine cost-effective resource allocation. Consequently, some researchers advocate for "debiasing"[27] or applying a "corrective approach"[28] to base social tariffs on corrected utilities. Bleichrodt et al[29] provide an illustration of how using biased utilities can distort allocation decisions. These authors found that the SG utility for EQ-5D-3L state 22322 was 0.59 before correction and 0.46 after applying PT-based corrections. Therefore, the benefit of a medical treatment that restores people to full health would be underestimated by nearly 30% if using the former value. On the contrary, if the corrected value was used, almost 30% more money on the treatment could be justified, which highlights the critical impact of bias correction.

The main objective of this article is to demonstrate that it is feasible to obtain an entire value set (using the EQ-5D-3L instrument) free of utility curvature and probability weighting biases, based on face-to-face interviews of general population. It is emphasized that we do not intend to propose a specific survey design as the standard for estimating corrected tariffs but rather we are opening a path to further elaboration and refinement of the "corrective approach" for its application in future estimations.

Although some previous large-scale attempts to correct for discounting EQ-5D tariffs exist,[26,30,31] no full time-preference corrected EQ-5D tariff has been estimated up to now. It has been done, however, by Jonker et al[32] for the Dutch SF-6D tariff. All of these works only consider discounting and participants were surveyed online.

Likewise, although Lipman et al,[33] indeed, demonstrate that it is feasible to apply a corrective approach (based on PT assumptions) with interviewer-assisted data collection with the general population, they interviewed only a nonrandom sample of 150 people.

This article presents the first EQ-5D-3L tariff corrected under non-EU (RDU specifically), following the approach used in an earlier study,[34] whose data are partially reported here. Two value sets are introduced that differ in the method used to elicit preferences for a large representative sample (N = 1676) of the Spanish general population. One of the methods (the traditional TTO) is riskless, whereas the other one (coined as the risky-TTO) is risky. Potential discrepancies across both methods cast a question as to whether utilities elicited in one decision context can be transferred to another one. Previous evidence on transferability of utility has been mostly based on testing external validity of TTO and SG.[10,35] This article provides more extensive evidence on this issue by comparing responses to riskless and risky TTOs, what constitutes the second aim of our article.

The article is organized as follows. Section 2 begins by providing an outline of RDU corrections, after which elicitation methods and the semiparametric correction procedure used are described. Next, the transferability test is explained. Lastly, the valuation survey and the econometric approach implemented to estimate EQ-5D-3L tariffs are presented. Section 3 reports the main findings of the study. Then, the main conclusions are summarized. Discussion closes the article.

Methods

Outline of RDU Corrections

We consider prospects denoted by *([Q_1, T_1]p; [Q_2, T_2])*, yielding outcome *(Q_1, T_1)* with probability *P* and outcome *(Q_2, T_2)* with probability *1-P*. Henceforth, any outcome *(Q, T)*, in which *Q* denotes health status and *T* life duration, is identified with the corresponding riskless prospect (1, *[Q, T]*). The remaining prospects are risky. Throughout the article, we assume that risky prospects are rank-ordered. That is, when we write ([Q_1, T_1], P;[Q_2, T_2])) we assume that *(Q_1, T_1) ≥ (Q_2, T_2)*.

In this article, both better (BTD) and worse (WTD) than death health states are considered, henceforth denoted as Q^+ and Q^-, respectively. It is also assumed that *U(Q, T)= H(Q)L(T)*, in which *U* is a real-valued function over outcomes *(Q, T)*, and *H* and *L* are utility functions over *Q* and *T*, respectively. Utility is linear in duration if *U(Q, T)= H(Q)T*. Nonlinear utility is also considered throughout the article by assuming a power specification defined by $L(T) = T^{\beta}$.[31,36] According to the usual scaling *H(FH)* = 1 and *U(Death)* = 0, in which *FH* stands for full health.

Under EU prospects *([Q_1, T_1], p; [Q_2, T_2])* are evaluated as:

$$pU(Q_1,T_1)+(1-p)U(Q_2,T_2) \tag{1}$$

RDU generalizes EU by allowing probability weighting. Under RDU prospects *([Q_1, T_1], p; [Q_2, T_2])* are evaluated as:

$$w(p)U(Q_1,T_1)+(1-w(p))U(Q_2,T_2) \tag{2}$$

in which *w* is a probability weighting function. This function satisfy *w(0)* = 0 and *w(1)* = 1 and is strictly increasing over the interval [0, 1].

PT generalizes RDU by allowing sign dependence, whereby outcomes in risky prospects are perceived as gains or losses relative to a reference point. In addition, individuals are assumed to be more sensitive to losses than to gains, a phenomenon known as loss aversion. As Wakker demonstrates, for 2-outcome prospects, such as *([Q_1, T_1], p; [Q_2, T_2])*, PT for gains aligns with RDU.[37]

Elicitation Methods and Semiparametric Correction Procedure

If Q^+ is considered BTD, the TTO asks for the duration T_{TTO} that leads to indifference between the outcome *(FH, T_{TTO})* and the outcome *(Q^+, T)*. Conversely, if Q^- is WTD, the TTO asks for the duration T^*_{TTO} that leads to indifference between the outcome *(Q^-, T-T^*_{TTO}; FH, T^*_{TTO})* and death. Duration *T* is set at 10 years.

For BTD states, if utility is linear in duration, the TTO indifference leads to:

$$H(Q^+)=\frac{T_{TTO}}{10} \tag{3}$$

In the case of states WTD, the framing used in the TTO is not symmetrical to that used for BTD states. Whereas for the former the more time is attached to *FH*, the lower is the valuation for Q^-, for the latter, the opposite occurs: the more time is attached to *FH*, the higher the utility is for Q^+. As argued elsewhere,[38] these procedural differences call into question the validity of aggregating BTD and WTD TTO values. Linear utility for Q^- is computed then as follows:

$$H(Q^-)=-T^*_{TTO}\,/\,(10-T^*_{TTO}) \tag{4}$$

Assuming that the utility for life duration fits a power specification, ie, $L(T) = T^{\beta}$, this leads under RDU to the following:

$$H(Q^+)=\left(\frac{T_{TTO}}{10}\right)^{\beta} \text{ and } H(Q^-)=-\left(T^*_{TTO}/(10-T^*_{TTO})\right)^{\beta} \tag{5}$$

If Q^+ is regarded as BTD, the risky-TTO (rTTO) asks for the duration T_{rTTO} that leads to indifference between the risky prospect *([FH, T_{rTTO}], P; [Death])* and the risky prospect *([Q^+, T], P; Death)*. If Q^- is WTD, then the rTTO asks for the duration T^*_{rTTO} that leads to indifference between *([FH, T^*_{rTTO}], P; [Death])* and *([FH, T], P; [Q^-, T])*. Throughout our elicitations, *T* = 10 years and *P* = .5.

Note that the framing used in the rTTO for states WTD asks for the indifference value T^*_{rTTO} in the same way as the rTTO version for states BTD. Setting *P* = .5 ensures that rTTO utilities are bounded between +1 and −1. This contrasts with TTO utilities (Eq. [4]), which are not bounded below 0 at −1. Although it is possible to rescale raw negative TTO utilities, for example, by applying the rescaling procedure used in the Measurement and Valuation of Health study that led to the EQ-5D-3L tariff for the United Kingdom,[2,38] the resulting rescaled TTO utilities would not be equivalent to rTTO utilities, which are censored at a utility of -1 by construction. Therefore, rather than rescaling negative TTO utilities, all TTO utilities lower than -1 will be censored to -1.

The evaluation of rTTO indifferences under RDU for Q^+, if utility is linear in duration, leads to the same utility as in the TTO (see Eq. [3]). If, on the contrary, we assume that $L(T) = T^\beta$, then:

$$H(Q^+) = \left(\frac{T_{rTTO}}{10}\right)^\beta \tag{6}$$

And the utility for Q^- is as follows:

$$H(Q^-) = \frac{w(0.5)\left((T^*_{rTTO})^\beta - (10)^\beta\right)}{(1-w(0.5))10} \tag{7}$$

The semiparametric method used to identify parameters b and $w^+(0.5)$ at the individual level, based on certainty equivalent (CE) elicitations, is that proposed by Miyamoto.[39] It needs substantially fewer questions than nonparametric approaches and therefore is less time consuming.[40] In addition, and in contrast to chained indifferences used in nonparametric methods (eg, Attema et al[25]), the CEs used here are not susceptible to error propagation, because they are not linked. Finally, this method needs not make parametric assumptions about probability weighting because it is only necessary to identify its value for $P = .5$.

The CE method asks for the duration T_{CE} that leads to indifference between the outcome (Q, T_{CE}) and the risky prospect $([Q, T_1], p; [Q, T_2])$, with $T_1 > T_2$, regardless of whether Q is regarded as BTD or WTD. Durations T_1 and T_2 are varied across the 6 CE questions as follows: CE1 = (T_1 = 8, T_2 = 0); CE2 = (10, 2); CE3 = (12, 4); CE4 = (16, 0); CE5 = (20, 4); CE6 = (24, 8). Throughout the measurements, probability was fixed at .5. We obtain 6 different CEs T_{CE} or T^*_{CE}, depending on whether the health state Q is regarded as BTD or WTD. The following equations are analyzed under RDU and solved by nonlinear regression analysis:

$$T_{CE} = \left(w^+(0.5)T_1^\beta + (1-w^+(0.5))T_2^\beta\right)^{1/\beta} \tag{8}$$

$$T^*_{CE} = \left((1-w^+(0.5))T_1^\beta + (w^+(0.5))T_2^\beta\right)^{1/\beta} \tag{9}$$

Note that all corrections made through Eq. (5)-(9) are health-state specific, which is a unique feature of our corrective approach.

The Transferability Test

Within the domain of states BTD, the TTO is an increasing monotonic transformation of the rTTO, in the sense that the former results from the latter by attaching all the probability mass to the most desirable outcome. According to our assumptions, utility for Q^+ remains the same irrespective of whether indifferences are analyzed through the TTO or the rTTO (Eq. [3] results). This equivalence does not change, although utility for duration is described by $L(T) = T^\beta$. Consequently, Eq. (5) (for H(Q+)) and Eq. (6) would have to provide the same result. Therefore, we can then state that utility is transferable across the TTO and the rTTO if for a same state BTD, the response given to the TTO is the same as that given to the rTTO.

Survey Design

The sample was representative of the Spanish adult general population in terms of age and sex (N = 1755). Surveys were conducted in 2 different regions (Andalucia and Region of Murcia) by 2 survey companies, using 20 interviewers for the task. The sample represents 2 different waves of the study, conducted in 2009 and 2019. All interviews were conducted through computer-assisted personal interviewing, with an average duration of 20 minutes.

The questionnaire was organized into 5 sections. Sections 1 (description of the health states valued directly and then rated on a visual analog scale, [VAS]), 3 (CE questions), and 5 (socio-demographic questions) were identical for all the respondents. Building on previous work by Lamers et al,[41] the set of health states selected for our study (see Table 1 in section Uncorrected Health-State Utilities) was the same as that proposed by Macran and Kind,[42] plus an additional state (ie, 13212). According to the Paris protocol[43] the 18 states were distributed into 9 balanced subsamples (N_i = 195) of 2 states each, anonymously labeled as X and W, being X better than W. The order of appearance of states was randomized in each of the valuation tasks. The order of sections 2 and 4, which contained TTO and rTTO questions, was varied at random from one interview to another. Health states were regarded as BTD or WTD throughout TTO and rTTO measurements from an initial choice between (Q, 10) and Death for the TTO and between *([FH, T10], 0.5; [Death])* and *([FH, 10], 0.5; [Q, 10])* for the rTTO.

TTO, rTTO, and CE elicited preferences using an iterative choice-based matching method called parameter estimation by sequential testing (PEST).[44] The logic of this procedure is to avoid that participants may realize the converging sequence toward indifference, which could induce strategic behavior (eg, always choosing the same alternative to shorten the task). In this way, the PEST method obscures the iteration routine, making the goal of reaching indifference less transparent to subjects.[45] Non-transparent techniques are less influenced by biases than transparent methods.[46] Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.05.004 illustrates how elicitation methods were presented to participants and describes the PEST method.

Modeling

Regression analyses were applied to predict the corrected Spanish EQ-5D-3L tariffs from the direct valuation of the 18 selected health states. We estimated 2 different EQ-5D-3L scoring algorithms based on corrected TTO and rTTO valuations (Eq. [5]-[9]).

Two additive models were estimated for each tariff, specifically the main effects (or "plain") model and the N3 model. We selected between them based on the criteria of consistency, goodness of fit, and parsimony.[2] The models were estimated using the random effects estimator, and the constant term was constrained to unity.[47] This ensures that utility of FH equals 1.

Results

Sample

A total of 79 individuals were excluded because of inconsistent responses. First, 28 respondents assigned valuations to the more severe state (W), higher than those assigned to the mildest state (X) for a same method; 8 out of those 28 were inconsistent in the VAS, 1 in the TTO and the remaining 19 in the rTTO task. Another 51 individuals regarded one of the health states as WTD (BTD) using 1 of the 3 methods and BTD (WTD) with another. Inconsistencies were most frequent in the comparison between TTO

Table 1. Uncorrected TTO and rTTO utilities and comparisons for BTD states (transferability test).

Health state	Mean (SD) utilities		Transf. test *P* values	
	TTO	rTTO	Wilcoxon signed-rank test	*t* test
11112	0.838 (0.225)	0.824 (0.206)	.245	.096
11113	0.626 (0.180)	0.636 (0.176)	.356	.044
11121	0.819 (0.163)	0.785 (0.213)	.941	.037
11131	0.593 (0.195)	0.560 (0.216)	1.000	.084
11133	0.286 (0.222)	0.248 (0.315)	.224	.469
11211	0.853 (0.114)	0.849 (0.113)	.225	.233
11312	0.489 (0.200)	0.511 (0.208)	.018	.001
12111	0.774 (0.173)	0.851 (0.154)	.000	.000
13212	0.392 (0.207)	0.371 (0.219)	.000	.015
13311	0.391 (0.227)	0.304 (0.280)	.000	.000
21111	0.814 (0.146)	0.816 (0.137)	.549	.648
22222	0.112 (0.138)	0.032 (0.453)	.699	.000
23232	−0.232 (0.423)	−0.149 (0.433)	.000	.000
32211	0.171 (0.372)	0.137 (0.396)	.000	.001
32223	−0.361 (0.429)	−0.239 (0.398)	.000	.000
32313	−0.250 (0.438)	−0.205 (0.426)	.000	.000
33323	−0.590 (0.434)	−0.374 (0.372)	.454	.039
33333	−0.755 (0.276)	−0.384 (0.350)	.500	.100

TTO indicates time trade-off; rTTO-N3, risky-TTO N3 model.

and rTTO, occurring for 38 individuals, followed by the comparison between rTTO and VAS, which occurred for 19 individuals, and in the comparison between TTO and VAS, it occurred for 13 respondents. It should be noted that some individuals exhibited inconsistencies in more than 1 comparison. The final sample consisted of 1676 individuals (Table 2), maintaining its representativeness.

Uncorrected Health-State Utilities

Raw utilities, as well as results of the transferability test, are shown in Table 1. Uncorrected TTO values range from −0.755 (state 33333) to 0.853 (state 11211). The lowest and the highest values for rTTO utilities are also for these same states (−0.384 and 0.849, respectively). Significant differences ($P < .05$) between raw TTO and rTTO utilities were found in 14 health states with the paired *t* test. (The states for which the null hypothesis cannot be rejected are 11112, 11131, 11211, and 21111.) These discrepancies are reduced to 10 states when the nonparametric Wilcoxon signed-rank test is used. (There is no significant difference for states 11112, 11113, 11121, 11131, 11211, 21111, 11133, and 22222.)

As is commonly observed with the TTO,[48] there is a significant concentration of responses at certain points within the utility distribution. For example, for the most severe health state 33333, two-thirds of the sample converged on the lowest possible value. Conversely, for the mildest health state 11112, 1 in 6 respondents converged on the highest possible utility value.

Lastly, as shown in Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.05.004, preferences remained stable across the same elicitation method (TTO and rTTO) over the years, between the 2 studies in which the data were collected.

Transferability

To test transferability (Table 1), comparisons between TTO and rTTO questions are restricted to those respondents who considered health states as BTD. Significant differences were observed for 12 out of 18 health states between TTO and rTTO questions (paired *t* test, $P < .05$). Therefore, the null hypothesis of equality is rejected for around two-thirds of the states. When the nonparametric test is used, a lower number of statistically significant discrepancies results, but they still affect nearly 50% of the states (8 out of 18).

Corrected Health-State Utilities

As noted above, optimal parameter estimates, β and $w(0.5)$, were identified from CE elicitations (Eq. [8]-[9]) at the individual level. These estimates were then used to correct TTO and rTTO utilities according to Eq. (5)-(7). Median estimates for the power function for life duration (Table 3) describe concavity, and $w(0.5) < 0.5$, suggesting that participants tended to underweight probability equals 0.5, consistent with previous evidence.[22,31]

Corrected utilities, accounting for both utility curvature and probability weighting, are shown in Table 4. The highest mean values correspond to the state 11211 (0.873 with the TTO method and 0.870 with the rTTO), whereas the state 33333 is scored the lowest (−0.772 and −0.415, respectively).

There are statistically significant differences for 11 health states between corrected TTO and rTTO valuations, as indicated by the *t* test ($P < .05$), accounting for approximately 61% of all the states. When the nonparametric test is used, significant differences emerge for 10 states.

Table 2. Characteristics of the sample.

Characteristics	*N* = 1676	%
Gender		
Female	844	50.36
Male	832	49.64
Age (years)		
18-29	368	21.96
30-41	450	26.85
42-53	384	22.91
54-65	279	16.65
>65	195	11.63
Marital status		
Single	607	36.22
Married or coupled	910	54.30
Separated, divorced, or widow	159	9.49
Number of children (mean)	0.83	
Educational level		
No studies	130	7.76
Primary	383	22.85
Secondary	668	39.86
Higher	495	29.53
Income level (Euros)		
Up to 900	133	7.94
901-1500	372	22.20
1501-2000	513	30.61
2001-3000	402	23.99
>3000	256	15.27
Smoker (%)	33.00	
Private medical insurance (%)	22.91	
Self-assessed health condition (EQ-5D)		
11111	1151	68.68
11121	198	11.81
11122	68	4.06
Other	262	15.45

EQ-5D Tariffs

Regression models based on TTO and rTTO data are presented in Table 5. These include 4 sets of estimates: 2 main effects or "plain" models (1 and 3) and 2 N3 models (2 and 4).

The coefficients of the models have the expected sign and are highly significant (paired *t* test, $P < .001$), except for the TTO N3 model. Moreover, all the estimates are consistent (ie, coefficients for level 3 are higher than for level 2). Overall, the largest deviations from full health concern mainly mobility and pain/discomfort dimensions.

It is apparent that almost all of the parameter estimates of models based on TTO data (1 and 2) are higher (in absolute value) than those for models based on rTTO data (3 and 4). The largest difference between parameter estimates of TTO and rTTO models was 0.182 and 0.167, respectively, for mobility level 3.

Regarding the selection of the best models, in the case of TTO models, the "plain" TTO model is preferred to model 2 because of its parsimony and consistency because it predicts slightly better without any variable with an implausible sign. For the rTTO models, the model with the N3 term significantly outperformed model 4, with a mean absolute error 26% lower. The comparison of the 2 selected models suggest quite similarity in terms of predictive validity. It cannot therefore be concluded that one clearly outperforms the other.

Table 6 shows some statistics for the value sets generated by our 2 selected models (TTO and rTTO-N3) in comparison with the existing Spanish tariff for the EQ-5D-3L estimated by Badía et al.[49]

Table 3. Parameter estimates under RDU.

Health state	Median (SD) estimates	
	β_RDU	w(0.5)
11112	0.577 (0.300)	0.410 (0.089)
11113	0.535 (0.316)	0.395 (0.102)
11121	0.584 (0.331)	0.422 (0.105)
11131	0.634 (0.253)	0.422 (0.104)
11133	0.581 (0.347)	0.448 (0.125)
11211	0.571 (0.350)	0.410 (0.099)
11312	0.632 (0.317)	0.414 (0.101)
12111	0.582 (0.361)	0.402 (0.103)
13212	0.605 (0.322)	0.420 (0.102)
13311	0.584 (0.338)	0.419 (0.106)
21111	0.570 (0.268)	0.391 (0.091)
22222	0.571 (0.250)	0.435 (0.116)
23232	0.890 (0.413)	0.472 (0.104)
32211	0.668 (0.440)	0.459 (0.116)
32223	0.729 (0.689)	0.498 (0.149)
32313	0.646 (0.721)	0.486 (0.148)
33323	0.890 (0.840)	0.484 (0.119)
33333	0.864 (0.679)	0.474 (0.144)

β indicates the parameter of the power utility function for life duration; w, the probability weighting parameter for $P = 0.5$.

Discussion

The main objective of this article was to demonstrate the feasibility of estimating a full value set for the EQ-5D-3L, free of discounting and probability weighting. This aim was achieved by applying the so-called "corrective approach"[28] using interviewing techniques, sample sizes, and sample representativeness equivalent to those in previous studies.[5,50]

Several researchers claim[27,28,51] that health resources allocation, a prescriptive task, should rely on data as descriptively valid as possible. Therefore, tools such as the EQ-5D-3L, recommended for use in cost-effectiveness analyses,[52] should yield unbiased utilities to prevent incorrect resource allocations.

The article presents 2 corrected value sets for the EQ-5D-3L estimated under more realistic assumptions than typically used in social tariffs estimations. These corrections utilize RDU,[17,20] with parameter estimates tailored at the individual level for distinct health states, which is a unique feature of this work.

The 2 corrected tariffs reported are based on preference elicitations performed by using 2 similar methods: the traditional riskless TTO and a novel risky-TTO (rTTO) procedure. Both methods are (at least under EU and RDU), indeed, logically equivalent for those measurements made with BTD health states. This allowed us to test transferability of valuations measured in a riskless context to a risky one. Our findings caution against routinely taking transferability of utility for granted.

We are aware that there are different possible approaches to correct health-state utilities under non-EU. Specifically, Lipman et al[23,33] present PT-based corrections under a different approach

Table 4. Corrected TTO and rTTO utilities under RDU.

Health state	Mean (SD) utilities		*P* values	
	TTO	rTTO	MWW test	*t* test
11112	0.864 (0.205)	0.853 (0.182)	.245	.146
11113	0.679 (0.188)	0.688 (0.184)	.356	.058
11121	0.853 (0.147)	0.822 (0.191)	.941	.036
11131	0.633 (0.203)	0.603 (0.211)	1.000	.095
11133	0.367 (0.266)	0.328 (0.336)	.092	.029
11211	0.873 (0.107)	0.870 (0.106)	.225	.297
11312	0.543 (0.212)	0.562 (0.217)	.018	.002
12111	0.801 (0.177)	0.866 (0.152)	.000	.000
13212	0.454 (0.223)	0.429 (0.239)	.000	.001
13311	0.458 (0.240)	0.376 (0.274)	.000	.000
21111	0.838 (0.142)	0.839 (0.136)	.549	.855
22222	0.154 (0.238)	0.171 (0.341)	.623	.342
23232	−0.262 (0.478)	−0.149 (0.478)	.000	.003
32211	0.230 (0.431)	0.245 (0.362)	.001	.477
32223	−0.360 (0.472)	−0.209 (0.399)	.000	.000
32313	−0.243 (0.488)	−0.135 (0.405)	.000	.005
33323	−0.611 (0.440)	−0.360 (0.328)	.001	.000
33333	−0.772 (0.378)	−0.415 (0.267)	.000	.000

MWW indicates Mann-Whitney-Wilcoxon; TTO, time trade-off; rTTO-N3, risky-TTO N3 model.

that offers some advantages over that implemented by us. On the one hand, they introduce a loss aversion index to losses in duration, offering a tractable equation for loss aversion in TTO measurements. Moreover, operationalizing loss aversion in this manner prevents variations in the reference point across BTD and WTD states. In contrast, our RDU-based corrections overlook loss

Table 5. Regression models based on corrected TTO and rTTO utilities.

Coefficients	(1) TTO	(2) TTO-N3	(3) rTTO	(4) rTTO-N3
Constant	1.000	1.000	1.000	1.000
Mobility 2	−0.223*	−0.210*	−0.166*	−0.198*
Mobility 3	−0.510*	−0.507*	−0.328*	−0.340*
Self-care 2	−0.175*	−0.181*	−0.195*	−0.178*
Self-care 3	−0.310*	−0.319*	−0.262*	−0.242*
Usual activities 2	−0.125*	−0.144*	−0.191*	−0.143*
Usual activities 3	−0.265*	−0.287*	−0.295*	−0.241*
Pain/discomfort 2	−0.197*	−0.186*	−0.174*	−0.194*
Pain/discomfort 3	−0.381*	−0.395*	−0.350*	−0.305*
Anxiety/depression 2	−0.163*	−0.170*	−0.139*	−0.121*
Anxiety/depression 3	−0.312*	−0.331*	−0.291*	−0.238*
N3		0.0399*		−0.103*
Obs	1676	1676	1676	1676
MAE	0.0321	0.0340	0.0345	0.0255
|pred. Error| < k				
k = 0.01	3.460	3.252	2.506	3.222
k = 0.05	14.707	15.006	14.409	14.230
k = 0.10	30.489	31.026	29.534	28.609

MAE indicates mean absolute error; TTO, time trade-off; rTTO-N3, risky-TTO N3 model.
*Paired *t* test, $P < .001$.

Table 6. Descriptive statistics of TTO, rTTO-N3, and Badía et al[49] tariffs' predicted values.

Descriptive statistics	TTO	rTTO-N3	Tariff by Badía et al
Mean	0.140	0.176	0.123
SD	0.318	0.274	0.329
Min	−0.702	−0.465	−0.653
p10	−0.271	−0.162	−0.273
p25	−0.085	−0.016	−0.096
p50	0.141	0.153	0.108
p75	0.363	0.344	0.286
p90	0.541	0.534	0.652
% negative values	33.33	26.75	37.45

TTO indicates time trade-off; rTTO-N3, risky-TTO N3 model.

aversion, implicitly assuming that all outcomes are perceived as gains. On the other hand, the nonparametric methodology by Lipman et al[22,23] avoids biases related to specific functional forms, whereas this article assumes a power functional for life duration.

Building on the above, other authors[16,51] have applied PT-based corrections exclusively for risky prospects, similarly omitting loss aversion in TTO measurements as done in our study. Likewise, support for using a power function has been reported elsewhere.[31] Furthermore, our semiparametric approach to estimating power parameters requires only the identification of the probability weighting function for a probability of .5, similar to the method used by Attema et al.[40] This approach reduces error propagation and is more efficient than nonparametric alternatives. Additionally, we utilized an iterative choice-based matching method to elicit preferences, concealing the convergence toward indifference, potentially minimizing inconsistencies compared with "transparent" methods.[46] Lastly, although TTO measurements are influenced by scale compatibility,[21] no correction has been applied to address this bias.

Our study faces various limitations. First, we acknowledge that, although there are other studies[41,53-55] apart from ours having estimated EQ-5D-3L tariffs with far fewer states than those (ie, 42) used in the Measurement and Valuation of Health study, it is undeniable that the more health states are valued, the smaller the risk of overfitting the models. Additionally, the application of RDU functionals to CE elicitations in our study depends on whether the health state is considered BTD or WTD, assuming a single duration (T = 10); therefore, maximum endurable time preferences[56] cannot be discarded. Lastly, if the TTO was affected by loss aversion, it would result in an upward bias, which could explain part of the differences between TTO and rTTO responses in the transferability test.

This article demonstrates the feasibility of implementing the corrective approach to generate corrected value sets, contributing to bridge the gap between "the current state-of-art in the literature and policy."[28] Our findings potentially pave the way for estimating a corrected EQ-5D-5L tariff for Spain and comparing it with the current standard.[57] Additionally, our findings reveal that utilities obtained in different contexts may not be directly interchangeable. Hence, it points out a significant research challenge: to ensure that value sets are not context dependent. Future research could explore this issue using mixed effects models to gauge the extent of transferability problems.

Finally, the rTTO method used in our study provides symmetrical framing for both BTD and WTD health states, representing an advantage over the traditional TTO approach. This new procedure, akin the "lead time" TTO,[58] is based on the comparison of 2 gambles, such as lottery equivalent methods.[59] An interesting line for future research would be the comparison of the rTTO and the composite TTO[60] used in the EQ-VT protocol, which suffers from limited discriminatory ability for WTD health states.[61]

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2024.05.004.

Article and Author Information

Accepted for Publication: May 7, 2024

Published Online: June 12, 2024

doi: https://doi.org/10.1016/j.jval.2024.05.004

Author Affiliations: Applied Economics Department, Faculty of Economics and Business, University of Murcia, Murcia, Región de Murcia, Spain (Abellan-Perpiñan); Applied Economics Department, Faculty of Economics and Business, University of Murcia, Murcia, Región de Murcia, Spain (Martinez-Perez); Applied Economics Department, Faculty of Economics and Business, University of Murcia, Murcia, Región de Murcia, Spain (Sanchez-Martinez); Department of Economics, School of Economics and Business, University of Navarra, Pamplona, Navarra, Spain (Pinto-Prades).

Correspondence: Jose-Maria Abellan-Perpiñan, PhD, Applied Economics Department, Faculty of Economics and Business, University of Murcia, Murcia, Región de Murcia, Spain. Email: dionisos@um.es

Author Contributions: *Concept and design:* Abellan-Perpiñan, Martinez-Perez, Sanchez-Martinez, Pinto-Prades
Acquisition of data: Abellan-Perpiñan, Martinez-Perez, Sanchez-Martinez, Pinto-Prades
Analysis and interpretation of data: Abellan-Perpiñan, Martinez-Perez, Sanchez-Martinez, Pinto-Prades
Drafting of the manuscript: Abellan-Perpiñan, Martinez-Perez, Sanchez-Martinez, Pinto-Prades
Critical revision of the paper for important intellectual content: Abellan-Perpiñan, Martinez-Perez, Sanchez-Martinez, Pinto-Prades
Statistical analysis: Martinez-Perez
Obtaining funding: Martinez-Perez, Sanchez-Martinez, Pinto-Prades
Supervision: Sanchez-Martinez

Funding/Support: Financial support for this study was provided entirely by a grant from the Spanish Ministry of Economy, Industry and Competitiveness, grant PID2019-104907GB-I00, and a grant from the Séneca Foundation (Science and Technology Agency of the Region of Murcia), grant 20825/PI/18. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Acknowledgment: The authors thank the reviewers for their comments, which have undoubtedly contributed significantly to improving the article.

REFERENCES

1. Pinto JL, Herrero C, Abellán JM. QALY-based cost effectiveness analysis. In: Adler MD, Fleurbaey M, eds. *The Oxford Handbook of Well-Being and Public Policy*. Oxford: Oxford University Press; 2016:160–192.
2. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
3. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21(2):271–292.

4. Tsuchiya A, Brazier J, Roberts J. Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. *J Health Econ*. 2006;25(2):334–346.
5. Xie F, Gaebel K, Perampaladas K, Doble B, Pullenayegum E. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Med Decis Mak*. 2014;34(1):8–20.
6. EuroQol Research Foundation. EQ-5D-3L user guide; Published 2018. https://euroqol.org/publications/user-guides. Accessed September 1, 2023.
7. Wang L, Poder TG. A systematic review of SF-6D health state valuation studies. *J Med Econ*. 2023;26(1):584–593.
8. Read JL, Quinn DM, Berrick HV, Fineberg HV, Weinstein ML. Preferences for health outcomes: comparison of assessment methods. *Med Decis Mak*. 1984;4(3):315–329.
9. Stiggelbout AM, Kiebert GM, Kievit J, Leer JWH, Stoter G, de Haes JCJM. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Med Decis Mak*. 1994;14(1):82–90.
10. Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *J Health Econ*. 1997;16(2):155–175.
11. Dyer JS, Sarin RK. Relative risk aversion. *Manag Sci*. 1982;28(8):875–886.
12. Gafni A, Birch S, Mehrez A. Economics, health, and health economics: HYEs versus QALYs. *J Health Econ*. 1993;12(3):325–339.
13. Bleichrodt H, Schmidt U. A context-dependent model of the gambling effect. *Manag Sci*. 2002;48(6):802–812.
14. Richardson J. Cost utility analysis: what should be measured? *Soc Sci Med*. 1994;39(1):7–22.
15. Wakker PP. Separating marginal utility and probabilistic risk aversion. *Theor Decis*. 1994;36(1):1–44.
16. Van Osch SM, Wakker PP, Van Den Hout WB, Stiggelbout AM. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Mak*. 2004;24(5):511–517.
17. Quiggin J. Theory of anticipated utility. *J Econ Behav Organ*. 1982;3(4):323–343.
18. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica*. 1979;47(2):263–291.
19. Tversky A, Kahneman D. Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertainty*. 1992;5(4):297–323.
20. Wakker PP, Stiggelbout AM. Explaining distortions in utility elicitation through the rank-dependent model for risky choices. *Med Decis Mak*. 1995;15(2):180–186.
21. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ*. 2002;11(5):447–456.
22. Bleichrodt H, Pinto JL. A parameter-free elicitation of the probability weighting function in medical decision analysis. *Manag Sci*. 2000;46(11):1485–1496.
23. Lipman SA, Brouwer WBF, Attema AE. QALYs without bias? Non-parametric correction of time trade-off and standard gamble weights based on prospect theory. *Health Econ*. 2019;28(7):843–854.
24. Bleichrodt H, Pinto JL. The validity of QALYs under nonexpected utility. *Econ J*. 2005;115(503):533–550.
25. Attema A, Bleichrodt H, Wakker P. A direct method for measuring discounting and QALYs more easily and reliably. *Med Decis Mak*. 2012;32(4):583–593.
26. Attema AE, Brouwer WB. Deriving time discounting correction factors for TTO tariffs. *Health Econ*. 2014;23(4):410–425.
27. Pinto-Prades J-L, Abellan-Perpiñan J-M. When normative and descriptive diverge: how to bridge the difference. *Soc Choice Welf*. 2012;38(4):569–584.
28. Lipman SA, Brouwer WBF, Attema AE. The corrective approach: policy implications of recent developments in QALY measurement based on prospect theory. *Value Health*. 2019;22(7):816–821.
29. Bleichrodt H, Abellan-Perpiñan JM, Pinto-Prades JL, Mendez-Martinez I. Resolving inconsistencies in utility measurement under risk: tests of generalizations of expected utility. *Manag Sci*. 2007;53(3):469–482.
30. Scalone L, Stalmeier PFM, Milani S, Krabbe PFM. Values for health states with different life durations. *Eur J Heal Econ*. 2015;16(9):917–925.
31. Craig BM, Rand K, Bailey H, Stalmeier PFM. Quality adjusted life-years without constant proportionality. *Value Health*. 2018;21(9):1124–1131.
32. Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Advocating a paradigm Shiftin health-state valuations: the estimation of time-preference corrected QALY tariffs. *Value Health*. 2018;21(8):993–1001.
33. Lipman SA, Attema AE, Versteegh MM. Correcting for discounting and loss aversion in composite time trade-off. *Health Econ*. 2022;31(8):1633–1648.
34. Abellan-Perpiñan JM, Martinez JE, Sanchez FI, Mendez I. *Debiasing EQ-5D tariffs. New estimation of the Spanish EQ-5D value set under nonexpected utility*. Documento de Trabajo; Published 2009.. http://www.centrodeestudiosandaluces.info/PDFS/E200906.pdf. Accessed September 1, 2023.
35. Abellan-Perpiñan JM, Bleichrodt H, Pinto-Prades JL. The predictive validity of prospect theory versus expected utility in health utility measurement. *J Health Econ*. 2009b;28(6):1039–1047.
36. Abellan-Perpiñan JM, Pinto-Prades JL, Mendez-Martinez I, Badia-Llach X. Towards a better QALY model. *Health Econ*. 2006;15(7):665–676.
37. Wakker PP. *Prospect Theory: for Risk and Ambiguity*. Cambridge, England: Cambridge University Press; 2010.
38. Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Econ*. 2006;15(4):393–402.
39. Miyamoto JM. Utility assessment under expected utility and rank-dependent utility assumptions. In: Chapman GB, Sonnenberg FA, eds. *Decision Making in Health Care*. Cambridge, England: Cambridge University Press; 2000.
40. Attema A, Brouwer W, l'Haridon O. Prospect theory in the health domain: a quantitative assessment. *J Heal Econ*. 2013;32(6):1057–1065.
41. Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJV. The Dutch tariff: results and arguments for an effective design for national. *Health Econ*. 2006;15(10):1121–1132.
42. Macran S, Kind P. *Valuing EQ-5D health states using a modified MVH protocol: preliminary results*. In: Badia X, Herdman M, Roset M, eds. 16th Plenary Meeting of the EuroQol Group, Sitges. Vol. 2000. Spain: Institut de Salut Pública de Catalunya; 1999 [Discussion Papers]:205-240.
43. Kind P. *A revised protocol for the valuation of health states defined by the EQ-5D-3L classification system - learning the lessons from the MVH study*. Paris: Presented at: 26th Plenary Meeting of the EuroQoL Group; 2009.
44. Luce RD. *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc; 2000.
45. Bostic R, Hernstein RJ, Luce RD. The effect on the preference reversal phenomenon of using choice indifferences. *J Econ Behav Organ*. 1990;13(2):193–212.
46. Pinto-Prades JL, Sánchez-Martínez FI, Abellán-Perpiñán JM, Martínez-Pérez JE. Reducing preference reversals: the role of preference imprecision and nontransparent methods. *Health Econ*. 2018;27(8):1230–1246.
47. Abellan-Perpiñan JM, Sanchez-Martinez FI, Martinez-Perez JE, Mendez-Martinez I. Lowering the 'floor' of the SF-6D scoring algorithm using a lottery equivalent method. *Health Econ*. 2012;21(11):1271–1285.
48. Stolk E, Ludwig K, Rand K, van Hout B, Ramos-Goñi JM. Overview, update, and lessons learned from the International EQ-5D-5L Valuation Work: version 2 of the EQ-5D-5L valuation protocol. *Value Health*. 2019;22(1):23–30.
49. Badía X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Mak*. 2001;21(1):7–16.
50. Devlin N, Brooks R. EQ-5D and the EuroQol Group: past, present and future. *Appl Health Econ Health Policy*. 2017;15(2):127–137.
51. Bleichrodt H, Pinto JL, Wakker PP. Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Manag Sci*. 2001;47(11):1498–1514.
52. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual; Published January 31, 2022. Updated October 31, 2023. https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741. Accessed September 1, 2023.
53. Chevalier J, Pouvourville G. Valuing EQ-5D using time trade-off in France. *Eur J Health Econ*. 2013;14(1):57–66.
54. Scalone L, Cortesi PA, Ciampichini R, et al. Italian population-based values of EQ-5D health states. *Value Health*. 2013;16(5):814–822.
55. Ferreira LN, Ferreira PL, Pereira LN, Oppe M. The valuation of the EQ-5D in Portugal. *Qual Life Res*. 2014;23(2):413–423.
56. Abellan-Perpiñan J-M, Martinez-Perez J-E, Pinto-Prades J-L, Sanchez-Martinez F-I. Testing nonmonotonicity in health preferences. *Med Decis Mak*. 2024;44(1):42–52.
57. Ramos-Goñi JM, Craig BM, Oppe M, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value Health*. 2018;21(5):596–604.
58. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. *Health Econ*. 2011;20(3):348–361.
59. McCord M, de Neufville R. Lottery equivalents: reduction of the certainty effect problem in utility assessment. *Manag Sci*. 1986;32(1):56–60.
60. Janssen BM, Oppe M, Versteegh MM, Stolk EA. Introducing the composite time trade-off: a test of feasibility and face validity. *Eur J Health Econ*. 2013;14(suppl 1):S5–S13.
61. Jakubczyk M, Lipman SA, Roudijk B, et al. Modifying the composite time trade-off method to improve its discriminatory power. *Value Health*. 2023;26(2):280–291.