

RESEARCH

Open Access



Scale and rate heterogeneity in the EQ-5D-5L valuation

Maksat Jumamyradov^{1*}, Benjamin M. Craig¹ and Michał Jakubczyk²

Abstract

Objectives To estimate values on a quality-adjusted life year (QALY) scale using individual preference evidence, choice analyses typically include ancillary parameters, such as scale factors and discount rates. These parameters potentially differ among respondents. In this study, we investigated how allowing heterogeneity in scale and rate affects the estimation of EQ-5D-5L values.

Methods Using the first wave of the 2016 EQ-5D-5L valuation study ($N = 1017$), we estimated a conditional logit (CL) model and three mixed logit models: random scale, random rate, and bivariate. Prior to the exploratory study, we hypothesized that scale and rate are correlated and that allowing heterogeneity in both parameters decreases the number of insignificant incremental effects. We confirmed the exploratory findings by re-estimating these models using paired comparison responses from a second wave ($N = 1229$).

Results Scale and rate exhibited significant heterogeneity and were positively correlated. As hypothesized, allowing this heterogeneity improved the face validity of the EQ-5D-5L value set by reducing the number of insignificant incremental effects (from 6 to 2 p -values > 0.05 ; out of 20). Nevertheless, the CL and bivariate mixed logit estimates are highly correlated and concordant (Pearson correlation coefficient of 0.897, Spearman correlation coefficient of 0.888, Lin's concordance coefficient of 0.763).

Conclusions Allowing this heterogeneity adds three parameters to the estimation (two variances and a correlation) and improves the face validity of the EQ-5D-5L values. This finding may influence experimental design and choice analysis in health valuation more generally.

Highlights

1. Allowing heterogeneity in scale and rate improves the estimation of EQ-5D-5L values in terms of face validity, namely, reducing the number of insignificant incremental effects.
2. From the high positive correlation between scale and rate, we can infer that people who heavily discount future outcomes are more sensitive to smaller differences in their net present value on a quality-adjusted life-year (QALY) scale.
3. Controlling for scale and rate heterogeneity slightly reduced the size of the incremental effects, raising the lower bound of the EQ-5D-5L values from -0.545 to -0.191.

Keywords Health valuation, EQ-5D-5L, Scale and discount rate

*Correspondence:

Maksat Jumamyradov
maksat@usf.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

In health valuation, the purpose is to estimate preference weights for health outcomes that represent societal values on a quality-adjusted life-year (QALY) scale. On a QALY scale, “immediate death” has a value of 0, and “Starting today 1 year with no health problem then die” has a value of 1. Apart from these two anchors, choice analyses often include ancillary parameters, such as scale factors and discount rates. The primary aim of this paper is to investigate how allowing heterogeneity in scale and rate affects the estimation of EQ-5D-5L values.

In a logistic regression, the scale parameter defines the proportional relationship between the value of the initial QALY and a change in the log-odds of choice. A smaller (larger) scale parameter implies that a larger (smaller) difference in value is necessary to achieve the same change in log-odds of choice. In other words, the scale parameter is an inverse measure of the size of the random component. Varying the scale parameter between individuals implies that some respondents have different sensitivities to the value of the initial QALY [1]. The sources of this scale heterogeneity may be related to their behavior (e.g., attention span) or preferences (e.g., connoisseur) [2]. In health preference research more generally, scaling parameters are estimated in analyses of willingness-to-pay (i.e., monetary scaling) and maximum acceptable risk (MAR).

Apart from the scale parameter, the value of a health outcome depends on temporal discounting. Starting in the 1970s, researchers characterized the value of quality-adjusted life span by simply multiplying quality of life by length of life (i.e., no discounting). However, in the late 2010s, it was shown that discounting may be incorporated into health valuations [3–5]. Discounting is widely accepted in economics and finance; however, some outcome researchers express health-state utilities anchored on “dead” and “full health” and do not account for temporal discounting. In economic evaluations more generally, the marginal utility of time is decreasing (i.e., each additional day is worth less than the prior day), so incorporating discounting into health valuation enhanced its coherence with microeconomic theory [6, 7]. More recently, Karim and colleagues showed how the discount rate may vary within and between latent classes [8]. The sources of rate heterogeneity may be related to the respondents’ perceptions of death (e.g., nontraders) or their marginal decrease in utility of life years.

Prior to the exploratory analysis, we hypothesized that by allowing individual-level randomness in these two ancillary parameters, the estimates of the EQ-5D-5L value set might improve in terms of face validity. The EQ-5D-5L descriptive system has five ordinal domains, each representing increasing severity of health problems.

Therefore, we assessed face validity by counting the number of insignificant incremental effects under alternative logit specifications estimated using a first survey wave. To complement this aim, we explored the variances and correlations of these parameters and their implications beyond health valuation.

As recommended by Craig, de Bekker-Grob, González Sepúlveda, and Greene, we confirmed the initial findings using a second wave [9]. The exploratory results led us to further hypothesize that scale and rate are positively correlated at the individual level. For example, the net present value (NPV) of the 10 QALYs depends on the discount rate, but the effect of NPV on the log-odds ratio depends on the scale parameter. Persons who discount heavily (lightly) may seem to be more (less) sensitive to differences in NPV, leading to a positive correlation. Analogously, a person who dislikes spicy foods may seem more sensitive to spice. Although this may now seem intuitive, to the best of our knowledge, no study has produced empirical evidence of this correlation.

The remainder of this paper is organized as follows. Section 2 describes the methods we used in this project, including the theoretical foundation, model specifications, exploratory (i.e., wave 1) and confirmatory (i.e., wave 2) data, and estimation techniques. In Sects. 3, 4 and 5, we provide the results, discussion and conclusions, respectively.

Methods

Random utility theory for paired comparisons

The theoretical framework of this choice analysis is based on random utility maximization (RUM) theory. According to RUM theory, the utility function $U_{itj} = V_{itj} + \varepsilon_{itj}$ of individual $i = 1, \dots, N$ for alternative $j = 1, \dots, J$ in choice situation $t = 1, \dots, T$ can be decomposed into a deterministic part of utility V_{itj} (representative utility) and a random part of utility ε_{itj} . In paired comparison modeling [10], individual i will choose an alternative j if and only if the probability that the utility associated with alternative j is higher than the utility of its alternative.

$$\begin{aligned}
 P_{itj} &= P(U_{itj} > U_{itk}), \forall k \neq j \\
 P_{itj} &= P(V_{itj} + \varepsilon_{itj} > V_{itk} + \varepsilon_{itk}), \forall k \neq j \\
 P_{itj} &= P(\varepsilon_{itk} - \varepsilon_{itj} < V_{itj} - V_{itk}), \forall k \neq j
 \end{aligned} \tag{1}$$

Choice probabilities are calculated based on a relative measure where the utility of one of the alternatives in the choice set is taken as a reference. To derive the choice probabilities, we need to make distributional assumptions about the random part of utility. The conditional logit (CL) model is derived under the assumption that ε_{itj} is independently and identically distributed (IID) with an extreme value type I (EV1) distribution [11–13]. As a

result, the difference between two IID EV1 random error terms ($\varepsilon_{itk} - \varepsilon_{itj}$) has a logistic distribution with scale parameter λ . This implies that the choice probabilities of the CL model can be expressed in terms of a logistic distribution with a cumulative distribution function

$$P_{itj} = \frac{1}{1 + \sum_{k=1}^J \exp[\lambda(V_{itk} - V_{itj})]}, \forall k \neq j \tag{2}$$

where λ is the scale parameter [14].

Scale and rate heterogeneity in health valuation

For this study, we extended the CL model (Eq. 2) for health valuation on a quality-adjusted life-year (QALY). By construction, the scale parameter is always positive, $\lambda = \exp(\mu)$, and represents the relationship between log-odds and the value of a health outcome V_{itj} on a QALY scale. We specify the value of a health outcome V_{itj} as a product of two values representing health V_{itj}^H and life years V_{itj}^Y :

$$V_{itj} = V_{itj}^H \times V_{itj}^Y \tag{3}$$

In this paper, we assume that the value of health $V_{itj}^H = 1 - \beta'x_{itj}$, where x_{itj} is a vector of 20 incremental indicators of health problems in mobility, self-care, usual activities, pain/discomfort and anxiety/depression (i.e., MO, SC, UA, PD, AD), and β is a vector of preference weights on a QALY scale. Its homogeneity is a simplifying assumption for the estimation of a single EQ-5D-5L value set that may be relaxed in future work.

More specifically, the value of the health profiles is parameterized using 20 incremental effects (i.e., 5 attributes with 4 levels each), where each effect is caused by a dummy variable representing an incremental change in the level of severity of an EQ-5D-5L attribute. Therefore, we can write

$$V_{itj}^H = 1 - \begin{pmatrix} \beta_1 MO_{12} + \beta_2 MO_{23} + \beta_3 MO_{34} + \beta_4 MO_{45} + \\ \beta_5 SC_{12} + \beta_6 SC_{23} + \beta_7 SC_{34} + \beta_8 SC_{45} + \\ \beta_9 UA_{12} + \beta_{10} UA_{23} + \beta_{11} UA_{34} + \beta_{12} UA_{45} + \\ \beta_{13} PD_{12} + \beta_{14} PD_{23} + \beta_{15} PD_{34} + \beta_{16} PD_{45} + \\ \beta_{17} AD_{12} + \beta_{18} AD_{23} + \beta_{19} AD_{34} + \beta_{20} AD_{45} \end{pmatrix} \tag{4}$$

As a criterion of face validity, all 20 incremental effects in vector β should be positive since they represent losses in value due to increases in the level of severity of a health condition from the full health profile [14].

For the value of life years V_{itj}^Y , the identity function is commonly assumed to be $V_{itj}^Y = Y_{itj}$, where Y_{itj} represents life years (i.e., no discounting). However, this functional form does not accurately represent the time preferences of the general population [4, 5]. Individuals usually

discount over time; i.e., future outcomes affect choices less than present outcomes. To allow for temporal discounting, we adapt the power function (see 4)

$$V_{itj}^Y = Y_{itj}^{\alpha_i}$$

where α_i is the individual-specific power. On a QALY scale, the value of time V_{itj}^Y equals 1 when Y_{itj} equals 1, regardless of the power α_i , and the identity function (i.e., no discounting) implies that the power is unity, $\alpha_i = 1$.

Apart from restricting the individual-specific scale parameter to be positive, $\lambda_i = \exp(\mu_i)$, we restricted the power α_i to the unit interval, $0 \leq \alpha_i \leq 1$. More specifically, we transform the power into a discount rate using the complementary log–log (CLL) function, $\alpha_i = \exp(-\exp(r_i))$ which is naturally bounded to the unit interval. At first glance, r_i has an inverse relationship with α_i , and a lower α_i implies greater discounting of life years; therefore, a higher rate r_i implies greater discounting. Future analyses may allow for negative discounting or alternative functional forms [15–17].

The bivariate distribution of the scale and rate among respondents

Due to limited panel evidence per respondent, it is not feasible to estimate individual-specific scales and rates as fixed effects (i.e., μ_i and r_i). Instead, we estimated a conditional logit (CL) model and three mixed logit models. First, we estimated the CL model under homogeneity ($\mu_i = \mu; r_i = r$). Under this specification, all respondents have the same scale parameter and discount rate. In the second and third specifications, we estimated the mixed logit models with random scale and random rate, respectively. We refer to these two mixed logit specifications as “univariate” models because each contains only one normally distributed random parameter.

Finally, in the fourth specification, we estimated a bivariate mixed logit model, including the mean and standard deviation of μ_i (i.e., μ and σ_μ , respectively) and r_i (i.e., r and σ_r , respectively), as well as their correlation. The ancillary parameters vary under a bivariate normal distribution and may be correlated. We assume that μ_i and r_i are normally distributed such that $(\mu_i, r_i) \sim N(\sigma_\mu^2, \rho, \sigma_r^2)$ where $\sigma_\mu^2 = Var(\mu_i)$, $\sigma_r^2 = Var(r_i)$ and $\rho = Corr(\mu_i, r_i)$.

To shed more light on this potential bias, we express the individual-specific ancillary component, $\lambda_i V_{itj}^Y$ (apart from the value of health), as an exponential regression with two ancillary parameters (an intercept μ_i and a coefficient α_i), where $Y_{itj} > 0$:

$$\lambda_i V_{itj}^Y = \exp(\mu_i) Y_{itj}^{\alpha_i} = \exp(\mu_i + \alpha_i \ln(Y_{itj}))$$

In this study, life years Y_{itj} range from 1 to 10 years; therefore, $\ln(Y_{itj})$ ranges from zero to 2.303. Given that $\ln(Y_{itj})$ is always positive, the ancillary component can increase through either ancillary parameter (μ_i or α_i). In econometric terms, $\ln(Y_{itj})$ is an instrumental variable needed to identify the two ancillary parameters.

Data

In 2016, 8,222 U.S. respondents (4074 in wave 1 and 4148 in wave 2) from all 50 states and Washington, D.C., completed an online survey that included 20 paired comparisons. The design of the paired comparisons was largely based on the EuroQol Valuation Technology (EQ-VT v1.0) protocols [18]. An example of the paired comparison conducted in the study is illustrated in Fig. 1. In this paper, we provide a general overview of the study. More details can be found in other studies [3, 19].

Each paired comparison is presented as a variation of health descriptions based on the EQ-5D-5L. The five dimensions (i.e., attributes) of the EQ-5D-5L are mobility, self-care, usual activities, pain/discomfort and anxiety/depression, where each dimension is characterized by five levels ranging from no problems (i.e., level 1) to slight, moderate, severe, and unable/extreme problems (i.e., level 5). For instance, the health description on the right side of Fig. 1 can be represented as a vector of five numbers 33333 since all five dimensions are at a moderate level. For each comparison, respondents were asked, “Which do you prefer?” regarding a pair of alternatives described using the EQ-5D-5L and lifespan attributes.

The online survey consisted of 3160 pairs, 1600 of which are efficient (or “quality only”) pairs and 1560 of which are quantity-quality pairs. In efficient pairs, both health descriptions consisted of varying levels of health problems with the same life years (e.g., 12345 vs 54321).

In the quantity-quality pairs, one of the health descriptions involves no health problems (i.e., 11111). Furthermore, 80 out of 1560 quantity-quality pairs included “immediate death”, which represents “dead” pairs, as one of the alternatives. The data were collected in two parts: an exploratory survey consisting of 1560 pairs and a confirmatory survey consisting of 1600 pairs. The survey data were collected at four temporal units (i.e., days, weeks, months, and years). This analysis included only the pairs with year units (1017 respondents in wave 1 and 1229 in wave 2) because the other pairs did not describe events after 1 year (i.e., discounting).

With the diversity of pairs, it is mathematically feasible to identify the scale and rate separately using either wave of this dataset. Imagine a paired comparison with identical lifespans. These pairs may identify differential scales within a population, μ_i . Imagine a paired comparison with differential life years. These pairs may identify differential scales, μ_i and rates, r_i . Apart from its pair types, this dataset is one of the largest national health valuation studies ever conducted [3], has both exploratory and confirmatory waves, and applied quota sampling at the pair level to assure that each pair had a minimum number of respondents along 18 demographic quotas.

Mixed logit and maximum simulated likelihood

To estimate the mixed logit models, the maximum likelihood (ML) estimator of parameter vector θ can be utilized when the density of dependent variable y_i conditional on a vector of independent variables x_i , $f(y_i|x_i, \theta)$, has a closed-form such that

$$\hat{\theta}_N = \operatorname{argmax}_{\theta} \sum_{i=1}^N \log f(y_i|x_i, \theta)$$

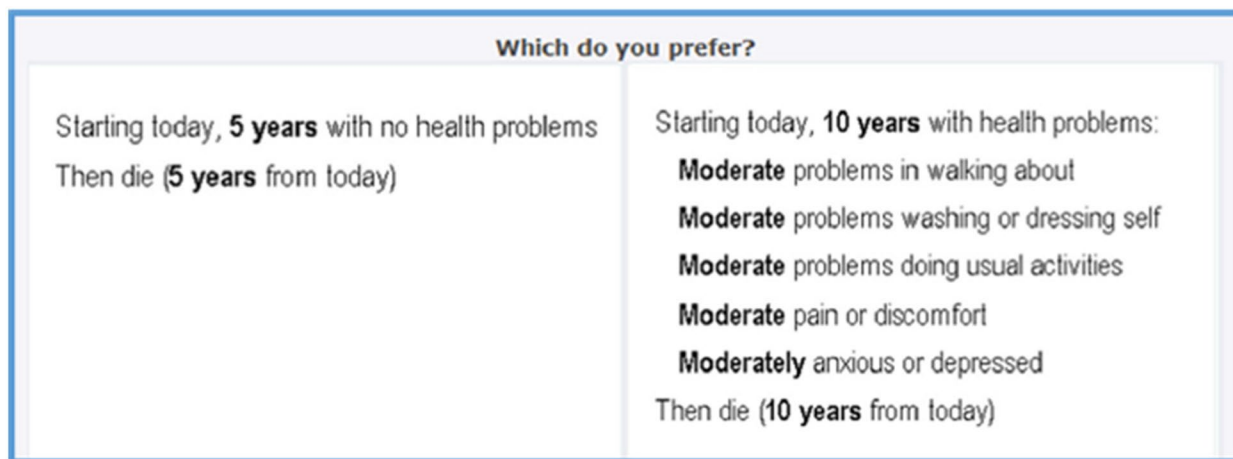


Fig. 1 Example of a paired comparison

where $i = 1, \dots, N$. However, ML is not feasible when $f(y_i|x_i, \theta)$ does not have a tractable closed-form. This can be because the density is specified only conditional on latent variables, which cannot be integrated out. Thus, the MSL estimator is a possible alternative [20, 21]. Suppose $\tilde{f}(y_i, x_i, u_i, \theta)$ is an unbiased simulator of the conditional density $f(y_i|x_i, \theta)$ such that

$$f(y_i|x_i, \theta) = E_u[\tilde{f}(y_i, x_i, u_i, \theta)|y_i, x_i]$$

where u_i is an individual-specific latent vector (μ_i and r_i) whose distribution is known and independent of (y_i, x_i) . Then, the MSL estimator of θ is defined as

$$\hat{\theta}_{SN} = \operatorname{argmax}_{\theta} \sum_{i=1}^N \log \left[\frac{1}{S} \sum_{s=1}^S \tilde{f}(y_i, x_i, u_i^s, \theta) \right]$$

where $u_i^s (s = 1, \dots, S)$ are drawn independently for each individual i from the distribution of u_i . The MSL estimator is obtained by replacing the intractable conditional p.d.f. $f(y_i|x_i, \theta)$ with its unbiased approximation based on the simulator $\tilde{f}(y_i, x_i, u_i^s, \theta)$. In this study, we estimate the mean and variance of each random parameter as well as their p -values [3].

In our MSL estimations of the three specifications of the mixed logit model, we use 250 Halton draws (i.e., $S = 250$) [22]. We used the MATLAB programming language for all estimations. More specifically, we began by estimating the CL comparator and three specifications using the wave 1 data, which helped us state our hypotheses more clearly. Afterwards, we re-estimated the models and tested these hypotheses using the wave 2 data. Furthermore, we compare the results between waves and models to assess how allowing heterogeneity in scale and rate affects the estimation of EQ-5D-5L values.

Results

In this section, we present the results for CL and mixed logit estimation using waves 1 and 2 separately. In Table 1, we compare the CL estimates with the mixed logit estimates where we allow correlations between μ_i and r_i . In Table 1A (Appendix), we present the univariate results for the mixed logit estimation with random scale and random rate separately.

Exploratory results

As shown in Table 1, the exploratory CL results produce three insignificant positive effects (p -value < 0.01; MO_{12} , UA_{45} and AD_{12}) and one insignificant negative effect for PD_{12} . There are also two additional effects with p -values between 1 and 5% (i.e., SC_{12} and UA_{12}). Furthermore, the CL results suggest that “immediate death” is

Table 1 Results for conditional and bivariate mixed logit models

N = 1017 & 1229	Conditional Logit		Bivariate Mixed Logit	
	Exploratory	Confirmatory	Exploratory	Confirmatory
MO_{12}	0.011	0.044 ^b	0.014 ^a	0.033 ^b
MO_{23}	0.044 ^b	0.050 ^b	0.044 ^b	0.064 ^b
MO_{34}	0.141 ^b	0.142 ^b	0.110 ^b	0.084 ^b
MO_{45}	0.117 ^b	0.122 ^b	0.069 ^b	0.034 ^b
SC_{12}	0.027 ^a	0.095 ^b	0.025 ^b	0.087 ^b
SC_{23}	0.038 ^b	0.016	0.025 ^b	0.045 ^b
SC_{34}	0.151 ^b	0.110 ^b	0.108 ^b	0.062 ^b
SC_{45}	0.153 ^b	0.147 ^b	0.098 ^b	0.090 ^b
UA_{12}	0.024 ^a	0.018	0.015 ^a	0.022 ^b
UA_{23}	0.027 ^b	0.011	0.024 ^b	0.027 ^b
UA_{34}	0.144 ^b	0.173 ^b	0.119 ^b	0.131 ^b
UA_{45}	0.030	0.099 ^b	0.027 ^b	0.056 ^b
PD_{12}	-0.002	-0.016	0.021 ^b	-0.006
PD_{23}	0.047 ^b	0.050 ^b	0.034 ^b	0.061 ^b
PD_{34}	0.225 ^b	0.215 ^b	0.161 ^b	0.153 ^b
PD_{45}	0.096 ^b	0.092 ^b	0.048 ^b	0.048 ^b
AD_{12}	0.020	-0.022	0.035 ^b	0.023 ^b
AD_{23}	0.100 ^b	0.046 ^b	0.058 ^b	0.045 ^b
AD_{34}	0.163 ^b	0.168 ^b	0.123 ^b	0.130 ^b
AD_{45}	0.033 ^b	-0.016	0.021 ^a	0.000
μ (mean)	0.401 ^b	0.062	0.988 ^b	0.704 ^b
r (mean)	-0.196 ^b	-0.634 ^b	-0.003	-2.051 ^b
μ (SD)			0.788 ^b	0.998 ^b
r (SD)			2.858 ^b	3.120 ^b
μ and r (corr)			0.836 ^b	0.912 ^b
$V(55555)$	-0.588 ^b	-0.545 ^b	-0.180 ^b	-0.191 ^b
LL	-12,225	-14,278	-10,249	-11,890

Notice that the scale parameter is equal to $\lambda = \exp(\mu)$, and the power is equal to $\alpha = \exp(-\exp(r))$ to bound the power between 0 and 1

^a, ^brepresent significance levels at the 5% and 1%, respectively

better than experiencing the worst possible EQ-5D-5L description for 1 year (i.e., $V(55555) = -0.588$). Since the estimated μ is 0.401, the scale parameter in the CL model is $\lambda = \exp(0.401) = 1.493$. Similarly, since the estimated r is -0.196, the power α in the CL model is $\exp(-\exp(-0.196)) = 0.439$.

In Table A1 (Appendix), we present the univariate results where we allow for random scale and random rate separately. The standard deviations of μ_i and r_i are 1.232 and 1.325, respectively, suggesting that scale and rate heterogeneity exist. However, the random scale model has insignificant effect (5 with p -value > 0.05), i.e., one more than the CL model, while the random rate model has same number of insignificant effects (4 with p -value > 0.05). In the exploratory results, allowing for one

random parameter increases the log-likelihood, but had little impact on the significance of the effects.

When we allow for heterogeneity in both ancillary parameters (Table 1), there are substantive improvements in the estimated incremental effects. In the bivariate mixed logit results, all 20 effects are positive and significant. The estimated standard deviations for μ_i and r_i are 0.788 and 2.858, respectively, which suggest that both the scale and rate parameters are heterogeneous. Furthermore, we find a strong correlation between μ_i and r_i , 0.836 (p -value < 0.01).

In the bivariate mixed logit model, “immediate death” is better than experiencing the worst possible EQ-5D-5L description for 1 year (i.e., $V(55555) = -0.180$); however, this value is closer to zero compared to the CL estimate. Apart from this difference in the lower bound, the twenty incremental effects are highly correlated and concordant between the CL and bivariate mixed logit estimates (Pearson correlation 0.970, Spearman correlation 0.916, Lin’s concordance 0.843). Furthermore, we computed the mean scale and power of the bivariate mixed logit as $\lambda = \exp(0.988) = 2.686$ and $\alpha = \exp(-\exp(-0.003)) = 0.369$. Therefore, the bivariate mixed logit model produce higher scale and lower power than the CL model.

Confirmatory results

The confirmatory CL results produce 6 insignificant incremental effects (i.e., SC_{23} , UA_{12} , UA_{23} , PD_{12} , AD_{12} , and AD_{45}), 3 of which are negative (i.e., PD_{12} , AD_{12} , and AD_{45}). Compared to the exploratory CL results, there are the same number of positive insignificant effects and 2 more negative effects. The confirmatory CL results suggest that the value of “immediate death” is better than experiencing the worst possible EQ-5D-5L description for 1 year (i.e., $V(55555) = -0.545$), which is slightly lower than the exploratory CL estimate (0.588). The estimated μ and r in the confirmatory CL are 0.062 and -0.634 , respectively. Therefore, the scale and power can be derived as $\lambda = \exp(0.062) = 1.064$ and $\alpha = \exp(-\exp(-0.634)) = 0.588$. Compared to the confirmatory CL results, the exploratory scale is larger (1.493), but its power is smaller (0.439).

When we allow for heterogeneity in scale and rate in the confirmatory analysis, there are substantive improvements in the estimated incremental effects. Specifically, in the bivariate mixed logit results, there is only 1 negative (i.e., PD_{12}) and 1 positive (i.e., AD_{45}) insignificant incremental effect. We computed the scale and power of the bivariate mixed logit model as $\lambda = \exp(0.704) = 2.022$ and $\alpha = \exp(-\exp(-2.051)) = 0.879$, respectively, which are higher than those of the CL model. The

estimated standard deviations for μ_i and r_i are 0.998 and 3.120, respectively, which suggests that both the scale and rate parameters are heterogeneous. Furthermore, we find a strong correlation between μ_i and r_i , 0.912 (p -value < 0.01).

In the bivariate mixed logit model, “immediate death” is better than experiencing the worst possible EQ-5D-5L description for 1 year (i.e., $V(55555) = -0.191$); however, this value is closer to zero compared to the CL estimate (-0.545). Apart from this difference in the lower bound, the twenty incremental effects are highly correlated and concordant between the CL and bivariate mixed logit estimates (Pearson correlation 0.897, Spearman correlation 0.888, Lin’s concordance 0.763).

Discussion

In this paper, we explored and confirmed heterogeneity in scale and rate, their correlation, and their effects on the estimation of EQ-5D-5L values. Allowing heterogeneity in scale and rate improved the EQ-5D-5L value set estimates in terms of face validity, namely, reducing the number of insignificant incremental effects.

A higher discount rate r_i implies that there is less variability in the net present value of life years. For instance, with a high discount rate, the value of 10 years decreases toward the value of 1 year. A higher scale implies that smaller differences in the value of health have a greater impact on log-odds. In other words, a larger scale parameter means more sensitivity. Since our results suggest that there is a high positive correlation between the scale parameter and the discount rate, we can infer that people who discount the future are more sensitive to smaller differences in the net present value of life years. This important finding may be confirmed in future health valuation studies.

In practical terms, allowing for scale heterogeneity implies that the analyst should also allow for rate heterogeneity (or vice versa) as well as estimate the correlation between scale and rate. However, no econometric package is currently available to facilitate this specification of the mixed logit, which may deter its uptake. In terms of the experimental design and blocking, future studies may assign “dying immediately,” episodes of one-year duration, and multi-year episodes to each respondent. This blocking can aid in the identification of scale and rate heterogeneity. If future studies block accordingly and such a package becomes available, reporting this correlation may become common practice in health valuation.

Although the twenty incremental effects are highly correlated and concordant between the CL and bivariate mixed logit estimates, controlling for scale and rate heterogeneity, reduced the size of the incremental effects, raising the lower bound of the EQ-5D-5L values from

-0.545 to -0.191. Although some effects decreased in size, the confirmatory bivariate estimation produced only two insignificant effects (PD_{12} and AD_{45}), which merits further discussion. The incremental effect PD_{12} represents the effect of the change from no to slight pain or discomfort. This effect is negative and insignificant in both the conditional logit models as well as the confirmatory bivariate mixed logit, which seems to suggest that U.S. adults are unwilling to sacrifice life years to relieve slight pain or discomfort. Further research is needed to verify this effect. The incremental effect AD_{45} represents the effect of the change from severely to extremely anxious or depressed. In two prior papers, Craig and colleagues [23, 24] showed that many U.S. adults prefer “extremely” over “severely” in this domain. This preference inversion contradicts the descriptive system and may be due to the diagnostic implications of severe mental health problems and/or the belief that moods may fluctuate between extrema under normal circumstances. The higher lower bound and two insignificant effects may accurately represent the EQ-5D-5L preferences of U.S. adults.

Although the incremental effects of the bivariate mixed logit appear to be better in terms of sign and significance, they are highly correlated with the CL estimates (Pearson correlation 0.897, Spearman correlation 0.888, Lin’s concordance 0.763). Figure 2 shows the 20 incremental effects from the confirmatory CL and bivariate mixed logit, where incremental effects are color-coded by dimension (i.e., MO: red, SC: green; UA: blue, PD: yellow, AD: black). The differences between the estimates seem to be larger among the more severe effects (from level 3 to 4 or from level 4 to 5).

Alternatively, some analysts may choose to use a hyperbolic discount function instead of a power function to allow for temporal discounting. Craig and colleagues [4] showed that decreasing the marginal value of life span under the assumption of power discounting provides better model fit than alternative functional forms. While Craig and colleagues [4] assumed a homogeneous discount rate (i.e., $r_i = r$), Jonker and colleagues [5] estimated the mixed logit model with a random hyperbolic discount rate and found strong evidence for non-linear time preferences. In this study, we extended both approaches and estimated a bivariate mixed logit model allowing a correlation between scale and power.

Future analyses may allow for heterogeneity in the incremental effects as well as scale and discount rate parameters, building from these findings. Before the estimation of such a complex model is attempted, we recommend that the authors conduct simulation analyses to verify that they can mitigate the simulation biases. For instance, Jumamyradov and colleagues conducted a simulation study and showed that the mixed logit model can produce biased results even when the model is correctly specified [25]. Nevertheless, we believe that our more parsimonious specification produced reliable results since we found a high correlation both in exploratory and confirmatory datasets.

There are three limitations in our analysis that we would like to mention. First, our mixed logit analysis is based on MSL estimation. This may be problematic for the bivariate specification because Jumamyradov and colleagues [25] showed that the MSL estimator of the mixed logit has difficulty estimating correlations and

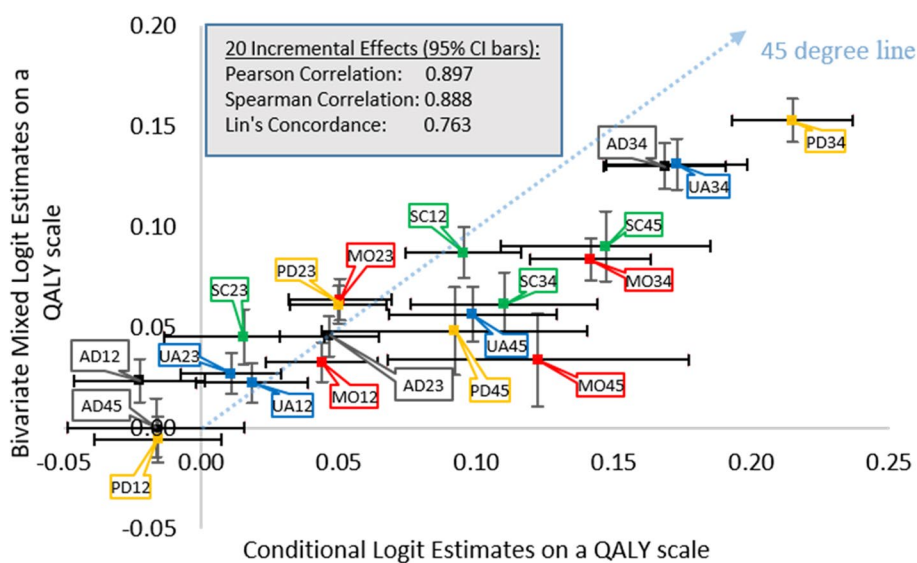


Fig. 2 Plot of 20 incremental effects for the conditional logit and bivariate mixed logit

may produce biased estimates even when correctly specified. Second, we assume only normally distributed random parameters and may consider other distributional assumptions in future research. Third, because of computational capacity constraints, we used only 250 Halton draws in our estimations, which is common place in the literature [22]. Some [26] have shown that increasing the number of Halton draws decreases the simulation bias for bivariate normal and bivariate Poisson-lognormal models.

Our study utilized EQ-5D-5L values from a United States-specific valuation study. While the data source was specific to the U.S., the underlying principles and results of this study are not confined to the U.S. context alone. The methodological approach and findings presented are designed to be broadly applicable and are likely to be generalizable to other settings. Similar methodologies can be applied to different populations and healthcare systems, reinforcing the validity of our approach across diverse settings.

We also would like to point that we acknowledge that the scale heterogeneity is a form of correlation among coefficients in mixed logit models [27, 28]. However, this is not relevant in our study since we are focusing on the correlation between the scale parameter and the discount rate.

Conclusion

Allowing heterogeneity in rate and scale added three parameters to the conditional logit model (two variances and a correlation) and greatly improved the face validity of the EQ-5D-5L values. We confirmed that persons who highly discount the future are more sensitive to differences in the net present value of QALYs. This intuitive pattern may be confirmed in future EQ-5D-5L valuation studies as well as influence experimental design and choice analysis in health preference research more generally.

Appendix

Table 1A. Results for univariate mixed logit models

N = 1017 & 1229	Random scale		Random rate	
	Exploratory	Confirmatory	Exploratory	Confirmatory
MO ₁₂	-0.012	0.019 ^b	0.003	0.025 ^b
MO ₂₃	0.049 ^b	0.076 ^b	0.045 ^b	0.064 ^b
MO ₃₄	0.087 ^b	0.093 ^b	0.103 ^b	0.082 ^b
MO ₄₅	0.104 ^b	0.054 ^b	0.049 ^b	0.028 ^a
SC ₁₂	0.016	0.057 ^b	0.022 ^b	0.097 ^b
SC ₂₃	0.023 ^a	0.027 ^b	0.020 ^b	0.025 ^b

N = 1017 & 1229	Random scale		Random rate	
	Exploratory	Confirmatory	Exploratory	Confirmatory
SC ₃₄	0.130 ^b	0.081 ^b	0.101 ^b	0.073 ^b
SC ₄₅	0.082 ^b	0.109 ^b	0.089 ^b	0.083 ^b
UA ₁₂	-0.001	-0.011	0.016 ^a	0.013 ^a
UA ₂₃	0.018 ^a	0.017 ^a	0.020 ^b	0.027 ^b
UA ₃₄	0.127 ^b	0.132 ^b	0.118 ^b	0.126 ^b
UA ₄₅	0.030 ^a	0.091 ^b	0.013	0.067 ^b
PD ₁₂	-0.007	-0.043 ^b	0.013	-0.020 ^b
PD ₂₃	0.035 ^b	0.046 ^b	0.029 ^b	0.062 ^b
PD ₃₄	0.155 ^b	0.175 ^b	0.140 ^b	0.142 ^b
PD ₄₅	0.124 ^b	0.065 ^b	0.058 ^b	0.034 ^b
AD ₁₂	0.009	-0.023 ^b	0.030 ^b	0.012
AD ₂₃	0.083 ^b	0.032 ^b	0.060 ^b	0.043 ^b
AD ₃₄	0.094 ^b	0.138 ^b	0.101 ^b	0.122 ^b
AD ₄₅	0.018 ^a	0.002	0.014	0.013
μ (mean)	0.194	0.228 ^b	1.275 ^b	0.899 ^b
r (mean)	-0.227 ^a	-0.483 ^b	0.422 ^b	-0.489 ^b
μ (SD)	1.232 ^b	1.136 ^b		
r (SD)			1.325 ^b	1.320 ^b
V(55555)	-0.162 ^b	-0.134 ^b	-0.043 ^b	-0.118 ^b
LL	-11964	-12989	-10666	-12279

Notice that the scale parameter is equal to μ and the power is equal to r to bound the power between 0 and 1

^a, ^b represent significance levels at 5% and 1%, respectively

Abbreviations

- MSL Maximum simulated likelihood
- ML Maximum likelihood
- QALY Quality-adjusted life year
- CL Conditional logit
- MO Mobility
- SC Self-care
- UA Usual activities
- PD Pain and discomfort
- AD Anxiety and depression
- EQ-VT EuroQol Valuation Technology
- CLL Complementary log-log
- MAR Maximum acceptable risk
- NPV Net present value
- EV1 Extreme value type 1
- IID Independently and identically distributed
- RUM Random utility maximization

Acknowledgements

The authors would like to thank the EuroQol Research Foundation for their support of Maksat Jumamyradov's dissertation, under awards 207-2020RA and 304-PHD. We also thank Drs. Murat Munkin and William Greene who advise Maksat on his dissertation.

Precis

We investigate how allowing heterogeneity in scale and rate affect the estimation of EQ-5D-5L values.

Authors' contributions

BMC and MJu were responsible for the concept and design of the manuscript. BMC and MJa were involved in data collection. BMC and MJu were involved

in analysis, interpretation of data, and drafting of the manuscript. All authors were involved in critical revision of paper for important intellectual content. MJu was responsible for the statistical analysis.

Funding

The EuroQol Research Foundation supported this project (No: 207-2020RA).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Jumamyradov reports grants from EuroQol Research Foundation, during the conduct of the study. Dr. Craig has nothing to disclose. Dr. Jakubczyk has nothing to disclose.

Author details

¹Department of Economics, University of South Florida, Tampa, FL, USA.

²Division of Decision Analysis and Support, SGH Warsaw School of Economics, Warsaw, Poland.

Received: 29 March 2024 Accepted: 30 June 2024

Published online: 13 July 2024

References

- Jonker MF. The garbage class mixed logit model: accounting for low-quality response patterns in discrete choice experiments. *Value Health*. 2022;25(11):1871–7.
- Groothuis-Oudshoorn CG, Flynn TN, Yoo HI, Magidson J, Oppe M. Key issues and potential solutions for understanding healthcare preference heterogeneity free from patient-level scale confounds. *Patient*. 2018;11(5):463–6.
- Craig BM, Rand K. Choice defines QALYs: a US valuation of the EQ-5D-5L. *Med Care*. 2018;56(6):529–36.
- Craig BM, Rand K, Bailey H, Stalmeier PF. Quality-adjusted life-years without constant proportionality. *Value Health*. 2018;21(9):1124–31.
- Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. *Value Health*. 2018;21(8):993–1001.
- Jonker MF, Bliemer MC. On the optimization of Bayesian D-efficient discrete choice experiment designs for the estimation of QALY tariffs that are corrected for nonlinear time preferences. *Value Health*. 2019;22(10):1162–9.
- Jonker MF, Norman R. 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 Econ*. 2022;31(2):431–9.
- Karim S, Craig BM, Tejada RA, Augustovski F. Preference heterogeneity in health valuation: a latent class analysis of the Peru EQ-5D-5L values. *Health Qual Life Outcomes*. 2023;21(1):1.
- Craig BM, de Bekker-Grob EW, González Sepúlveda JM, Greene WH. A guide to observable differences in stated preference evidence. *Patient*. 2022;15(3):329–39.
- McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. *Frontiers in Econometrics*. New York, USA: Academic Press; 1974. p. 105–42.
- Revelt D, Train K. Mixed logit with repeated choices: households' choice of appliance efficiency level. *Rev Econ Stat*. 1998;80(4):647–57.
- Brownstone D, Train K. Forecasting new product penetration with flexible substitution patterns. *J Econom*. 1998;89(1–2):109–29.
- McFadden D, Train K. Mixed MNL models of discrete response. *J Appl Economet*. 2000;15(5):447–70.
- David HA. *The method of paired comparisons*. London, UK: Griffin; 1963.
- Attema AE, Bleichrodt H, L'haridon O, Peretti-Watel P, Seror V. Discounting health and money: New evidence using a more robust method. *J Risk Uncertain*. 2018;56:117–40.
- Lipman SA, Attema AE, Versteegh MM. Correcting for discounting and loss aversion in composite time trade-off. *Health Econ*. 2022;31(8):1633–48.
- Jonker MF, Donkers B. Interaction effects in health state valuation studies: an optimal scaling approach. *Value in Health*. 2023;26(4):554–6.
- Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goñi JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics*. 2016;34(10):993–1004.
- Jakubczyk M, Craig BM, Barra M, et al. Choice defines value: a predictive modeling competition in health preference research. *Value Health*. 2018;21(2):229–38.
- Gourieroux C, Monfort A. Simulation based inference in models with heterogeneity. *Annales d'Economie et de Statistique*. 1991;20(21):69–107.
- Gourieroux C, Monfort A. *Simulation-based econometric methods*. New York, USA: Oxford University Press; 1996.
- Palma MA, Vedenov DV, Bessler D. The order of variables, simulation noise, and accuracy of mixed logit estimates. *Empir Econ*. 2020;58(5):2049–83.
- Craig BM, Pickard AS, Rand-Hendriksen K. Do health preferences contradict ordering of EQ-5D labels? *Qual Life Res*. 2015;24:1759–65.
- Craig BM, Monteiro AL, Herdman M, Santos M. Further evidence on EQ-5D-5L preference inversion: a Brazil/US collaboration. *Qual Life Res*. 2017;26:2489–96.
- Jumamyradov M, Craig BM, Munkin M, and Geene W. Comparing the mixed logit estimates and true parameters under informative and uninformative heterogeneity: a simulated discrete choice experiment. *Comput Econ*. Accepted 2024. <https://doi.org/10.1007/s10614-024-10637-x>.
- Jumamyradov M, Munkin MK. Biases in Maximum Simulated Likelihood Estimation of Bivariate Models. *J Econom Methods*. 2021;11(1):55–70.
- Hess S, Rose JM. Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation*. 2012;39:1225–39.
- Hess S, Train K. Correlation and scale in mixed logit models. *J Choice Model*. 2017;23:1–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.