Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Themed Section: Evolution of EuroQoL

Effect of Health State Sampling Methods on Model Predictions of EQ-5D-5L Values: Small Designs Can Suffice

Zhihao Yang, MSc [1,2,], Nan Luo, PhD [3], Gouke Bonsel, PhD [1,4], Jan Busschbach, PhD [1], Elly Stolk, PhD [4]*

[1]*Erasmus University Rotterdam, Rotterdam, The Netherlands;* [2]*Guizhou Medical University, Guiyang, China;* [3]*National University of Singapore, Singapore, Singapore;* [4]*The EuroQol Office, Rotterdam, The Netherlands*

ABSTRACT

Objective: The current five-level EQ-5D (EQ-5D-5L) valuation protocol requires the valuation of 86 states. It has been demonstrated that the selection of empirically valued health states affects the extrapolated values in three-level EQ-5D (EQ-3D-3L). In this investigation, we aim to compare the performance of the current EQ-5D-5L valuation design with other designs. **Study Design:** 1603 university students participated in a valuation study using a visual analog scale (VAS) to produce values for all EQ-5D-5L states. Different designs were generated to test their prediction accuracy. **Methods:** Subsamples of the dataset were used to mimic data obtained from a particular design; the remaining dataset was used as the validation set. In addition to EuroQol Group Valuation Technology (EQ-VT) design, alternative subsamples and designs were created using random, orthogonal, and "optimizing D-efficiency" sampling methods. The root mean squared error (RMSE) was used as the measure of prediction accuracy. **Results:** The EuroQol Group Valuation Technology (EQ-VT) design showed an average RMSE of 3.44 on EQ-VAS, for all 3125 health states combined. Notably, a 25-state orthogonal design performed similarly to the EQ-VT design, with a smaller RMSE of 3.40, and was thus the most efficient design. One caveat with respect to the orthogonal design was that it did not predict the mild states well. **Conclusions:** Our study supports the EQ-VT design. Smaller designs were identified with similar overall prediction accuracy. It is worth investigating whether issues with misprediction of mild states can be resolved, as the use of smaller size designs would reduce the cost of the valuation of EQ-5D-5L considerably.

Keywords: EQ-5D-5L, misprediction, orthogonal design, value set

Introduction

The EQ-5D is a health-related quality of life (HRQoL) questionnaire widely used in health economic, clinical, and population health studies. EQ-5D has two validated versions, which both compromise five dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. The three-level EQ-5D (EQ-5D-3L) describes each dimension at three levels (roughly corresponding to no, moderate, and extreme problems) and the five-level EQ-5D (EQ-5D-5L) expands its descriptions to five levels (roughly corresponding to no, slight, moderate, severe, and extreme problems) [1]. Compared to EQ-5D-3L, the descriptive richness of the EQ-5D-5L is an advantage when the goal is to understand the health state of a respondent, but potentially complicates the development of value sets. Through a valuation study, all health state "values" (some prefer "utilities" or "index values") can be derived from the corresponding value set. These values indicate how desirable the health states are. Performing such a valuation study for EQ-5D-5L is a challenge in terms of the trade-off between feasibility and validity. EQ-5D-5L defines 3125 states, which ideally should all be valued, but that is infeasible under standard conditions. Hence, in practice, only a subset of the health states is directly valued, and from this subset the values of all health states can be estimated through statistical modeling. Value sets for Short-Form Six-Dimension (SF-6D) and EQ-5D-3L have also been developed using this statistical modeling approach [2].

Selecting the subset of health states for direct valuation ("the empirical state set") is an important design matter for valuation studies and it is still evolving. For EQ-5D-3L, the Measurement and Valuation of Health (MVH) study protocol containing an empirical set of 42 EQ-5D-3L health states is most widely used [3]. Without applying explicit statistical considerations, the MVH study oversampled mild and commonly seen health states [4]. For EQ-5D-5L, the current valuation protocol was built on the results of several iterative pilot studies [2,5,6]. It was decided that the number of

[*] *Address correspondence to:* Zhihao Yang, Na 2019, Erasmus Medical Center, 80 Wytemaweg, 3015 CN Rotterdam, the Netherlands.
E-mail: z.yang@erasmusmc.nl

https://doi.org/10.1016/j.jval.2018.06.015

states in the design should be somewhere between 80 and 100, as the EQ-5D-5L main effects model has 21 parameters (five health dimensions × four dummy variables for severity levels + intercept). By ensuring that the total number of health states was four times larger than the number of parameters in the main effects model, multilevel modeling could be applied, that is, a random coefficient model to account for the effects of individual background variables [5]. Next, the number of health states to be valued by a single respondent was maximized at 10. To arrive at around 80 health states, a blocked design (10 blocks, each block with 10 states) was used, employing a balanced selection of states with respect to their utility values. Hence each block was planned to include the pits state, that is, the most severe health state (55555) and one of the five very mild states (21111, 12111, 11211, 11121, and 11112). This left eight unique states per block, in total 80 health states, to be defined; these were randomly selected out of the remaining 3118 health states. The selection of the 80 health states for the protocol was based on Monte Carlo simulations to predict the prior values obtained from the multinational pilot study, instead of choosing predominantly mild states [6]. The "optimal" set of 80 states was selected on the mean squared error (MSE) between the prior parameters and estimated parameters from a "main effects" model, and level balance, but without making orthogonality an explicit criterion [5].

Using an EQ-5D-3L visual analog scale (VAS) saturated dataset (a dataset in which the values of all 243 health states are known), two studies investigated the effect of health-state selection on prediction adequacy [4,7]. Both studies found that by improving the statistical efficiency of the design, the number of health states in the empirical state set in a valuation study could be reduced without loss of precision or validity [7,8]. In particular, the orthogonal design appeared ideal as it possessed two statistical properties: level balance and orthogonality (i.e., level pair balance) [8]. As the EQ-5D-5L empirical state set of 86 states (also known as the "EQ-VT set") was selected without constraints concerning orthogonality, the design choice of EQ-5D-5L may have suffered from misprediction effects, as found in some design choices of EQ-5D-3L [8].

Furthermore, while larger designs may be favored, given the advantages that they offer in the context of model exploration, we note that published EQ-5D-5L value sets have never used models with more than 22 parameters, leaving a surplus of 64 degrees of freedom [9]. This indicates that there could be redundancy in the current design, but we must proceed with caution when we aim to investigate this. In EQ-5D-3L, we have seen that a reduced design with 17 states from the original MVH design (42 states) introduced large prediction errors in the final value set [8]. Nevertheless, using a small design could reduce the cost of a valuation study and increase the feasibility of such a study for countries with limited resources. Hence, for any given degree of prediction accuracy, the smallest design with the least number of health states to be directly valued is sought, so that the cost of a valuation study can be minimized.

In this article, we revisit the EQ-VT design through two research questions:

1. Is there a more efficient (thus less costly) empirical set of health states than the current 86 EQ-VT set to derive an equally valid EQ-5D-5L value set?
2. Because 86 states in the EQ-VT design were divided into 10 blocks, and the pits state and 5 mild states were oversampled given that they were in all the blocks, what was the impact on prediction performance of oversampling these particular states in the current EQ-5D-5L design?

To address these questions, we collected values for all 3125 EQ-5D-5L health states in a dedicated direct EQ-VAS valuation study. This saturated VAS dataset enabled us to compare the prediction performance of the 86 health states subset with any alternative subset of health states. VAS was used in this research for its simplicity and VAS values served as proxies for time trade-off (TTO) values. Even though TTO (a trade-off exercise involving duration) and VAS (a direct scaling exercise) are two different tasks [10], they are both used to elicit cardinal preference data on the same object. Moreover, from experience, we know that a VAS data set can be close to its TTO counterpart [11–14]. Nevertheless, the results of this research should be seen in the light of the assumption that the selection artifacts are independent of the valuation methods employed.

Methods

Protocol to Collect the Saturated VAS Dataset

The current EQ-VT protocol requires each health state to have at least 100 observations so that the estimate of the (mean) value of each health state is sufficiently precise [5]. Adopting this sample size requirement, we obtained a saturated dataset by inviting 1600 university students as respondents, each of whom provided VAS values for approximately 197 health states: (1600 students × 197 health states/respondent)/3125 states = 100 observations/health state. We divided 3123 health states into 16 blocks using a stratified random selection process so that each block contained around 197 states (11111 and 55555 were presented in all blocks). For details of the data collection protocol, which aimed at an equivalent response burden across respondents (see the Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.06.015).

We organized 16 sessions of group interviews. Around 100 students were recruited to participate in each session, and each student received a randomly chosen block of health states. Each student received 100 RMB (equivalent of €15) as an incentive payment.

Tested Designs

After we obtained the empirical values for all 3125 health states, we tested how well the EQ-VT set with 86 health states and other candidate health state sets predicted the values for all 3125 EQ-5D-5L health states. In short, subsamples of the dataset were drawn to mimic the data obtained using a particular design, then a model was applied to estimate all 3125 health states, and finally these predictions were compared with the empirical values.

Using the EQ-VT 86 states set as a reference selection, we investigated the performance of orthogonal, random, or D-efficient designs of different sizes (number of health states in the subset). We started the size selection at 25 health states because this was the smallest size for orthogonal design in a five-factor five-level classification system (main effects modeling only). For each design, size selections of 25, 50, 75, 100, and 200 health states were created. For each design of a different size, 100 variants were produced.

Both the orthogonal design and D-efficient design are standard design choices in conjoint analysis and both designs aim to optimize statistical efficiency [15]. An orthogonal design defines an empirical state set, which satisfies the criterion that all severity levels and all severity level combinations (to a defined degree of level interaction: 2e or 3e, etc.) are equally prevalent and therefore balanced [16]. An orthogonal design is not always available as some combinations of dimension levels are not feasible (in the case of EQ-5D, the combination of "unable to walk" with "no problems in usual activities" appears to conflict). Alternatively, D-efficient design can be used. A D-efficient design aims at minimizing the geometric mean of the eigenvalues given $|(X'X) - 1|1 - p$ [15] from the empirical state set, taking into account level balance. Hence, a D-efficient design is efficient as the matrix of the vector of parameter estimates in a least squares analysis is proportional to $|(X'X) - 1|1 - p$, which is minimized [15]. In our study, orthogonal designs were provided by N-gene [17] and D-efficient designs generated through Stata 14.0 by selecting the 100 most D-efficient designs from 5000 random candidates. The Stata code can be found in the supplementary materials. For comparison, we created a series of random designs, imposing the restriction that the design should be severity balanced. For this purpose we first computed the "misery index," which is the sum score of the digits that represent the EQ-5D health states: 54321 = 5+4+3+2+1 = 15. We then classified all 3125 states into 5 misery index groups (≤10, 11–13, 14–16, 17–19, ≥20) and randomly selected health states from each group. Hence, across empirical sets, balance was present in terms of the number of health states in each of the five "misery strata." It should be noted that there are also other designs (e.g., Bayesian), which take both prior information and statistical efficiency into consideration.

Analysis

First, to obtain some insight into the data, we described the saturated dataset by plotting the relation between the mean VAS values of all health states to their misery index scores and showed the distribution of all observations along the VAS scale.

The performance of the different principles in selecting health states was quantified through computation of the root mean squared error (RMSE) as the primary measure of prediction performance (the higher, the worse the performance). For each design, an ordinary least squares (OLS) main effects model was used to fit the model for the empirical data of that particular design. In this article, we fitted the model using individual-level data (100 raw VAS observations per state) [7,18,19]. In the main effects model, the VAS value of a health state was explained by 20 dummy variables and one intercept. For each dimension (MO for mobility, SC for self-care, UA for usual activity, PD for pain or discomfort, AD for anxiety or depression), four dummy variables were used to represent the deviation from level one to the other four levels; for example, MO_3 takes one if the health state has a problem in the third level of mobility, and takes 0 if otherwise [19].

$$\text{VAS value} = \alpha + \beta_1 MO_2 + \beta_2 MO_3 + \beta_3 MO_4 + \beta_4 MO_5 + \beta_5 SC_2 + \beta_6 SC_3 + \beta_7 SC_4 + \beta_8 SC_5 + \beta_9 UA_2 + \beta_{10} UA_3 + \beta_{11} UA_4 + \beta_{12} UA_5 + \beta_{13} PD_2 + \beta_{14} PD_3 + \beta_{15} PD_4 + \beta_{16} PD_5 + \beta_{17} AD_2 + \beta_{18} AD_3 + \beta_{19} AD_4 + \beta_{20} AD_5 + \varepsilon. \quad (1)$$

In the modeling, 100 observations per state were used across all design choices. This meant that all data were used, except for 11111 and 55555 because these states were sampled in every block of the questionnaire. To avoid "overweighting" 11111 and 55555, the number of observations was limited to 100.

To answer our first research question, we summarized the RMSE of all designs (orthogonal, random, and D-efficient, all with different sizes and 100 variants), using a boxplot to combine the results of the simulations per specific design. The RMSE of the EQ-VT design was added in the boxplot as a reference. We defined as the most efficient design the one that systematically achieved the lowest RMSE relative to sample size. The most efficient design was reported in detail, with further descriptive tables including comparisons with the EQ-VT design.

To test our second research question, we fitted the model using weighted OLS regression: two times for the five mildest states and 10 times for the pits state 55555, as undertaken in the EQ-VT protocol. The comparison was made with the EQ-VT design with an equal 100 observations for all 86 states. Similarly, we examined how adding the five mildest states and the pits state in the most efficient design identified from the above analysis would impact on the misprediction. In the detailed comparison of the most efficient designs, we reported on the RMSE separately for the empirical state set only, on the validation state set only, and for all 3125 states combined. We also considered whether prediction error depended on health state severity. For this purpose, we categorized the values into 10 groups along the VAS scale: <30, ≥30 and <35, ≥35 and <40, ≥40 and <45, ≥45 and <50, ≥50 and <55, ≥55 and <60, ≥60 and <65, ≥65 and <70, ≥70. Finally, we estimated the number of health states with large prediction errors, defined by the absolute error (AE): AE > 5 and AE > 10. For reference, we listed the following for 10 random health states: observed mean VAS values, standard error, 95% confidence interval, and predicted VAS value.

Results

Description of the Saturated Dataset

In total, 1603 students participated in the study and finished the valuation task. This resulted in 100 observations for all states except 11111 and 55555, which each had 1600 observations.

The misery index for the EQ-5D-5L ranged from 5 (state 11111) to 25 (state 55555), but the number of different health states with the same misery index ranged from 1 to 381. In Figure 1, for each

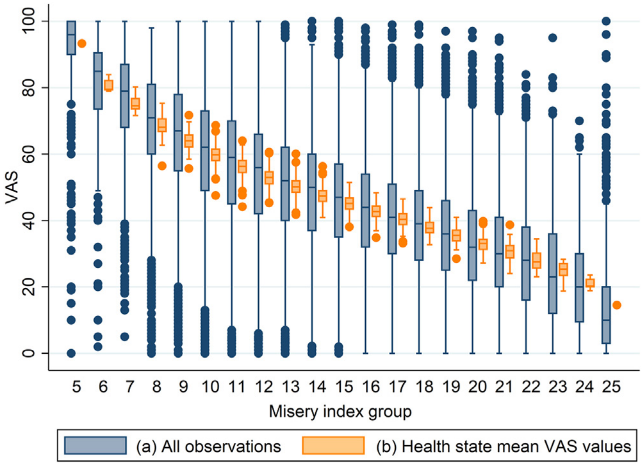

Fig. 1 – Empirical VAS values and mean VAS values of all health states by misery index. The box plot was sorted on the misery index group. It should be noted that one misery index value could result from more than one health state. The "All observations" is based on all VAS observations; the "Health state mean VAS values" is based on the mean VAS values of health states. VAS, visual analog scale.

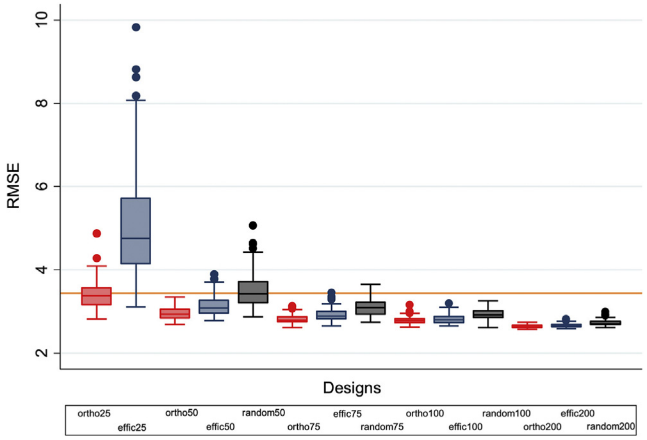

Fig. 2 – Boxplot showing the variations of different designs' RMSE of the predictions for 3125 health states. The reference line is EQ-VT design with RMSE = 3.44. The random design with 25 states was excluded because of a large RMSE. EQ-VT, EuroQol Group Valuation Technology; RMSE, root mean squared error; ortho, orthogonal design; effic, D-efficient design; random, random design.

misery index score (5–25) the following were plotted separately: (a: All observations) its relationship with all VAS value observations of a given misery index (blue boxplot), and (b: Health state mean VAS values) its relationship with the mean VAS value per health state with that misery index (orange boxplot). For comparison, we put (a) and (b) side by side in Figure 1. In (a) an outlier was interpreted as one observation; in (b) an outlier was interpreted as the average value of one health state. On average, the value of 11111 was around 90, and the value of 55555 was around 10, which together represented the range of the values. The health state values decreased along the misery index, as expected. Detailed descriptions concerning the quality of the saturated dataset can be found in the Appendix (see Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.06.015).

Comparison of Design Performance

We summarize design performance in Figure 2 using boxplots. The boxplots show the median RMSE for designs of each type and the variance observed across the 100 variants. The reference line represents the EQ-VT design. This design has no variance, as it is fixed by protocol. The EQ-VT design performed well with the RMSE = 3.44, but we can also note that all the other designs of the same size performed even better. When the sample size was limited below 50, the orthogonal design performed better than the D-efficient and random designs. In designs of a size = 75, the orthogonal and D-efficient designs performed similarly, whereas the random selection design was slightly worse. Random selection designs had many more variations. Noticeably, the small orthogonal design with 25 states on average performed about as well as the EQ-VT design. Other designs of size 25 performed poorly. The random designs with 25 states were not plotted in Figure 2 as their RMSE = 7.65 was beyond the range of the Y-axis. The D-efficient design with 25 health states performed the worst among all plotted designs.

When inspecting the outlier variants in the orthogonal design, we noticed that the outliers were mainly due to the inclusion of state 11111. Given the favorable outcomes for the small orthogonal design, in the following analysis we compared this in detail with the standard EQ-VT design.

In Table 1, we report the RMSEs for the empirical health state set, the validation health state set, and all health states taken together, for the small orthogonal design and the EQ-VT design, and the variants of both designs in adding, weighting, and removing five mild states and the pits state. Excluding the five mildest states and the pits state in EQ-VT, or restricting the design to an orthogonal design only, improved the overall RMSE. Furthermore, overrepresenting the five mildest states and the pits state following the current EQ-VT protocol increased the overall RMSE.

Figure 3 shows that the EQ-VT set predicted evenly along the scale when the five mildest states were included. In contrast, removal of the five mildest states from the EQ-VT set and/or restriction to only a small orthogonal design improved the fit for severe states but increased mispredictions for mild states.

Table 1 – RMSE by empirical/validation state set for EQ-VT design and 25 orthogonal design

	No. of states	Empirical state set	Validation state set	All 3125 states
EQ-VT protocol (weighted for pits & 5 mildest)	86	2.69	3.69	3.66
EQ-VT protocol	86	2.65	3.45	3.44
EQ-VT protocol (excluding pits & 5 mildest)	80	2.39	3.02	3.00
25 orthogonals	25	1.03	3.41	3.40
25 orthogonals (extending pits & 5 mildest)	31	2.61	3.88	3.87

EQ-VT, EuroQol Group Valuation Technology; RMSE, root mean squared error.
*The italic design was repeated more than 100 times.

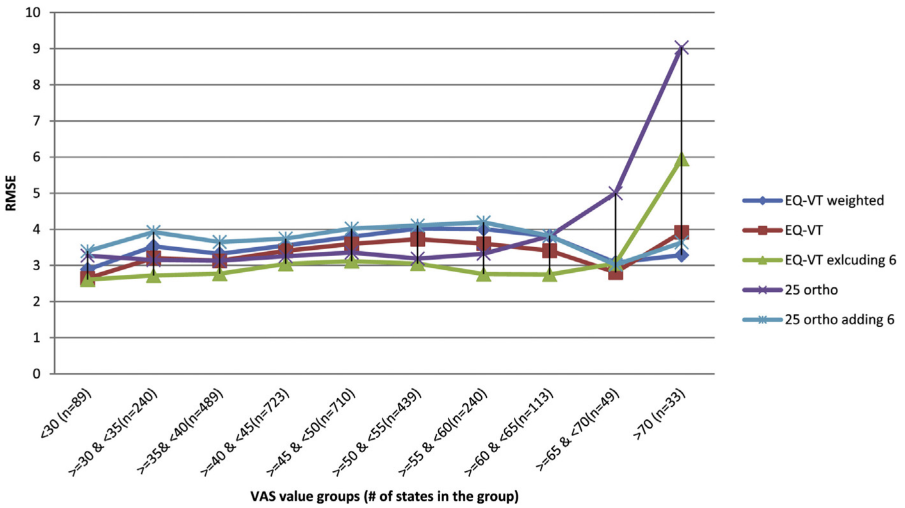

Fig. 3 – RMSE over VAS values for EQ-VT design and 25 orthogonal design. ortho, orthogonal design; EQ-VT, EuroQol Group Valuation Technology design; RMSE, root mean squared error; VAS, visual analog scale.

Figure 4 shows that large mispredictions occurred least frequently in the orthogonal designs, regardless of size.

Table 2 lists the observed VAS values and predicted VAS values of a random set of 10 health states with different severity levels.

Discussion

We obtained a saturated dataset that allowed for head-to-head comparison of different principles in the selection of health states in valuation studies. We found that the EQ-VT design performed well in terms of misprediction effects measured by the overall RMSE. In addition, we observed that designs with fewer states can perform as well as the EQ-VT design if they are constructed with attention to their statistical properties. The orthogonal design with 25 states performed closely to the standard EQ-VT with 86 states in terms of overall RMSE. Importantly, values generated on the basis of a small orthogonal design with 25 states contained fewer large mispredictions (defined by AE > 5 & AE > 10) than the values generated on the basis of the EQ-VT. Both designs provided sufficient prediction accuracy, which was below the oft-used minimum important difference (MID) [20,21]. To answer our first research question, the small orthogonal design with 25 states was the most efficient design we identified.

A caveat to the use of the small orthogonal design lies in the large mispredictions in the mild states (VAS value >70) compared to EQ-VT. There are several possible explanations here. First, this could be a consequence of underrepresenting the mild states in orthogonal designs compared to the EQ-VT design (note that in a small orthogonal design with 25 states, only 1 or no health state is mild). Thus, to address our second research question, by giving the five mildest states more weight in the blocked EQ-VT design or by extending a small orthogonal design with the five mildest states, the predictions for mild states improved, at the price of increased mispredictions for the moderate/severe states. Second, we did not take account of the consideration that the mean values for mild states could be seen as censored at one [22], and

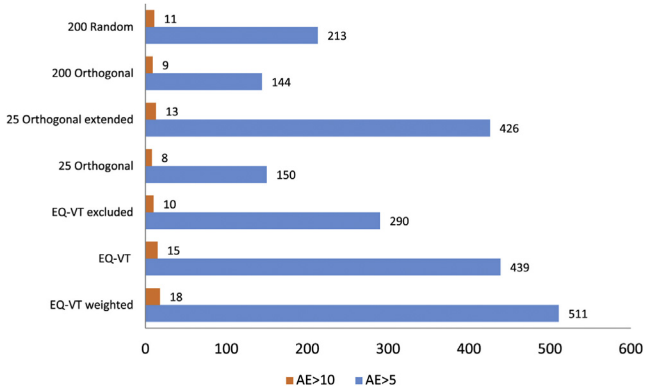

Fig. 4 – Count of large misprediction errors (AE > 5, AE > 10). *Random design and orthogonal design with 200 states were added for reference. AE, absolute error.

Table 2 – Observed values and predicted values for 10 random health states

Health state	Observed value	SE	95% CI	Predicted by orthogonal* (mean, SE)	Predicted by EQ-VT
21112	73.7	1.5	70.7–76.6	67.2, 3.1	71.7
13112	68.0	1.6	64.9–71.1	65.3, 2.9	70.5
23113	61.3	1.5	58.2–64.3	58.8, 2.6	60.4
25511	51.9	1.8	48.3–55.4	50.2, 2.5	53.2
14334	43.0	1.8	39.6–46.5	45.0, 2.1	45.0
44513	41.0	1.5	38.1–43.9	40.9, 2.3	42.3
13455	35.5	1.7	32.2–38.8	37.6, 2.0	37.1
24445	35.0	1.6	31.9–38.0	33.0, 2.0	33.3
45354	30.5	1.5	27.6–33.3	28.9, 2.1	26.1
55555	14.5	0.4	13.8–15.2	20.2, 2.3	18.2

CI, confidence interval; SE, standard error; EQ-VT, EuroQol Group Valuation Technology.
* For the large orthogonal designs with 100 variants, the averaged predicted means and the SEs were estimated.

that the main effects model did not capture all effects on valuations [23]. Moreover, the models that we used could introduce further bias, as they do not consider the possible heteroskedastic nature of the data [22,24]; that is, severe states have more variance than the mild states. It is possible that these issues also affect VAS data differently from they affect TTO data, because VAS data are characterized by relatively low values for mild states, translating into a large intercept. Hence, while awaiting better understanding and modeling of the upper part of the scale in general, consideration could be given to the use of small orthogonal designs extended with the five mildest states if the resulting values are predominantly used for the “better” half of the health states.

Another important finding was that the performance of the orthogonal designs depended on inclusion of state 11111. Because of the nonadditivity of domains in the upper part of the scale, a gap usually exists between 11111 and all other states. In this saturated dataset, the value of 11111 was 90.48, and the next highest value was 83.93 for 11121. Thus, the value of the state 11111 could not be derived from the value impacts of level 1 of the 5 dimensions in non-11111 states, and conversely the impact of level 1 in general (in non-11111 states) would be mispredicted if it primarily relied on the empirical value of 11111. As the output of design generators like N-gene could permit any translation of the basic permutation scheme, researchers could opt for a variant without state 11111. In addition, this upper gap issue (11111 effects) of a VAS exercise may have disappeared in TTO data as 11111 is the reference state (no need to value and have a theoretical value of 1) and the gap effect is then translated into the model intercept.

Better performance for statistically efficient designs was similar to the results found in previous EQ-5D-3L studies [3,7,8]. Although we conclude that, in using a main effects model, an orthogonal design is stable and efficient, the D-efficient design is a good alternative when an unrestricted orthogonal design is deemed inapplicable. Theoretically, the more the D-efficient design achieves level balance and orthogonality, the more efficient it is [15]. Hence, compared to the orthogonal design, which already optimizes statistical efficiency, D-efficient design may need more states (to compensate for the loss in efficiency) to achieve the same prediction accuracy. Similar to EQ-5D, the valuation of other HRQoL instruments such as SF-6D and HUI may also benefit from using statistically efficient designs. Further research is required.

Some general limitations apply. First, we used a saturated VAS dataset to mimic the design choices in EQ-5D-5L valuation studies that use TTO as their elicitation method. Raw VAS values do not have ratio properties. If we assume a (monotonic) linear relation between VAS and TTO [12,14], then we would also expect our conclusions to be valid for TTO. Nevertheless, it should be noted that TTO data display more heteroscedasticity between states and more heterogeneity between respondents, and thus we may expect to use more states or observations in a TTO valuation study. Second, there may be a blocking effect as we divided all 3125 states into 16 blocks when collecting the saturated dataset. While this essentially suggests a two-level analysis, we assumed there was no such effect. Third, we used university students as respondents, who have limited experience in health problems and whose preferences may be more homogeneous. This may have led to smaller RMSE compared to studies using the general public as respondents, but this is a minor issue as the purpose of this study was to test hypotheses rather than to generate value sets.

Our results inspire faith in the design of the EQ-VT for current EQ-5D-5L valuation studies [2,5]. We noted that small orthogonal designs with 25 states performed almost as well as other designs but produced biased estimates for mild states. Further research with respect to this phenomenon, and strategies to avoid it, are warranted because of the potential benefits that can be reaped from adopting small designs. That is, employing a small orthogonal design with 25 states (or 31, if extended to add the 5 mildest states and the pits state) could reduce sample size requirements by more than 50%. Future research should also investigate the validity of orthogonal designs using TTO data.

Supplementary Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2018.06.015.

REFERENCES

[1] Luo N, Liu G, Li M, et al. Estimating an EQ-5D-5L value set for China. Value Health 2017;20(4):662–9.

[2] Oppe M, Rand-Hendriksen K, Shah K, et al. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics 2016;34(10):993–1004.

[3] Lamers LM, McDonnell J, Stalmeier PF, et al. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006;15(10):1121–32.

[4] Yang Z, Luo N, Bonsel G, et al. Selecting health states for EQ-5D-3L valuation studies: statistical considerations matter. Value Health 2018;21(4):456–61.

[5] Oppe M, Hout B. The 'power' of eliciting EQ-5D-5L values. EuroQol Working Paper Series, 2017.

[6] Oppe M, Devlin NJ, van Hout B, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health 2014;17(4):445–53.

[7] Bonsel G, Oppe M, janssen B. Unexpected large misspecification effects of health profiles selection and interaction analysis to obtain a value function from unsaturated valuation datasets, using the standard EuroQol approach. EuroQol Plenary Meeting 2014 Discussion Papers, 2014.
[8] Yang Z, Luo N, Busschbach J, Stolk E. Using orthogonal design in selecting health states for the construction of EQ-5D-3L value set. Value Health 2016;19(7):A386.
[9] Kim SH, Ahn J, Ock M, et al. The EQ-5D-5L valuation study in Korea. Qual Life Res 2016;25(7):1845–52.
[10] Robinson A, Dolan P, Williams A. Valuing health status using VAS and TTO: what lies behind the numbers? Soc Sci Med 1997;45(8):1289–97.
[11] Sun S, Chen J, Kind P, et al. Experience-based VAS values for EQ-5D-3L health states in a national general population health survey in China. Qual Life Res 2015;24(3):693–703.
[12] Bernert S, Fernandez A, Haro JM, et al. Comparison of different valuation methods for population health status measured by the EQ-5D in three European countries. Value Health 2009;12(5):750–8.
[13] Badia X, Herdman M, Roset M, Ohinmaa A. Feasibility and validity of the VAS and TTO for eliciting general population values for temporary health states: a comparative study. Health Serv Outcomes Res Methodol 2001;2:51–65.
[14] Craig BM, Busschbach JJ, Salomon JA. Modeling ranking, time trade-off, and visual analog scale values for EQ-5D health states: a review and comparison of methods. Med Care 2009;47(6):634–41.
[15] Kuhfeld WF, Tobias RD, Garratt M. Efficient experimental-design with marketing-research applications. J Marketing Res 1994;31(4):545–57.
[16] Hedayat A, Sloane NJA, Stufken J. Orthogonal Arrays: Theory and Applications. New York: Springe-Verlag; 1999.
[17] CMP Ltd. Ngene 1.1.2 User Manual & Reference Guide 2014. Available from: https://dl.dropboxusercontent.com/u/9406880/NgeneManual112.pdf.
[18] Busschbach J, McDonnell J, Hout B. Testing different parametric relations between the EuroQol health description and health valuations in students. EuroQol Plenary Meeting 1996 Discussion Papers, 1996.
[19] Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35(11):1095–108.
[20] Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes 2007;5:70.
[21] Coteur G, Feagan B, Keininger DL, Kosinski M. Evaluation of the meaningfulness of health-related quality of life improvements as assessed by the SF-36 and the EQ-5D VAS in patients with active Crohn's disease. Aliment Pharmacol Ther 2009;29(9):1032–41.
[22] Feng Y, Devlin NJ, Shah KK, et al. New methods for modelling EQ-5D-5L value sets: an application to English data. Health Econ 2018;27(1):23–38.
[23] Versteegh M, Vermeulen M, Evers MAA, et al. Dutch tariff for the five-level version of EQ-5D. Value Health 2016;19(4):343–52.
[24] Devlin NJ, Shah KK, Feng Y, et al. Valuing health-related quality of life: an EQ-5D-5L value set for England. Health Econ 2018;27(1):7–22.