This is a repository copy of *Using the online elicitation of personal utility functions approach to derive a patient-based 5-level version of EQ-5D value set: a study in 122 patients with rheumatic diseases from Germany*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/232049/

Version: Published Version

Article:

Schneider, P., Blankart, K., Brazier, J. et al. (2 more authors) (2024) Using the online elicitation of personal utility functions approach to derive a patient-based 5-level version of EQ-5D value set: a study in 122 patients with rheumatic diseases from Germany. Value in Health, 27 (3). pp. 376-382. ISSN: 1098-3015

https://doi.org/10.1016/j.jval.2023.12.009

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

ScienceDirect

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

Preference-Based Assessments

Using the Online Elicitation of Personal Utility Functions Approach to Derive a Patient-Based 5-Level Version of EQ-5D Value Set: A Study in 122 Patients With Rheumatic Diseases From Germany

Paul Schneider, PhD, Katharina Blankart, PhD, John Brazier, PhD, Ben van Hout, PhD, Nancy Devlin, PhD

ABSTRACT

Objectives: Traditional preference elicitation methods, such as discrete choice experiments or time trade-off, usually require large sample sizes. This can limit their applicability in patient populations, where recruiting enough participants can be challenging.
The objective of this study was to test a new method, called the Online elicitation of Personal Utility Functions (OPUF) approach, to derive an EQ-5D-5L value set from a relatively small sample of patients with rheumatic diseases.

Methods: OPUF is a new type of online survey that implements compositional preference elicitation techniques. Central to the method are 3 valuation steps: (1) dimension weighting, (2) level rating, and (3) anchoring. An English demo version of the OPUF survey can be accessed at https://valorem.health/eq5d5l. From the responses, a personal EQ-5D-5L utility function can be constructed for each participant, and a group-level value set can be derived by aggregating model coefficients across participants.

Results: A total of 122 patients with rheumatic disease from Germany completed the OPUF survey. The survey was generally well received; most participants completed the survey in less than 20 minutes and were able to derive a full EQ-5D-5L value set. The precision of mean coefficients was high, despite the small sample size.

Conclusions: Our findings demonstrate that OPUF can be used to derive an EQ-5D-5L value set from a relatively small sample of patients. Although the method is still under development, we think that it has the potential to be a valuable preference elicitation tool and to complement traditional methods in several areas.

Keywords: EQ-5D-5L, MCDA, preference elicitation, rheumatic disease, utility values, value set.

VALUE HEALTH. 2024; 27(3):376–382

Introduction

In some countries, including Germany and Sweden, health technology assessment agencies recommend that health economic evaluations should be informed by preferences of patients.[1,2] This means that health-related quality of life (HRQoL) weights (utility values or value sets) for computing quality-adjusted life-years (QALYs) are to be derived either from patients with a particular disease (patient-based value set) or from individuals who assess their own (impaired) health (experience-based value set). In most countries, however, health technology assessment agencies use utility values derived from the preferences of the general public. But even then, decision makers may—and probably should—be keen to also consider the patients' perspective when making decisions about the allocation of limited healthcare resources.[3-9]

Eliciting preferences from patients to derive utility values, however, can be difficult.[6,10] Traditional elicitation methods, such as discrete choice experiments (DCEs), time trade-off (TTO), or standard gamble, require large sample sizes[11,12]: several hundred participants are commonly needed to estimate a reliable preference model. Recruiting such a large number of patients for a study can be difficult, and in many cases (eg, rare diseases), it will not be feasible at all. This limits the availability of quantitative evidence on patient preferences and thus the use of patient preferences to inform health policy decision making.

Recently, a new method, called Online elicitation of Personal Utility Functions (OPUF), was developed.[13,14] It implements multiple compositional preference elicitation techniques into an adaptive and easy-to-use online tool. OPUF allows constructing preferences for small groups and even on the individual person level.

Here, we report the results of a valuation study that was conducted to test the feasibility of using OPUF to elicit 5-level version of the EQ-5D-5L health state preferences from patients with rheumatic diseases in Germany. We demonstrate how the

1098-3015/Copyright © 2024, International Society for Pharmacoeconomics and Outcomes Research, Inc. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

new method allows constructing a value set based on a sample of just 122 participants.

Methods

Respondents

The valuation study was conducted in Germany between May and July 2022. Participants were recruited through the German Rheumatism Association (Deutsche Rheuma-Liga e.V., DRL). The invitation to participate in our study was distributed to their members through a newsletter and social media. Participants were offered a financial incentive of €5. The survey was open to anyone who identified as a patient with a rheumatic disease. We did not specify any exclusion criteria. This patient group was selected after consultation with the Institute for Quality and Efficiency in Health Care in Germany, and because a DCE-based EQ-5D-5L value set for a similar patient group was available.[1]

EQ-5D-5L

The EQ-5D-5L is the most widely used generic measure of HRQoL.[15,16] It consists of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each of which has 5 levels (no problems, slight problems, moderate problems, severe problems, and extreme problems). EQ-5D-5L health states are commonly denoted by a 5-digit code, representing the respective dimension-levels. The system can describe a total of 3125 health states, ranging from “11111” (full health, no health problems) to “55555” (the worst state with extreme problems on all 5 dimensions). Although population-based EQ-5D-5L value sets currently exist for 25 countries (more studies are presently ongoing),[17] patient-based value sets are scarce.[1,18]

OPUF—General Method

OPUF is a web-based version of the Personal Utility Function (PUF) approach, originally proposed by Devlin et al.,[14] which, in turn, is based on methods developed in the context of multi-criteria decision analysis and multiattribute value theory. Although PUF was designed as an interviewer-led, face-to-face approach, with deliberative and reflective components, OPUF combines multiple preference elicitation methods into a relatively short, easy-to-navigate online survey.[19-21]

The main difference between OPUF and traditional health valuation methods is its underlying paradigm: OPUF is a compositional, instead of a decompositional, preference elicitation technique. Although DCE or TTO require participants to evaluate entire health states, from which partial values for dimension-level coefficients are subsequently inferred (=decompositional approach), in the OPUF method, partial values are elicited directly from the participants. The process broadly consists of 3 steps:

First, criteria weighting, which determines the relative importance of the different dimensions of the HRQoL measure, on a scale from 0 to 100.

Second, level rating, which determines, within each dimension, the relative importance of any intermediate levels (eg, slight, moderate, and severe problems walking about), on a scale that is anchored at the worst (unable to walk about = 0) and the best level (no problems walking about = 100).[22]

Finally, anchoring, which is an additional step that allows mapping the values obtained in the previous steps on to the QALY scale, which is anchored at full health (=1) and dead (=0).

From the responses, an additive preference model, containing 1 coefficient for each dimension level, can then be derived for each participant by multiplying the level ratings with the respective dimension weights and rescaling the resulting values to the QALY scale using the anchoring factor.

This model can then be used to derive values for any health state by multiplying the dimension-level coefficients with the respective dimension-levels of the health state and summing up the values and subtracting the sum from 1.

A more detailed description of the OPUF method,[13] a formal description of method used to construct personal preference models from individual level responses, and a simple example can be found in the Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.12.009.

OPUF - Implementation

For this study, we adapted a previous (English) version of the EQ-5D-5L OPUF survey tool and translated it into German.[13] The survey was built using modern JavaScript frameworks (Vue.js for the front-end; Node.js for the back-end). The different valuation tasks were tested in online interviews with a small group of lay persons and experts and piloted in a small study with 25 participants recruited though prolific.[23]

A demo version of an OPUF survey (translated into English) is available at: https://valorem.health/eq5d5l.

In the following, we provide a brief overview of the different valuation steps and how they were implemented in the study.

Warm-up

At the beginning of the survey, each participant was asked to report their own health state using the EQ-5D-5L and rate their subjective level of health using the EQ-VAS. This was done not only to assess the health of the participants but also to familiarize participants with the instrument.

Criteria weighting

First, participants were shown a list of the worst problems of each dimension. Depending on their choice, the respective dimension was set to 100 in the subsequent task, in which participants had to complete a swing rating exercise. The task was to assign values between 0 and 100 to swings from the worst to the best level on each dimension. The 100 points, assigned to the most important dimension, were fixed, and used as a yard stick to help participants to determine the relative importance of the other 4 dimensions. The order of the dimensions was randomized.

Level rating

For each dimension, the participants were asked to place the 3 intermediate levels (slight, moderate, and severe problems) on a scale from 0 (no problems) to 100 (extreme problems). For this, participants had to “drag and drop” labels with the respective level description onto the scale. This method avoids any anchoring effects. The order of the 5 dimensions was randomized.

Anchoring

The anchoring task consists of 2 steps. The task begins with a pairwise comparison between the objectively worst EQ-5D-5L health state “55555” (option A) and “being dead” (option B). Depending on their stated choice, participants got to see different tasks.

Option A. If a participant preferred “55555” over “being dead,” they were asked to locate the position of “55555” on a visual analog scale between “No health problems” (=100) and “being dead” (=0). The selected value was then used as the anchor point for the PUF.

Option B. If participants preferred "being dead" over "55555," a binary search algorithm was initiated, in which the state that was shown as option A adaptively changed, to find the health state that they considered to be equivalent to "being dead."[14,24] For this, all 3125 EQ-5D-5L health states were ranked from the best to the worse, based on the participant's responses to the previous tasks. After the first comparison ("55555" vs "being dead"), the algorithm selects the median state (which may be different for each participant). Depending on the participant's subsequent choices, the algorithm then jumps up or down in half interval steps (bisection method). After 6 iterations, the rank of the equal-to-dead state is identified with a maximum error of ±49 ranks, and the search ends. The normalized utility value of the equal-to-dead state is then used to rescale and anchor the PUF.

Demographic questions and feedback

At the end of the survey, we asked for basic demographic information and rheumatic diseases diagnoses. Participants were also invited to share feedback on the survey and to make suggestions for improvement.

Data Analysis and Modeling

Participants' responses were analyzed on multiple levels.

First, we assessed the raw response data of the 3 valuation steps separately.

Second, we constructed personal EQ-5D-5L utility functions for each participant. The utility functions were specified as additive models with 20 coefficients—4 for each of the 5 dimensions, representing the utility decrement associated with levels 2 to 5. The models were constructed using the procedure described above. For a detailed description of the procedure used to construct the personal preference models from individual level responses, and a simple example, see Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.12.009.

Finally, we aggregated the PUFs model coefficients to derive a preference function for the group as a whole. This was done by averaging the coefficients across all participants. The aggregate preference function was then used to generate an EQ-5D-5L value set, that is, QALY-weights for all 3125 EQ-5D-5L health states.

All analyses were conducted in R 4.2.1 (R Foundation, Vienna, Austria).

Engagement

One indicator of engagement that we assessed was the time participants spent on completing the survey. Furthermore, the levels of the EQ-5D-5L instrument have a predefined order: slight problems, for example, (weakly) dominates moderate problems. We can, therefore, utilize the level rating task to check participants' understanding and their engagement with the task by assessing the frequency of implausible level ratings, which means ratings that violate the correct order of the levels.

DCE Hold-Out Tasks

Participants completed 3 forced choice DCE holdout tasks. The tasks were generated adaptively, based on participants' PUF. For each participant, health states were generated to create choice sets in which the utility difference between the 2 states was around 0.05 (hard), 0.1 (moderate), and 0.25 (easy) on the personal utility scale. The generation of the health states was a random process and did not account for whether a particular health state was plausible or realistic. The order of the DCEs was randomized. Trivial choices, involving dominated or dominating alternatives, were excluded. We predicted participants' choices in the DCE hold-out tasks, based on PUFs, and then compared those with the observed choices.

Feedback

At the end of the survey, participants were invited to share feedback on the survey and make suggestions for improvement. A formal qualitative analysis of the responses is beyond the scope of this paper. However, we performed a crude thematic analysis to assess how the survey was received and whether any potential issues were raised, which would indicate problems with the response data.

Results

Sample Demographic and Health Characteristics

A total of 122 participants completed the survey between May and July 2022. Most participants were female (n = 111, 91%). Further demographic information is provided in Table 1.

Participants reported various rheumatic disease diagnoses. The most common conditions were rheumatoid arthritis (n = 72, 59%), osteoarthritis (n = 31, 25%), psoriatic arthritis 21 (17%), fibromyalgia (n = 21, 17%), and chronic pain syndrome (n = 19, 16%). Sixty-four (52%) participants reported more than 1 condition. Further details can be found in the Appendix (see Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.12.009).

Corresponding to the prevalent health conditions, many participants reported poor EQ-5D-5L health states. Only 1 (1%) participant was in full health, and 22 (18%) reported having only mild health problems, whereas 41 (34%) reported having severe or extreme problems on at least 1 EQ-5D dimension. The most frequently affected dimension was pain/discomfort (n = 118, 97%) with a mean (SD) level severity score of 2.9 (0.78), followed by usual activities (n = 104, 85%), with a score of 2.5 (0.79), anxiety/depression (n = 90, 74%) with a score of 2.2 (0.95), mobility (n = 87, 71%) with a score of 2.2 (0.97), and lastly self-care (n = 52, 43%), with a score of 1.6 (SD = 0.79). The mean (SD) and median (interquartile range [IQR]) EQ VAS was 61.1 (18.14) and 61.5 (48.0-75.0), respectively.

OPUF Survey Results

Criteria weighting

The most important EQ-5D dimension was pain/discomfort (mean = 90.1, SD = 16.4), followed by usual activities (mean = 87.6, SD = 17.1), self-care (mean = 85.8, SD = 20.2), and mobility (mean = 84.9, SD = 15.9). The least important dimension was anxiety/depression (mean = 72.9, SD = 27.7).

Level rating

Level ratings were similar across all 5 dimensions. The mean ratings for the intermediate levels, slight, moderate, and severe problems, were around 75, 50, and 22, respectively. The ratings for the best and the worst level were fixed at 100 and 0. Full results of the level ratings are provided in Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.12.009.

Anchoring

The majority of participants (n = 89, 73%) indicated that they would prefer "being dead" over the worst EQ-5D-5L health state ("55555"). The mean (SD) and median (IQR) utility values for the worst health state were −0.32 (0.52) and −0.26 (0.1; −0.65), respectively. A total of 11 (9%) participants had utility scores below −1, with 1 participant having a utility score as low as −1.9 for state "55555."

Table 1. Sample characteristics.

Group	N (%)
Age	
18-29	18 (14.8%)
30-39	23 (18.9%)
40-49	20 (16.4%)
50-59	36 (29.5%)
60-69	18 (14.8%)
70+	7 (5.7%)
Sex	
Female	111 (91%)
Male	9 (7.4%)
Other	2 (1.6%)
Highest secondary education degree	
University entrance qualification	67 (54.9%)
Entrance qualification for universities of applied sciences	25 (20.5%)
Intermediate secondary education	22 (18%)
Basic secondary education	4 (3.3%)
None/other	4 (3.2%)
Total number of participants	122 (100%)

Personal and group-level utility functions

For each participant, we successfully constructed a personal EQ-5D-5L utility function, using their individual responses from the criteria weighting, the level rating, and the anchoring task.

Model coefficients were aggregated across participants, by means of averaging, to obtain a utility function that reflects the preferences of the group of patients with rheumatic diseases as a whole. Table 2 below shows the resulting mean coefficient estimates and bootstrapped 95% confidence intervals.

The reported coefficients can be used to construct utility values for any of the 3125 EQ-5D-5L health states. For example, the utility value for the health state "12345" is $1 - (0 + 0.066 + 0.141 + 0.222 + 0.222) = 0.35$. The utility values of states "22222," "33333," "44444," and "55555" are 0.69, 0.35, −0.05, and −0.32, respectively, to give just a few more examples.

A simplified visualization of the group value set—compared with all 122 individual patient value sets—is provided in Figure 1. The graphs illustrate that, although the study was conducted in a population of patients with similar health problems, the personal value sets showed a considerable degree of heterogeneity.

Engagement

On average, participants spent about 17 minutes (SD = 16) on the OPUF survey. The median completion time was 13.6 minutes (IQR 9.9-18.7). The fastest participant completed the survey in 4.5 minutes, which is still within reasonable time limits.

Using the 15 individual level rating tasks as a means to assess participants' understanding and attention, we found that participants were generally able to rate the intermediate levels in a consistent way across the 5 EQ-5D dimensions. Only 13 participants (11%) made 2 or more errors, whereas 10 participants (8%) made 1 error, and 99 participants (81%) made no errors. The results suggest a high level of engagement and understanding of the level rating task.

DCE Hold-Out Tasks

Overall, the constructed PUFs predicted participants' observed choices in the 3 hold-out DCE tasks with an accuracy of 67.8%. The predictive accuracy varied by difficulty: the accuracy was 60.7% for the hard DCE task (with a utility difference of 0.05 between the 2 states in the choice set), 61.5% for the medium task (0.1 difference), and 81.1% for the easy task (0.25 difference). Thirty-nine participants (32%) made no "error," 52 (43%) made 1, 27 (22%) made 2, and 4 (3%) made 3 "errors."

Feedback

Of the 122 participants, 49 (40%) sent feedback or suggestions to improve the survey. The median word count was 19 (IQR: 10-30). The crude thematic analysis identified 5 main themes, in which most of the responses could be categorized (responses could be categorized in multiple themes)—see list below. No major issues were identified, but some suggestions for improvement were made.

1. Introspection/reflection: 20 participants (41%) found the survey interesting or thought-provoking. Some also reported that the survey helped them to better understand their own priorities.
2. Difficulties: 14 participants (29%) made suggestions for improvement or reported difficulties with certain aspects of the survey, including the navigation on specific tasks, instructions for the level rating task, and the handling of the sliders.
3. Overall assessment: 8 participants (16%) submitted an overall evaluation of the survey, which ranged from "very good" ("sehr gut") to "so-so" ("geht so").
4. Unrealistic states: 5 participants (10%) commented on the DCE task and noted that some states were unrealistic or implausible.
5. Other: 8 participants (16%) provided comments that were not directly related to the survey.

Discussion

Main Findings

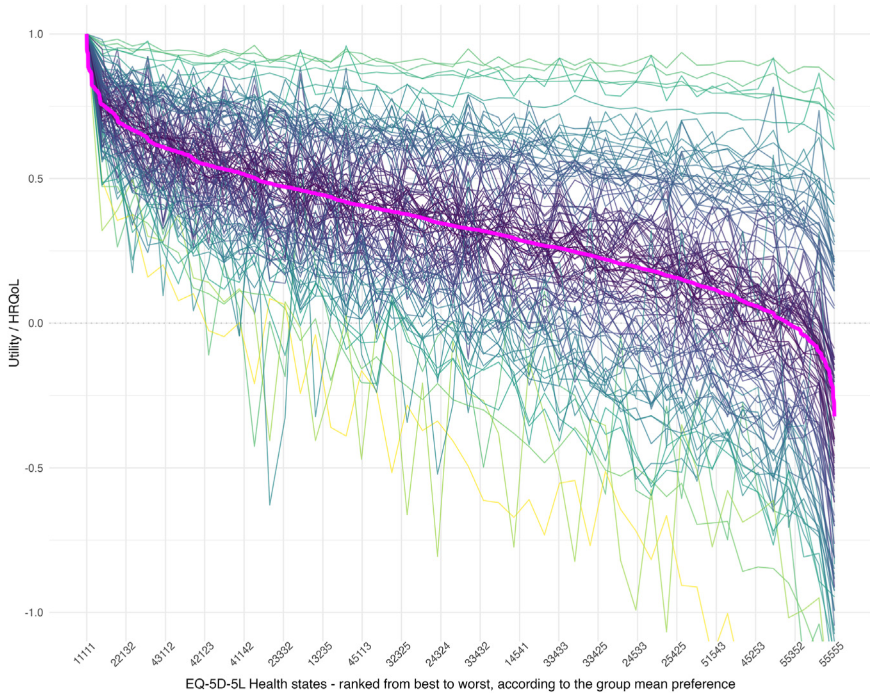
Based on a sample of 122 participants, we were able to construct an EQ-5D-5L value set for patients with rheumatic disease from Germany. Despite the small sample size, the precision of

Table 2. EQ-5D-5L value set based on the preference data of 122 patients with rheumatic diseases from Germany.

	Mobility	Self-care	Usual activities	Pain/discomfort	Anxiety/depression
lvl2	0.063 (0.055-0.072)	0.066 (0.056-0.078)	0.066 (0.058-0.075)	0.061 (0.053-0.070)	0.054 (0.047-0.062)
lvl3	0.129 (0.117-0.142)	0.133 (0.119-0.146)	0.141 (0.128-0.155)	0.136 (0.123-0.150)	0.115 (0.102-0.128)
lvl4	0.208 (0.190-0.227)	0.212 (0.192-0.232)	0.227 (0.207-0.248)	0.222 (0.202-0.242)	0.176 (0.159-0.194)
lvl5	0.269 (0.248-0.290)	0.270 (0.247-0.293)	0.280 (0.258-0.303)	0.281 (0.261-0.301)	0.222 (0.202-0.243)

Note. Shown are mean coefficients and bootstrapped 95% confidence intervals, based on 10 000 iterations.

Figure 1. Simplified illustration of the aggregate group preference (thick line) and the personal utility functions of all 122 participants. Shown are the utility values for a sample of 50 health states, ranked from the best on the left to the worst on the right (according to the aggregate group preference). The colors of the lines representing personal preference functions indicate their Euclidean distance from the average preference: purple (smaller distance) to yellow (greater distance). Utility values below −1 are not shown.



mean coefficient estimates was high, and the resulting model was internally consistent. The reported health state utilities could be readily used in health economic evaluations to gain insight into the value of different treatments for rheumatic diseases from the patients' perspective.

This is, to our knowledge, the first study to apply the OPUF method in a population of patients with rheumatic diseases—or any patient population, for that matter. So far, OPUF has only been used in smaller pilot studies and in samples of the general population.[13,14] The results show that the method is feasible and acceptable. The feedback we received further suggests that the OPUF survey was generally well received; many participants even found it to be interesting and thought-provoking. Notwithstanding, several participants also reported difficulties with the navigation, handling, or instructions on specific tasks. These issues should be addressed in future studies.

The ability to create value sets for relatively small patient groups may seem more or less relevant, depending on the context: if health policy decisions are to be informed solely by a single value set, based on preferences of the general population, other, more established methods may be better suited.[3-5] If, however, patient preferences are to be taken into account as the primary source of information, as a supplement, or simply to gain a better understanding of the preferences of a particular patient group, the OPUF method may offer a valuable alternative.

Comparison With Previous Studies

Although the OPUF method is still under development and further refinement may be needed, our findings can be compared with the results from previous study. Only very recently, Ludwig et al[1] conducted a DCE study to elicit EQ-5D-5L health state preferences from 453 patients with rheumatic arthritis in Germany. The ranking of the dimensions, as well as the relative coefficient values reported in their article, are considerably different from our results. The most important dimension for the patients in the study by Ludwig et al was self-care (level 5 coefficient = 0.364), followed by mobility (0.355), pain/discomfort (0.339), anxiety/depression (0.330), and lastly usual activities (0.272). In our study, in contrast, we found that the most important dimension was pain/discomfort (level 5 coefficient = 0.281), followed by usual activities (0.280), self-care (0.270), mobility (0.269), and anxiety/depression (0.222). Differences in the absolute coefficient values should be interpreted with caution because patient preferences were estimated on a latent DCE scale in the study by Ludwig et al[25] and then anchored using the pits state value (=−0.661) from the official German EQ-5D-5L value set. It should also be noted that their final model contains 3 logical inconsistencies in the level ordering. Within the pain/discomfort dimension, for example, the level 2 coefficient has a higher value than the level 3 coefficient (0.121 vs 0.089).

There are mainly 2 possible explanations for these differences. First, Ludwig et al used a different method, namely, DCE, to elicit preferences. Second, the characteristics of the patient sample were considerably different from our sample in terms of age, sex, and the reported health conditions.

A direct comparison between our results and the official German value set[25] may be difficult to interpret, because of the differences in both, the valuation method (OPUF vs EQ-VT), as well as the study population (general population vs patients with rheumatic diseases). It seems noteworthy, however, that despite the considerable difference in sample sizes, both studies achieved similar precision for the coefficient estimates. The official German EQ-5D-5L valuation study, which used the EQ-VT protocol and included 1158 participants, reported standard errors around mean estimates that ranged from 0.008 to 0.011. This corresponds to 95% confidence interval widths of 0.024 to 0.043 (calculated by multiplying the standard error by 3.92). These intervals are comparable to the 95% confidence intervals we achieved in our study with a sample of 122 participants.

Strengths and Limitations

Our study has several strengths but also limitations that should be considered when interpreting the results.

First of all, we would like to note that any comparison of our findings with results from other studies should take into account the differences in the valuation method. It may well be the case that value sets generated with OPUF systematically differ from value sets generated with other methods. Some indication for this can be seen in the fact that the PUFs we constructed were not good predictors of participants' choices in DCE hold-out tasks. Only for "easy" choice sets, with a utility difference of 0.25 between health states, we observed a good predictive accuracy of 81%. When the utility difference was 0.1 or 0.05, accuracy dropped to about 61%, with only marginal differences between the 2 levels, which was unexpected.

Discrepancies between OPUF and DCE are not necessarily surprising. It is well established that different valuation methods tend to produce different results.[26,27] Moreover, OPUF is based on a different theoretical framework than the commonly used TTO and DCE methods; it is a compositional preference elicitation technique, in which participants evaluate each dimension and each dimension-level individually (in comparison with some yardstick).[19] This approach has several advantages, as demonstrated in this study, but it also requires making stronger assumptions about the underlying preference structure. In particular, the OPUF method assumes an additive model, which may not always be appropriate.

Decompositional methods, such as DCE or TTO, on the other hand, involve holistic evaluations of entire health states. This requires more participants, and often more intricate statistical modeling,[28] but in principle, it is able to accommodate more complex, non-additive preference structures and, most importantly, accounting for interaction terms. However, in practice, few studies have actually done so, and it was found that interaction terms generally do not markedly improve model fit.[29]

We think that neither approach can be said to be inherently superior to the other. To reiterate a common refrain, there is no gold standard. Choosing a preference elicitation method that is appropriate for a given context is crucial but can also be challenging.[20,30] Some general guidance and important factors to consider have recently been proposed by the IMI PREFER Consortium.[31]

The OPUF method may be particularly well suited when it is difficult to recruit a large sample of participants (note, using DCE, it would most likely not have been feasible to construct an EQ-5D-5L value set for 122 patients). Moreover, the fact that dimensions are evaluated individually may make the OPUF method useful for valuing instruments that are more complex than the EQ-5D-5L (such as the EQ-HWB[32])—when there many dimensions, it can be difficult for participants to evaluate all dimensions simultaneously.

From a practical perspective, the OPUF approach is very flexible. It can be adapted to different settings and instruments, is easy to implement as a stand-alone online survey, and can be completed in a short time; most participants completed the survey in less than 20 minutes. This makes it also comparably less expensive and thus potentially attractive for researchers in resource-constrained settings.

Notwithstanding the potential advantages of the OPUF method noted above, it is important to acknowledge that OPUF is still under development—this applies to both, the actual methods being used and their implementation in a modern web-based interface. Although completion times and error rates indicated a good level of engagement, the feedback we received suggests that further refinement may be needed to ensure that all participants can complete the survey without difficulties. Moreover, the evenly spaced splits across the 100-0 scale observed in the level rating task, and the relatively homogeneous (compared with other valuation studies) weights applied to the 5 EQ-5D dimensions may suggest that some participants did not pay attention, did not fully understand the tasks, or applied simple heuristics to answer the questions. Although this is not uncommon in self-complete online surveys, future studies should explore ways to provide additional support and guidance to increase participant engagement and understanding. Finally, because this study was conducted online and participants were recruited through a patient organization, participants in our study are unlikely to be representative of patients with rheumatic diseases in Germany.

Conclusion

This study demonstrates, for the first time, that it is possible to use the OPUF approach to derive an internally consistent EQ-5D-5L value set from a relatively small sample of patients with rheumatic diseases. Our results show that the OPUF method is feasible and the feedback we received further suggests that it was generally well received by the participants. Although OPUF is still under development, we think that it has the potential to complement traditional preference elicitation methods, especially in situations where it is difficult to recruit a large sample of participants.

Author Disclosures

Links to the disclosure forms provided by the authors are available here.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2023.12.009.

Article and Author Information

Accepted for Publication: December 16, 2023

Published Online: January 27, 2024

doi: https://doi.org/10.1016/j.jval.2023.12.009

Author Affiliations: ScHARR, University of Sheffield, Sheffield, England, UK (Schneider, Brazier, van Hout); CINCH, University of Duisburg/Essen, Essen, Germany (Schneider, Blankart); Valorem Health, Bochum, Germany (Schneider); Open Health, York, England, UK (van Hout); University of Melbourne, Melbourne, VIC, Australia (Devlin).

Correspondence: Paul Schneider, Valorem Health, Bochum, Germany. Email: schneider@valorem.health

Author contributions: *Concept and design:* Schneider, Blankart, Brazier
Acquisition of data: Schneider, Blankart
Analysis and interpretation of data: Schneider, Blankart, Brazier, van Hout, Devlin

Drafting of the manuscript: Schneider
Critical revision of the paper: Schneider, Blankart, Brazier, van Hout, Devlin
Administrative, technical, or logistic support: Blankart, Brazier
Supervision: Blankart, Brazier, van Hout, Devlin
Dr. Devlin is an Editor for *Value in Health* and had no role in the peer-review process of this article.

Funding/Support: This work was supported by the Wellcome Trust (108903/Z/19/Z), the University of Sheffield, and the CINCH Health Economics Research Center, at the University of Duisburg/Essen.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The authors would like to thank the German Rheumatism Association (Deutsche Rheuma-Liga e.V., DRL) and in particular Jürgen Clausen, for their support in recruiting participants for this study. We would also like to thank Jörg Lauterberg, Sarah Mostardt, and Anja Schwalm from the Institute for Quality and Efficiency in Health Care (IQWiG) for their useful comments on the study design. Finally, we are grateful to all participants for their time and effort.

Ethical Approval: The study was approved by the ethics committee of the University of Sheffield, United Kingdom and the University of Duisburg/Essen, Germany. All participants provided informed consent.

Data and Code Availability: The data and source code of the survey, as well as the analysis scripts, have been uploaded onto Github. We are happy to share access to the respective repositories upon request. Please contact the corresponding author for more information.

REFERENCES

1. Ludwig K, Ramos-Goñi JM, Oppe M, Kreimeier S, Greiner W. To what extent do patient preferences differ from general population preferences? *Value Health*. 2021;24(9):1343–1349.
2. Burström K, Teni FS, Gerdtham UG, et al. Experience-based Swedish TTO and VAS value sets for EQ-5D-5L health states. *Pharmacoeconomics*. 2020;38(8):839–856.
3. de Bekker-Grob EW, Berlin C, Levitan B, et al. Giving patients' preferences a voice in medical treatment life cycle: the PREFER public–private project. *Patient*. 2017;10(3):263–266.
4. Hoos A, Anderson J, Boutin M, et al. Partnering with patients in the development and lifecycle of medicines: a call for action. *Ther Innov Regul Sci*. 2015;49(6):929–939.
5. Bouvy JC, Cowie L, Lovett R, Morrison D, Livingstone H, Crabb N. Use of patient preference studies in HTA decision making: a NICE perspective. *Patient*. 2020;13(2):145–149.
6. Dirksen CD. The use of research evidence on patient preferences in health care decision-making: issues, controversies and moving forward. *Expert Rev Pharmacoecon Outcomes Res*. 2014;14(6):785–794.
7. Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The role of condition-specific preference-based measures in health technology assessment. *Pharmacoeconomics*. 2017;35(suppl 1):33–41.
8. Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol Assess*. 2014;18(9):1–224.
9. Versteegh M, Brouwer W. Patient and general public preferences for health states: a call to reconsider current guidelines. *Soc Sci Med*. 2016;165:66–74.
10. Van Overbeeke E, Janssens R, Whichello C, et al. Design, conduct, and use of patient preference studies in the medical product life cycle: a multi-method study. *Front Pharmacol*. 2019;10:1395.
11. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *Patient*. 2015;8(5):373–384.
12. Oppe M, Devlin NJ, Hout B van, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health*. 2014;17(4):445–453.
13. Schneider PP, Hout B van, Heisen M, Brazier J, Devlin N. The online elicitation of personal utility functions (OPUF) tool: a new method for valuing health states. *Wellcome Open Res*. 2022;7:14.
14. Devlin NJ, Shah KK, Mulhern BJ, Pantiri K, van Hout B. A new method for valuing health: directly eliciting personal utility functions. *Eur J Health Econ*. 2019;20(2):257–270.
15. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
16. Richardson JRJ, Mckie JR, Bariola EJ. Multiattribute utility instruments and their use. In: *Encyclopedia of Health Economics*. Vol 2. Amsterdam, the Netherlands: Elsevier; 2014:341–357.
17. Devlin N, Pickard S, Busschbach J. The development of the EQ-5D-5L and its value sets. In: *Value Sets for Eq-5D-5L*. Cham, Switzerland: Springer; 2022:1–12.
18. Leidl R, Reitmeir P. An experience-based value set for the EQ-5D-5L in Germany. *Value Health*. 2017;20(8):1150–1156.
19. Marsh K, IJzerman M, Thokala P, et al. Multiple criteria decision analysis for health care decision making—emerging good practices: report 2 of the ISPOR MCDA emerging good practices task force. *Value Health*. 2016;19(2):125–137.
20. Keeney RL, Raiffa H, Rajala DW. Decisions with multiple objectives: preferences and value trade-offs. *IEEE Trans Syst Man Cybern*. 1979;9(7):403-403.
21. Edwards W, Barron FH. SMARTS and SMARTER: improved simple methods for multiattribute utility measurement. *Organ Behav Hum Decis Processes*. 1994;60(3):306–325.
22. Michel YA, Augestad LA, Barra M, Rand K. A Norwegian 15D value algorithm: proposing a new procedure to estimate 15D value algorithms. *Qual Life Res*. 2019;28(5):1129–1143.
23. Palan S, Schitter C. Prolific. Ac—a subject pool for online experiments. *J Behav Exp Fin*. 2018;17:22–27.
24. Sullivan T, Hansen P, Ombler F, Derrett S, Devlin N. A new tool for creating personal and social EQ-5D-5L value sets, including valuing "dead.". *Soc Sci Med*. 2020;246:112707.
25. Ludwig K, Graf von der Schulenburg J, Greiner W. German value set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36(6):663–674.
26. Robinson A, Spencer AE, Pinto-Prades JL, Covey JA. Exploring differences between TTO and DCE in the valuation of health states. *Med Decis Making*. 2017;37(3):273–284.
27. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford, UK: OXFORD University Press; 2017.
28. Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Med Care*. 2017;55(7):e51.
29. Poudel N, Fahim S, Qian J, Garza K, Chaiyakunapruk N, Ngorsuraches S. Methodological similarities and variations among EQ-5D-5L value set studies: a systematic review. *J Med Econ*. 2022;25(1):571–582.
30. Belton V, Stewart T. *Multiple Criteria Decision Analysis: an Integrated Approach*. New York, NY: Springer Science and Business Media; 2002.
31. The PREFER Consortium. PREFER recommendations - why, when and how to assess and use patient preferences in medical product decision-making. Zenodo. https://zenodo.org/records/6491042. Accessed September 4, 2022.
32. Brazier J, Peasgood T, Mukuria C, et al. The EQ-HWB: overview of the development of a measure of health and wellbeing and key results. *Value Health*. 2022;25(4):482–491.