VALUE IN HEALTH 20 (2017) 1150–1156

Available online at www.sciencedirect.com

ScienceDirect

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

An Experience-Based Value Set for the EQ-5D-5L in Germany

Reiner Leidl, Dr. (Prof.)[1,2,*], *Peter Reitmeir, Dr.*[1]

[1]*Institute for Health Economics and Health Care Management, HelmholtzZentrum München, Neuherberg, Germany;* [2]*Munich Center of Health Sciences, Ludwig-Maximilians University, Munich, Germany*

ABSTRACT

Objective: Valuation of health states provides a summary measure useful to health care decision makers. Results may depend on whether the currently experienced health state or a hypothetical health state is being evaluated. This study derives a value set for the EuroQoL Five-Dimensional Five-Level Questionnaire (EQ-5D-5L) by focusing on the individual's current experience. **Data and Methods:** Data include four pooled population surveys of the general German population in 2012–2015 ($N = 8114$). For valuation, a visual analogue scale (VAS) was used. Six specifications of a generalized linear model with binomial error distribution and constraint parameter estimation were analyzed. In each 1000 simulation runs, models were cross-validated after splitting the sample into an estimation part and a validation part. Predictive accuracy was measured by mean absolute error and sum of squared errors. **Results:** The models rendered a consistent set of parameters. With regard to predictive accuracy, the model considering all problem levels within the five dimensions and the highest problem level reached performed best overall. **Discussion:** Estimation proved to be feasible. Predictive accuracy exceeded that of a similar, experience-based value set for the EQ-5D-3L. Compared with a Dutch value set for the EQ-5D-5L derived for hypothetical health states, experienced values tended to be slightly lower for mild health states and substantially higher for severe health states. Clinical relevance and usefulness of the value set remain to be determined in future studies. **Conclusions:** For decision makers who prioritize patient-relevant benefit, the experience-based value set provides a novel option to summarize health states, reflecting how health states experienced are valued in a population.

Keywords: EuroQoL, patient reported outcome, quality of life, valuation.

Introduction

In the economic analysis of health-related quality of life, value sets or tariffs are important tools. They provide a standard by which described health states can be valued with reference to a population. The most widely used concept aims to generate a utility-based value set (UBVS), which reflects the preferences of the population surveyed and is often used to calculate quality-adjusted life years (QALYs) [1,2]. For measurement, respondents are asked to make a choice from descriptions of hypothetical health states, and the time trade-off method is mostly used for elicitation. A key normative point in the UBVS is whether general public preferences or individual patient preferences should be used in the valuation [3]. The UBVS is very successful in supporting decisions on allocating health care resources, for example, considering its long-lasting use by the UK National Institute for Health and Care Excellence (NICE) [4]. In contrast, with regard to the health states to be assessed, critics have stated that "NICE should value real experiences over hypothetical opinions" [5]. Swedish regulators have suggested especially considering patient preferences in the assessment of health care intervention [6], and German regulators are legally required to consider the benefits to patients in drug reimbursement decisions [7]. Thus, utilities derived from trade-offs between hypothetical states may not provide optimal evidence to support all types of health care decisions.

This study started from a normative position which requires patient-relevant benefit to be shown for measuring effects. For quality of life, this position implies that the patient should both describe and evaluate his or her own, currently experienced health state. From such valuations, an experience-based value set (EBVS) can be estimated. The EBVS is a population standard of average perception of a person's own current health state. By taking the described normative position, a decision maker may find it useful to check an EBVS and, eventually, adapt the results from studies in which respondents or valuation methods did not meet requirements, for example, because these results originated from a UBVS or came from another country. A comparison of the results of an EBVS with those of other approaches will indicate whether the choice of the valuation approach significantly affected the results. For patients receiving total hip replacement in the United Kingdom and Germany, a study analyzed quality of

---

The authors have indicated that they have no conflicts of interest with regard to the content of this article.

* *Address correspondence to:* Reiner Leidl, Institute for Health Economics and Health Care Management, HelmholtzZentrum München, Postfach 1129, 85758 Neuherberg, Germany.

E-mail: Leidl@helmholtz-muenchen.de..

http://dx.doi.org/10.1016/j.jval.2017.04.019

life 6 months after surgery [8]. No difference was found when both groups matched by propensity score were compared using the traditional UK utility tariff for both. Change of perspective and use of a German EBVS for both showed significantly higher quality of life for the UK study patients. Another important feature is that the EBVS reports a population average of how respondents value their experienced health states. Use of the EBVS as a benchmark thus turns it into a diagnostic tool. These average population valuations can be compared with direct patient valuations, and significant differences indicate that further explanation is required. In a clinical study, patients with acute heart disease were found to evaluate their health states much lower than the EBVS benchmark immediately after discharge from the acute care setting, but the valuation differences disappeared after 3 weeks of rehabilitation [9]. This was an indication of the shock experienced close to the acute event, and this was not captured by the descriptive part of the instrument used, the EuroQoL Five-dimensional Three-Level questionnaire (EQ-5D-3L). Finding subgroups of patients with systematic differences from EBVS population averages may help identify groups with special medical needs.

Thus, EBVSs refer to current, experienced health states of respondents and provide a summary index for quality of life from a population perspective. In a narrow definition, an EBVS cannot include the state of being dead. However, an EBVS has also been evaluated by using the time trade-off method [10]. This means that trade-offs with the hypothetical state of being dead have been elicited. Use of direct valuation of experienced health on a visual analogue scale (VAS) for the EBVS does not specify a choice between health states and excludes the valuation of the state of being dead. As a result, the scale used cannot be anchored by setting the value of the state of being dead zero, and comparability of the scales between respondents is restricted. In spite of that, the psychometric characteristics of direct, non-anchored VAS valuations have been found to be acceptable, performing equally in terms of content validity and reliability and, in part, better than choice-based methods [9,11]. By multiplying the values of an EBVS by the duration of the respective health state, quality adjusted life time can be calculated by the EBVS. Because of the conceptual differences between the UBVS and the EBVS, it must be made clear that quality adjustment in the EBVS differs from the QALY concept.

The valuation approaches discussed here all refer to the descriptive basis of the EQ-5D, a widely used generic quality of life instrument with five dimensions. The EQ-5D-3L version with three problem levels covers 243 health states. In this version, VAS values were estimated at the population level in previous studies [12,13], but EBVSs have only been developed for Germany, Sweden, and China [10,14,15]. These EVBSs have also been used and tested in a number of chronic diseases [16–21] and in elective surgery [22,23]. Meanwhile, a more differentiated version, the EQ-5D-5L, has been developed, covering five problem levels and a total of 3125 health states [24]. An EBVS for the EQ-5D-5L, to the best of our knowledge, has yet to be constructed.

Focusing on the respondents' currently experienced health, this study aimed to develop an EBVS for the EQ-5D-5L for Germany. By extending the econometric approach that had been developed for the EQ-5D-3L [14], several model specifications were derived and tested with regard to predictive performance. For all health states, values were finally compared with those of a traditional UBVS.

Data and Methods

This study considers the criteria established by the Checklist for Reporting Valuation Studies of the EQ-5D [25]. The EQ-5D-5L used in this study has been added to four-yearly surveys of the general German population. The surveys were conducted by the Wort and Bild Verlag from 2012 to 2015. Informed consent was obtained from all individual participants included in the study. The surveys are representative of the general population aged 14 years and older. The EQ-5D-5L questions, including the VAS, were administered by written questionnaire. From 2012 to 2014, the surveys produced 6074 respondents; the methods and results have been presented in detail elsewhere [26]. The 2015 survey was conducted in the same way, adding another 2040 respondents and generally stable baseline results; for detail, see the study by Huber et al. [27]. EQ-5D-5L data from all four surveys were pooled for this study.

The VAS was used for direct valuation. For pragmatic purposes, all VAS models and results were reported on the 0–1 scale instead of the original 0–100 scale. To generate the EBVS, VAS scores for all health states were evaluated by using a valuation index (IDX) based on a generalized linear model. IDX was assumed to be binomially distributed, described by a probability of success (P value) and 100 experiments. Although EQ-5D data were collected once from each individual, these assumptions could theoretically be made by considering the following: In 100 experiments, a person valuing his or her health state at P would be given a random number from the (0–1) range of valuation that refers to a defined health state. The person would then be asked whether his or her health state is at least as good as the state presented. The share of experiments in which the respondent is expected to agree is, then, P. These distributional assumptions keep predictions on the 0–1 scale and ensure that exact estimates are achieved at the two extremes of the range. To estimate an EBVS for the EQ-5D-3L version, this type of model has been found to be better in terms of consistency and performance compared with the ordinary least-squares regression model and the scale-transformed regression model [14].

The overall level of problems reported may impact the health state valuation. To overcome this, a *maximum problem level* (MPL) concept has been introduced and defined as the highest problem level reported. *MPL2/3/4/5* is defined as at least one slight/moderate/severe/extreme problem reported, respectively. Based on the MPL, MPL2 is set to 1 if the MPL is ≥ 2, MPL3 is set to 1 if the MPL is ≥ 3, and so on. The MPL concept is used within the EQ-5D dimensions in a similar manner. For example, *MO2/3/4/5* is defined as at least one slight/moderate/severe/extreme problem, respectively, that is reported with regard to mobility. Defining the remaining four dimensions, respectively, SC2 to SC5 are introduced for self-care, UA2–UA5 for usual activities, PD2–PD5 for pain and discomfort, and AD2–AD5 for anxiety and depression. Each of these parameters reflects the additional decrement in valuation when the problem level reported increases by one unit. For a correct ordinal ranking of the IDX values, the model is then estimated under the constraint that, except for the intercept, all parameter estimates must be non-positive. For the generalized linear model in equation 1, the logit function is used as a link function between the linear predictor η and the expected *P* value.

$$
\begin{aligned}
& IDX \cdot 100 \sim Bin(100, p) \\
& \eta = \beta_{00} + \beta_{01}MPL2 + \beta_{02}MPL3 + \beta_{03}MPL4 + \beta_{04}MPL5 \\
& \quad + \beta_{1}MO2 + \beta_{2}SC2 + \beta_{3}UA2 + \beta_{4}PD2 + \beta_{5}AD2 \\
& \quad + \beta_{6}MO3 + \beta_{7}SC3 + \beta_{8}UA3 + \beta_{9}PD3 + \beta_{10}AD3 \\
& \quad + \beta_{11}MO4 + \beta_{12}SC4 + \beta_{13}UA4 + \beta_{14}PD4 + \beta_{15}AD4 \\
& \quad + \beta_{16}MO5 + \beta_{17}SC5 + \beta_{18}UA5 + \beta_{19}PD5 + \beta_{20}AD5 = \beta X \\
& Logit(E(IDX|X)) = Logit(p) = \ln\left(\frac{p}{1-p}\right) = \eta \qquad (1)
\end{aligned}
$$

With regard to the MPL, three models can now be specified: In model 1, the MPL is not considered at all. Thus, in the first line of the η equation, only β_{00} is used, leading to a 21-parameter model. Model 2 considers the existence of at least one problem. Thus, it only uses $\beta_{00} + \beta_{01}$MPL2 in the first line of the η equation, leading to a 22-parameter model. Model 3 considers each problem level reported separately, thus using all parameters in the η equation, rendering a 25-parameter model. MPL enters model 3 as follows: If, for example, no problems except one moderate problem are reported across all dimensions, MPL2 and MPL3 are set to 1, and MPL4 and MPL5 are set to 0.

In all three models, problem levels are considered to represent categories. The models are thus labeled “categorical.” In sensitivity analysis, it was assumed that all increases of one problem level within a dimension are the same and that the five problem levels can be represented by a single linear variable (“score”). Results for the “linear” models 4 to 6, corresponding to categorical models 1 to 3, are provided in the Appendix.

Model results were cross-validated by using 1000 independent simulations. For each simulation, the whole sample was split into two, estimating the model in the first subsample and comparing the results with those of the second. This split was determined so that a health state with at least four observations in the pooled sample had a 95% probability of being included in the estimation sample. To fully account for respondents with different MPLs, MPL1 was defined as an additional group that included all respondents who did not report any problem at all. Cross-validation was reported for subgroups with MPL1, MPL2, MPL3, and MPL4. Accuracy in predicting actual valuations was measured by mean absolute error (MAE) and by the sum of squared errors (SSE), which attaches a higher weight to larger deviations. Models were compared across simulation runs for MAE, SSE, and the mean rank among the tested models, including the linear models of the sensitivity analysis. The best performing model was selected, and its estimations were compared with observed valuations by using MAE, also illustrating the impact of the number of health states observed. As a first step to explore how a UBVS for the EQ-5D-5-5L might structurally differ from the EBVS developed here, as a German tariff is not yet available [28], the Dutch tariff, which is based on composite time trade-off and a Tobit model with constraints, was graphically compared for all 3125 health states [29].

Calculations were conducted by using the procedure NLMIXED from the statistics software package SAS, version 9.3. For optimization, the maximization of the likelihood function was chosen under the above-mentioned constraint. Further processing of results was performed with Microsoft Excel 2016.

Results

The total pooled sample included 8114 respondents and 316 health states reported (see supplementary data file S2). VAS valuations observed ranged from 0 to 1. Of the respondents, 62.2% did not report any problem (health state 11111). An MPL of 2/3/4/5 was reported by 23.1%, 11.1%, 3.0%, and 0.5%, respectively. The most restricted individual reported extreme problems in all dimensions except “anxiety/depression with a severe problem. Respondents’ VAS valuations were distributed widely across health states of all MPLs. With a higher MPL, the number of health states in the EBVS that had higher valuations declined (Fig. S1; see supplementary file S1).

In terms of parameter estimates, all three models were found to be qualitatively similar, rendering significant results for most of the parameters. In the dimensions of “pain/discomfort” or “anxiety/depression,” the impact of problems reported on the index was substantially reduced when the problem level occurring in all of the dimensions was considered in the model. No additional impact was estimated for the extreme problem level in mobility and in usual activities, and for the severe problem level in self-care (Table 1). Taking model 3 as an example, valuation for each additional problem level reported in the five dimensions continued to decline, with the strongest for pain and discomfort, giving a valuation of –0.18 when problems increased from severe to extreme (Table 2). To derive the value for a described health state, all parameters to estimate η had to be summed up and then retransformed by $e^{\mathrm{Sum}}/(1+e^{\mathrm{Sum}})$. For an individual reporting no problems except extreme pain, model 3 would thus require summing up of parameters for the intercept, for MPL2, MPL3, MPL4, and MPL5, as well as for PD2, PD3, PD4, and PD5, rendering a value of 0.2336 and, retransformed, a value of 0.55814. The linear models also yielded significant parameter estimates for sensitivity analysis (Table S1).

For cross-validation runs, the sample was randomly split into 6075 respondents for model estimation and 2039 to test the model’s predictive performance. Compared with the total sample, the number of respondents in the subgroups declined with increasing MPL; SSE as a cumulative indicator also declined, whereas MAE increased in both level and dispersion except for the large group not reporting any problem at all. Model 3, which integrated all problem levels separately, performed best with regard to all indicators with one exception: In the group of respondents reporting, at most, a slight problem (MPL2), model 3 was outpaced by model 1 not integrating any problem level, with regard to MAE and both rank indicators (Table 3). The same structure was found in the linear models, which, overall, did not perform any better compared with the categorical models. An exception among linear models, model 1, again performed better in the subgroup of respondents reporting, at most, a slight problem (Table S2).

As a result, categorical model 3 was selected for the value set. The exact values for all health states are provided in supplementary data file S2. Overall, MAE for this model was 0.076. For 112 health states observed more than twice, which included 96.9% of individuals in the sample, MAE dropped to 0.039. For 59 health states that were observed more than five times, including 94.3% of individuals, MAE dropped further to 0.026. Thus, the correlation between the predicted valuations and the observed valuations improves when focusing on health states that were observed more frequently (Fig. 1). There was no indication of systematic bias, although the fact that predictive accuracy may decline in very bad health states cannot be excluded.

The Dutch UBVS was compared with the German EBVS for mean valuations of all health states: The two value sets did not produce the same valuation for any single health state. In 544 health states, the UBVS exceeded the EBVS; in the remaining 2581 health states, it was the opposite. More systematic differences are revealed by distinguishing values for 243 health states without any severe or extreme problems reported (MPL <4) and values for the remaining health states with at least one severe or extreme problem reported (MPL >3). For the first group, a strong correlation could be observed, but a somewhat higher level of values was observed in the Dutch tariff. For the group reporting at least one severe or extreme problem, the opposite was oberved; the bulk of observations was above the line of equality with lower values for the Dutch tariff, and the more severe the health states, the greater the difference was (Fig. 2).

Discussion

By extending an estimation approach developed earlier, this study tested six model specifications (including the sensitivity analyses) to estimate an EBVS for the EQ-5D-5L in Germany, relying on VAS valuation. The approach ensures that the estimation stays within the defined scale and that parameter estimates are consistent. Only the categorical models fully match the scaling properties of the problem levels surveyed in the EQ-5D-5L questionnaire. Constraint estimation of these models leads to a zero estimate for three out of the 20 parameters for

Table 1 – Three categorical models to estimate the experience-based value set, with different ways to integrate maximum problem level (MPL).

		Model 1: No maximum problem level included			Model 2: Just MPL2 included			Model 3: MPL2, MPL3, MPL4 and MPL5 included		
		β	SE	*P*	β	SE	*P*	β	SE	*P*
	Intercept	2.3754	0.005	< 0.001	2.4365	0.005	< 0.001	2.4365	0.005	< 0.001
Maximum problem level	MPL2	–	–	–	–0.5422	0.015	< 0.001	–0.5199	0.015	< 0.001
	MPL3	–	–	–	–	–	–	–0.1857	0.017	< 0.001
	MPL4	–	–	–	–	–	–	–0.1305	0.029	< 0.001
	MPL5	–	–	–	–	–	–	0.0000[*]	–	–
Mobility	MO2	–0.3455	0.011	< 0.001	–0.2852	0.011	< 0.001	–0.2812	0.011	< 0.001
	MO3	–0.0953	0.014	< 0.001	–0.1144	0.014	< 0.001	–0.0519	0.015	< 0.001
	MO4	–0.0757	0.022	< 0.001	–0.0778	0.022	< 0.001	–0.0219	0.028	0.430
	MO5[*]	0.0000[*]	–	–	0.0000[*]	–	–	0.0000[*]	–	–
Self-care	SC2	0.0000[*]	–	–	–0.0088	0.014	0.535	–0.0071	0.014	0.618
	SCDF3	–0.1003	0.021	< 0.001	–0.1082	0.022	< 0.001	–0.1122	0.022	< 0.001
	SC4[*]	0.0000[*]	–	–	0.0000	–	–	0.0000[*]	–	–
	SC5	–0.5721	0.110	< 0.001	–0.5341	0.110	< 0.001	–0.4887	0.109	< 0.001
Usual activity	UA2	–0.2489	0.011	< 0.001	–0.2525	0.012	< 0.001	–0.2387	0.012	< 0.001
	UA3	–0.1828	0.017	< 0.001	–0.1789	0.017	< 0.001	–0.1682	0.017	< 0.001
	UA4	–0.2521	0.031	< 0.001	–0.2748	0.031	< 0.001	–0.2899	0.031	< 0.001
	UA5[*]	0.0000[*]	–	–	0.0000[*]	–	–	0.0000[*]	–	–
Pain/discomfort	PD2	–0.8547	0.009	< 0.001	–0.4543	0.014	< 0.001	–0.4621	0.014	<0.001
	PD3	–0.2901	0.011	< 0.001	–0.3238	0.011	< 0.001	–0.2031	0.016	< 0.001
	PD4	–0.2748	0.022	< 0.001	–0.2819	0.022	< 0.001	–0.2355	0.027	< 0.001
	PD5	–0.4777	0.065	< 0.001	–0.4651	0.065	< 0.001	–0.4661	0.064	< 0.001
Anxiety/depression	AD2	–0.3708	0.009	< 0.001	–0.2097	0.010	< 0.001	–0.2085	0.010	< 0.001
	AD3	–0.1195	0.015	< 0.001	–0.1359	0.015	< 0.001	–0.05	0.017	< 0.001
	AD4	–0.1990	0.031	< 0.001	–0.2007	0.031	< 0.001	–0.15	0.034	< 0.001
	AD5	–0.4348	0.070	< 0.001	–0.4070	0.069	< 0.001	–0.32	0.069	< 0.001

MPL is the maximum problem level reported, here across all dimensions. For example, MPL2 indicates that at least one problem is reported. The other dummy variables refer to the respective problem level reported in each single dimension. For a calculation example, see text.

SE, standard error.

[*] No significance testing because of the estimation under constraints.

Table 2 – Valuation decrements for increasing problem levels in categorical model 3, by EQ-5D dimension.

	Mobility	Self-care	Usual activity	Pain/discomfort	Anxiety/depression
Slight problem	–0.08	–0.05	–0.08	–0.11	–0.07
Moderate problem	–0.12	–0.09	–0.13	–0.18	–0.11
Severe problem	–0.14	–0.11	–0.21	–0.25	–0.15
Extreme problem	–0.19	–0.25	–0.27	–0.43	–0.28

Note: Even though the parameter estimates for MO5, SC4, and UC5 are zero, valuation declines for these problem increases because MPL increases.

problem levels in the five dimensions. Yet, these three problem levels will still decrease valuation if their occurrence increases the MPL variable, thus underlining the relevance of the MPL experienced. The "linear" models employ substantially fewer parameters but require the additional assumption that problem levels reflect a linear scale. In a structural sensitivity analysis, these models were also found to have useful results, although they did not show better performance.

The sample size, which was a strength of this study, allowed for comprehensive cross-validation to assess predictive accuracy. This included consideration of model performance in subgroups defined by the MPLs reported. Results were mostly consistent across validation indicators and subgroups, with the categorical model accounting for all MPLs showing the best overall performance. This best model achieved an MAE of 0.076 (0.076–0.079), which is somewhat better than the 0.093 (0.090–0.097) that had been achieved in the external validation of the EBVS developed for the EQ-5D-3L [14]. Although the 5L version of the EQ-5D is more differentiated than the 3L version, it also has to be taken into account that the sample size for developing the 5L EBVS is about four times as large as that for the 3L EBVS.

Finally, a comparison with a Dutch UBVS revealed important systematic differences. The UBVS had slightly higher values than the EBVS for relatively mild health states and clearly lower values once severe or extreme problems occur. The difference between the UBVS and the EBVS increases with increasing severity of the health states. This finding resembles reports on the EQ-5D-3L indicating that in the UBVS valuations strongly decrease as soon as a problem level of three is observed, thus creating a gap compared with valuations where this is not the case [30]. The normative discussion regarding whose preferences should be used in valuation underlines this finding, pointing out as a main argument for patient valuations that the general public may overestimate the impact of poor health [3]. Several other factors may contribute to the differences observed in the present study, which, in part, also restricted comparability. There were differences in scale and valuation method. The UBVS was anchored for the state of being dead and comprised 484 health states with negative values. The scale of the EBVS was not as wide, as it was not anchored for the state of being dead, but the VAS values were all non-negative. Furthermore, the samples referred to different national populations and to hypothetical health states in the Dutch case and experienced ones in the German case. The relative contributions of these factors could not be quantified in this study. Moreover, to what extent preferences on severe hypothetical health states would tend to exaggerate what was being observed in terms of experience and to what extent respondents adapted to the severe health states that they experienced remain to be analyzed. Yet the impact of the difference in valuation functions found is obvious: The greater decrements in valuation of the very severe health states in the UBVS will attribute a larger effect to interventions that can protect against these very severe health states. This observation corresponds to the findings stressed in a study comparing a Swedish EBVS with a UK UBVS, which concluded that when used for allocation decisions, the UBVS tends to prioritize interventions improving quality of life, whereas the EBVS tends to prefer

Table 3 – Cross-validation of models using mean absolute deviation (MAE) and sum of squared errors (SSE), by maximum problem level (MPL) reported.

Maximum problem level	Model	MAE	Rank, MAE	SSE	Rank, SSE
All	1	0.077 (0.077–0.080)	5.00 (5–5)	21.30 (21.11–22.92)	5.00 (5–5)
All	2	0.076 (0.076–0.079)	2.09 (2–3)	20.89 (20.71–22.44)	2.07 (2–2)
All	3	0.076 (0.076–0.079)	1.14 (1–3)	20.74 (20.55–22.36)	1.09 (1–1)
MPL1	1	0.067 (0.067–0.070)	5.00 (5–5)	10.33 (10.30–10.64)	4.90 (5–5)
MPL1	2	0.066 (0.066–0.069)	2.50 (2.5–2.5)	10.26 (10.22–10.60)	2.53 (2.5–2.5)
MPL1	3	0.066 (0.066–0.069)	2.50 (2.5–2.5)	10.26 (10.22–10.60)	2.53 (2.5–2.5)
MPL2	1	0.079 (0.078–0.087)	2.21 (2–5)	4.85 (4.71–6.07)	1.38 (1–5)
MPL2	2	0.081 (0.080–0.086)	3.91 (3–4)	4.86 (4.74–5.95)	2.86 (1–3)
MPL2	3	0.080 (0.080–0.085)	2.85 (1–3)	4.83 (4.71–5.93)	1.97 (2–2)
MPL3	1	0.109 (0.104–0.110)	5.00 (5–5)	4.24 (3.94–4.28)	5.01 (5–5)
MPL3	2	0.105 (0.101–0.106)	2.06 (2–3)	3.93 (3.73–3.95)	2.08 (2–3)
MPL3	3	0.105 (0.101–0.105)	1.21 (1–3)	3.86 (3.70–3.88)	1.12 (1–2)
MPL4,5	1	0.135 (0.135–0.139)	4.92 (5–5)	1.88 (1.84–2.25)	3.13 (3–5)
MPL4,5	2	0.133 (0.133–0.138)	2.14 (2–4)	1.85 (1.80–2.20)	2.14 (2–4)
MPL4,5	3	0.131 (0.131–0.137)	1.16 (1–3)	1.79 (1.74–2.16)	1.17 (1–3)

MPL is the maximum problem level reported, here across all dimensions. MPL1, 2, 3, and 4,5 refer to the respective subgroups. MPL1 is defined as the state without any problem. Group sizes: all individuals 8114; MPL1, 5050; MPL2, 3064; MPL3, 1186; MPL4,5, 325. The ranking indicators refer to both categorical and linear models; ranks thus range from 1 to 6 with 1 being the best. In parentheses, the range between the 5th and 95th percentile is given.

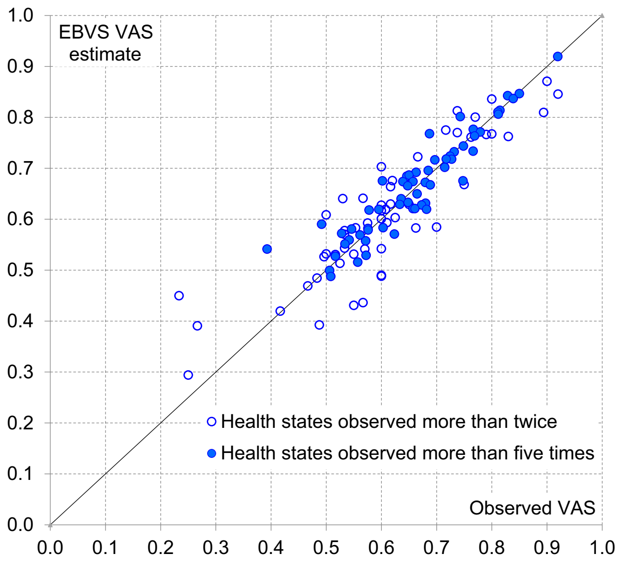

Fig. 1 – Comparing estimates of the experience-based value set (EBVS) and observed visual analogue scale (VAS). Out of 316 health states observed, 112 were observed more than twice and a subgroup of these, 59 health states, were observed more than five times. Pearson correlations between mean values estimated and observed are 0.904 and 0.930, respectively. The diagonal line indicates equality between estimated and observed values.

interventions increasing survival [31]. Clearly, decision makers should be aware that the choice of the valuation approach may affect their decisions. This choice may have to consider normative or legal requirements. In addition, the choice may also be influenced by the type of decision to be made, for example, whether resources are to be allocated according to *ex ante* preferences or whether patient management is to be improved according to the benefits experienced.

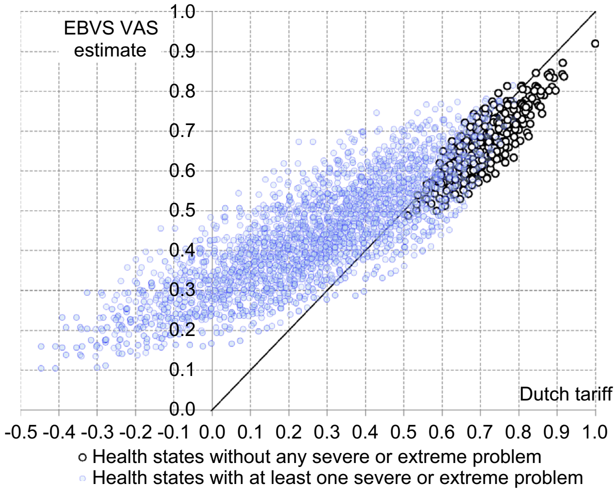

Fig. 2 – Comparing values for all EQ-5D-5L health states for the Dutch tariff based on hypothetical states (23) and the German experience-based value set (EBVS). Number of health states without any severe or extreme problem: 243; health states with at least one severe or extreme problem: 2882; total number of health states: 3125. The diagonal line indicates equality between the values of the two value sets.

There were several limitations to this study. First, only about 10% of all possible EQ-5D-5L health states were reported in the sample. The more severe states were observed less frequently and with higher variance of valuation. However, variances were not used in the EBVS, and parameter estimates were unbiased. Four representative yearly surveys provided a substantial base of health states, but institutionalized populations, including hospital patients, were less likely to be included. To what extent the EBVS adequately reflects the valuation of severe diseases remains to be analyzed. This leads to the second, more general, limitation. The aim of the EBVS concept is to reflect patient-relevant benefit, and whether or not this can be adequately achieved for the EQ-5D-5L, as developed in Germany, remains to be shown by taking into consideration the psychometric properties of the EBVS in clinical application. A major issue for use in patient management is whether or not estimates of patient valuations are accurate enough for a clinical assessment, at least at a group level. Further analysis is needed here: Compared with minimal clinically important differences (MCIDs) for using the EQ-5D-5L and VAS valuation, MAEs were found to be somewhat below the MCIDs reported for patients who had experienced stroke [32] but slightly above the MCIDs—although within confidence intervals—reported for patients with chronic obstructive pulmonary disease [33]. Accordingly, it has to be considered, by clinical application, whether or not accuracy is considered sufficient to properly reflect effectiveness. Clinical relevance and accuracy also have to be considered when using the UBVS approach. Third, other topics for future research include methodologic work on disentangling the effects of diverging populations, valuation methods, and valuation targets, namely, hypothetical or experienced health states, or, when it comes to the use of value sets in decision making, even such issues as the strategic behaviors of respondents.

Conclusions

The EBVS concept differs from the UBVS concept and the QALY approach. The EBVS approach may inform decision makers who prioritize patient-relevant benefit but who may not require information for QALY maximization. In the appraisal of quality of life, when direct patient valuation relevant to the decision context is missing, an EBVS offers a novel tool to test whether context-specific valuation would matter and to indicate valuation results from an average population experience. The focus on experience may bring together clinical and economic assessments of quality of life. Compared with a previous EBVS for the EQ-5D-3L, the predictive accuracy of the EBVS for the EQ-5D-5L could be further increased, even for the substantially larger number of health states. Before using the new EBVS for the 5L version in a clinical context, it is strongly recommended that its respective psychometric properties be considered. This indicates an important research need, including comparison of the performance of the EBVS with that of the UBVS. Besides contributing to the evidence on quality of life in the assessment of interventions, the EBVS also offers new ways to integrate quality of life into patient management. It provides a benchmark that reflects the valuation of experienced health from a population perspective. Patient groups that systematically differ in valuation can thus be identified. An underestimation by the EBVS may indicate opportunities for good patient management, whereas an overestimation may help identify risk groups of patients with specific medical needs.

Acknowledgment

The data for this study were provided by Wort & Bild Verlag (WuB), Baierbrunn, Germany. The authors are especially grateful to Martin Vogelmann for adding the EQ-5D questions to the WuB surveys for our research. An earlier version of this article was presented to the 2016 Conference of the European Association of Health Economics in Hamburg, Germany. Helpful comments by session participants as well as by René Stark, MD MPH; Manuel Huber, MSc; HelmholtzZentrum München; Matthias Arnold, MSc MBR, and Dipl. Hdl.; and Florian Kirsch, MBR, Ludwig-Maximilians University, Munich, are gratefully acknowledged.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2017.04.019 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value Health 2009;12:S5–9.

[2] Szende A, Janssen B, Cabases J. Self-reported population health: an international perspective based on EQ-5D. The Netherlands: Springer Open, 2014.

[3] Versteegh MM, Brouwer WB. Patient and general public preferences for health states: a call to reconsider current guidelines. Soc Sci Med 2016;165:66–74.

[4] Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess 2015;19:1–503.

[5] Dolan P. NICE should value real experiences over hypothetical opinions. Nature 2009;462:35.

[6] The Dental and Pharmaceutical Benefits Agency (TLV). General guidelines for economic evaluations from the Pharmaceutical Benefits Board (LFNAR 2003:2). Stockholm, Sweden: the Dental and Pharmaceutical Benefits Agency, 2003.

[7] Arzneimittel-Nutzenbewertungsverordnung. The Ordinance on the Benefit Assessment of Pharmaceuticals, § 2(3). Part I Nr. 68, p. 2324. Bonn, Germany: Bundesgesetzblatt, 2010.

[8] Vogl M, Leidl R, Plötz W, et al. Comparison of pre-and post-operative health-related quality of life and length of stay after primary total hip replacement in matched English and German patient cohorts. Qual Life Res 2015;24:513–20.

[9] Leidl R, Schweikert B, Hahmann H, et al. Assessing quality of life in a clinical study on heart rehabilitation patients: how well do value sets based on given or experienced health states reflect patients' valuations? Health Qual Life Outcomes 2016;14:48.

[10] Burström K, Sun S, Gerdtham U-G, et al. Swedish experience-based value sets for EQ-5D health states. Qual Life Res 2014;23:431–42.

[11] Leidl R, Reitmeir P, König H-H, et al. The performance of a value set for the EQ-5D based on experienced health states in patients with inflammatory bowel disease. Value Health 2012;15:151–7.

[12] Kind P. Valuing EQ-5D health states—a VAStly simpler solution? In: Busschbach J, Rabin R, De Charro F, eds., Proceedings of the 24th Scientific Plenary Meeting of the EuroQoL Group—, September 13–15, 2009; Kijkduin-The Hague, The Netherlands, Rotterdam: EuroQoL Group Executive Office, 2009.

[13] Rand-Hendriksen K, Augestad LA, Kristiansen IS, et al. Comparison of hypothetical and experienced EQ-5D valuations: relative weights of the five dimensions. Qual Life Res 2012;21:1005–12.

[14] Leidl R, Reitmeir P. A value set for the EQ-5D based on experienced health states. Pharmacoeconomics 2011;29:521–34.

[15] 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;24:693–703.

[16] Hunger M, Sabariego C, Stollenwerk B, et al. Validity, reliability and responsiveness of the EQ-5D in German stroke patients undergoing rehabilitation. Qual Life Res 2012;21:1205–16.

[17] Wacker ME, Jörres RA, Karch A, et al. Relative impact of COPD and comorbidities on generic health-related quality of life: a pooled analysis of the COSYCONET patient cohort and control subjects from the KORA and SHIP studies. Respi Res 2016;17:81.

[18] Kiadaliri AA, Eliasson B, Gerdtham U-G. Does the choice of EQ-5D tariff matter? A comparison of the Swedish EQ-5D-3L index score with UK, US, Germany and Denmark among type 2 diabetes patients. Health Qual Life Outcomes 2015;13:145.

[19] Cooper A, Wallman J, Gülfe A. What passes for good? Experience-based Swedish and hypothetical British EuroQol 5-Dimensions preference sets yield markedly different point estimates and patient acceptable symptom state cut-off values in chronic arthritis patients on TNF blockade. Scand J Rheumatol 2016;45:470–3.

[20] Gülfe A, Wallman JK, Kristensen LE. EuroQol-5 dimensions utility gain according to British and Swedish preference sets in rheumatoid arthritis treated with abatacept, rituximab, tocilizumab, or tumour necrosis factor inhibitors: a prospective cohort study from southern Sweden. Arthritis Res Ther 2016;18:51.

[21] Little MRP, Peters A, Leidl R. Does experience matter when valuing health? A comparison of EQ-5D value tariffs in a German population study. Value Health 2014;14:364–71.

[22] Vogl M, Wilkesmann R, Lausmann C, et al. The impact of preoperative patient characteristics on the cost-effectiveness of total hip replacement: a cohort study. BMC Health Serv Res 2014; 14:342.

[23] Nemes S, Burström K, Zethraeus N, et al. Assessment of the Swedish EQ-5D experience-based value sets in a total hip replacement population. Qual Life Res 2015;24:2963–70.

[24] 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;20:1727–36.

[25] Xie F, Gaebel K, Perampaladas K, et al. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. Med Decis Making 2014;34:8–20.

[26] Huber MB, Reitmeir P, Vogelmann M, et al. EQ-5D-5L in the general German population: comparison and evaluation of three yearly cross-section surveys. Int J Environ Res Public Health 2016;13:343.

[27] Huber MB, Felix J, Vogelmann M, et al. Health-related quality of life of the general German population in 2015: results from the EQ-5D-5L. Int J Environ Res Public Health 2017;14:426.

[28] Ludwig K, von der Schulenburg JG, Greiner W. Valuation of the EQ-5D-5L with composite time trade-off for the German population—an exploratory study. Health Qual Life Outcomes 2017;15:39.

[29] Versteegh MM, Vermeulen KM, Evers SM, et al. Dutch tariff for the five-level version of EQ-5D. Value Health 2016;19:343–52.

[30] Parkin D, Devlin N, Feng Y. What determines the shape of an EQ-5D index distribution? Med Decis Making 2016;36:941–51.

[31] Aronsson M, Husberg M, Kalkan A, et al. Differences between hypothetical and experience-based value sets for EQ-5D used in Sweden: implications for decision makers. Scand J Public Health 2015;43: 848–854.

[32] Chen P, Lin K-C, Liing R-J, et al. Validity, responsiveness, and minimal clinically important difference of EQ-5D-5L in stroke patients undergoing rehabilitation. Qual Life Res 2016;25:1585–96.

[33] Nolan CM, Longworth L, Lord J, et al. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. Thorax 2016;71:493–500.