



## ORIGINAL PAPER

# A comparison of the responsiveness of EQ-5D-5L and the QOLIE-31P and mapping of QOLIE-31P to EQ-5D-5L in epilepsy

Ben F. M. Wijnen<sup>1,2,9</sup> · Iris Mosweu<sup>3</sup> · Marian H. J. M. Majoie<sup>2,4,5,6</sup> · Leone Ridsdale<sup>7</sup> · Reina J. A. de Kinderen<sup>1,8</sup> · Silvia M. A. A. Evers<sup>1,8</sup> · Paul McCrone<sup>3</sup>

Received: 26 April 2017 / Accepted: 25 August 2017 / Published online: 4 September 2017  
© The Author(s) 2017. This article is an open access publication

## Abstract

**Objective** To investigate the responsiveness of and correlation between the EQ-5D-5L and the QOLIE-31P in patients with epilepsy, and develop a mapping function to predict EQ-5D-5L values based on the QOLIE-31P for use in economic evaluations.

**Methods** The dataset was derived from two clinical trials, the ZMILE study in the Netherlands and the SMILE study in the UK. In both studies, patients' quality of life using the EQ-5D-5L and QOLIE-31P was measured at baseline and 12 months follow-up. Spearman's correlations, effect sizes (EF) and standardized response means (SRM) were calculated for both the EQ-5D-5L and QOLIE-31P domains and sub scores. Mapping functions were derived using ordinary least square (OLS) and censored least absolute deviations models.

**Results** A total of 509 patients were included in this study. Low to moderately strong significant correlations were found between both instruments. The EQ-5D-5L showed high ceiling effects and small EFs and SRMs, whereas the QOLIE-31P did not show ceiling effects and also showed small to moderate EFs and SRMs. Results of the different mapping functions indicate that the highest adjusted  $R^2$  we were able to regress was 0.265 using an OLS model with squared terms, leading to a mean absolute error of 0.103. **Conclusions** Results presented in this study emphasize the shortcomings of the EQ-5D-5L in epilepsy and the importance of the development of condition-specific preference-based instruments which can be used within the QALY framework. In addition, the usefulness of the constructed mapping function in economic evaluations is questionable.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10198-017-0928-0) contains supplementary material, which is available to authorized users.

✉ Ben F. M. Wijnen  
b.wijnen@maastrichtuniversity.nl

<sup>1</sup> Department of Health Services Research, CAPHRI School of Public Health and Primary Care, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

<sup>2</sup> Department of Research and Development, Epilepsy Centre Kempenhaeghe, Heeze, The Netherlands

<sup>3</sup> King's Health Economics (KHE), Institute of Psychiatry, Psychology and Neuroscience at King's College London, London, UK

<sup>4</sup> Department of Neurology, Academic Centre for Epileptology, Epilepsy Centre Kempenhaeghe and Maastricht University Medical Centre, Maastricht, The Netherlands

<sup>5</sup> School of Mental Health and Neuroscience, Maastricht University Medical Center, Maastricht, The Netherlands

<sup>6</sup> School of Health Professions Education, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

<sup>7</sup> Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience at King's College London, London, UK

<sup>8</sup> Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands

<sup>9</sup> Duboisdomein 30, 6229 GT Maastricht, The Netherlands

**JEL Classification** D610 Allocative Efficiency · Cost-Benefit Analysis

## Introduction

Epilepsy is a disorder of the brain, characterized by recurrent seizures. Seizure episodes are a result of excessive electrical discharges in a group of brain cells. Different parts of the brain can be the site of such discharges. These discharges result in a variety of clinical manifestations, depending on where they occur in the brain. The clinical manifestations can vary from the briefest lapses of attention or muscle jerks to severe and prolonged convulsions [1].

In economic evaluation, both in general and in the field of epilepsy, the quality adjusted life year (QALY) is routinely used as a summary measure of health outcome for economic evaluations, which incorporates the impact on both the quantity and quality of life (QoL). For example, the use of QALYs is required by the National Institute for Health and Clinical Excellence (NICE) in England and Wales [2] and the Healthcare institute in the Netherlands [3] for an intervention to be reimbursed. The utility part of QALYs requires health state values as QALYs are calculated based on the time spent in a specific health state multiplied by the corresponding utility of that health state. Commonly used measures to include in the QALYs are generic utility measures, such as the EuroQol 5 dimensions 5 levels (EQ-5D-5L) [4, 5], Short Form 6 dimensions (SF-6D) [6] and the Health Utility Index (HUI) [7]. Generic utility instruments are designed to be applicable in a large variety of conditions.

However, there are instances, especially in clinical research, where some generic utility measures fail to capture changes that, even if small, are important to patients. Some studies use condition-specific or condition-specific utility measures to address this limitation. It is suggested that these instruments are likely to be more responsive than generic instruments, whose strengths include breadth and applicability across conditions and interventions [8]. The responsiveness of an instrument is likely to be dependent on several factors such as the nature of the condition and the domains included in the instrument. For example, the EuroQol-5D-3L (i.e. a generic utility instrument) has been demonstrated to correlate in a moderately to good way with criterion measures in patients with chronic low back pain [9]. In contrast, the EQ-5D-3L was deemed unsuitable for people with dementia, leading to the development of a condition-specific questionnaire (DEMQOL) [10]. In epilepsy, the EQ-5D-3L has been shown to correlate well with another generic quality of life instrument, the 15D-instrument [11]. However, in patients with newly diagnosed

focal epilepsy, the EuroQol-5D-3L was compared to an epilepsy-specific instrument (NEWQOL-6D) and was shown to be less responsive than the NEWQOL-6D [12]. Selai et al. [13] examined the use of the EQ-5D-3L in people with epilepsy and concluded that adaptation, seizures, and the stigma of epilepsy considerably impair quality of life but are not captured using the EQ-5D-3L, which limit its applicability [13].

Wiebe et al. [8] evaluated 43 randomized controlled trials which used generic and specific QoL instruments and concluded that specific instruments are more responsive than generic tools. Furthermore, they stress that investigators may come to misleading conclusions by using generic instruments. However, condition-specific measures lack cross-program comparability. Furthermore, if a condition-specific quality of life instrument were used for the calculation of QALYs, the valuation set should be constructed according to the same principles as generic utility measures (i.e. the multi attribute utility theory [14]), which is often not the case. An alternative option would be to derive well-conducted and validated mapping functions to map condition-specific outcomes to generic utilities. A mapping function is a regression equation used to predict values of, in this case, a generic utility instrument, using scores/values from a condition-specific instrument as regressors (also known as ‘cross-walking’) [15]. Albeit not resolving issues regarding insensitivity of generic instruments, mapping is a solution which enables health state utilities to be predicted when no preference-based measure has been included in the study [15, 16]. Such mapping functions are supposed to yield utility values comparable generic instruments [16]. However, the performance of a mapping function is dependent on and requires a degree of overlap between both measures and that the two measures are administered on the same population [15, 17]. The aim of this study is to compare the EQ-5D-5L and an often used condition-specific QoL instrument, the Quality of Life in Epilepsy-Patients-Weighted 31p (QOLIE-31P) [18]. The objective of this study is to investigate the correlation between and the responsiveness of the EQ-5D-5L and the QOLIE-31P in patients with epilepsy. In addition, we aim to develop a mapping function to predict EQ-5D-5L values based on the QOLIE-31P for use in economic evaluations.

## Methods

### QOLIE-31p

The QOLIE-31-P is a condition-specific QoL instrument which consists of 38 items assessing 7 domains of epilepsy: seizure worry, overall QOL, emotional well-being, energy-fatigue, cognitive functioning including memory,

medication effects, social functioning and an overall score. In addition, for each domain, questions regarding how much distress a person feels about problems and worries related to epilepsy are included. Each domain is scored on a scale ranging from 0 to 100. Afterwards a final score can be calculated using weights derived from the amount of distress related to each domain. The final score ranges from 0 to 100, in which higher values indicate a better QoL [19].

### EQ-5D-5L

The EQ-5D-5L is generally used as a generic QoL instrument which consists of five dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression, each of which can have one of five responses [4, 5] (e.g. no pain, slight pain, moderate pain, severe pain and extreme pain). This measure produces a possible 3125 distinct health states ranging from 11111 (full health) to 55555 (worst). The EQ-5D-5L was valued using both the Dutch and the UK tariffs [5, 20, 21].

### Data set

The SMILE study data [22, 23] and the ZMILE study data [24] were used for the analyses. Both studies examined the (cost-) effectiveness of a self-management program for patients with epilepsy. Follow-up data was available for 12 months in both studies. Inclusion criteria for both studies were similar (i.e. epilepsy diagnosis, prescribed antiepileptic drugs, no severe psychiatric disorders, being able to participate and benefit from group sessions). However, the SMILE study included patients from age  $\geq 16$  years whereas the ZMILE study included patients aged  $\geq 18$  years, and patients in the SMILE group were also screened to have had at least two seizures in the 12 months before inclusion. Patients with complete data for each of the measures across each time point were included.

For the direct response mapping, the data set was randomly split (using the “approximately 50% of the cases” function in SPSS) into two separate data sets: (1) the “estimation sample” ( $N = 283$ ), which was used to derive the mapping functions; (2) the “validation sample” ( $N = 224$ ) which was used to validate the mapping functions.

### Responsiveness analyses

Descriptive analyses are presented for patient characteristics. To measure concurrent validity (i.e. the strength of the relationship between measures of the same concept) Spearman’s correlation was calculated between the domains and total scores (i.e. utilities) of the EQ-5D-5L

and the QOLIE-31P. Spearman’s correlation was used due to the skewed nature of the data, especially EQ-5D-5L utilities. Strong correlations indicate that the preference-based measures are assessing related constructs. Correlations are considered weak if scores are less than 0.3, moderate if scores are between 0.3 and less than 0.7, and strong if scores are 0.7 or higher [25]. Bonferroni correction was applied to account for multiple testing (i.e. adjustment of  $p$ -values) [26].

To determine the predictive validity or responsiveness of both instruments (i.e. the ability of an instrument to detect relevant changes in QoL over time) the standardized response mean (SRM) and effect size (EF) were calculated. The SRM is a standard indicator of change across measures and time points and was calculated by  $SRM = (M1 - M2) / (SD1 - SD2)$ , where  $M1$  is the mean pre-assessment and  $M2$  is the mean post-assessment, and  $SD1$  and  $SD2$  are the standard deviations of both assessments [25]. SRMs of less than 0.2 are considered small, 0.5 moderate, and 0.8 large [25, 27]. The EF is calculated as the difference between follow-up and baseline divided by the standard deviation of the group’s baseline scores. The SRM and EF were calculated for those patients amongst whom a change in health state was observed between baseline and follow-up. Floor and ceiling effects were examined. For each questionnaire the proportion of respondents with a minimum score (referred to as ‘floor effects’) or a maximum score (referred to as ‘ceiling effects’) was calculated. If a large proportion of the population is at the floor (lowest possible score) or ceiling (highest possible score), then this impairs the ability of the measure to pick up decreases or increases in QoL, respectively [12]. The EF and the SRM are the most common measures for responsiveness. Positive values reflect (standardized) improvements in the number of standard deviations of the baseline scores (EF) or the score differences (SRM) (i.e. unit-free) [28].

### Mapping approach

To estimate EQ-5D-5L utilities based on the QOLIE-31P, direct response mapping was used to regress QOLIE-31P scores to EQ-5D-5L utilities. In direct mapping, a regression equation is used to predict the values of the EQ-5D-5L using scores/values from the QOLIE-31P as regressors. Next, the coefficients of the model are used to carry out the conversion from the source measure to the target measure in the required dataset [29]. Spearman’s correlations of the independent variables were used to determine whether there was collinearity between independent variables, which would then be removed from the analyses. A collinearity threshold of  $>0.70$  was used [30].

Ordinary least-squares (OLS) and censored least absolute deviations (CLAD) regression was used to estimate the

model. The OLS is the most commonly used model in mapping studies [15, 31]. However, it is unable to restrict for the range of values and may lead to implausible predicted values (e.g. EQ-5D values above 1). The CLAD model was therefore used as it has the ability to account for censored or bounded data. In addition, it is robust to heteroscedasticity and can also be used for skewed data [15, 32].

As the aim of this study was to derive a predictive model, all items (domains) were included in the model despite their significance level, which is often considered best practice [33–35].

Furthermore, no attempt was made to predict the individual EQ-5D-5L dimensions separately as this has been shown to be a less efficient strategy or to give similar results in terms of prediction [36].

For both the OLS and the CLAD model, specifications of the mapping functions were constructed as proposed by Brazier et al. (2010) [15]. We started with a simple additive model by predicting EQ-5D values from the total QOLIE-31P scores including age, gender, employment, and living arrangements (model 1). Next, the EQ-5D-5L values were predicted from the 7 QOLIE-31P dimension sub scores (model 2). To relax the assumptions of the simple additive model, squared terms for dimension sub scores were included in the model (model 3) [15]. As suggested by Brazier et al. (2010) only significant squared terms were included in the model to reduce the number of variables [15].

The predictive validity of the mapping models was assessed by using: (1) the goodness of fit as assessed using adjusted/pseudo R-squared (OLS and CLAD) in the estimation sample; and (2) the predictive performance of the models in the validation sample was assessed using the mean absolute error (MAE).

All analyses were done in STATA 15 (StataCorp, College Station, TX, USA).

## Results

The dataset consisted of 509 patients of which 102 patients were recruited for the ZMILE study and 407 for the SMILE study. In total, 53.0% of the patients were female and the majority of the patients were aged between 25 and 44 years old. Most of the patients had a household or lived with others (73.1%) and 51.3% of the patients were unemployed. Mean quality of life according to the EQ-5D-5L was 0.86 and mean condition-specific quality of life according to the QOLIE-31P was 65.82. More detail regarding the characteristics of the population(s) is reported in Table 1.

## Validity and responsiveness

An assessment of the strength of the relationship between the EQ-5D-5L and the QOLIE-31P based on Spearman's correlation coefficient, showed moderately strong significant correlations between both instruments for the total score (Table 2). Only a few statistically significant correlations were found between the sub scores of the QOLIE-31P and the sub scores of the EQ-5D-5L. All sub scores of the QOLIE-31P did significantly correlate with the total EQ-5D-5L scores. At baseline and 12 months follow-up ceiling effects on the EQ-5D-5L were substantial with 37.8 and 33.5% of the patients reporting the maximum score. No ceiling effects were found for the QOLIE-31P.

Details regarding the EF and SRM are presented in Table 3. EFs and SRMs all appear to be relatively small. Both the EF and SRM estimates are smaller for the EQ-5D-5L than the QOLIE-31P. For the EQ-5D-5L, values range from  $-0.017$  to  $0.043$  for the EF and from  $-0.023$  to  $0.025$  for the SRM which would be considered small. The EF and SRM values for the QOLIE-31P range from  $0.082$  to  $0.290$  (EF) and from  $0.07$  to  $0.270$  (SRM), which would be regarded as small to moderate.

## Mapping functions

The EQ-5D index scores had a somewhat bimodal distribution, and the distribution of the QOLIE-31P index scores were normally distributed (see Online Supplementary Materials 1). The inclusion of age was shown to have a significant effect on the prediction of EQ-5D-5L scores. All other demographic variables were excluded from the analyses. In addition, there was a significant effect associated with country (i.e. SMILE or ZMILE dataset). Hence, age and country were included in all mapping functions.

For the OLS mapping functions, model 3 performed best with an MAE of 0.103 and an adjusted  $R^2$  of 0.265. Inclusion of age significantly improved the model, hence a model without age was only constructed for OLS. For the CLAD mapping functions, CLAD model 3 performed best with a MAE of 0.097 and a pseudo  $R^2$  of 0.160. It should be noted, however, that including squared terms only improved adjusted/pseudo  $R^2$  values and only marginally improved MAE in the estimation sample (see Table 4). All models predicted values above 1 (full health), of which OLS model 2 was closest to 1 with maximum values of 1.020. A graphical representation of the model fits is presented in Fig. 1.

When assessing the MAE in the validation sample, using the mapping functions derived from the estimation sample, OLS model 1 performed best with a MAE of 0.114. Likewise, for the CLAD mapping functions, model 1 performed best with a MAE value of 0.109 (see Table 5).

**Table 1** (Baseline) characteristics of the population

Characteristics	ZMILE sample ( <i>n</i> = 102)	SMILE sample ( <i>n</i> = 407)	Total ( <i>n</i> = 509)
Gender			
Male	50 (49.0%)	185 (45.5%)	235 (46.2%)
Female	52 (51.0%)	219 (53.8%)	271 (53.2%)
Missing values	0	3	
Age in years			
16–24	17 (16.7%)	46 (11.3%)	63 (12.4%)
25–44	42 (41.2%)	194 (47.7%)	236 (46.4%)
45–64	37 (36.2%)	142 (34.9%)	179 (35.2%)
≥65	6 (5.9%)	22 (5.4%)	28 (5.5%)
Missing values	0	3	
Living arrangements, <i>n</i> (%)			
Household/living with others	67 (65.7%)	305 (74.9%)	372 (73.1%)
Living alone	26 (25.5%)	95 (23.4%)	121 (23.8%)
Other arrangements	4 (3.9%)	4 (1%)	8 (1.6%)
Missing values	5	3	
Employment			
Not unemployed	54 (52.9%)	207 (50.9%)	261 (51.3%)
Specifically employed or student	44 (43.1%)	197 (48.4%)	241 (47.4%)
Missing values	4	3	
Quality of life			
EQ-5D-5L baseline	0.83	0.87	0.86
EQ VAS	74.77	67.00	68.53
QOLIE-31P baseline	64.74	66.05	65.82

**Table 2** Spearman's correlation coefficients between QOLIE-31P values and EQ-5D-5L values

	EQ-5D mobility	EQ-5D self- care	EQ-5D usual activities	EQ-5D pain	EQ-5D anxiety and depression	Total EQ-5D score
QOLIE-31P energy	−0.2330	−0.1896	−0.1991	−0.2689*	−0.2293	0.4499*
QOLIE-31P mood	−0.1252	−0.0285	−0.1725	−0.1743	−0.3823*	0.4881*
QOLIE-31P daily activities	−0.2350	−0.2138	−0.2557*	−0.2644*	−0.2422*	0.4335*
QOLIE-31P cognition	−0.1041	−0.0607	−0.1906	−0.1661	−0.2539*	0.2948*
QOLIE-31P medication effects	−0.0569	−0.1061	−0.1744	−0.1969	−0.2304	0.2730*
QOLIE-31P seizure worry	−0.1099	−0.0802	−0.1502	−0.1901	−0.3057*	0.3906*
QOLIE-31P overall QOL	−0.1794	−0.1620	−0.2245	−0.1808	−0.3142*	0.4049*
Total QOLIE-31P score	−0.2000	−0.1476	−0.2855*	−0.3171*	−0.3354*	0.5653*

\* Significant correlation at 5% level

Given the higher adjusted/pseudo  $R^2$  values of model 3 compared to model 1 for both OLS and CLAD, and given the lower MAE for OLS model 3 compared to CLAD model 3 in the validation sample, the best mapping function would be OLS model 3. The regression coefficients for this model are presented in Online Supplementary Materials 1.

## Discussion

The aim of this study was to compare the responsiveness of the QOLIE-31P and the EQ-5D-5L in epilepsy and to predict EQ-5D-5L values based on QOLIE-31P scores with the development of a mapping function. Looking at concurrent validity, although the correlations were significant,

**Table 3** Standardized response means for QOLIE-31P and EQ-5D-5L

	Mean difference between BS and FU12M	Effect size	Standardized response mean
Total EQ-5D-5L score	-0.004	-0.017	-0.023
EQ-5D mobility	0.276	0.010	0.003
EQ-5D self-care	0.305	0.010	0.003
EQ-5D usual activities	2.600	0.043	0.025
EQ-5D pain	0.619	0.012	0.006
EQ-5D anxiety	-1.847	-0.017	-0.019
Total QOLIE-31P score	2.414	0.187	0.212
QOLIE-31P energy	1.875	0.110	0.080
QOLIE-31P mood	1.906	0.082	0.070
QOLIE-31P daily activities	8.317	0.284	0.270
QOLIE-31P cognition	3.196	0.131	0.125
QOLIE-31P medication effects	5.878	0.177	0.178
QOLIE-31P seizure worry	7.696	0.290	0.254
QOLIE-31P overall QOL	2.963	0.128	0.116

EQ-5D-5L EuroQol 5 dimensions with 5 levels, BS baseline, FU12M follow-up measurement at 12 months

**Table 4** Summary of observed and predicted values for all models in estimation dataset ( $N = 283$ )

	Observed EQ-5D utility	Predicted EQ-5D utilities			
		Total QOLIE-31P scores	Total QOLIE-31P and age and country	QOLIE-31P domain scores and age and country	Domain scores and squared terms and age and country
		OLS model 0	OLS model 1	OLS model 2	OLS model 3
Mean	0.867	0.868	0.868	0.871	0.872
Minimum	0.055	0.668	0.673	0.690	0.647
Maximum	1.000	1.026	1.020	1.082	1.086
MAE		0.110	0.107	0.103	0.103
Adjusted $R^2$		0.151	0.180	0.211	0.265
			CLAD model 1	CLAD model 2	CLAD model 3
Mean	0.867	-	0.915	0.916	0.920
Minimum	0.055	-	0.711	0.697	0.698
Maximum	1.000	-	1.072	1.15	1.119
MAE		-	0.099	0.099	0.097
Pseudo $R^2$		-	0.116	0.129	0.160

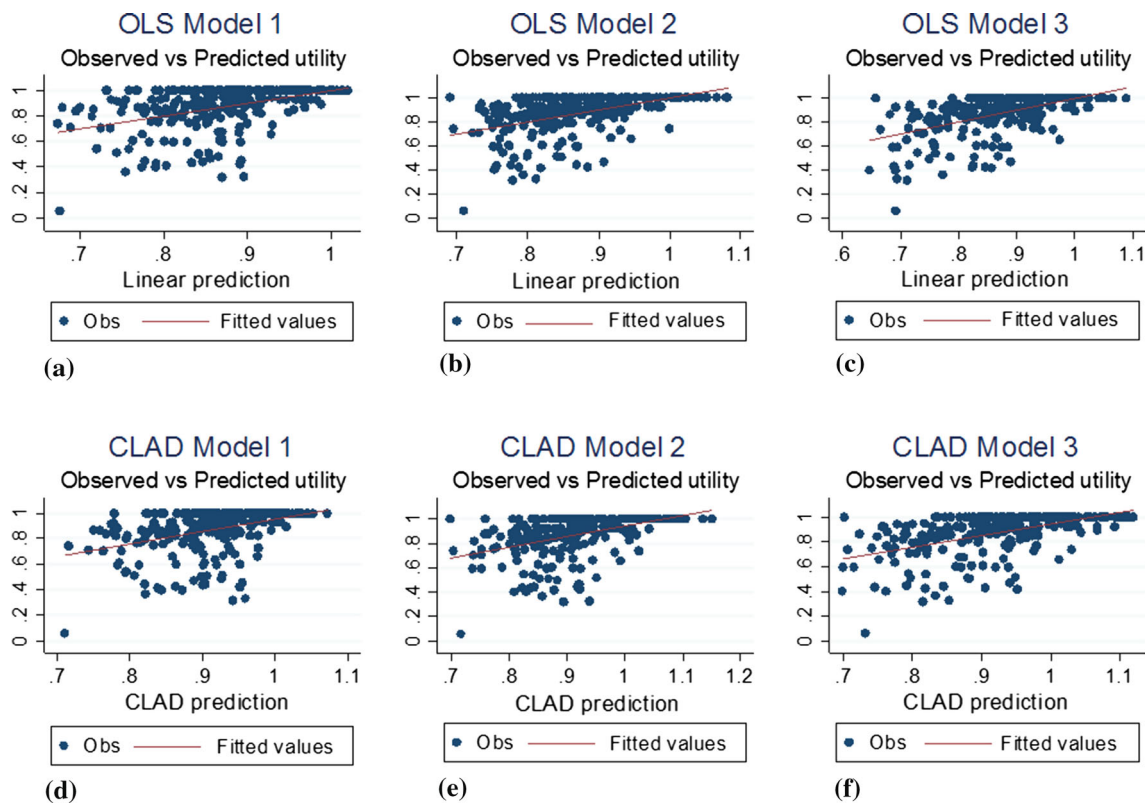
MAE mean absolute error, OLS ordinary least squares, CLAD censored least absolute deviations

the strength was only moderate between both instruments when looking at the total score. This may imply that both instruments are only measuring the same concept to some extent and impose conceptual differences.

The EQ-5D-5L showed substantially high ceiling effects and was demonstrated to have rather small EFs and SRMs, whereas the QOLIE-31P did not show ceiling effects and had small to moderate EFs and SRMs. This indicates that the QOLIE-31P has an overall higher responsiveness based on EF, SRM and ceiling effects. Furthermore, this study provides a mapping function which can be used in (future) economic evaluations to map QOLIE-31P data to EQ-5D-5L values.

The relatively small EFs and SRMs may be explained by a lack of responsiveness of both instruments, especially the EQ-5D-5L. However, part of these small estimates can be explained by the fact that the patients' health state did not change much over time. The EF and SRM were calculated in all patients who had any change during follow-up, due to the lack of a known clinically meaningful difference for both instruments; this led to an underestimation of the EF and SRM. The EQ-5D-5L, however, performed substantially worse than the QOLIE-31P (i.e. lower estimated EFs and SRMs).

Using different mapping functions, the highest adjusted  $R^2$  we were able to regress was 0.265 using an OLS model



**Fig. 1** Scatter plots comparing observed vs predicted EQ-5D-5L values for OLS (a–c), CLAD (d–f)

**Table 5** Summary of observed and predicted values for all models in validation dataset ( $N = 224$ )

	Observed EQ-5D utility	Predicted EQ-5D utilities		
		Total QOLIE-31P and age and country	QOLIE-31P domain scores and age and country	Domain scores and squared terms and age and country
		OLS model 1	OLS model 2	OLS model 3
Mean	0.863	0.865	0.865	0.865
Minimum	0.054	0.649	0.685	0.629
Maximum	1.000	1.039	1.100	1.084
MAE		0.114	0.116	0.118
		CLAD model 1	CLAD model 2	CLAD model 3
Mean	0.863	0.912	0.907	0.911
Minimum	0.054	0.682	0.688	0.699
Maximum	1.000	1.062	1.136	1.120
MAE		0.109	0.117	0.119

with squared terms, which led to a MAE of 0.103. Overall, this model performed best given the results within the estimation and validation sample. Although theoretically preferred, the use of a CLAD model did not perform better than the OLS model, especially in the validation sample. Mixed results have been reported in studies using CLAD models [17], with some concluding that CLAD improved the model fit [37, 38] and others concluding that the improvement of CLAD over OLS was small or did not

have an impact [39]. The adjusted/pseudo  $R^2$  values found in this study were relatively small, which is not uncommon. In a review of Brazier et al. (2010), it was found that models mapping a generic instrument onto a generic preference-based measure achieved an adjusted  $R^2$  of more than 0.5 within sample. However, in studies examining the fit of functions mapping from condition-specific to generic measures, results were more variable ranging from 0.17 to 0.51 [15]. In addition, errors were often larger for models

mapping a generic measure onto a generic preference-based measure [15, 17]. Likewise, the usefulness of our mapping function in economic evaluations is questionable given the relatively large mean absolute error and poor model fit.

Another way of mapping would be to use a model to predict responses of each of the five dimensions of the EQ-5D-5L from the QOLIE-31P (sub) scores; so-called indirect response mapping or response mapping models [40]. As the purpose of the mapping part of this study was to derive a regression function, this method was not applied. In addition, as mentioned above, indirect response mapping has been shown to be a less efficient strategy or to give similar results in terms of prediction [36].

The use of mapping to derive EQ-5D-5L values is fundamentally limited by the degree of overlap between two instruments [17]. Although several studies reported limitations with generic preference-based quality of life instruments regarding their responsiveness and ability to discriminate between health states (e.g. McTaggart-Cowan et al. [41]), the use of generic preference-based instruments is mandatory in most national guidelines for pharmacoeconomic evaluations, for example in the UK and the Netherlands [3, 42]. However, given the limited responsiveness, low correlations, and the poor model fit of the mapping functions it may be argued that there is a need for the development of condition-specific preference-based measures for patients with epilepsy. General (non-preference based) condition-specific instruments, such as the QOLIE-31P, are an important source of evidence; however, their use in economic evaluation is severely limited because they were not designed for this purpose and, unless they are preference-based, they theoretically cannot be used to calculate quality adjusted life years (QALYs) [14, 43]. Several attempts have been made to derive condition-specific quality of life instruments to calculate (condition-specific) QALYs, such as the development of new instruments [44, 45] or the development of condition-specific preference-based measures from existing instruments [46, 47]. Of course, such a measure could not be the sole outcome of interest for economic evaluations, as they lack the comparability between conditions, a distinct advantage of generic instruments [48]. We agree with Brazier et al. (2010), that development of a condition-specific preference-based instrument should not be seen as an alternative to generic preference-based measures, but rather as a supplement [43]. Condition-specific preference-based instruments may have an important role in ensuring that the benefits of health-care interventions are adequately reflected in QALY estimates for economic evaluations in all conditions [49].

This study is subject to several limitations. First, we only investigated the use of a few mapping models,

whereas a wide variety of models exist, such as GLM or Tobit models. Furthermore, other correlation coefficients may have been used, such as polychoric correlation coefficients. However, given the marginal differences between the models used in this study, the model fit is not likely to be improved substantially. In addition, structural equation modelling could be used to analyze the structural relationship between EQ-5D-5L and QOLIE-31P and latent constructs. Second, our estimations are based on a pooled data set containing data from both the UK and the Netherlands. Although inclusion criteria for both studies were similar and there was no significant difference between the countries regarding regression estimates, this may have introduced extra heterogeneity within the data. Lastly, the pooled dataset was divided into an estimation sample and a validation sample. This has the advantage that it assesses the mapping function by its prime purpose; however, it reduces the sample size of the estimation sample. The use of the whole sample for the model estimations, however, did not substantially improve the model(s).

## Conclusion

There was a low to moderate correlation between the sub scores and total scores of the EQ-5D-5L and the QOLIE-31P. Both the EF and SMRs were relatively low, especially for the EQ-5D-5L. Mapping functions to regress QOLIE-31P values to EQ-5D-5L values did not show an optimal fit with relatively low adjusted  $R^2$  values. The results presented in this study may emphasize the importance of the development of condition-specific preference-based instruments which can be used within the QALY framework and hence be incorporated as an important supplement in economic evaluations. The development of such condition-specific preference-based quality of instruments can ensure that the benefits of health-care interventions are adequately reflected in QALY estimates for economic evaluations not only in epilepsy but for all conditions.

**Acknowledgements** The views and opinions expressed are those of the authors and do not necessarily reflect those of the HTA programme, the UK NIHR, NHS or the Department of Health. We would like to thank Dr. M. Dzingina (King's College London) for his help with the analyses of the results.

## Compliance with ethical standards

**Funding** This study was funded by the Netherlands Organization for Health Research and Development (ZonMw), Grant application no. 836011018. Data collection in the SMILE trial was funded by the National Institute for Health Research [Health Technology Assessment (HTA) programme] (Project no. 09/165/01).



**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- World Health Organization. Epilepsy. (2016). <http://www.who.int/mediacentre/factsheets/fs999/en/>
- Whitehead, S.J., Ali, S.: Health outcomes in economic evaluation: the QALY and utilities. *Br. Med. Bull.* **96**(1), 5–21 (2010). doi:[10.1093/bmb/ldq033](https://doi.org/10.1093/bmb/ldq033)
- Nederland, Z.: Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Zorginstituut Nederland, Dieren (2015)
- Brooks, R.: EuroQol: the current state of play. *Health policy (Amsterdam, Netherlands)*. **37**(1), 53–72 (1996)
- Devlin N, Shah K, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England (2016)
- Brazier, J., Roberts, J., Deverill, M.: The estimation of a preference-based measure of health from the SF-36. *J. Health Econ.* **21**(2), 271–292 (2002). doi:[10.1016/S0167-6296\(01\)00130-8](https://doi.org/10.1016/S0167-6296(01)00130-8)
- Horsman, J., Furlong, W., Feeny, D., Torrance, G.: The health utilities index (HUI®): concepts, measurement properties and applications. *Health Qual. Life Outcomes*. **1**, 54 (2003). doi:[10.1186/1477-7525-1-54](https://doi.org/10.1186/1477-7525-1-54)
- Wiebe, S., Guyatt, G., Weaver, B., Matijevic, S., Sidwell, C.: Comparative responsiveness of generic and specific quality-of-life instruments. *J. Clin. Epidemiol.* **56**(1), 52–60 (2003)
- Soer, R., Reneman, M.F., Speijer, B.L., Coppes, M.H., Vroomen, P.C.: Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. *Spine J. Off. J. North Am. Spine Soc.* **12**(11), 1035–1039 (2012). doi:[10.1016/j.spinee.2012.10.030](https://doi.org/10.1016/j.spinee.2012.10.030)
- Smith, S.C., Lamping, D.L., Banerjee, S., Harwood, R., Foley, B., Smith, P. et al.: Measurement of health-related quality of life for people with dementia: development of a new instrument (DEM-QOL) and an evaluation of current methodology. *Health Technol. Assess. (Winchester, England)*. **9**(10), 1–93, iii–iv (2005)
- Stavem, K., Bjørnæs, H., Lossius, M.I.: Properties of the 15D and EQ-5D utility measures in a community sample of people with epilepsy. *Epilepsy Res.* **44**(2–3), 179–189 (2001). doi:[10.1016/S0920-1211\(01\)00201-7](https://doi.org/10.1016/S0920-1211(01)00201-7)
- Mulhern, B., Pink, J., Rowen, D., Borghs, S., Butt, T., Hughes, D., et al.: Comparing generic and condition-specific preference-based measures in epilepsy: EQ-5D-3L and NEWQOL-6D. *Val. Health.* (2016). doi:[10.1016/j.jval.2016.03.1860](https://doi.org/10.1016/j.jval.2016.03.1860)
- Selai, C.E., Elstner, K., Trimble, M.R.: Quality of life pre and post epilepsy surgery. *Epilepsy Res.* **38**(1), 67–74 (2000)
- Von Neumann, J., Morgenstern, O.: *Theory of games and economic behavior*. Princeton university press, Princeton (2007)
- Brazier, J.E., Yang, Y., Tsuchiya, A., Rowen, D.L.: A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur. J. Health Econ.* **11**(2), 215–225 (2010). doi:[10.1007/s10198-009-0168-z](https://doi.org/10.1007/s10198-009-0168-z)
- Versteegh, M.M., Leunis, A., Uyl-de Groot, C.A., Stolk, E.A.: Condition-specific preference-based measures: benefit or burden? *Val. Health.* **15**(3), 504–513 (2012)
- Longworth, L., Rowen, D.: Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Val. Health J. Int. Soc. Pharmacoecon. Outcomes Res.* **16**(1), 202–210 (2013). doi:[10.1016/j.jval.2012.10.010](https://doi.org/10.1016/j.jval.2012.10.010)
- Cramer, J.A., Perrine, K., Devinsky, O., Bryant-Comstock, L., Meador, K., Hermann, B.: Development and cross-cultural translations of a 31-item quality of life in epilepsy inventory. *Epilepsia*. **39**(1), 81–88 (1998)
- Cramer, J.A., Van Hamme, G.: Maintenance of improvement in health-related quality of life during long-term treatment with levetiracetam. *Epilepsy Behav. E&B.* **4**(2), 118–123 (2003)
- Versteegh, M.M., Vermeulen, K.M., Evers, S.M.A.A., de Wit, G.A., Prenger, R., Stolk, E.A.: Dutch tariff for the five-level version of EQ-5D. *Val. Health.* **19**(4), 343–352 (2016). doi:[10.1016/j.jval.2016.01.003](https://doi.org/10.1016/j.jval.2016.01.003)
- Feng, Y., Devlin, N., Koonal, S., Brendan, M., van Hout, B.: New methods for modelling EQ-5D-5L value sets: an application to English data. Office of Health Economics Research (2016)
- Kralj-Hans, I., Goldstein, L.H., Noble, A.J., Landau, S., Magill, N., McCrone, P., et al.: Self-management education for adults with poorly controlled epilepsy (SMILE (UK)): a randomised controlled trial protocol. *BMC Neurol.* **14**, 69 (2014). doi:[10.1186/1471-2377-14-69](https://doi.org/10.1186/1471-2377-14-69)
- Magill, N., Ridsdale, L., Goldstein, L.H., McCrone, P., Morgan, M., Noble, A.J., et al.: Self-management education for adults with poorly controlled epilepsy (SMILE (UK)): statistical, economic and qualitative analysis plan for a randomised controlled trial. *Trials*. **16**, 269 (2015). doi:[10.1186/s13063-015-0788-9](https://doi.org/10.1186/s13063-015-0788-9)
- Leenen, L.A., Wijnen, B.F., de Kinderen, R.J., Majoie, M.H., van Heugten, C.M., Evers, S.M.: Cost-effectiveness of a multi-component intervention for adults with epilepsy: study protocol of a Dutch randomized controlled trial (ZMILE study). *BMC Neurol.* **14**, 255 (2014). doi:[10.1186/s12883-014-0255-3](https://doi.org/10.1186/s12883-014-0255-3)
- Cohen, J.: *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates, Hillsdale (1988)
- Rosner, B.: Multiple comparisons: Bonferroni approach. *Fund. Biostat* **5** (2000)
- Kazis, L.E., Anderson, J.J., Meenan, R.F.: Effect sizes for interpreting changes in health status. *Med. Care* **27**(3), S178–S189 (1989)
- Angst, F., Verra, M.L., Lehmann, S., Aeschlimann, A.: Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. *BMC Med. Res. Methodol.* **8**(1), 26 (2008). doi:[10.1186/1471-2288-8-26](https://doi.org/10.1186/1471-2288-8-26)
- Khan, K.A., Petrou, S., Rivero-Arias, O., Walters, S.J., Boyle, S.E.: Mapping EQ-5D utility scores from the PedsQL™ generic core scales. *Pharmacoeconomics*. **32**(7), 693–706 (2014). doi:[10.1007/s40273-014-0153-y](https://doi.org/10.1007/s40273-014-0153-y)
- Tabachnick, B.G., Fidell, L.S., Osterlind, S.J.: *Using multivariate statistics* (2001)
- Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernandez Alava, M., Mukuria, C., 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. (Winchester, England)* **18**(9), 1–224 (2014). doi:[10.3310/hta18090](https://doi.org/10.3310/hta18090)
- Teckle, P., McTaggart-Cowan, H., Van der Hoek, K., Chia, S., Melosky, B., Gelmon, K., et al.: Mapping the FACT-G cancer-specific quality of life instrument to the EQ-5D and SF-6D. *Health Qual. Life Outcomes*. **11**(1), 203 (2013). doi:[10.1186/1477-7525-11-203](https://doi.org/10.1186/1477-7525-11-203)
- Wu, S., Harris, T.J., McAuley, K.B.: The use of simplified or misspecified models: linear case. *Can. J. Chem. Eng.* **85**(4), 386–398 (2007). doi:[10.1002/cjce.5450850401](https://doi.org/10.1002/cjce.5450850401)
- Shmueli, G.: To explain or to predict? *Stat. Sci.* 289–310 (2010)

35. Greenberg, E., Parks, R.P.: A predictive approach to model selection and multicollinearity. *J. Appl. Econ.* **12**(1), 67–75 (1997)
36. Crott, R., Briggs, A.: Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *Eur. J. Health Econ.: HEPAC Health Econ. Prevent. Care.* **11**(4), 427–434 (2010). doi:[10.1007/s10198-010-0233-7](https://doi.org/10.1007/s10198-010-0233-7)
37. Sullivan, P.W., Ghushchyan, V.: Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. *Med. Decis. Making Int. J. Soc. Med. Decis. Making.* **26**(4), 401–409 (2006). doi:[10.1177/0272989x06290496](https://doi.org/10.1177/0272989x06290496)
38. Payakachat, N., Summers, K.H., Pleil, A.M., Murawski, M.M., Thomas 3rd, J., Jennings, K., et al.: Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. *Qual. Life Res. Int. J. Qual. Life Aspects Treatment Care Rehabil.* **18**(7), 801–813 (2009). doi:[10.1007/s11136-009-9499-6](https://doi.org/10.1007/s11136-009-9499-6)
39. Kontodimopoulos, N., Aletras, V.H., Paliouras, D., Niakas, D.: Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. *Val. Health J. Int. Soc. Pharmacoecon. Outcomes Res.* **12**(8), 1151–1157 (2009). doi:[10.1111/j.1524-4733.2009.00569.x](https://doi.org/10.1111/j.1524-4733.2009.00569.x)
40. Dakin, H.: Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health Qual. Life Outcomes.* **11**, 151 (2013). doi:[10.1186/1477-7525-11-151](https://doi.org/10.1186/1477-7525-11-151)
41. McTaggart-Cowan, H.M., Marra, C.A., Yang, Y., Brazier, J.E., Kopec, J.A., FitzGerald, J.M., et al.: The validity of generic and condition-specific preference-based instruments: the ability to discriminate asthma control status. *Qual. Life Res.* **17**(3), 453–462 (2008). doi:[10.1007/s11136-008-9309-6](https://doi.org/10.1007/s11136-008-9309-6)
42. Williams, I., McIver, S., Moore, D., Bryan, S.: The use of economic evaluations in NHS decision-making: a review and empirical investigation. *Health Technol. Assess. (Winchester, England).* **12**(7), iii, ix–x, 1–175 (2008)
43. Brazier, J., Tsuchiya, A.: Preference-based condition-specific measures of health: what happens to cross programme comparability? *Health Econ.* **19**(2), 125–129 (2010). doi:[10.1002/hec.1580](https://doi.org/10.1002/hec.1580)
44. Torrance, G., Keresteci, M., Casey, R., Rosner, A., Ryan, N., Breton, M.: Development and initial validation of a new preference-based disease-specific health-related quality of life instrument for erectile function. *Qual. Life Res.* **13**(2), 349–359 (2004)
45. Revicki, D.A., Leidy, N.K., Brennan-Diemer, F., Sorensen, S., Toggias, A.: Integrating patient preferences into health outcomes assessment: the multiattribute asthma symptom utility index. *Chest* **114**(4), 998–1007 (1998)
46. Brazier, J., Czoski-Murray, C., Roberts, J., Brown, M., Symonds, T., Kelleher, C.: Estimation of a preference-based index from a condition-specific measure: the King's health questionnaire. *Med. Decis. Making Int. J. Soc. Med. Decis. Making.* **28**(1), 113–126 (2008). doi:[10.1177/0272989x07301820](https://doi.org/10.1177/0272989x07301820)
47. Yang, Y., Tsuchiya, A., Brazier, J., Young, Y. (eds.): Deriving a preference-based measure for health from the AQLQ. Health Economists Studying Group Meeting (HESG), London (2006)
48. Donaldson, C., Atkinson, A., Bond, J., Wright, K.: Should QALYs be programme-specific? *J. Health Econ.* **7**(3), 239–257 (1988). doi:[10.1016/0167-6296\(88\)90027-6](https://doi.org/10.1016/0167-6296(88)90027-6)
49. Brazier, J.E., Rowen, D., Mavranouzouli, I., Tsuchiya, A., Young, T., Yang, Y., et al.: Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol. Assess. (Winchester, England).* **16**(32), 1–114 (2012). doi:[10.3310/hta16320](https://doi.org/10.3310/hta16320)