

RESEARCH

Open Access



# Predicting EQ-5D-5L crosswalk from the PROMIS-29 profile for the United Kingdom, France, and Germany

Christoph Paul Klapproth<sup>1\*</sup> , J. van Bebber<sup>1</sup>, C. J. Sidey-Gibbons<sup>2</sup>, J. M. Valderas<sup>3,4</sup>, A. Leplege<sup>5,6</sup>, M. Rose<sup>1,7</sup> and F. Fischer<sup>1</sup>

## Abstract

**Background:** EQ-5D health state utilities (HSU) are commonly used in health economics to compute quality-adjusted life years (QALYs). The EQ-5D, which is country-specific, can be derived directly or by mapping from self-reported health-related quality of life (HRQoL) scales such as the PROMIS-29 profile. The PROMIS-29 from the Patient Reported Outcome Measures Information System is a comprehensive assessment of self-reported health with excellent psychometric properties. We sought to find optimal models predicting the EQ-5D-5L crosswalk from the PROMIS-29 in the United Kingdom, France, and Germany and compared the prediction performances with that of a US model.

**Methods:** We collected EQ-5D-5L and PROMIS-29 profiles and three samples representative of the general populations in the UK (n = 1509), France (n = 1501), and Germany (n = 1502). We used stepwise regression with backward selection to find the best models to predict the EQ-5D-5L crosswalk from all seven PROMIS-29 domains. We investigated the agreement between the observed and predicted EQ-5D-5L crosswalk in all three countries using various indices for the prediction performance, including Bland–Altman plots to examine the performance along the HSU continuum.

**Results:** The EQ-5D-5L crosswalk was best predicted in France ( $nRMSE_{FRA} = 0.075$ ,  $nMAE_{FRA} = 0.052$ ), followed by the UK ( $nRMSE_{UK} = 0.076$ ,  $nMAE_{UK} = 0.053$ ) and Germany ( $nRMSE_{GER} = 0.079$ ,  $nMAE_{GER} = 0.051$ ). The Bland–Altman plots show that the inclusion of higher-order effects reduced the overprediction of low HSU scores.

**Conclusions:** Our models provide a valid method to predict the EQ-5D-5L crosswalk from the PROMIS-29 for the UK, France, and Germany.

## Key points

- We provide mapping from PROMIS-29 profile to EQ-5D-5L crosswalk in the United Kingdom, France, and Germany.
- Due to the country specificity of health state utility, mapping algorithms for health state utility should not be generalized across countries.
- The application of polynomial regression models that account for non-linearity improves the prediction performance, in particular for poorer health states.
- The application of foreign models should be avoided.

\*Correspondence: christoph-paul.klapproth@charite.de

<sup>1</sup> Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité - Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany

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



## Background

Quality-adjusted life years (QALYs) are routinely used in cost-utility analyses (CUA) to evaluate the economic effectiveness of health care innovations or interventions [1]. QALYs are of particular importance in health technology assessments (HTAs) [2]. For example, the National Institute of Health and Clinical Excellence (NICE) in England and Wales has endorsed QALYs to compare health care interventions from an economic perspective [1]. In light of budget constraints in publicly funded health care systems, QALYs serve as a benchmark for the allocation of scarce resources in a way that maximizes utility to individuals and to society [2].

A QALY is defined as the product of the number of life years and a health state utility (HSU) score that represents the value of a particular health state. HSU values can at best achieve a value of 1 (full health). A value of 0 is considered dead and health states with a negative value are considered worse than dead. Individual HSU scores are patient-reported, generic, preference-based measures of health-related quality of life (HRQoL) [3]. The most frequently used generic HRQoL measure is the EuroQoL EQ-5D-5L crosswalk differentiating 3125 (i.e.,  $5^5$ ) health states. The EQ-5D-5L crosswalk is the default HSU score for economic evaluations demanded by HTA agencies such as NICE [4–7].

The Patient Reported Outcome Measurement Information System (PROMIS), on the other hand, is increasingly used internationally to measure clinical and condition-specific, non-preference HRQoL for its favourable psychometric properties: high validity, high reliability, high precision, and flexible administration [8, 9]. PROMIS is a common metric for a large variety of different health domains, aiming at comprehensive assessment, standardization and integration of different measures and items. It constitutes a collection of generic and condition-specific, non-preference-based patient reported outcome measures (PROMs) that have been developed using item response theory (IRT) [10]. For each PROM, so-called item banks have been developed comprising items that are highly informative regarding the PROM to be measured and that do not function substantially different across the most prominent demographic groups (e.g., women and men) [11, 12]. These item banks can be used to develop tailored short forms or for computerized adaptive testing (CAT) [13]. PROMIS overcomes significant limitations of legacy instruments such as ceiling effects and is, being translated to many languages and showing invariance to nationality, becoming the international reference measurement approach to PROMs [9, 14–16].

For economic evaluations, the preference-based EQ-5D-5L crosswalk is best obtained directly using the

EQ-5D-5L questionnaire. If direct assessment is not available, a common strategy is to estimate HSU scores by using a mapping algorithm from a non-preference-based PROM such as PROMIS [14, 17–20]. Little consensus exists on which mapping method is the most appropriate. In a recent systematic review, 147 studies mapping the EQ-5D were identified [17]. In more than 75% ordinary least squares (OLS) linear regression was used. Although OLS linear regression showed robust results compared to alternative methods, it has several drawbacks [21, 22]: First, predicted HSU scores may fall outside the possible range of the metric (i.e., values greater than one). Second, the relationship between non-preference-based PROM and HSU might be non-linear, meaning that the impact of health domains differs across the HSU continuum [22].

As PROMIS is increasingly used in clinical, non-preference HRQoL measurement and the EQ-5D-5L crosswalk is the required HSU for economic evaluations, developing a mapping between these two would open the perspective to use PROMIS for economic evaluations. As both are multidimensional generic HRQoL measures covering similar dimensions or domains (EQ-5D mobility and EQ-5D self-care versus PROMIS physical function, EQ-5D pain/discomfort versus PROMIS pain interference, EQ-5D anxiety/depression versus PROMIS anxiety or PROMIS depression, EQ-5D usual activities versus PROMIS ability to participate in social roles and activities), we can reasonably assume conceptual overlap, as previous mappings have as well [19, 20].

Mapping PROMIS to EQ-5D-5L crosswalk also opens a perspective for the use of other PROMs in economic evaluations: Because of its invariance property, PROMIS domains can also be measured using items from a different condition-specific measure that is anchored to the PROMIS metrics. For example, items from self-reported anxiety measured by MASQ, PANAS and GAD-7 are anchored on the PROMIS Anxiety metric [23]. Items from the BDI-2, CES-D, and PHQ-9, measuring depression, are anchored on the PROMIS Depression metric [24]. Therefore, mapping from PROMIS T-scores to EQ-5D-5L crosswalk enables the mapping of a broad range of PROMs to the EQ-5D-5L crosswalk via PROMIS.

Using OLS linear regression on US data, Revicki (2009) estimated a model to predict the former EQ-5D version, the EQ-5D-3L index value, from five PROMIS T-scores [19]: physical function, fatigue, pain interference, anxiety, and depression. For this PROMIS domain model, Revicki reports that approximately 57% (adjusted  $R^2$ ) of the variance in EQ-5D-3L index value can be explained by the variables in the model, and the intraclass correlation coefficient (ICC) is 0.73. Furthermore, 95% of all the residuals are between  $-0.20$  (2.5%) and  $0.15$  (97.5%). The relatively small width of these so-called empirical limits

of agreement (LoA) is indicative of an appropriate fitted model. However, Revicki also reported that the model does not work very well for low levels of health (EQ-5D-3L index value < 0.40). Revicki used the EQ-5D-3L questionnaire and applied the US EQ-5D-3L value set by Shaw (2005) [25]. As health preferences differ between countries, the EQ-5D-3L index value is country-specific [26, 27]. Revicki's model can therefore only be used to predict the EQ-5D-3L index value from PROMIS in the US.

Therefore, the primary aim of this study is to develop mapping functions from PROMIS-29 to the EQ-5D-5L crosswalk for the UK, France, and Germany so that PROMIS can be used for economic valuations in these countries. For each health domain, we explored the form of its relationship with the EQ-5D-5L crosswalk and examined whether these relationships would be the same across the three countries under investigation. Also, we aimed at improving prediction performance by including higher order coefficients. Furthermore, we investigated whether the optimal models would be structurally equivalent across countries and compared prediction performance of our models to Revicki's model.

## Methods

### Samples

Data were collected online by an independent polling company (Ipsos) in April and May 2015. Quota sampling was employed to obtain samples representative of the general population with respect to sex, age, occupation, region, and population density of the UK ( $n = 1509$ ), France ( $n = 1501$ ), and Germany ( $n = 1502$ ). Sample weights were calculated using the random iterative method (RIM) to match the latest data available in each country (census 2011 for the UK and Germany, census 2012 for France).

Participation in our general population samples was voluntary and data protection laws obeyed by Ipsos. If a respondent chose to drop out at some point, the data given until that point was not included. As skipping items was not possible, there were no missing data.

### Measures

#### *PROMIS domains and item banks*

We used the PROMIS-29 v2.0 Profile to assess seven core domains of health, each assessed with four items: physical function, fatigue, pain interference, anxiety, depression, sleep disturbance, and the ability to participate in social roles and activities (referred to as participation in the remainder of this article) plus the visual analogue scale (VAS) expressing pain intensity on a scale ranging from 0 to 10 [28]. PROMIS-29 has, compared to other short forms, enough items to achieve a sufficient degree

of precision while maintaining a reasonable response burden. Items are measured on five levels (e.g. "never", "rarely", "sometimes", "often", "always" or "not at all", "a little bit", "somewhat", "quite a bit", "very much") and refer to the past 7 days (except physical function). Answers yield a number from one to five, which, once fed into the online PROMIS converter (<http://www.healthmeasures.net/score-and-interpret/calculate-scores>), give one correspondent PROMIS T-Score ( $M = 50 \pm SD = 10$ ) per domain with the US general population as a reference. Note that due to the invariance property of IRT, T-Scores obtained from the PROMIS-29 are on the same metric as the scores Revicki used in his analysis, though these scores were generated using different items. For desirable constructs (e.g., physical function), higher T-scores indicate better health, whereas for undesirable domains (e.g., depression), higher T-scores indicate poorer health states.

The psychometric properties of the PROMIS-29 profile, including evidence of construct and criterion validity, have been reported elsewhere [29–32]. An earlier analysis of the data used in this study revealed that scores on the seven health domains of the PROMIS-29 are measurement invariant across the UK, France, and Germany except for one item [33].

#### *EQ-5D-5L crosswalk value set*

The EuroQoL EQ-5D is a standardized patient-reported HRQoL questionnaire, measuring five health dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Its original version, the EQ-5D-3L, differentiates 3 levels per domain, defining  $3^5$  or 243 health states. Its revised version, the EQ-5D-5L, has five levels: "No problems" (or 1), "Slight problems" (2), "Moderate problems" (3), "Severe problems" (4), and "Extreme problems" (5), defining  $5^5$  or 3125 different health states. We chose the EQ-5D-5L questionnaire because it can differentiate more health states and is more sensitive. Each health state is assigned a HSU by different value sets, reflecting the preferences of the general population in the respective countries. For many countries, there is not yet a value set for the 5L version. An EQ-5D-5L crosswalk value set was developed for the purpose of using 3L value sets for health states described by the 5L version. We used these EQ-5D-5L crosswalk value sets as they are available for all three countries of our samples [4, 26, 34, 35].

The maximum HSU for the best health state of 11111 is 1.00 or "full health" while 0.00 is considered "dead". The minimum HSU of the worst health state of 55555 is negative, considered "worse than dead":  $-0.594$  in the UK,  $-0.530$  in France, and  $-0.205$  in Germany [26].

## Statistical analysis

### *Relationships among individual health domains and health state utility across the UK, France, and Germany*

To obtain a first impression of the form of the relationships among individual health domains and HSU and to judge whether the relationships are stable across the three countries under investigation, we plotted the seven domain scores against HSU in the UK, France, and Germany.

### *Optimal models for predicting health state utility in the three countries*

We applied stepwise regression with backward selection to find the best models to predict the EQ-5D-5L crosswalk for the UK, France, and Germany, starting with full models that incorporated linear, quadratic, and cubic effects for all seven PROMIS-29 domains. We included polynomials up to the third degree as we expected that such polynomials can more flexibly fit the observed data, e.g. in case of nonlinear relationships between predictors and outcome. We used raw polynomials for linear, quadratic and cubic effects in order

also determined the width between the 95% empirical limits of agreement and compared them to the 95% theoretical limits of agreement (i.e.,  $\pm 1.96 \times \text{SD}(\text{residuals})$ ). To check the prediction performance along the HSU continuum, Bland–Altman plots were used.

We use cross-validation to check for overfitting [42]. With this in-sample cross-validation technique, the initial dataset is randomly split into 10 subsamples of approximately equal size. One of these subsamples is kept for validation, while the other nine subsamples are used for parameter estimation. This process is repeated ten times, and the results are averaged across repetitions. Overfitting would show when a model's nRMSE is substantially smaller than the average nRMSE of the models of the 10 subsamples.

We used *R* version 3.4.1, IBM SPSS Statistics version 23, and Microsoft Excel version 15 to run the analyses.

### *Impact of misspecified mapping functions on the prediction performance*

To the best of our knowledge, as of December 2020, the mapping function by Revicki was the only one available for predicting the EQ-5D-3L index value from the PROMIS-29 T-scores [19]:

$$\begin{aligned} EQ - 5D = & 1.0266 + 0.0077 \times \text{Physical functioning} - 0.0021 \times \text{fatigue} \\ & - 0.0040 \times \text{Pain interference} - 0.0023 \times \text{anxiety} - 0.0022 \times \text{depression} \end{aligned}$$

to obtain coefficients which can be used for prediction independently.

Because sociodemographic factors such as age and sex are known to be useful in predicting HSU, they were also entered as possible predictors [17]. The PROMIS pain intensity VAS was not included as pain is already covered by the pain interference domain, which proved to be superior than the VAS [36]. Also, while all other domains comprise of 4 items, the pain intensity domain within PROMIS-29 has only this single item, not measured on a T-Score metric.

The Bayesian information criterion (BIC) was used to steer the inclusion and exclusion of predictors in the stepwise regression analyses [37]. We chose nRMSE and nMAE as measures of the prediction precision and bias as they are preferred over either  $R^2$  or BIC used by Revicki [19, 38]. The nRMSE is the normalized root of the sum of the squared residuals between observed and predicted scores and the nMAE is the normalized mean absolute error of the absolute residuals. Both are normalized with respect to the different scale ranges of the EQ-5D-5L crosswalk in the UK, France, and Germany [39–41]. We

We were interested in quantifying the detrimental effect of applying this foreign mapping function to the data collected in Europe. Note that application of Revicki's model to the data collected in the UK, France and Germany (1) disregards the country specificity of any version of the EQ-5D, (2) does not utilize the potential predictive value of the two PROMIS-29 health domains not used by Revicki, (3) does not take higher-order effects into account, and in combination with the foregoing, (4) disregards country dependency of the form of relationships (i.e., the specific values of the regression coefficients used).

Because we were also interested in which factor is mainly responsible for the differences in prediction performance, we moved stepwise from Revicki's model to our models as follows: First, we used the five health domains of Revicki's model, but with regression coefficients optimized towards the data collected in each country separately. Second, we investigated the incremental value of adding either sleep disturbance, participation, or both to the prediction equation. Third, we allowed for incorporation of quadratic and/or cubic effects.

## Results

### Sample characteristics

We only briefly summarize the most important differences between the three samples here. The interested reader is referred to Table 4 (See “Appendix” section) for a comprehensive overview of the marginal distributions of sex, age, educational level, occupational status, and income in the three samples. Participants in the German sample (mean age = 50.0 years old) were slightly older than participants in the French (48.4 years old) and UK samples (47.8 years old). Participants in the German sample were more likely to have a low educational background (23.4%) than participants in the French (7.6%) and UK samples (8.1%). Participants in the French sample were more likely to be unemployed/inactive (48.4%) than participants in the German (41.5%) and UK samples (39.4%).

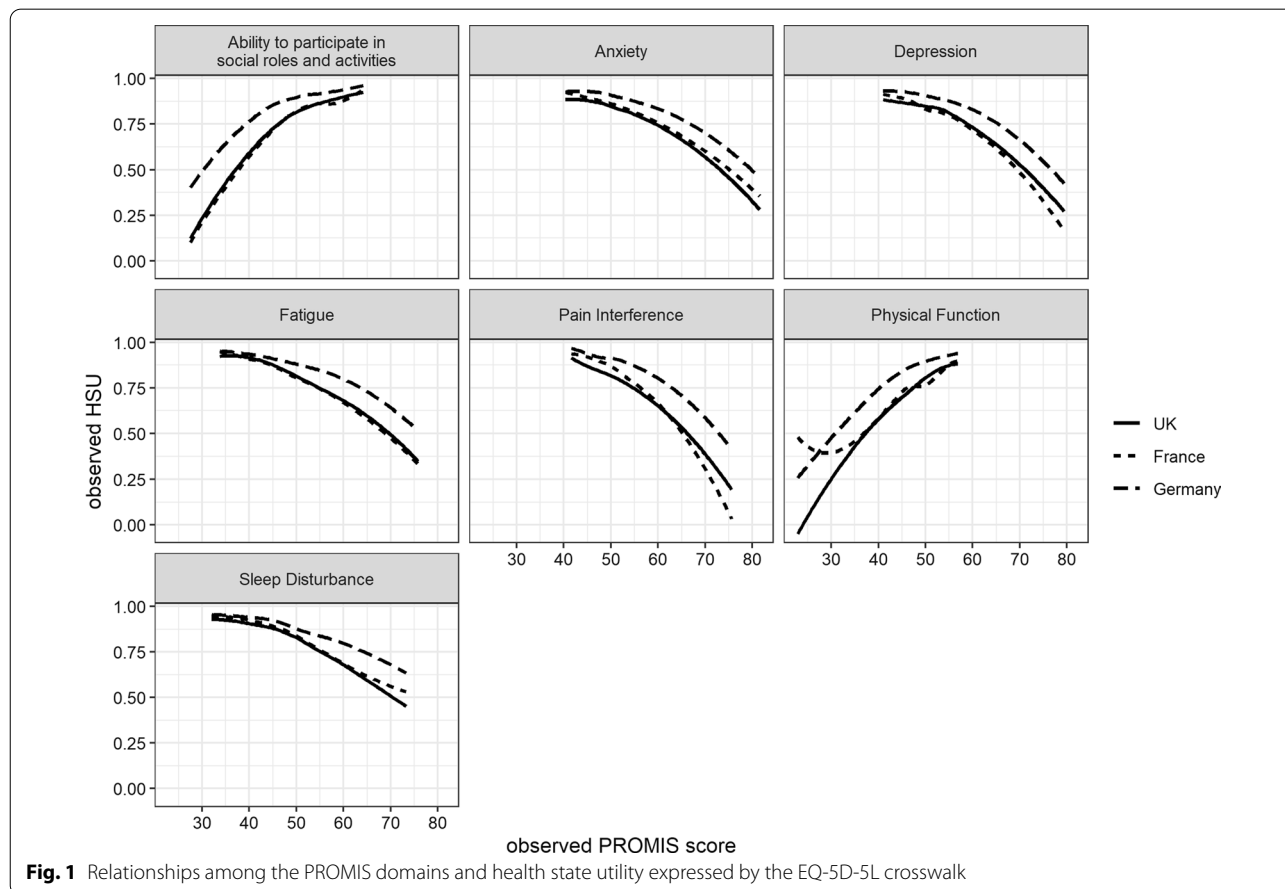
### Relationships among individual health domains and health state utility across the UK, France, and Germany

The relationships among the seven PROMIS domains and HSU expressed by the EQ-5D score in the three European countries are displayed in Fig. 1.

A number of conclusions can be drawn from Fig. 1. First, with the exception of low levels of physical functioning in France, the relationships among the seven PROMIS domains and HSU are comparable across the three European countries. Second, most of the curves are not simple straight lines and are slightly curvilinear, indicating that changes at severer levels have a greater impact on HSU. Third, all the relationships are in accordance with theoretical expectations. Higher values on the positive PROMIS domains (participation and physical function) correspond to higher HSU values, and higher values on the five negative PROMIS domains correspond with lower HSU values. Fourth, participation and physical function seem to have the strongest relationship with HSU because these curves are the steepest.

### Optimal models for predicting health state utility in the three countries

Recall that we used stepwise regression with backward selection to find optimal models for predicting HSU for the UK, France, and Germany. The primary models thus comprised linear, quadratic, and cubic effects for each PROMIS domain plus effects for age and sex. Effects that



**Table 1 Coefficients of the optimal models for the United Kingdom, France, and Germany**

	UK			France			Germany		
	Regression coefficient	Standardized regression coefficient	SE	Regression coefficient	Standardized regression coefficient	SE	Regression coefficient	Standardized regression coefficient	SE
Constant	2.288E-0		7.874E-1	2.910E-0		5.665E-1	-1.181E-0		3.047E-1
Age	9.590E-4	0.069	2.032E-4	-1.372E-3	-0.107	1.903E-4			
Anxiety	1.120E-2	0.499	2.951E-3						
Pain interference	-1.773E-1	-7.479	4.27E-2						
Physical function	5.354E-2	1.881	5.24E-3	-3.027E-1	-9.807	3.202E-2			
Depression							7.425E-3	0.404	1.664E-3
Participation	1.334E-2	0.573	4.027E-3	9.415E-2	3.719	2.660E-2	8.834E-2	4.915	1.963E-2
Anxiety <sup>a</sup>	-1.227E-4	-0.604	2.758E-5						
Pain interference <sup>a</sup>	3.042E-3	13.970	7.651E-4	2.122E-4	0.900	4.059E-5			
Physical function <sup>a</sup>	-4.853E-4	-1.566	5.544E-5	7.506E-3	22.864	7.839E-4	5.596E-4	2.581	6.114E-5
Sleep disturbance <sup>a</sup>				-2.390E-5	-0.088	4.542E-6	-1.763E-5	-0.097	3.415E-6
Participation <sup>a</sup>	-1.061E-4	-0.460	3.785E-5	-1.706E-3	-7.104	5.465E-4	-1.733E-3	-9.850	4.073E-4
Anxiety <sup>b</sup>							-1.480E-7	-0.070	5.293E-8
Depression <sup>b</sup>	-3.453E-7	-0.145	5.665E-8	-3.487E-7	-0.121	5.494E-8	-8.951E-7	-0.421	1.991E-7
Fatigue <sup>b</sup>				-2.456E-7	-0.088	5.782E-8			
Pain interference <sup>b</sup>	-1.769E-5	-6.852	4.460E-6	-3.697E-6	-1.270	5.046E-7	-7.808E-7	-0.421	4.198E-8
Sleep disturbance <sup>b</sup>	-1.860E-7	-0.059	5.000E-8						
Physical function <sup>b</sup>				-5.805E-5	-12.841	6.167E-6	-6.865E-6	-2.300	8.279E-7
Participation <sup>b</sup>				1.026E-5	3.471	3.670E-6	1.113E-5	4.998	2.763E-6

Coefficients are displayed as negative exponentials with four digits, beginning with the first non-zero digit of the coefficient. HSU is expressed on a scale ranging from -0.594 (UK), -0.53 (France), and -0.205 (Germany) to 1, and the PROMIS domains are expressed as T-scores (M = 50). All the coefficients displayed differ significantly from zero at  $p < 0.01$

did not significantly improve the prediction performance were sequentially removed from these models. The coefficients of the final models to optimally estimate the EQ-5D-5L crosswalk from PROMIS-29 for the UK, France, and Germany can be found in Table 1.

The (unstandardized) regression coefficients of Table 1 can be used to compute the EQ-5D-5L crosswalk from the PROMIS T-scores:  $EQ-5D = Constant + Coefficient (Age) \times Age + Coefficient (Anxiety) \times T-score (Anxiety) + \dots + Coefficient (Participation^3) \times (T-score (Participation))^3$ . However, interpretation of the regression coefficients needs to take into account two specifics of polynomial regression models.

First, the regression coefficients of the higher-order effects appear to be much smaller than those for the linear effects, as the values of the predictor variables (with  $M = 50$ ) are taken to the power of two for the quadratic effects ( $M^2 = 2,500$ ) and to the power of three for the cubic effects ( $M^3 = 125,000$ ). Hence, coefficients have a substantially larger impact on the scale of the criterion.

Second, the single standardized regression coefficients shown in Table 1 should not be used to infer the form of the relationship between the individual health

domains and the EQ-5D-5L crosswalk because we have up to three effects (linear, quadratic, and cubic) in each health domain, and the relationship thus must be described by the summed effect of all three effects. Furthermore, not all coefficients are in agreement with Fig. 1 which plotted the relationship of a single health domain to the EQ-5D-5L crosswalk, irrespective of the values in all the other health domains. Instead, the regression coefficients are optimal given the effect of all the other effects already taken into account (stepwise procedure), which also explains why the final models in the three countries are so different. Age, for example, has a positive effect on HSU in the UK, a negative effect on HSU in France, and no effect on HSU in Germany. Although out of the 23 possible predictors twelve (UK and France) and ten (Germany) were kept in the final models, only four effects were consistently chosen across countries: the linear effect of participation, the quadratic effect of physical functioning, and cubic effects of depression and pain interference.

The prediction performance of these models is summarized in Table 2. HSU expressed by the EQ-5D-5L crosswalk can be best mapped from the PROMIS-29 in France

**Table 2 Prediction performance of the optimal models for the United Kingdom, France, and Germany and results of the tenfold cross-validation**

	nRMSE	Mean nRMSE (CV)	SD nRMSE (CV)	nMAE	Mean nMAE (CV)	SD nMAE (CV)	95% theoretical LoA	95% empirical LoA
UK	0.076	0.077	0.0083	0.053	0.054	0.0046	±0.25	−0.20;0.17
France	0.075	0.076	0.0062	0.052	0.053	0.0041	±0.23	−0.19;0.17
Germany	0.079	0.080	0.0096	0.051	0.051	0.0037	±0.19	−0.16;0.13

*nRMSE* normalized root mean square error, *nMAE* normalized mean absolute error, *LoA* levels of agreement, *CV* cross-validation, *SD* standard deviation, *UK* United Kingdom

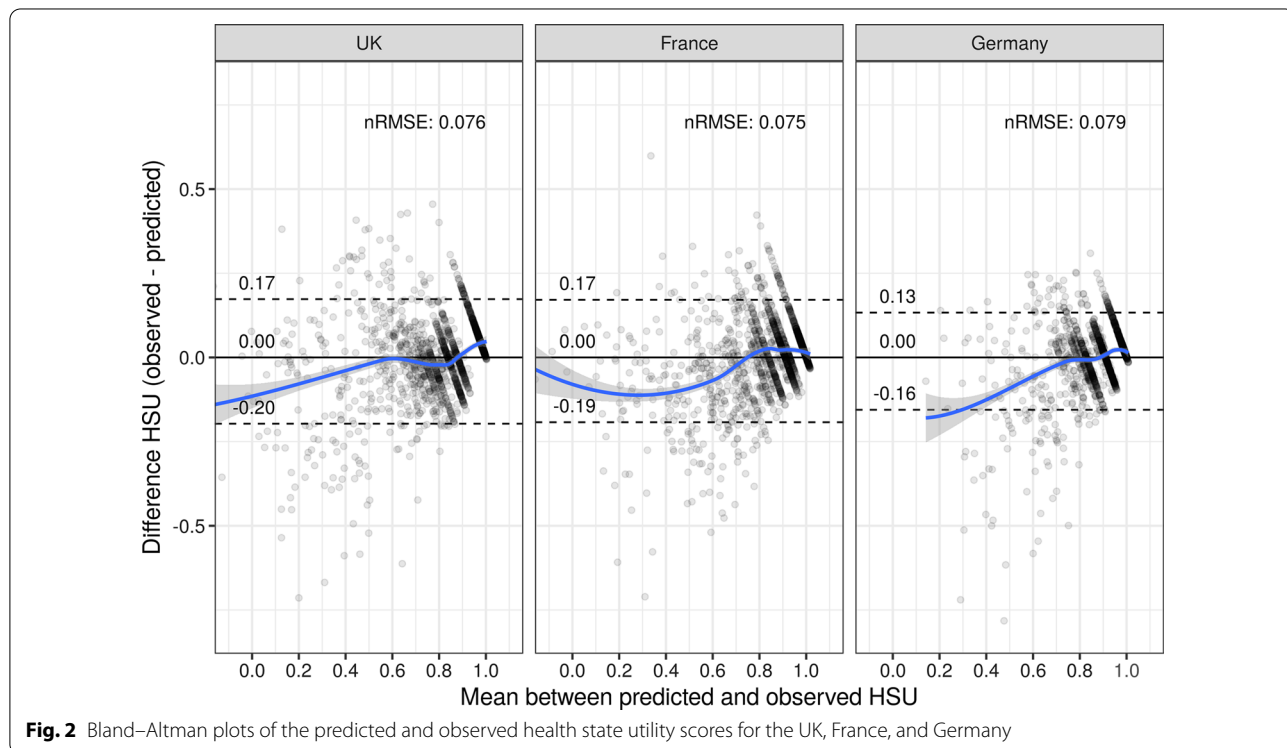
( $nRMSE_{FRA} = 0.075$ ,  $nMAE_{FRA} = 0.052$ ), followed by the UK ( $nRMSE_{UK} = 0.076$ ,  $nMAE_{UK} = 0.053$ ) and Germany ( $nRMSE_{GER} = 0.079$ ,  $nMAE_{GER} = 0.051$ ). Furthermore, for all three countries, the widths of the empirical limits of agreement are always smaller than the widths of the theoretical limits of agreement. All models were confirmed by tenfold cross-validation, having a marginally smaller nRMSE and nMAE compared the mean nRMSE and mean nMAE, respectively, of the 10 models of the cross-validation subsamples.

The prediction performances of the final models along the HSU continuum are depicted in the Bland–Altman plots in Fig. 2. Note that especially in the German sample, there are not many respondents with low HSU (EQ-5D-5L crosswalk < 0.2). Furthermore, prediction performance appears to be slightly better for high levels

of HSU (EQ-5D-5L crosswalk > 0.8) than for intermediate or low HSU.

**Impact of misspecified mapping functions on the prediction performance**

The differences in the prediction performances between the applications of Revicki’s model versus our models are depicted in Table 3. The application of Revicki’s model to the European data would systematically underestimate the EQ-5D-5L crosswalk for the UK (−0.10) and for France (−0.09) but not for Germany. The prediction performance of Revicki’s model is the best in Germany, and the differences in the prediction performances between Revicki’s and our mapping functions are smaller in Germany than for the UK or for France, as indicated by the values of the nRMSE, nMAE, and empirical LoAs.



**Fig. 2** Bland–Altman plots of the predicted and observed health state utility scores for the UK, France, and Germany

**Table 3** The detrimental effect of using Revicki's model to predict the EQ-5D-5L crosswalk from the PROMIS-29 for the United Kingdom, France, and Germany

	R <sup>2</sup> <sub>adj</sub>	ICC	Bias	nRMSE	nMAE	95% theoretical LoA	95% empirical LoA
<i>France</i>							
Revicki	0.61	0.78	-0.09	0.112	0.072	-0.38; 0.20	-0.38; 0.08
Polynomial regression	0.72	0.85	0.00	0.075	0.052	±0.23	-0.19; 0.17
<i>Germany</i>							
Revicki	0.53	0.73	0.00	0.091	0.058	-0.22; 0.22	-0.18; 0.14
Polynomial regression	0.64	0.80	0.00	0.079	0.051	±0.19	-0.16; 0.13
<i>UK</i>							
Revicki	0.68	0.82	-0.10	0.113	0.075	-0.39; 0.19	-0.39; 0.07
Polynomial regression	0.74	0.86	0.00	0.076	0.053	±0.25	-0.20; 0.17

UK United Kingdom, *adj* adjusted, *ICC* intraclass correlation coefficient, *nRMSE* normalized root mean squared error, *nMAE* normalized mean absolute error, *LoA* levels of agreement

The last step was to investigate which factor was mainly responsible for the observed differences in the prediction performances between Revicki's and our models. The results of the application of country-specific regression coefficients for the five health domains specified by Revicki (first alternative model; M1), the incorporation of sleep disturbance and/or participation (M2c), or the incorporation of quadratic and cubic trends into the five-domain model specified by Revicki (M3) are shown in Fig. 3. The average prediction performance ( $nRMSE_{UK}=0.082$ ,  $nRMSE_{FRA}=0.085$ , and  $nRMSE_{GER}=0.087$ ) mainly improves by incorporating country-specific regression coefficients into the five health domain models specified by Revicki. However, neither this model (M1) nor the incorporation of sleep disturbance and/or participation (M2c) improves the prediction performance for low levels of HSU, but the incorporation of quadratic and cubic effects (M3) does improve the prediction performance for low levels of HSU. That is, overprediction of HSU is clearly reduced by adding these higher-order effects to the three regression equations.

## Discussion

### Summary of main findings

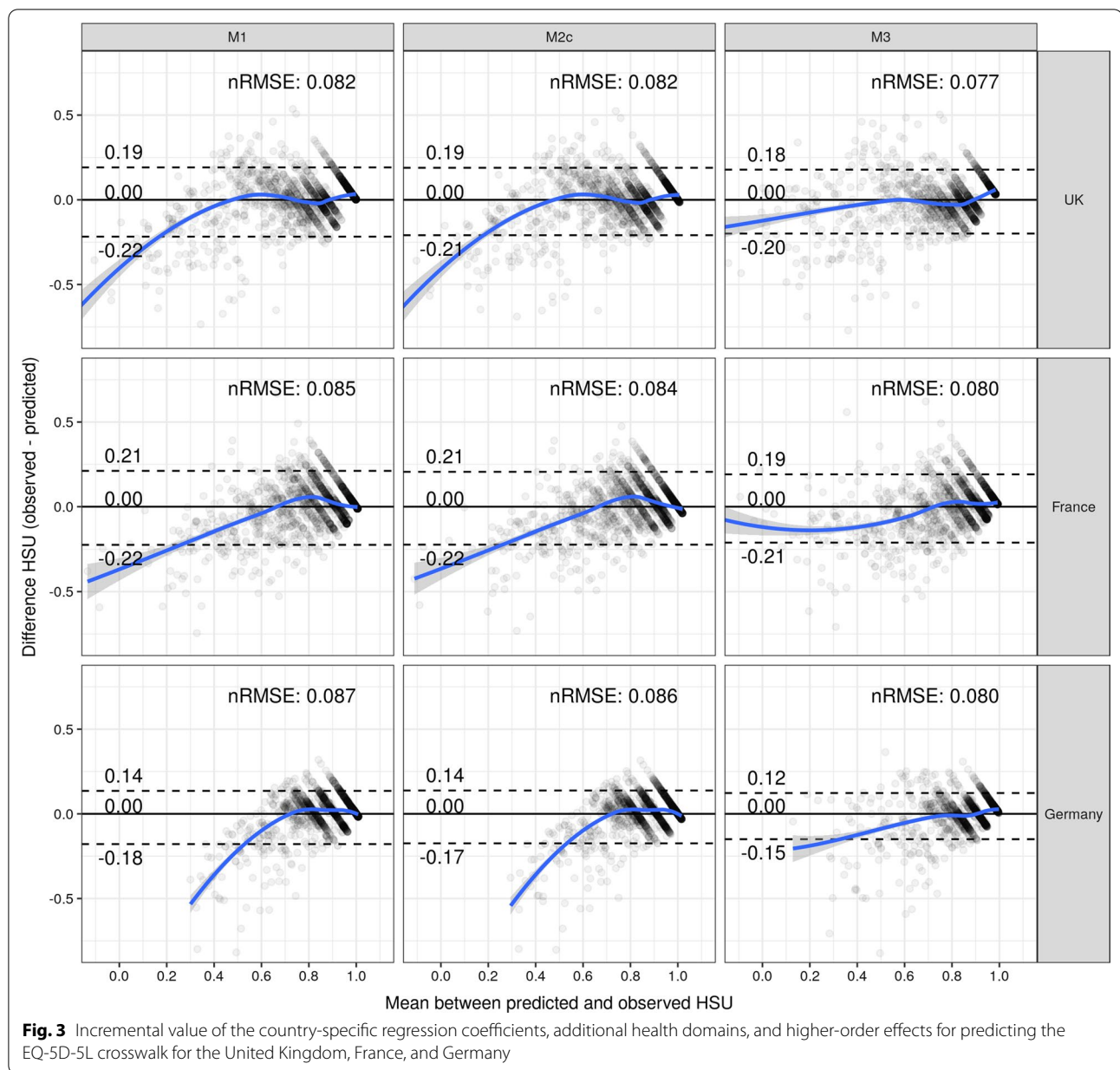
We developed optimal models for mapping the EQ-5D-5L crosswalk from the PROMIS-29 in the UK, France, and Germany. Furthermore, we showed that the incorporation of higher-order effects into the regression equations substantially reduced overestimation of low HSU. The EQ-5D-5L crosswalk can therefore now be predicted from PROMIS-29 in three major European countries for QALY in CUA for HTA assessments, enabling the use of PROMIS for economic evaluations in Europe. This is of practical importance since HTA agencies demand the EQ-5D-5L crosswalk as HSU for QALY and PROMIS is

more frequently used in clinical, non-preference HRQoL. We believe our models are highly applicable achieving a good degree of precision, also in lower spectrums of health, while at the same time avoiding high complexity with a manageable number of predictors. Our results in terms of the nRMSE and nMAE perform very well compared to what is usually reported for mapping algorithms [17, 43–47].

The major comparator to our models is Revicki's OLS linear US model, the only one predicting the EQ-5D-3L index value from PROMIS-29. All our models perform better in terms of R-squared and ICC while the LoA were comparable. Revicki did neither report MAE nor RMSE. Furthermore, Revicki's uses the former version of the EQ-5D, the EQ-5D-3L with the US value set as target measure, while we use the EQ-5D-5L crosswalk value sets from the UK, France, and Germany, respectively. We demonstrated that the application of Revicki's US model to European data will yield biased results, especially for poor health states. However, this model performs well in upper ranges of health. One might therefore consider using a foreign model with domestic data as a second-best option to predict the EQ-5D-5L crosswalk for QALY in CUA if a country-specific mapping algorithm is not available, especially in a group of healthier patients. This decision might make sense, for example, when using our German model for Austrian data in or using Revicki's US model for Canadian data, since in both cases, cultural proximity can reasonably be assumed.

Apart from Revicki's model predicting the EQ-5D-3L index value from PROMIS-29, there is also another model of his, predicting the EQ-5D-3L index value from PROMIS Global Health (GH) items, using linear regression in a US sample [19]. Thompson (2017) mapped PROMIS-GH to the EQ-5D-3L index value in a US sample applying linear and equipercentile equating, treating





PROMIS-GH items as categorical variables [20]. So compared to our models, both models differ in respect of population, source measure, and target measure: They use the US value set for the EQ-5D-3L index value while we use the EQ-5D-5L crosswalk for the UK, France, and Germany, respectively. Thompson’s models additionally differs in the mapping method applied. In terms of R-squared, our model for Germany performs at least as good and our models for the UK and France perform better than both Revicki’s and Thompson’s PROMIS-GH models. In terms of MAE, all our models perform better.

Despite Thompson’s the different method, low EQ-5D-3L index values were still overestimated [20]. Both studies did not report a RMSE.

Generally however, researchers should be aware that the consequences of working with a suboptimal mapping algorithm can be substantial: incremental cost-effectiveness ratio (ICER) of costs per QALY can differ between British pound sterling (GBP) 18,000 and GBP 32,000 depending on what mapping algorithm is used [48]. NICE has adopted a threshold of GBP 30,000 per QALY representing the public’s maximum additional willingness to

pay for a new treatment or a new drug compared to the existing standard of care [49]. Consequently, imprecise mapping methods have a great impact on CUA in HTA assessments and consequently on what innovations are made available to patients.

### Strengths and limitations

This study was conducted using three large samples representative of the general population in three European countries. To ensure comparability, the sampling strategies were the same across countries. This strength of our study is directly related to its foremost weakness: Severe health states are not frequently observed in the general population, and the proposed models therefore rely on few observations for low HSU. Furthermore, our models allowed judgement of the incremental value of incorporating two additional health domains and higher-order effects for HSU prediction.

Finally, some authors have argued against OLS regression as a type of mapping method even though, as outlined above, it is the most widely used method. First, arguments against that method are due to the phenomenon of regression to the mean. Second, linear regression models tend to predict HSU score greater than one, which is a value that is impossible by definition of HSU [22]. In our study, the risk of predicting HSU values greater than one is circumvented by incorporation of non-linear trends.

### Directions for future research and the PROMIS preference score (PROPr) for QALYs

Our mapping functions should be confirmed to samples with a greater frequency of low HSU. Therefore, we are planning to replicate our findings with data collected from spine patients who were assessed before surgery. It would also be interesting whether regressing the EQ-5D dimensions on the PROMIS domain scores first and then calculating the EQ-5D-5L crosswalk from the regressed EQ-5D dimensions has incremental value [50].

PROMIS data can also be used to estimate a new preference-based HSU score: Hanmer developed the PROMIS Preference Score (PROPr) to compute HSU for QALYs directly from 7 PROMIS health domains: cognition, depression, fatigue, pain, physical function, sleep disturbance, and participation [51–55]. Note that these 7 PROMIS domains are not equivalent with those 7 domains from the PROMIS-29 profile (anxiety is missing in the PROPr, while cognition is missing in the PROMIS-29) [25, 53, 56, 57].

The PROPr could potentially be used instead of the EQ-5D-5L crosswalk in CUA. Since many European HTA authorities such as NICE specifically demand the use of the EQ-5D-5L crosswalk to measure HSU in CUA, mapping the PROMIS-29 to the EQ-5D-5L crosswalk will still be needed [49]. Also, as of December 2020, there is no PROPr value set for European preferences [53, 54].

### Conclusion

Our mapping functions can be used to predict the EQ-5D-5L crosswalk from the PROMIS-29 for CUA in HTA for the UK, France and Germany. The inclusion of polynomial regression terms decreases the prediction bias for lower HSU.

Our results support the assertion that mapping functions are country-specific. The application of Revicki's model to the data collected in the three European countries leads to biased HSU estimates for the UK and France and to less precise estimates in all three countries. Estimation of country-specific regression coefficients for the five health domains identified by Revicki strongly improves the average prediction performance but does not remedy the overestimation of low HSU.

### Abbreviations

3L: 3 levels; 5L: 5 levels; BDI-2: Beck depression inventory version 2; BIC: Bayesian information criterion; CAT: Computerized adaptive testing; CES-D: Center for Epidemiologic Studies Depression Scale; CUA: Cost-utility analysis; CV: Cross validation; EQ-5D: EuroQoL 5 dimensions; FRA: France; GAD-7: Generalized anxiety disorder 7; GBP: British pound sterling; GER: Germany; GH: Global Health; HRQoL: Health-related quality of life; HSU: Health state utility; HTA: Health technology assessments; ICC: Intraclass correlation coefficient; ICER: Incremental cost-effectiveness ratio; LoA: Limits of agreement; MASQ: Mood and anxiety symptom questionnaire; NICE: National Institute of Health and Clinical Excellence; (n)MAE: (Normalized) mean absolute error; (n)RMSE: (Normalized) root mean square error; M: Mean; OLS: Ordinary least squares; PANAS: Positive and negative affect schedule; PHQ-9: Patient-health questionnaire-9; PROM: Patient reported outcome measure; PROMIS: Patient reported outcome measurement information system; PROMIS-29 v2.0: PROMIS profile 29 version 2.0; PROPr: PROMIS preference score; QALY: Quality-adjusted life years; RIM: Random iterative method; SD: Standard deviation; SG: Standard gamble; TTO: Time trade-off; UK: United Kingdom of Great Britain and Northern Ireland; US: United States of America; VAS: Visual Analogue Scale.

### Acknowledgements

Not applicable.

### Authors' contributions

CPK and FF conceived the study; CPK, JVB, and FF planned and conducted data analysis; the manuscript was drafted by CPK and JVB and CJS-G, JMV, AL, MR, and FF revised it. All authors read and approved the final manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL. This study was funded by the Centre Virchow-Villerme.

**Availability of data and materials**

Data is available on reasonable request.

**Ethical approval and consent to participate**

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. Participation was voluntary.

**Consent for publication**

Not applicable.

**Competing interests**

All authors declare that they have no competing interests.

**Author details**

<sup>1</sup> Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité - Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. <sup>2</sup> Department of Symptom Research, MD Anderson Cancer Center, University of Houston, Houston, TX, USA. <sup>3</sup> Health Services and Policy Research Group, University of Exeter, Exeter, UK. <sup>4</sup> NIHR Peninsula Collaboration for Leadership in Applied Health Research and Care, Exeter, UK. <sup>5</sup> APEMAC, EA 4360, Paris Descartes University, Paris, France. <sup>6</sup> Département d'Histoire et de Philosophie des Sciences, Laboratoire SPHERE, UMR 7219, CNRS-Université Paris Diderot - Sorbonne Paris Cité, Paris, France. <sup>7</sup> Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, USA.

**Appendix**

See Table 4.

**Table 4 Sociodemographic characteristics of the samples**

Variable	Value	UK (%)	France (%)	Germany (%)	p value
Sex	Male	49.0	48.3	48.5	0.918
	Female	51.0	51.7	51.5	
Age	18–29	19.8	18.6	16.4	0.027
	30–39	16.8	16.3	14.4	
	40–49	18.2	18.5	18.3	
	50–59	16.1	17.4	18.5	
	60–69	13.9	14.3	13.4	
	70+	15.2	15.1	19.0	
Education	Low	8.1	7.6	23.4	<0.001
	Medium	48.6	43.2	48.2	
	High	43.3	49.2	28.4	
Occupation	Managers and professionals	21.2	10.8	12.5	<0.001
	Technicians, clerks, service workers	25.1	22.9	28.9	
	Workers, elementary occupations, armed forces	14.3	17.8	17.1	
	Inactive or unemployed	39.4	48.4	41.5	
Income	Lower income	16.8	12.8	19.2	<0.001
	Lower-middle income	20.2	19.5	18.0	
	Higher-middle income	30.7	27.2	21.8	
	Higher income	21.0	24.4	24.2	
	Prefer not to answer	11.3	16.2	16.8	

UK United Kingdom

Received: 11 June 2020 Accepted: 18 November 2020  
Published online: 17 December 2020

## References:

- Weinstein MC, Torrance G, McGuire A. QALYs: the basics. *Value Heal (Internet)*. 2009;12(Suppl. 1):S5–9. <https://doi.org/10.1111/j.1524-4733.2009.00515.x>.
- Klarman HE, Francis JO, Rosenthal GD. Cost effectiveness analysis applied to the treatment of chronic renal disease. *Med Care (Internet)*. 1968;6(1):48–54. Available from: <http://www.jstor.org/stable/3762651>.
- Valderas JM, Alonso J. Patient reported outcome measures : a model-based classification system for research and clinical practice. *Qual Life Res*. 2008;17:1125–35.
- Rabin R, Oemar M, Oppe M, Janssen B, Herdman M. EQ-5D-5L user guide version 2.1. 2015;(April):28. Available from: [http://www.euroqol.org/fileadmin/user\\_upload/Documenten/PDF/Folders\\_Flyers/EQ-5D-5L\\_UserGuide\\_2015.pdf](http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf).
- Greiner W, Weijnen T, Nieuwenhuizen MN, Oppe S, Badia X, et al. A single European currency for EQ-5D health states. *Eur J Heal Econ*. 2003;4:222–31.
- Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20:1727–36.
- Devlin N, Krabbe P. The development of new research methods for the valuation of EQ-5D-5L. *Eur J Heal Econ*. 2013;14(Suppl.):1–3.
- Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *J Rheumatol*. 2009;36(9):2061–6.
- Alonso J, Bartlett SJ, Rose M, Aaronson NK, Chaplin JE, Efficace F, et al. The case for an international patient-reported outcomes measurement information system (PROMIS<sup>®</sup>) initiative. *Health Qual Life Outcomes*. 2013;11(1):1–5.
- Embretson SE, Reise SP. *Item response theory for psychologists*. London: Psychology Press; 2013.
- PROMIS Cooperative Group. PROMIS<sup>®</sup> Instrument Maturity Model [Internet]. 2012. p. 1–4. Available from: [http://www.healthmeasures.net/images/PROMIS/PROMISStandards\\_Vers\\_2\\_0\\_MaturityModelOnly\\_508.pdf](http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers_2_0_MaturityModelOnly_508.pdf).
- Rupp AA, Zumbo BD. Understanding parameter invariance in unidimensional IRT models. *Educ Psychol Meas*. 2006;66(1):63–84.
- Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and promis: assessment of physical function. *J Rheumatol*. 2014;41(1):153–8.
- Hays RD, Revicki DA, Feeny D, Fayes P, Spritzer KL, Cella D. Using linear equating to map PROMIS global health items and the PROMIS-29 V2.0 profile measure to the health utilities index mark 3. *Pharmacoeconomics*. 34(10):1015–22.
- Terwee CB, Roorda LD, De Vet HCW, Dekker J, Westhovens R, Van Leeuwen J, et al. Dutch-Flemish translation of 17 item banks from the Patient-Reported Outcomes Measurement Information System (PROMIS). *Qual Life Res*. 2014;23(6):1733–41.
- Oude Voshaar MAH, ten Klooster PM, Taal E, Krishnan E, van de Laar MAFJ. Dutch translation and cross-cultural adaptation of the PROMIS<sup>®</sup> physical function item bank and cognitive pre-test in Dutch arthritis patients. *Arthritis Res Ther (Internet)*. 2012;14(2):R47. Available from: <http://arthritis-research.com/content/14/2/R47>.
- Mukuria C, Rowen D, Harnan S, Rawdin A, Wong R, Ara R, et al. An updated systematic review of studies mapping (or cross-walking) measures of health-related quality of life to generic preference-based measures to generate utility values. *Appl Health Econ Health Policy (Internet)*. 2019;17(3):295–313. <https://doi.org/10.1007/s40258-019-00467-6>.
- Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health Qual Life Outcomes (Internet)*. 2013;11(1):151. <https://doi.org/10.1186/1477-7525-11-151>.
- Revicki DA, Kawata AK, Harnan N, Chen W-H, Hays RD, Cella D. Predicting EuroQol (EQ-5D) scores from the patient-reported outcomes measurement information system (PROMIS) global items and domain item banks in a United States sample. *Qual Life Res*. 2009;18(6):783–91.
- Thompson NR, Lapin BR, Katzan IL. Mapping PROMIS global health items to EuroQol (EQ-5D) utility scores using linear and equipercentile equating. *Pharmacoeconomics*. 2017.
- Crott R. Direct mapping of the QLQ-C30 to EQ-5D preferences: a comparison of regression methods. *Pharmacoeconomics Open*. 2018;2(2):165–77.
- Hernández Alava M, Wailoo AJ, Ara R. Tails from the peak district: Adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value Heal (Internet)*. 2012;15(3):550–61. <https://doi.org/10.1016/j.jval.2011.12.014>.
- Schalet BD, Cook KF, Choi SW, Cella D. Establishing a common metric for self-reported anxiety: linking the MASQ, PANAS, and GAD-7 to PROMIS anxiety. *J Anxiety Disord (Internet)*. 2014;28(1):88–96.
- Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychol Assess*. 2014;26(2):513–27.
- Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43(3):203–20.
- Van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Heal (Internet)*. 2012;15(5):708–15. <https://doi.org/10.1016/j.jval.2012.02.008>.
- Lamu AN, Chen G, Gamst-Klaussen T, Olsen JA. Do country-specific preference weights matter in the choice of mapping algorithms? The case of mapping the Diabetes-39 onto eight country-specific EQ-5D-5L value sets. *Qual Life Res (Internet)*. 2018;27(7):1801–14. <https://doi.org/10.1007/s11136-018-1840-5>.
- Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, et al. PROMIS<sup>®</sup> adult health profiles: efficient short-form measures of seven health domains. *Value Heal*. 2019;22(5):537–44.
- Choi SW, Reise SP, Pilkonis PA, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Qual Life Res*. 2010;19(1):125–36.
- Hinchcliff M, Beaumont JL, Thavarajah K, Varga J, Chung A, Podluszky S, et al. Validity of two new patient-reported outcome measures in systemic sclerosis: patient-reported outcomes measurement information system 29-item health profile and functional assessment of chronic illness therapy-dyspnea short form. *Arthritis Care Res (Hoboken) (Internet)*. 2011 Nov;63(11):1620–8. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/22034123>.
- Beaumont JL, Cella D, Phan AT, Choi S, Liu Z, Yao JC. Comparison of health-related quality of life in patients with neuroendocrine tumors with quality of life in the general US population. *Pancreas*. 2012;41(3):461–6.
- Yount SE, Beaumont JL, Chen S-Y, Kaiser K, Wortman K, Van Brunt DL, et al. Health-Related Quality of Life in Patients with Idiopathic Pulmonary Fibrosis. *Lung (Internet)*. 2016 Apr;194(2):227–34. <https://doi.org/10.1007/s00408-016-9850-y>.
- Fischer F, Gibbons C, Coste J, Valderas JM, Rose M, Leplège A. Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany. *Qual Life Res (Internet)*. 2018;27(4):999–1014. <https://doi.org/10.1007/s11136-018-1785-8>.
- Martí-Pastor M, Pont A, Ávila M, Garin O, Vilagut G, Forero CG, et al. Head-to-head comparison between the EQ-5D-5L and the EQ-5D-3L in general population health surveys. *Popul Health Metr*. 2018;16(1):1–11.
- Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. *Pharmacoeconomics (Internet)*. 2018;36(6):675–97. <https://doi.org/10.1007/s40273-018-0623-8>.
- Bernstein DN, Kelly M, Houck JR, Ketz JP, Flemister AS, DiGiovanni BF, et al. PROMIS pain interference is superior vs. numeric pain rating scale for pain assessment in foot and ankle patients. *Foot Ankle Int*. 2019;40(2):139–44.
- Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods (Internet)*. 2012/02/06. 2012;17(2):228–43. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/22309957>.
- Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Heal Econ*. 2010;11:215–25.

39. Lamu AN. Does linear equating improve prediction in mapping? Cross-walking MacNew onto EQ-5D-5L value sets. *Eur J Heal Econ* [Internet]. 2020;21(6):903–15. <https://doi.org/10.1007/s10198-020-01183-y>.
40. Lamu AN, Olsen JA. Testing alternative regression models to predict utilities: mapping the QLQ-C30 onto the EQ-5D-5L and the SF-6D. *Qual Life Res* (Internet). 2018;27(11):2823–39. <https://doi.org/10.1007/s11136-018-1981-6>.
41. Gamst-Klaussen T, Lamu AN, Chen G, Olsen JA. Assessment of outcome measures for cost–utility analysis in depression: mapping depression scales onto the EQ-5D-5L. *BJPsych Open*. 2018;4(4):160–6.
42. Blum A, Kalai A, Langford J. Beating the holdout: bounds for KFold and progressive cross-validation. *COLT*. 1999;203–8.
43. Collado-Mateo D, Chen G, Garcia-Gordillo MA, Iezzi A, Adsuar JC, Olivares PR, et al. Fibromyalgia and quality of life: mapping the revised fibromyalgia impact questionnaire to the preference-based instruments. *Health Qual Life Outcomes*. 2017;15(114):1–9.
44. Marriott E-R, van Hazel G, Gibbs P, Hatswell AJ. Mapping EORTC-QLQ-C30 to EQ-5D-3L in patients with colorectal cancer. *J Med Econ* (Internet). 2017;20(2):193–9. <https://doi.org/10.1080/13696998.2016.1241788>.
45. Ameri H, Yousefi M, Yaseri M, Nahvijou A, Arab M, Akbari Sari A. Mapping EORTC-QLQ-C30 and QLQ-CR29 onto EQ-5D-5L in Colorectal Cancer Patients. *J Gastrointest Cancer*. 2019 ([Epub ahead of print]).
46. Beck AJCC, Kieffer JM, Retèl VP, van Overveld LFI, Takes RP, van den Brekel MWM, et al. Mapping the EORTC QLQ-C30 and QLQ-H&N35 to the EQ-5D for head and neck cancer: can disease-specific utilities be obtained? *PLoS ONE*. 2019;14(12):1–16.
47. Yang F, Wong CKH, Luo N, Piercy J, Moon R, Jackson J. Mapping the kidney disease quality of life 36-item short form survey (KDQOL-36) to the EQ-5D-3L and the EQ-5D-5L in patients undergoing dialysis. *Eur J Heal Econ* [Internet]. 2019;20(8):1195–206. <https://doi.org/10.1007/s10198-019-01088-5>
48. Pennington B, Davis S. Mapping from the health assessment questionnaire to the EQ-5D: the impact of different algorithms on cost-effectiveness results. *Value Heal* (Internet). 2014;17(8):762–71. <https://doi.org/10.1016/j.jval.2014.11.002>
49. NICE. Guide to the methods of technology appraisal (internet). NICE Guidelines. 2013. Available from: nice.org.uk/process/pmg9.
50. Ali FM, Kay R, Finlay AY, Piguet V, Kupfer J, Dalgard F, et al. Mapping of the DLQI scores to EQ-5D utility values using ordinal logistic regression. *Qual Life Res*. 2017;26(11):3025–34.
51. Hanmer J, Feeny D, Fischhoff B, Hays RD, Hess R, Pilkonis PA, et al. The PROMIS of QALYs. *Health Qual Life Outcomes* (Internet). 2015;15–7. <https://doi.org/10.1186/s12955-015-0321-6>.
52. Hanmer J, Cella D, Feeny D, Fischhoff B, Hays RD, Hess R, et al. Selection of key health domains from PROMIS® for a generic preference-based scoring system. *Qual Life Res*. 2017;2017:1–9.
53. Hanmer J, Dewitt B. The development of a preference-based scoring system for PROMIS® (PROPr): A Technical Report Version 1.4. 2017.
54. Hanmer J, Cella D, Feeny D, Fischhoff B, Hays RD, Hess R, et al. Evaluation of options for presenting health-states from PROMIS® item banks for valuation exercises. *Qual Life Res* (Internet). 2018;27(7):1835–43. <https://doi.org/10.1007/s11136-018-1852-1>.
55. Dewitt B, Feeny D, Fischhoff B, Cella D, Hays RD, Hess R, et al. Estimation of a preference-based summary score for the patient-reported outcomes measurement information system: the PROMIS®-Preference (PROPr) scoring system. *Med Decis Mak*. 2018;38(6):683–98.
56. Chevalier J, De Pouvourville G. Valuing EQ-5D using time trade-off in france. *Eur J Heal Econ*. 2013;14(1):57–66.
57. Hanmer J, Cella D, Feeny D, Fischhoff B, Hays RD, Hess R, et al. Selection of key health domains from PROMIS® for a generic preference-based scoring system. *Qual Life Res*. 2017;26(12):3377–85.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

