

# Predicting EQ-5D Index Scores from the PROMIS-29 Profile for the United Kingdom, France, and Germany

Christoph Paul Klapproth (✉ [christoph-paul.klapproth@charite.de](mailto:christoph-paul.klapproth@charite.de))

Charite Universitätsmedizin Berlin <https://orcid.org/0000-0001-6057-5920>

**Jan van Bebber**

Charite Universitätsmedizin Berlin

**Christopher J. Gibbons**

Harvard University

**Jose M Valderas**

University of Exeter

**Leplege Alain**

Universite Paris Diderot Sorbonne

**Matthias Rose**

Charite Universitätsmedizin Berlin

**Felix Fischer**

Charite Universitätsmedizin Berlin

---

## Research

**Keywords:** models for predicting EQ-5D, Patient Reported Outcome Measures Information System, UK, France, Germany

**Posted Date:** June 17th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-34792/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published on December 17th, 2020. See the published version at <https://doi.org/10.1186/s12955-020-01629-0>.

# Abstract

*Background:* EQ-5D health utility (HU) scores are commonly used in health economics to compute quality-adjusted life years (QALYs). EQ-5D scores, which are 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 for predicting EQ-5D scores 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=1,509), France (n=1,501), and Germany (n=1,502). We used stepwise regression with backward selection to find the best models to predict the EQ-5D score from all seven PROMIS-29 domains. We investigated the agreement between the observed and predicted EQ-5D scores in all three countries using various indices for the prediction performance, including Bland-Altman plots to examine the performance along the HU continuum.

*Results:* The EQ-5D index scores were best predicted in Germany ( $RMSE_{GER} = 0.10$ ,  $MAE_{GER} = 0.06$ ), followed by France ( $RMSE_{FR} = 0.11$ ,  $MAE_{FR} = 0.08$ ) and the UK ( $RMSE_{UK} = 0.12$ ,  $MAE_{UK} = 0.09$ ). The Bland-Altman plots show that the inclusion of higher-order effects reduced the underprediction of low HU scores.

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

## Key Points

- We provide mapping from PROMIS-29 profile to EQ-5D index value in the United Kingdom, France, and Germany
- Due to the country specificity of health utility, mapping algorithms for health 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.

## 1 Introduction

### 1.1 The concepts of quality-adjusted life years, health utility, and the EuroQoL

Quality-adjusted life years (QALYs) are routinely used in cost-utility analyses to evaluate the economic effectiveness of health care innovations or interventions(1). QALYs are of particular importance in health technology assessments (HTAs)(2). Both the National Institute of Health and Clinical Excellence (NICE) in England and Wales and the US Panel on Cost-Effectiveness in Health and Medicine have 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 utility (HU) score that represents the value of a particular health state. HU 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 HU scores are patient-reported, preference-based ratings of health-related quality of life (HRQoL)(3). The most frequently used HRQoL measure, the EuroQoL EQ-5D-5L (EQ-5D), covers the following five domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression(4–7). Each of these domains is rated on a five-point scale, thus differentiating 3125 (i.e.,  $5^5$ ) health states. In valuation surveys in the general population of 10 different countries, these health states were ranked and linked to a single EQ-5D index value, expressing country-specific valuations of HRQoL(8–12).

## **1.2 Indirect derivation of individual EQ-5D index values by mapping and the Patient Reported Outcomes Measurements Information System (PROMIS)**

EQ-5D index values for individuals are best obtained directly using the EQ-5D-3L or EQ-5D-5L questionnaire. If direct assessment is not feasible, a common strategy is to estimate HU scores by using a “mapping” or “crosswalk” algorithm from a non-preference-based patient-reported outcome measure (PROM)(13,14). Little consensus exists on which linking method is the most appropriate choice. In a recent systematic review, 147 studies mapping the EQ-5D index values were identified(13). In more than 75% of all mappings in this review, ordinary least squares (OLS) linear regression was used. Although OLS linear regression showed robust results compared to alternative methods, it has several drawbacks(15,16). First, predicted HU 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 HU scores might be non-linear, meaning that the impact of symptoms and/or health domains differs across the HU continuum(16).

Developing a mapping algorithm to link the health domains of the Patient Reported Outcome Measurement Information System (PROMIS) to the EQ-5D index value is of special importance because PROMIS is increasingly used due to its favourable psychometric properties. It constitutes a collection of generic and condition-specific, non-preference-based PROMs that have been developed using item response theory (IRT)(17). 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)(18,19). These item banks can be used to develop tailored short forms or for computerized adaptive testing (CAT)(20). PROMIS overcomes significant limitations of legacy instruments such as ceiling effects and is becoming the reference measurement approach to PROMs(21). Due to the invariance property of PROMIS-29 domains, for each health domain, PROMIS scores are obtained on the same metric, regardless of which item sets have been utilized(18,19).

This property is possible even if a measure has been used that is only linked to one of the PROMIS metrics. Respondents’ item answers can still be placed on that PROMIS health domain. For example, self-reported anxiety measured by MASQ, PANAS and GAD-7 is linked to the PROMIS Anxiety metric(22). Depressive symptoms measured by BDI-2, CES-D, and PHQ-9 can be expressed on the PROMIS Depression metric(23). Therefore,

mapping from PROMIS T-scores to EQ-5D creates the potential to link a broad range of PROMs to HU expressed by the EQ-5D.

Using OLS linear regression on data collected in the US, Revicki (2009) estimated a model to predict EQ-5D index scores from five PROMIS T-scores in the US(24):

For this PROMIS domain model, Revicki reports that approximately 57% (adjusted  $R^2$ ) of the variance in EQ-5D index scores 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 this equation does not work very well for low levels of health (EQ-5D < 0.40). Revicki used the EQ-5D-3L questionnaire and applied the US EQ-5D-3L value set by Shaw (2015)(25). EQ-5D index values and mappings are country-specific(8,26). Revicki's model can therefore only be used to predict EQ-5D scores in the US.

### 1.3 Aims of this study and research questions

As EQ-5D scores are known to be country-specific, the primary aim of this study was to develop mapping functions to link PROMIS-29 to EQ-5D index values for the UK, France, and Germany. For each health domain, we explored the form of its relationship with HU expressed by the EQ-5D and examined whether these relationships would be the same across the three countries under investigation. Furthermore, we investigated whether the optimal models would be structurally equivalent across countries and compared the prediction performance of the final models to the prediction performance of Revicki's model.

## 2 Methods

### 2.1 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 the marginal distributions of sex, age, occupation, region, and population density of the UK (n=1,509), France (n=1,501), and Germany (n=1,502). 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).

We only briefly summarize the most important differences between the three samples here. The interested reader is referred to Table A.1 (Appendix) 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%).

As participants could only proceed through the survey by answering each item, there were no missing data.

### 2.2 Measures

## *PROMIS domains and item banks*

We used the PROMIS-29 v2.0 Profile to assess seven core domains of health: physical function, fatigue, pain, anxiety, depression, sleep disturbance, and the ability to participate in social roles and activities (referred to as participation in the remainder of this article)(27). The visual analogue scale (VAS) item expressing pain intensity on a scale ranging from 0 to 10 was not used in this study. Each domain is assessed with four items, and the domain scores are expressed as T-scores ( $M = 50$  &  $SD = 10$ ) 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(28–31). 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(32). Hence, the predictor scores of self-reported health that we used in this study are invariant with respect to nationality.

## *EQ-5D-5L*

The EQ-5D-5L is a standardized, patient-reported, and preference-based instrument to measure generic health[3–8]. Five health dimensions are involved: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension of the EQ-5D-5L has five levels (i.e., response options): “No problems” (or 1), “Slight problems” (2), “Moderate problems” (3), “Severe problems” (4), and “Extreme problems” (5). These define  $5^5$  or 3125 different health states. The value assigned to each of these health states is determined by so-called value sets, developed by EuroQoL using time trade-off (TTO) and visual analogue scale (VAS) as preference elicitation methods(4,8). The maximum value for a health state is 1.00 or “full health”. The minimum value depends on the value set applied and can be negative, then considered “worse than dead”. For example, a pattern of 11111 is translated to a health state value of 1, while the pattern 54545 may correspond to -0.2. Note that persons in different countries value health states differently, so the EQ-5D index value is country-specific(8,9,11,12,25).

EQ-5D index values can be derived from EQ-5D-5L using either the crosswalk to the 3L value set or using the new 5L value sets(8). Crosswalks to the 3L value sets are available for ten countries, including the US, the UK, France, and Germany(4,8). A 5L value set is available for Germany(12). There is also one for England, which is not equivalent to the UK, and none yet for France(9,10). We therefore used the 3L crosswalk set for all three samples, thereby ensuring comparability among our samples and to Revicki’s model, which used the 3L value set for the US(8,24,25).

## **2.3 Statistical analysis**

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

To obtain a first impression of the form of the relationships among individual health domains and HU and to judge whether the relationships are stable across the three countries under investigation, we plotted the seven

domain scores against health utility in the UK, France, and Germany.

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

We applied stepwise regression with backward selection to find the best models to predict the EQ-5D score for the UK, France, and Germany, starting with full models that incorporated linear, quadratic, and cubic effects for the same seven PROMIS domains as Revicki. Because sociodemographic factors such as age and sex are known to be useful in predicting HU, they were also entered as possible predictors(13).

The Bayesian information criterion (BIC) was used to steer the inclusion and exclusion of predictors in the stepwise regression analyses(33). To minimize the risk of significance by chance, for each model estimated, we used 10-fold cross-validation(34). 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.

The root mean square error (RMSE) and the mean absolute error (MAE) were used as measures of the prediction precision. Note that we deliberately chose to use different criteria than those used by Revicki because measures of precision and bias, such as the RMSE and the MAE, are preferred over either  $R^2$ -based or information-based (AIC and BIC) criteria(35). In addition, we determined the width between the 95% empirical limits of agreement and compared them to the 95% theoretical limits of agreement (i.e.,  $\pm 1.96 * SD(\text{residuals})$ ). To check the prediction performance along the HU continuum, especially for low levels of HU, Bland-Altman plots were used. We used *R* version 3.4.1, IBM SPSS Statistics version 23, and Microsoft Excel version 15 to run the analyses.

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

To the best of our knowledge, as of February 2020, the mapping function reported by Revicki was the only one available for predicting EQ-5D scores from the PROMIS-29(24). Hence, 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 (i) disregards the country specificity of the EQ-5D, (ii) does not utilize the potential predictive value of the PROMIS-29 health domains not used by Revicki, (iii) does not take higher-order effects into account, and in combination with the foregoing, (iii) 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 (M3).

## 3 Results

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

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

**Figure 1.** Relationships among the PROMIS domains and health utility expressed by the EQ-5D score

A number of general conclusions can be drawn from Figure 1. First, with the exception of low levels of physical functioning in France, the relationships among the seven PROMIS domains and HU are comparable across the three European countries employed in our study. 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 HU. Third, all the relationships are in accordance with theoretical expectations. Higher values on the positive PROMIS domains (participation and physical function) correspond to higher HU values, and higher values on the five negative PROMIS domains correspond with lower HU values. Fourth, participation and physical function seem to have the strongest relationship with HU because these curves are the steepest.

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

Recall that we used stepwise regression with backward selection to find optimal models for predicting the EQ-5D scores 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 did not significantly improve the prediction performance were sequentially removed from these models. The final models to optimally estimate the EQ-5D score by the PROMIS-29 for the UK, France, and Germany can be found in Table 1 below (the standardized coefficients are in parentheses). All the models were confirmed by 10-fold cross-validation.

**Table 1.** Optimal models for the UK, France, and Germany

|                                | UK                  | France               | Germany             |
|--------------------------------|---------------------|----------------------|---------------------|
| Constant                       | 2.288E-0            | 2.910E-0             | -1.181E-0           |
| Age                            | 9.590E-4 (0.069)    | -1.372E-3 (-0.107)   |                     |
| Anxiety                        | 1.120E-2 (0.499)    |                      |                     |
| Pain Interference              | -1.773E-1 (-7.479)  |                      |                     |
| Physical Function              | 5.354E-2 (1.881)    | -3.027E-1 (-9.807)   |                     |
| Depression                     |                     |                      | 7.425E-3 (0.404)    |
| Participation                  | 1.334E-2 (0.573)    | 9.415E-2 (3.719)     | 8.834E-2 (4.915)    |
| Anxiety <sup>2</sup>           | -1.227E-4 (-0.604)  |                      |                     |
| Pain Interference <sup>2</sup> | 3.042E-3 (13.970)   | 2.122E-4 (0.900)     |                     |
| Physical Function <sup>2</sup> | -4.853E-4 (-1.566)  | 7.506E-3 (22.864)    | 5.596E-4 (2.581)    |
| Sleep Disturbance <sup>2</sup> |                     | -2.390E-05 (-0.088)  | -1.763E-05 (-0.097) |
| Participation <sup>2</sup>     | -1.061E-4 (-0.460)  | -1.706E-3 (-7.104)   | -1.733E-3 (-9.850)  |
| Anxiety <sup>3</sup>           |                     |                      | -1.480E-07 (-0.070) |
| Depression <sup>3</sup>        | -3.453E-07 (-0.145) | -3.487E-07 (-0.121)  | -8.951E-07 (-0.421) |
| Fatigue <sup>3</sup>           |                     | -2.456E-07 (-0.088)  |                     |
| Pain Interference <sup>3</sup> | -1.769E-05 (-6.852) | -3.697E-06 (-1.270)  | -7.808E-07 (-0.421) |
| Sleep Disturbance <sup>3</sup> | -1.860E-07 (-0.059) |                      |                     |
| Physical Function <sup>3</sup> |                     | -5.805E-05 (-12.841) | -6.865E-06 (-2.300) |
| Participation <sup>3</sup>     |                     | 1.026E-05 (3.471)    | 1.113E-05 (4.998)   |

*Coefficients are displayed as negative exponentials with four digits, beginning with the first non-zero digit of the coefficient. HU 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  $p < .01$ . The standardized regression coefficients are displayed within parentheses.*

The unstandardized coefficients displayed in Table 1 can be used to compute EQ-5D scores from the PROMIS T-scores. 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 (quadratic and cubic effects) appear to be much smaller than those for the linear effects, as the values of the predictor variables (with mean=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 score 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 the coefficients are in agreement with Figure 1. In Figure 1, we plotted the relationship of a single health domain with the EQ-5D score, irrespective of the values in all the other health domains. Instead, the regression coefficients displayed in Table 1 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 HU in the UK, a negative effect on HU in France, and no effect on HU 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 below. HU expressed by the EQ-5D score can be best mapped from the PROMIS-29 in Germany ( $RMSE_{GER} = 0.10$ ,  $MAE_{GER} = 0.06$ ), followed by France ( $RMSE_{FR} = 0.11$ ,  $MAE_{FR} = 0.08$ ) and the UK ( $RMSE_{UK} = 0.12$ ,  $MAE_{UK} = 0.09$ ). 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.

**Table 2.** Prediction performance of the optimal models for the UK, France, and Germany

|                | RMSE | MAE  | 95% theoretical LoA | 95% empirical LoA |
|----------------|------|------|---------------------|-------------------|
| <b>UK</b>      | 0.12 | 0.09 | $\pm 0.25$          | -0.20; 0.17       |
| <b>France</b>  | 0.11 | 0.08 | $\pm 0.23$          | -0.19; 0.17       |
| <b>Germany</b> | 0.10 | 0.06 | $\pm 0.19$          | -0.16; 0.13       |

RMSE: root mean squared error; MAE: mean absolute error; 95% theoretical and empirical levels of agreement.

The prediction performances of the final models along the HU continuum are depicted in the Bland-Altman plots below. Note that especially in the German sample, there are not many respondents with low health utility (EQ-5D < 0.2). Furthermore, prediction performance appears to be slightly better for high levels of HU (EQ-5D > 0.8) than for intermediate or low HU.

**Figure 2.** Bland-Altman plots of the predicted and observed health utility scores for the UK, France, and Germany

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

The differences in the prediction performances between the application of Revicki’s model versus our models are depicted in Table 3 below. The application of Revicki’s model to the data collected in Europe would systematically underestimate the HU score for the UK (-0.10) and for France (-0.09) but not for Germany. As was the case for our models, 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 function are smaller in Germany than for the UK or for France, as indicated by the values of the RMSE, MAE, and empirical LoAs.

**Table 3.** The detrimental effect of using Revicki’s model to map the EQ-5D score from the PROMIS-29 for the UK, France, and Germany

|                |                       | $R^2_{adj}$ | ICC  | Bias  | RMSE | MAE  | 95% theoretical LoA | 95% empirical LoA |
|----------------|-----------------------|-------------|------|-------|------|------|---------------------|-------------------|
| <b>France</b>  | Revicki               | 0.61        | 0.78 | -0.09 | 0.17 | 0.11 | -0.38; 0.20         | -0.38; 0.08       |
|                | Polynomial Regression | 0.72        | 0.85 | 0.00  | 0.11 | 0.08 | ± 0.23              | -0.19; 0.17       |
| <b>Germany</b> | Revicki               | 0.53        | 0.73 | 0.00  | 0.11 | 0.07 | -0.22; 0.22         | -0.18; 0.14       |
|                | Polynomial Regression | 0.64        | 0.80 | 0.00  | 0.10 | 0.06 | ± 0.19              | -0.16; 0.13       |
| <b>UK</b>      | Revicki               | 0.68        | 0.82 | -0.10 | 0.18 | 0.12 | -0.39; 0.19         | -0.39; 0.07       |
|                | Polynomial Regression | 0.74        | 0.86 | 0.00  | 0.12 | 0.09 | ± 0.25              | -0.20; 0.17       |

The last step in our analyses 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 Figure 3 below. The average prediction performance ( $RMSE_{UK}=.13$ ,  $RMSE_{FR}=.13$ , and  $RMSE_{GER}=.10$ ) 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 HU, but the incorporation of quadratic and cubic effects (M3) does improve the prediction performance for low levels of HU. That is, underprediction of these health states is clearly reduced by adding these higher-order effects to the three regression equations.

**Figure 3.** Incremental value of the country-specific regression coefficients, additional health domains, and higher-order effects for predicting the EQ-5D score for the UK, France, and Germany

## 4 Discussion

### 4.1 Summary of main findings

In this paper, we developed optimal models for mapping the EQ-5D index values from the PROMIS-29 for the UK, France, and Germany. In contrast, with Revicki’s model, which was optimized towards valuations of health states in the US, our models can be used to optimally predict HU expressed by the EQ-5D score for the UK, France, and Germany. Furthermore, we showed that the incorporation of higher-order effects into the regression equations substantially reduced the underestimation of low health utilities. The EQ-5D index value can therefore now be predicted from the PROMIS-29 in three major European countries for use in economic evaluations of health interventions. Our results in terms of the RMSE and MAE are well within the limits of what is usually reported for

mapping algorithms(13,36–40). The global underestimation of the predicted EQ-5D values in OLS has also been reported in dialysis patients(41).

We furthermore demonstrated that the application of a foreign model, in this case, the application of a 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 if a country-specific mapping algorithm is not available. 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.

However, researchers should be aware that the consequences of working with a suboptimal mapping algorithm derived abroad can be substantial: Pennington and Davis demonstrated that the 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(42). NICE has adopted a threshold of GBP 30,000 per QALY(43) representing the public's maximum additional willingness to pay for a new treatment or a new drug compared to the existing standard of care. Consequently, imprecise mapping methods have a great impact on what innovations are made available to patients.

## **4.2 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 health states. Furthermore, our models allowed judgement of the incremental value of incorporating two additional health domains and higher-order effects for HU 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 HU score greater than one, which is a value that is impossible by definition of HU(16). In our study, the risk of predicting HU values greater than one is circumvented by incorporation of non-linear trends.

## **4.3 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 health states. Therefore, we are planning to replicate our findings with data collected from spine patients who were assessed before surgery. In this study, we will also investigate whether regressing the EQ-5D dimensions on the PROMIS domain scores first and then calculating the EQ-5D index values from the EQ-5D dimensions has incremental value(44).

PROMIS data can also be used to estimate a new preference-based HU score: Hanmer developed the PROMIS Preference Score (PROPr) to compute HU for QALYs directly from 7 PROMIS health domains: cognition, depression, fatigue, pain, physical function, sleep disturbance, and ability to participate in social roles and activities (in this paper, this domain is referred to as participation)(45–49). Note that these 7 PROMIS domains are not entirely equivalent with those 7 domains from the PROMIS-29 profile (anxiety is missing in the PROPr,

while cognition is missing in the PROMIS-29). The PROPr was valuated in US preferences using the standard gamble method (SG), while the EQ-5D uses TTO(10,25,47,50).

The PROPr could potentially be used instead of the EQ-5D index in cost-effectiveness analyses. Since many European HTA authorities such as NICE specifically demand the use of the well-established EQ-5D index value to measure HU in cost-effectiveness analyses, mapping the PROMIS-29 to the EQ-5D will still be needed(43). Furthermore, although PROMIS is being used more frequently in the US, it has not yet attained a dominant role in HRQoL assessments in Europe. In addition, as of February 2020, there is no PROPr value set for European preferences(47,48).

## 5 Conclusion

Our mapping functions can be used to predict EQ-5D index values from the PROMIS-29 for cost-utility analyses in health technology assessments in the UK, France and Germany. The inclusion of polynomial regression terms decreases the prediction bias for lower health states.

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 HU 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 underprediction of low health states.

## Declarations

Compliance with ethical standards

Funding: This study was funded by the Centre Virchow-Villerme.

Conflict of interest: Authors declare that they have no conflict of interest.

Ethical approval: 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: Informed consent was obtained from all individual participants included in the study.

## References

1. Weinstein MC, Torrance G, McGuire A. QALYs: The basics. *Value Heal* [Internet]. 2009;12(SUPPL. 1):S5–9. Available from: <http://dx.doi.org/10.1111/j.1524-4733.2009.00515.x>
2. 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>

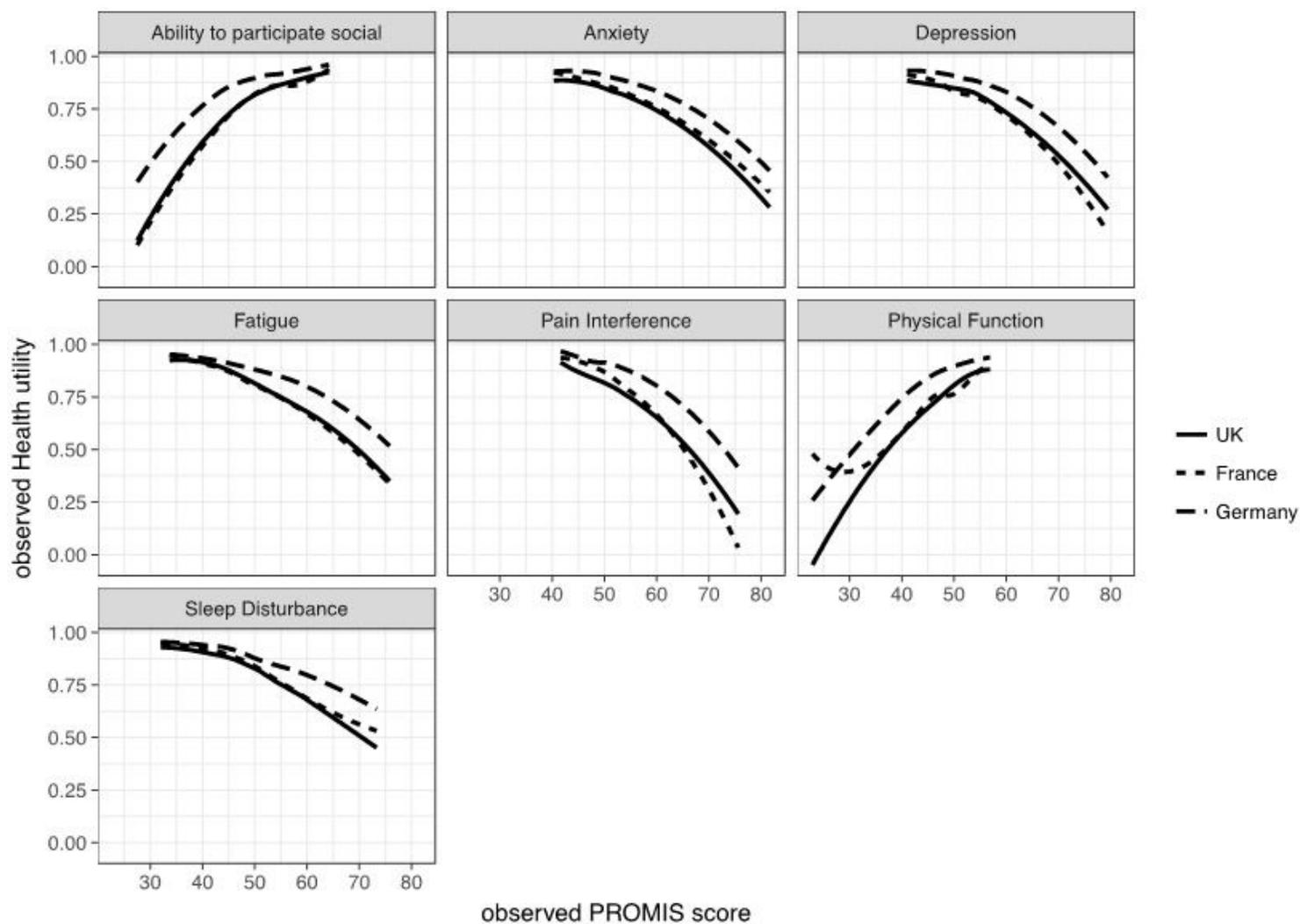
3. Valderas JM, Alonso Jo. Patient reported outcome measures : a model-based classification system for research and clinical practice. *Qual Life Res.* 2008;(17):1125–35.
4. Rabin R, Oemar M, Oppe M, Janssen B, Herdman M. EQ-5D-5L User Guide Version 2.1. 2015;(April):28. Available from: [http://www.euroqol.org/fileadmin/user\\_upload/Documenten/PDF/Folders\\_Flyers/EQ-5D-5L\\_UserGuide\\_2015.pdf](http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf)
5. Greiner W, Weijnen T, Nieuwenhuizen M, N, Oppe S, Badia X, et al. A single European currency for EQ-5D health states. *Eur J Heal Econ.* 2003;(4):222–31.
6. 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.
7. 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.
8. 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. Available from: <http://dx.doi.org/10.1016/j.jval.2012.02.008>
9. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Heal Econ (United Kingdom).* 2018;27(1):7–22.
10. Chevalier J, De Pouvourville G. Valuing EQ-5D using Time Trade-Off in France. *Eur J Heal Econ.* 2013;14(1):57–66.
11. Pickard AS, Law EH, Jiang R, Pullenayegum E, Shaw JW, Xie F, et al. United States Valuation of EQ-5D-5L Health States Using an International Protocol. *Value Heal [Internet].* 2019;22(8):931–41. Available from: <https://doi.org/10.1016/j.jval.2019.02.009>
12. Ludwig K, Schulenburg JG Von Der, Greiner W, Ludwig K. German Value Set for the EQ-5D-5L. *Pharmacoeconomics [Internet].* 2018;36(6):663–74. Available from: <https://doi.org/10.1007/s40273-018-0615-8>
13. 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. Available from: <https://doi.org/10.1007/s40258-019-00467-6>
14. 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. Available from: <http://hqlo.biomedcentral.com/articles/10.1186/1477-7525-11-151>
15. Crott R. Direct Mapping of the QLQ-C30 to EQ-5D Preferences: A Comparison of Regression Methods. *Pharmacoeconomics - Open.* 2018;2(2):165–77.
16. 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. Available from: <http://dx.doi.org/10.1016/j.jval.2011.12.014>
17. Embretson SE, Reise SP. *Item Response Theory For Psychologists.* Psychology Press; 2013.
18. PROMIS Cooperative Group. PROMIS ® 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)

19. Rupp AA, Zumbo BD. Understanding parameter invariance in unidimensional IRT models. *Educ Psychol Meas.* 2006;66(1):63–84.
20. 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.
21. Hays RD, Revicki DA, Feeny D, Fayers 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.
22. 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].* 2013/12/01. 2014 Jan;28(1):88–96. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24508596>
23. 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):513–527.
24. Revicki DA, Kawata AK, Harnam 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.
25. 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.
26. 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. Available from: <http://dx.doi.org/10.1007/s11136-018-1840-5>
27. Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, et al. PROMIS® Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. *Value Heal.* 2019;22(5):537-544.
28. 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.
29. 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>
30. 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.
31. 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. Available from: <https://doi.org/10.1007/s00408-016-9850-y>
32. 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. Available from: <http://dx.doi.org/10.1007/s11136-018-1785-8>
33. 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 Jun;17(2):228–43. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/22309957>
34. Blum, A., Kalai, A., Langford J. Beating the Holdout: Bounds for KFold and Progressive Cross-Validation. *COLT*. 1999;203–208.
  35. Brazier JE, Yang Y, Tsuchiya A, Rownen 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.
  36. 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.
  37. 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. Available from: <https://www.tandfonline.com/doi/full/10.1080/13696998.2016.1241788>
  38. 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]).
  39. Beck AJCC, Kieffer JM, Retèl VP, van Overveld LFJ, 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.
  40. 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. Available from: <https://doi.org/10.1007/s10198-019-01088-5>
  41. Yang F, Devlin N, Luo N. Impact of mapped EQ-5D utilities on cost-effectiveness analysis: in the case of dialysis treatments. *Eur J Heal Econ [Internet]*. 2019;20(1):99–105. Available from: <http://dx.doi.org/10.1007/s10198-018-0987-x>
  42. 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. Available from: <http://dx.doi.org/10.1016/j.jval.2014.11.002>
  43. NICE. Guide to the Methods of Technology Appraisal [Internet]. NICE Guidelines. 2013. Available from: [nice.org.uk/process/pmg9](http://nice.org.uk/process/pmg9)
  44. 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.
  45. 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. Available from: <http://dx.doi.org/10.1186/s12955-015-0321-6>
  46. 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;1–9.
  47. Hanmer J, Dewitt B. The Development of a Preference-based Scoring System for PROMIS® (PROPr): A Technical Report Version 1.4 [Internet]. 2017. Available from: <https://creativecommons.org/licenses/by-nc-sa/4.0/>
  48. 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. Available from: <http://dx.doi.org/10.1007/s11136-018-1852-1>

49. 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.
50. 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.

## Figures



**Figure 1**

Relationships among the PROMIS domains and health utility expressed by the EQ-5D score

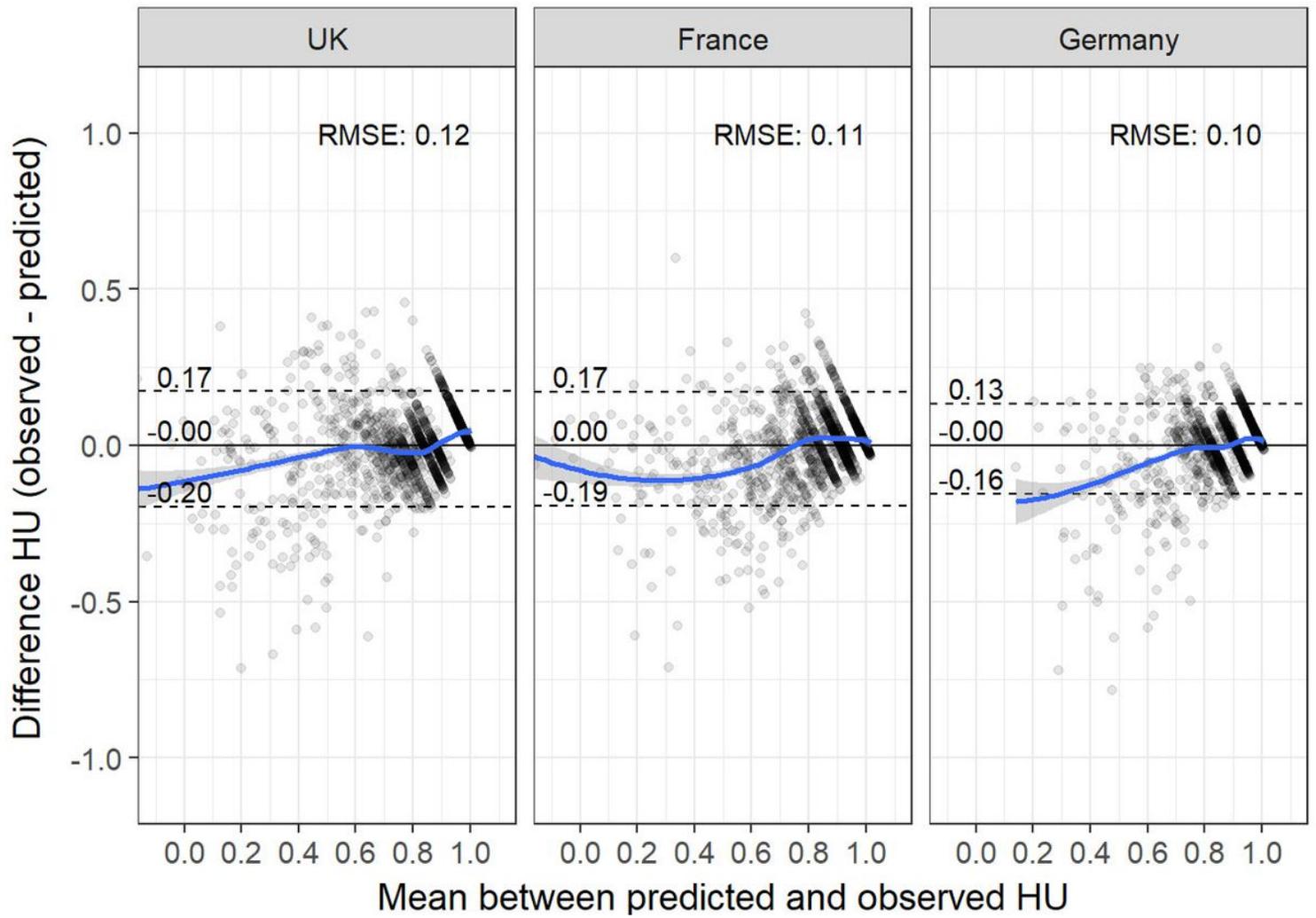
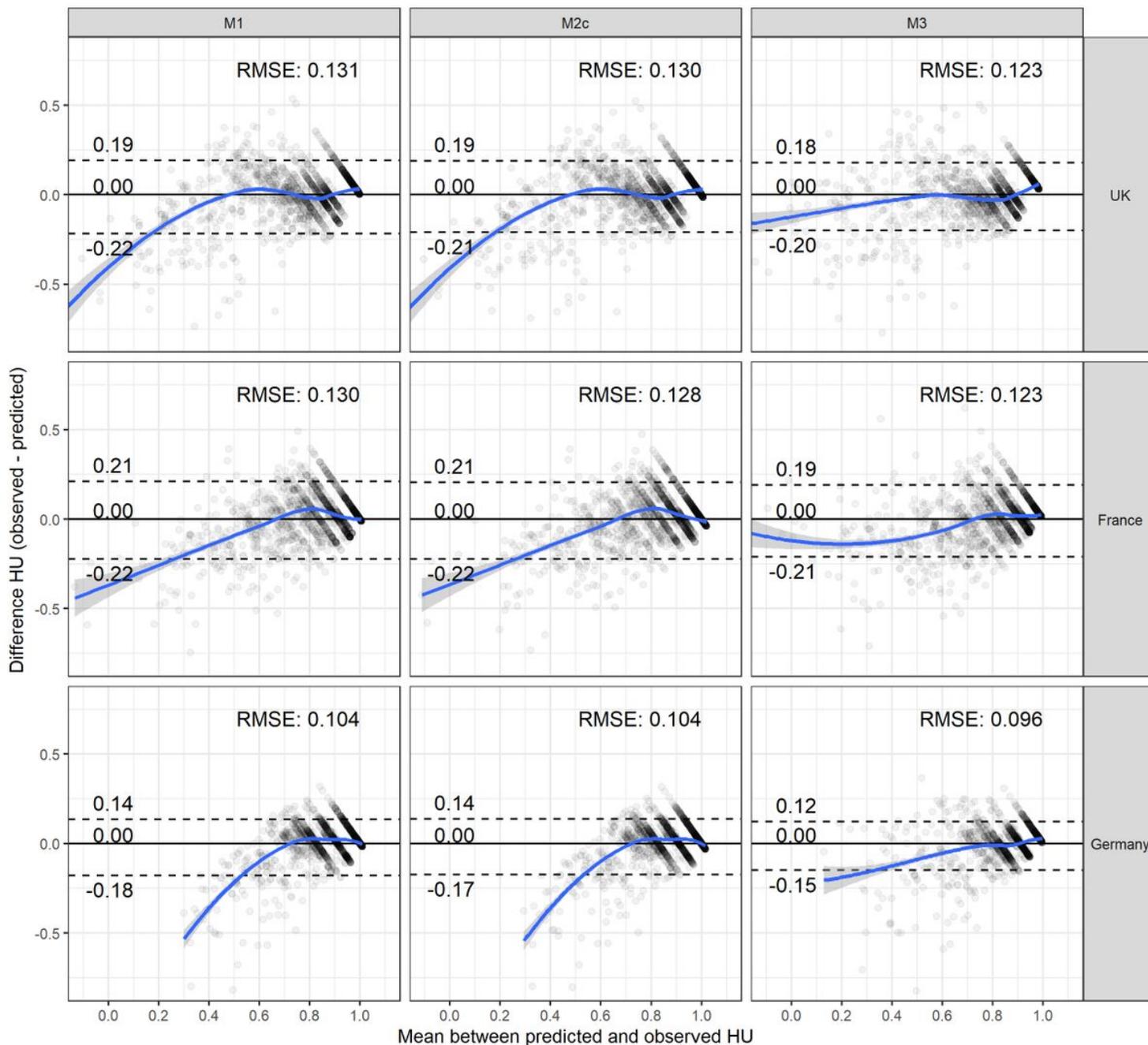


Figure 2

Bland-Altman plots of the predicted and observed health utility scores for the UK, France, and Germany



**Figure 3**

Incremental value of the country-specific regression coefficients, additional health domains, and higher-order effects for predicting the EQ-5D score for the UK, France, and Germany

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableA1.docx](#)