



Mapping from SIBDQ to EQ-5D-5L for patients with inflammatory bowel disease

Isa Maria Steiner¹ · Bernd Bokemeyer² · Tom Stargardt¹

Received: 7 September 2022 / Accepted: 31 May 2023 / Published online: 27 June 2023
© The Author(s) 2023

Abstract

Objective Clinical studies commonly use disease-specific measures to assess patients' health-related quality of life. However, economic evaluation often requires preference-based utility index scores to calculate cost per quality-adjusted life-year (QALY). When utility index scores are not directly available, mappings are useful. To our knowledge, no mapping exists for the Short Inflammatory Bowel Disease Questionnaire (SIBDQ). Our aim was to develop a mapping from SIBDQ to the EQ-5D-5L index score with German weights for inflammatory bowel disease (IBD) patients.

Methods We used 3856 observations of 1055 IBD patients who participated in a randomised controlled trial in Germany on the effect of introducing regular appointments with an IBD nurse specialist in addition to standard care with biologics. We considered five data availability scenarios. For each scenario, we estimated different regression and machine learning models: linear mixed-effects regression, mixed-effects Tobit regression, an adjusted limited dependent variable mixture model and a mixed-effects regression forest. We selected the final models with tenfold cross-validation based on a model subset and validated these with observations in a validation subset.

Results For the first four data availability scenarios, we selected mixed-effects Tobit regressions as final models. For the fifth scenario, mixed-effects regression forest performed best. Our findings suggest that the demographic variables age and gender do not improve the mapping, while including SIBDQ subscales, IBD disease type, BMI and smoking status leads to better predictions.

Conclusion We developed an algorithm mapping SIBDQ values to EQ-5D-5L index scores for different sets of covariates in IBD patients. It is implemented in the following web application: <https://www.bwl.uni-hamburg.de/hcm/forschung/mapping.html>.

Keywords Short Inflammatory Bowel Disease Questionnaire · SIBDQ · EQ-5D · EQ-5D-5L · Mapping · Preference-based measure

JEL Classification I1 · C3

Introduction

The prevalence of inflammatory bowel disease (IBD), including Crohn's disease (CD), ulcerative colitis (UC) and indeterminate colitis (IC), has been rising in high-income countries since the mid-twentieth century [1]. In Europe,

approximately 0.4% of the population has an IBD [2]. First diagnosis of the disease usually occurs in early adulthood [3], and there is currently no cure for it. The goal of treatment is to maintain health-related quality of life and avoid disability, which implies inducing remission in the short term and maintaining it in the long term [4, 5].

The number of treatment options for IBD patients has increased substantially in recent years. In particular, advanced medical therapies, such as biologics, have facilitated the management of patients with complicated disease courses [6]. Many of the new treatment options are, however, very costly, raising questions about their affordability for health care systems. Health economic evaluations are

✉ Isa Maria Steiner
isa.maria.steiner@uni-hamburg.de

¹ Hamburg Center for Health Economics, University of Hamburg, Esplanade 36, 20354 Hamburg, Germany

² Interdisziplinäres Crohn Colitis Centrum Minden, Märchenweg 17, 32429 Minden, Germany

a useful tool for weighing the health outcomes of different treatment options against their associated costs. The most frequently used health measure for this purpose is the quality-adjusted life-year (QALY), which seeks to reflect both the length and quality of a patient's life [7].

Calculating QALYs requires preference-based measures, such as the EQ-5D-5L index score, as input parameters [8]. Because such index scores are generic utility measures, they can be compared across different disease types. This generalisability often comes at the price of a lower sensitivity than disease-specific quality of life measures [9–11], which is why the latter are often preferred in clinical trials. However, the results of clinical studies that report only disease-specific measures are of limited use for cost-utility analysis. In such situations, mapping from the disease-specific quality of life measure to a utility index score can help overcome this problem.

In the context of IBD, the Inflammatory Bowel Disease Questionnaire (IBDQ) and its short version, the Short Inflammatory Bowel Disease Questionnaire (SIBDQ), are the most widely used disease-specific quality of life instruments for adult patients [12]. While a mapping already exists for the IBDQ [13], to our knowledge, no study has been published so far that maps the SIBDQ to a utility index score. Our aim was to close this gap by developing an algorithm for mapping values from the SIBDQ to EQ-5D-5L index scores. To do so, we applied regression and machine learning models for different sets of covariates, thereby taking differences in data availability into account. Moreover, we developed and provide an online tool that allows for a user-friendly application of the final mapping algorithm.

Methods

This mapping study complies with the Mapping onto Preference-Based Measures Reporting Standards (MAPS) checklist [14]. Details are provided in Online Appendix 2.

Data

Our study is based on data from the German multicentre randomised controlled trial CED_{Bio-Assist}. The trial investigated the impact of introducing regular appointments with an IBD nurse specialist on patients' quality of life and other outcome parameters among IBD patients in addition to standard care with biologics. It was implemented under real-life conditions.

Ethical approval was obtained. The registration ID in the German register for clinical trials is DRKS00020265. To be eligible for inclusion, patients had to be diagnosed with an IBD, at least 18 years old and either receiving an ongoing biologic therapy or scheduled to start a biologic therapy after

inclusion. Inclusion took place between 6 January 2020 and 18 January 2021. Patient-reported outcomes and disease-related information were collected via patient and physician questionnaires at baseline, as well as 6, 12 and 18 months after baseline. This time interval corresponded to the usual distance between appointments in standard care of IBD patients in Germany. Remote and on-site monitoring took place to optimise data quality.

The data set contained 3999 observations of 1066 patients. After we deleted observations with missing values for any of the variables considered in the study, the data set contained 3856 observations of 1055 patients. The data was then randomly split into a model subset (80%) with 3066 observations of 844 patients and a validation subset (20%) with 790 observations of 211 patients.

SIBDQ

The SIBDQ is the short version of the 32-item IBDQ [15, 16]. It consists of 10 items belonging to four different domains: systemic, social, bowel and emotional. Each item is answered on a Likert-type response scale with 1 and 7 representing the lowest and highest quality of life, respectively. We defined the general score as the mean of the scores of all 10 items. Moreover, we calculated subscales for each of the four domains as the mean of the scores of the respective items. The SIBDQ is available in multiple languages, including English, German, Spanish, Portuguese, French and Mandarin. Validation studies showed good reliability, validity and responsiveness [12, 17–19].

EQ-5D-5L

The EQ-5D-5L is a generic instrument that measures health-related quality of life. Its major components are five Likert-type response scales between 1 and 5 in each dimension. There are three functional dimensions (mobility, self-care and usual activities) and two somatic symptom dimensions (pain/discomfort and anxiety/depression) [20]. The responses to the five scales can be translated into a single utility index score. This index score is based on a country-specific value set, which is usually obtained with a discrete choice experiment, time trade-off or standard gamble approach. The German value set, which is the basis for our analysis, ranges from -0.661 to 1 [21]. In general, the distribution of the index score is skewed to the left with many observations being close or equal to 1 [22]. While there are still few studies assessing the responsiveness of the EQ-5D-5L, its reliability and validity have been assessed in multiple publications, suggesting that it can be applied in a wide range of populations and settings [23].

Covariates considered for mapping

The main goal in mapping studies is to estimate the relationship between two quality of life measures. However, there are other factors that might affect health-related quality of life and therefore improve predictions and generalisability when included in the estimation process [24]. In the case of IBD, female gender and age were found to be negatively associated with EQ-5D-5L index scores in a representative population sample in Germany [25]. Moreover, the body mass index (BMI) is considered to affect health-related quality of life, especially in the presence of obesity [26]. Smoking status has different implications across IBD subgroups: while continued smoking is a risk factor for adverse outcomes in CD patients, it may be protective against adverse outcomes in UC patients [27]. Finally, previous studies indicate that quality of life may improve over time after an IBD has been diagnosed [9].

To allow for mapping depending on data availability, we assumed five different scenarios: (1) Only the general SIBDQ score is mapped to EQ-5D-5L; (2) The general SIBDQ score and the demographic variables age and gender are available for mapping; (3) Additionally, the patient's BMI, smoking status and type of IBD (i.e., CD vs. UC or IC) are known; (4) In addition to the variables included in scenario 3, the number of years since the patient's first IBD diagnosis is known; (5) In addition to the variables included in scenario 4, the SIBDQ subscales are available for mapping.

Estimated models

Since our data contained repeated measurements, we estimated the models considered for mapping with random intercepts by individual. The following four approaches were used to map the SIBDQ to the EQ-5D-5L: First, we estimated linear mixed-effects regressions (LMER) because linear regression is the most common approach in studies mapping disease-specific measures to EQ-5D index scores [28]. Second, we ran a mixed-effects Tobit regression, which is another common approach for estimating utility index scores [28, 29]. Tobit regression is an extension of classical regression approaches that was designed for dealing with limited dependent variables [30]. Third, we ran an adjusted limited dependent variable mixture model (ALDVMM), which was specifically developed to estimate regressions with the EQ-5D index as dependent variable [31]. The ALDVMM was estimated with one component. The lower and upper limits were set to -0.661 and 0.974 , as they represent the lowest and second highest value in the German value set respectively. In contrast to the other models, we did not estimate random intercepts for ALDVMM, as this option is currently neither available in the R nor in the Stata package.

We did, however, estimate cluster-robust standard errors, as suggested by Oliveira Gonçalves et al. [32]. Finally, we estimated our mapping algorithm using a mixed-effects regression forest (MERF). MERF consist of a random effect and a fixed part of the model, where the latter is estimated with a regression forest [33, 34]. Regression forest is an approach from machine learning that does not require an assumption about the distribution of the dependent variable or the functional form of the relationship between outcome and explanatory variables. It consists of many regression trees that split random samples of the data into different segments and make local predictions of the outcome variable. The final prediction for a given set of predictors is then obtained by averaging the predictions across multiple regression trees [35]. Online Appendix 2 contains a more detailed description of the estimated models.

Because it is not easy to decide whether and in which form specific variables should be included in regression models, e.g., whether quadratic, cubic or interaction terms are sensible, we specified LMER, Tobit and ALDVMM for each scenario in two different ways: the first (LMER), third (Tobit) and fifth (ALDVMM) method consisted of parametrising the model for each data availability scenario without polynomials and interaction terms. As a second (LMER'), fourth (Tobit') and sixth (ALDVMM') method, we applied the 'lasso' technique developed by Tibshirani et al. [36] to select variables that included third order polynomials of all continuous variables, as well as interaction terms of all variables. Thus, the models OLS', Tobit' and ALDVMM' were specified with the variables selected by the lasso. We did not use lasso in addition to the last method (MERF) because the latter already conducts variable selection by definition. As a result, we estimated seven models (LMER, LMER', Tobit, Tobit', ALDVMM, ALDVMM' and MERF) for each of the five sets of covariates as described above (see Table 1).

Model selection and validation

For model selection and estimation, only the model subset was used. We selected one final model for each covariate scenario based on tenfold cross-validation, which has become increasingly popular for mapping studies in recent years (e.g., [37–41]). We thus randomly split the sample into 10 folds, i.e., equally sized parts of the data set. Subsequently, we estimated each of the models described in Table 1 based on the combined data of nine folds. We calculated performance measures based on the tenth fold, called the test set. Because there is no single measure that is clearly preferred in the literature to evaluate the performance of a model, we calculated multiple measures. More specifically, we chose the mean squared error (MSE), the mean absolute error (MAE) and the coefficient of determination R^2 . We

Table 1 Description of the estimated models

Covariates	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Method 7
	LMER	LMER'	Tobit	Tobit'	ALDVMM	ALDVMM'	MERF
1 General SIBDQ	Model 1.1	Model 2.1	Model 3.1	Model 4.1	Model 5.1	Model 6.1	Model 7.1
2 General SIBDQ, age, gender	Model 1.2	Model 2.2	Model 3.2	Model 4.2	Model 5.2	Model 6.2	Model 7.2
3 General SIBDQ, age, gender, BMI, smoking status, IBD type	Model 1.3	Model 2.3	Model 3.3	Model 4.3	Model 5.3	Model 6.3	Model 7.3
4 General SIBDQ, age, gender, BMI, smoking status, IBD type, years since first diagnosis	Model 1.4	Model 2.4	Model 3.4	Model 4.4	Model 5.4	Model 6.4	Model 7.4
5 General SIBDQ, age, gender, BMI, smoking status, IBD type, years since first diagnosis, SIBDQ subscales	Model 1.5	Model 2.5	Model 3.5	Model 4.5	Model 5.5	Model 6.5	Model 7.5

ALDVMM adjusted limited dependent variable mixture model, *BMI* body mass index, *IBD* inflammatory bowel disease, *LMER* linear mixed-effects regression, *MERF* mixed-effects regression forest, *SIBDQ* Short Inflammatory Bowel Disease Questionnaire

also considered the range of predicted values to compare different models. The process was iterated until each of the 10 folds had served as a test set once. We then compared performance measures and selected the best performing model for each scenario, which subsequently was estimated using all observations in the model subset. A model was considered best if it outperformed the other models on at least two of the three performance measures. For a more detailed description of the estimation procedure of predicted utilities and a definition of the performance measures, refer to Online Appendix 2.

As a robustness check, we ran the best-performing models without some of the variables added at earlier stages (e.g., a scenario (5) model was run excluding age and gender). In the last step, we validated the predictions of these models using the validation subset.

All statistical analyses were performed in R version 4.2.3. The code can be found in Online Appendix 1. Moreover, we programmed a web application using the R package Shiny [42], for which more information is provided in Online Appendix 2.

Results

Description of the sample and conceptual overlap

Table 2 summarises the characteristics of the 1055 patients in the data set at baseline. Of these patients, 592 were diagnosed with CD and 448 with UC. The remaining 15 patients were diagnosed with IC, i.e., they may have presented with symptoms of both CD and UC. The mean age was 40.97 years and 52.42% of all IBD patients were female, which is comparable to other studies in Germany [43, 44]. On average, patients' first IBD diagnosis dated back 12.15 years. The mean BMI of 25.63 was similar to the BMI of the total population in Germany [45]. Smoking was more

Table 2 Patient characteristics at baseline

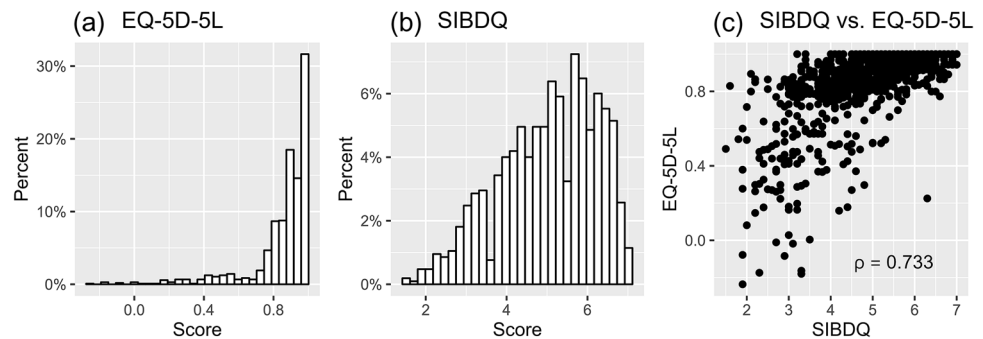
	CD (n=592)	UC (n=448)	IBD (n=1055)
	Ø (sd)/share	Ø (sd)/share	Ø (sd)/share
Age in years	40.73 (13.59)	41.32 (14.00)	40.97 (13.90)
Gender			
Female	57.77%	44.87%	52.42%
Male	42.23%	55.13%	47.58%
Currently smoking	24.16%	8.05%	17.17%
BMI	25.30 (5.33)	26.08 (5.42)	25.63 (5.36)
Years since initial diagnosis	13.48 (10.78)	10.48 (9.15)	12.15 (10.21)
EQ-5D-5L index score	0.84 (0.20)	0.88 (0.16)	0.86 (0.19)
SIBDQ			
General score	4.97 (1.18)	5.08 (1.24)	5.02 (1.21)
Bowel subscale	5.19 (1.26)	5.18 (1.39)	5.18 (1.32)
Emotional subscale	4.70 (1.42)	4.89 (1.37)	4.78 (1.40)
Social subscale	5.54 (1.58)	5.59 (1.65)	5.55 (1.61)
Systemic subscale	4.48 (1.43)	4.72 (1.45)	4.58 (1.45)

BMI body mass index, *CD* Crohn's disease, *IBD* inflammatory bowel disease, *SIBDQ* Short Inflammatory Bowel Disease Questionnaire, *sd* standard deviation, *UC* ulcerative colitis

common in patients with CD than in patients with UC, a finding that has also been found in other studies [46].

The EQ-5D-5L index had a mean of 0.86 and a standard deviation of 0.19. The values ranged from -0.236 to 1. Thus, no patient reached the lowest score possible. There was a strong ceiling effect, with 24.59% of all patients having an EQ-5D-5L index score of 1. The distribution had a skewness of -2.58 and a kurtosis of 10.75, indicating that it was highly left skewed with few observations in the lower part. This is shown in Fig. 1a. Overall, the distribution was similar to that in other studies reporting EQ-5D-5L values for IBD patients [47–50].

Fig. 1 **a** Distribution of the EQ-5D-5L index score at baseline, **b** Distribution of the general SIBDQ score at baseline, **c** Scatterplot of the general SIBDQ score against the EQ-5D-5L index score at baseline; *SIBDQ* Short Inflammatory Bowel Disease Questionnaire, ρ Spearman correlation coefficient



The general SIBDQ score had a mean value of 5.02 and a standard deviation of 1.21, which is comparable to the values in other studies reporting this measure [50–53]. With the score ranging from 1.5 to 7, the lowest possible value was not reached. Approximately 1% of all patients had a general SIBDQ score of 7. Thus, the ceiling effect was smaller than that in the EQ-5D-5L. The distribution had a skewness of -0.49 , meaning that it was only slightly skewed to the left. This can be seen in Fig. 1b. Figure 1c shows a scatterplot of the general SIBDQ score against the EQ-5D-5L index score at baseline. The Spearman correlation coefficient of 0.73 was significant ($p < 0.001$), indicating a strong positive relationship between the two measures.

Mapping algorithms

Table 3 shows the performance measures of the different models after tenfold cross-validation. The average predicted minimum (maximum) of LMER and LMER' was above 0.55 (1.02). Tobit, Tobit', ALDVMM, ALDVMM' and MERF had a similar predicted range, with an average predicted minimum (maximum) slightly below 0.5 (1.0) across most models. Model 7.5 (MERF with all considered covariates) resulted in the widest range with an average lowest (highest) predicted value of 0.38 (0.99). For data availability scenarios 1–4, Tobit performed the best in terms of MSE, MAE and R^2 . For data availability scenario 5, MERF yielded the best performance measures.

The introduction of more covariates did not always improve predictions. As can be seen from our robustness checks in Online Appendix 2, performance measures improved when we did not include age and gender in the final models. Moreover, excluding years since initial diagnosis led to better results in terms of MSE, MAE and R^2 .

Based on this, we chose the following three final models:

- Final model 3.1, which is a mixed-effects Tobit regression with the general SIBDQ as the only covariate, was chosen for data availability scenarios 1 (only the general

SIBDQ score is available) and 2 (general SIBDQ score, age and gender are available).

- Final model 3.3', which is a mixed-effects Tobit regression with the covariates general SIBDQ score, BMI, smoking status and disease type, was chosen for data availability scenarios 3 (general SIBDQ score, BMI, smoking status and disease type) and 4 (BMI, smoking status, disease type and disease duration).
- Final model 7.5', which is a MERF estimated with the covariates general SIBDQ score, BMI, smoking status, disease type and SIBDQ subscales, was chosen for data availability scenario 5.

In Fig. 2, the EQ-5D-5L utility index scores predicted by our three final models are plotted against observed scores in the validation subset. While the Spearman correlation coefficient resulting from model 7.5' was lower than those obtained from the other final models, MSE, MAE and R^2 improved consistently with the number of covariates. Moreover, all performance measures were better than those obtained from tenfold cross-validation. Regression results and variable importance measures of the three final models can be found in Online Appendix 2.

The following link leads to the web application, which allows our final models to be applied for mapping: <https://www.bwl.uni-hamburg.de/hcm/forschung/mapping.html>.

Discussion

The aim of our study was to develop a mapping algorithm that allows EQ-5D-5L estimates for health economic evaluation to be obtained from the disease-specific SIBDQ. To do so, we compared different direct mapping approaches, including LMER, mixed-effects Tobit regression, ALDVMM and MERF for different sets of covariates. We used data from 1055 adult IBD patients who had enrolled in a German randomised controlled trial.

Overall, there was a relatively strong correlation between SIBDQ and EQ-5D-5L, resulting in a Spearman correlation coefficient of 0.733 at baseline. We concluded that there was

Table 3 Performance measures from tenfold cross-validation

Model	Method	Performance measures						Predicted range	
		MSE		MAE		R2		Min	Max
		Mean	sd	Mean	sd	Mean	sd	Mean	Mean
model 1.1	LMER	0,0184	0,0065	0,0807	0,0097	0,4039	0,0889	0,5641	1,0356
model 2.1	LMER'	0,0184	0,0065	0,0807	0,0097	0,4039	0,0889	0,5641	1,0356
model 3.1	Tobit	0,0175	0,0062	0,0763	0,0089	0,4287	0,1015	0,4870	0,9926
model 4.1	Tobit'	0,0175	0,0062	0,0763	0,0089	0,4287	0,1015	0,4870	0,9926
model 5.1	ALDVMM	0,0176	0,0061	0,0806	0,0092	0,4253	0,1089	0,4583	0,9862
model 6.1	ALDVMM'	0,0176	0,0061	0,0806	0,0092	0,4253	0,1089	0,4583	0,9862
model 7.1	MERF	0,0176	0,0062	0,0773	0,0084	0,4267	0,1067	0,4262	0,9878
model 1.2	LMER	0,0183	0,0065	0,0810	0,0097	0,4042	0,0931	0,5627	1,0478
model 2.2	LMER'	0,0184	0,0066	0,0808	0,0098	0,4029	0,0910	0,5647	1,0371
model 3.2	Tobit	0,0175	0,0062	0,0767	0,0093	0,4288	0,1076	0,4850	0,9945
model 4.2	Tobit'	0,0176	0,0063	0,0764	0,0091	0,4277	0,1037	0,4879	0,9928
model 5.2	ALDVMM	0,0176	0,0061	0,0807	0,0094	0,4251	0,1140	0,4588	0,9886
model 6.2	ALDVMM'	0,0176	0,0062	0,0807	0,0093	0,4245	0,1107	0,4594	0,9865
model 7.2	MERF	0,0179	0,0065	0,0786	0,0094	0,4193	0,1046	0,4747	0,9890
model 1.3	LMER	0,0182	0,0063	0,0815	0,0095	0,4085	0,0935	0,5615	1,0587
model 2.3	LMER'	0,0184	0,0065	0,0807	0,0097	0,4039	0,0889	0,5641	1,0356
model 3.3	Tobit	0,0173	0,0061	0,0766	0,0091	0,4367	0,1073	0,4845	0,9956
model 4.3	Tobit'	0,0175	0,0062	0,0763	0,0089	0,4287	0,1015	0,4870	0,9926
model 5.3	ALDVMM	0,0173	0,0060	0,0804	0,0093	0,4327	0,1132	0,4616	0,9899
model 6.3	ALDVMM'	0,0176	0,0062	0,0807	0,0093	0,4241	0,1114	0,4595	0,9865
model 7.3	MERF	0,0178	0,0063	0,0812	0,0091	0,4205	0,0986	0,4849	0,9695
model 1.4	LMER	0,0182	0,0063	0,0816	0,0095	0,4077	0,0936	0,5610	1,0593
model 2.4	LMER'	0,0184	0,0065	0,0807	0,0097	0,4039	0,0889	0,5641	1,0356
model 3.4	Tobit	0,0173	0,0061	0,0765	0,0091	0,4365	0,1071	0,4831	0,9957
model 4.4	Tobit'	0,0175	0,0062	0,0763	0,0089	0,4287	0,1015	0,4870	0,9926
model 5.4	ALDVMM	0,0173	0,0061	0,0803	0,0093	0,4327	0,1130	0,4605	0,9900
model 6.4	ALDVMM'	0,0176	0,0061	0,0806	0,0092	0,4253	0,1089	0,4583	0,9862
model 7.4	MERF	0,0174	0,0060	0,0791	0,0091	0,4318	0,1103	0,4275	0,9911
model 1.5	LMER	0,0176	0,0061	0,0798	0,0089	0,4277	0,0904	0,5560	1,0515
model 2.5	LMER'	0,0178	0,0064	0,0787	0,0092	0,4233	0,0878	0,5634	1,0270
model 3.5	Tobit	0,0167	0,0058	0,0758	0,0086	0,4531	0,1046	0,4830	0,9958
model 4.5	Tobit'	0,0170	0,0061	0,0754	0,0088	0,4450	0,1021	0,4942	0,9926
model 5.5	ALDVMM	0,0167	0,0057	0,0793	0,0089	0,4515	0,1100	0,4625	0,9900
model 6.5	ALDVMM'	0,0170	0,0060	0,0794	0,0090	0,4444	0,1084	0,4671	0,9858
model 7.5	MERF	0,0163	0,0057	0,0756	0,0084	0,4661	0,1112	0,3745	0,9905

Best performing models for each set of covariates are marked in grey

MAE mean absolute error, *MSE* mean squared error, *LMER* linear mixed effects regression, *ALDVMM* adjusted limited dependent variable mixture model, *MERF* Mixed Effects Regression forest

sufficient conceptual overlap between these two measures to develop a mapping algorithm.

Tobit regression performed best in four out of five considered data availability scenarios, while MERF had the

best performance in the most complex data availability scenario. Specifying the regression models based on lasso often resulted in all covariates other than the general SIBDQ score being excluded from the model in most cases. Thus,

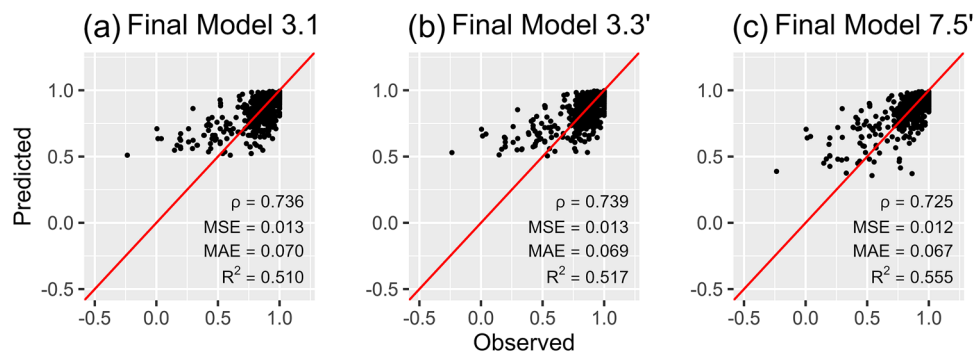


Fig. 2 Scatterplots of the observed vs. predicted EQ-5D-5L utility index scores in the validation subset for the final models, **a** Mixed-effects Tobit regression with general SIBDQ score only, **b** Mixed-effects Tobit regression with general SIBDQ score, BMI, smoking

status and disease type; **c** Mixed-effects regression forest with general SIBDQ score, BMI, smoking status, disease type and SIBDQ subscales; ρ Spearman correlation coefficient, *MAE* mean absolute error, *MSE* mean squared error

the performance measures were mostly equal to the bivariate specification with the general SIBDQ score as the only explanatory variable, except when SIBDQ subscales were used for mapping.

The performance measures of the final models in the validation subset were better than those obtained from tenfold cross-validation, indicating that the model did not overfit the data.

The most important predictor for EQ-5D-5L in final model 7.5' was the general SIBDQ score, followed by the SIBDQ Emotional and Social subscales. These reduced impurity by 15.5, 13.4 and 8.2, respectively. Our findings suggest that the mapping was not improved by the demographic variables age and gender, nor by information about time since initial diagnosis. Compared to other variables, the variable importance of disease subgroup, which reduced impurity by only 0.32, was low. We therefore concluded that providing a unique mapping algorithm for CD, UC and IC patients rather than estimating single mappings by subgroup would be sufficient.

Our results also indicate that using MERF to predict utility scores may improve outcomes compared to classical approaches when many covariates are available for mapping. Employing machine learning techniques for mapping is still uncommon [29]. However, some mapping studies have taken this path in recent years [41, 54, 55]. Because the core strength of machine learning is prediction, mapping studies appear to be a suitable application field for it. One could argue that the main focus of mapping algorithms should be prediction accuracy instead of interpretability, because no causal relationships are addressed. Moreover, practicability can be provided by embedding the model into a user-friendly environment, as we did with our web application. We would therefore welcome an increased use of machine learning methods in other mapping studies.

This study has several important limitations. First, we estimated the utility index score directly, because a crosswalk approach would have required a larger data set to ensure a sufficient share of the 3125 attainable health states of the EQ-5D-5L. Thus, conceptual dimensions of the SIBDQ are related to general population preferences based on the conceptual dimensions of the EQ-5D-5L. This may compromise the conceptual clarity of the relationship between the two measures [56].

Second, while the upper limit of OLS predictions exceeded the theoretical range, Tobit, ALDVM and MERF tended to underpredict high health states. The lower part of the distribution, on the other hand, was systematically overpredicted by all models. The main explanation for these outcomes in the lower half of the distribution is that there were few observations with low health states in our data. However, the problem of over- and underprediction is common in mapping studies [24].

Third, the data set used to validate the final models was not independent of the data set used for model selection and estimation, because they were both subsamples of patients participating in the same RCT. Moreover, splitting the data set into a model and a validation subset came at the cost of a reduced sample size, which is why ISPOR guidelines currently do not recommend this approach [24]. However, in machine learning literature it is common practice to apply this form of sample splitting in order to obtain an honest evaluation of the final model performance [57]. In our case, we considered splitting the sample as appropriate since the size of the model subset was still comparable to or even larger than samples used in other mapping studies (e.g. [37]).

Fourth, the inclusion criteria of the clinical trial stipulated that patients must be adults and receiving or eligible for biologic therapy. This implies that our sample consisted exclusively of patients with a complicated disease course.

Although the distribution of age, sex, SIBDQ and EQ-5D-5L was similar to that reported in studies with different inclusion criteria, the mapping may not be suitable for paediatric IBD patients or IBD patients receiving conventional medical treatment. Research exploring the performance of the mapping for different groups of patients is needed in order to decide whether our mapping algorithm can be recommended as a general approach to predict EQ-5D-5L values with the SIBDQ.

Conclusion

We developed and provide an algorithm that maps SIBDQ values to EQ-5D-5L index scores for different sets of covariates. An online tool to use the mapping algorithm in research practice is available via the following link: <https://www.bwl.uni-hamburg.de/hcm/forschung/mapping.html>. Our study targets situations in which utility index scores are not directly available. However, measuring and reporting EQ-5D-5L index values directly should be the preferred option. Moreover, our results suggest that machine learning methods may be superior to traditional regression approaches for mapping applications of this nature.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10198-023-01603-9>.

Funding Open Access funding enabled and organized by Projekt DEAL. Funding was received by the German Federal Joint Committee (Gemeinsamer Bundesausschuss). Moreover, we acknowledge financial support from the Open Access Publication Fund of Universität Hamburg.

Declarations

Conflict of interest The randomised controlled trial was funded by the Innovationsfonds of the German Federal Joint Committee (Gemeinsamer Bundesausschuss).

Ethics approval The local ethics review board at the Universität zu Lübeck approved the protocol for this study (approval number: 19-360).

Consent to participate and for publication The manuscript does not contain individual person's data in any form.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Kaplan, G.G., Windsor, J.W.: The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 56–66 (2021). <https://doi.org/10.1038/s41575-020-00360-x>
- Ng, S.C., Shi, H.Y., Hamidi, N., Underwood, F.E., Tang, W., Benchimol, E.I., Panaccione, R., Ghosh, S., Wu, J.C.Y., Chan, F.K.L., Sung, J.J.Y., Kaplan, G.G.: Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet Lond. Engl.* **390**, 2769–2778 (2017). [https://doi.org/10.1016/S0140-6736\(17\)32448-0](https://doi.org/10.1016/S0140-6736(17)32448-0)
- Shivashankar, R., Tremaine, W.J., Harmsen, W.S., Loftus, E.V.: Incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota from 1970 through 2010. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **15**, 857–863 (2017). <https://doi.org/10.1016/j.cgh.2016.10.039>
- Torres, J., Bonovas, S., Doherty, G., Kucharzik, T., Gisbert, J.P., Raine, T., Adamina, M., Armuzzi, A., Bachmann, O., Bager, P., Biancone, L., Bokemeyer, B., Bossuyt, P., Burisch, J., Collins, P., El-Hussuna, A., Ellul, P., Frei-Lanter, C., Furfaro, F., Gingert, C., Gionchetti, P., Gomollon, F., González-Lorenzo, M., Gordon, H., Hlavaty, T., Juillerat, P., Katsanos, K., Kopylov, U., Krustins, E., Lytras, T., Maaser, C., Magro, F., Kenneth Marshall, J., Myrelid, P., Pellino, G., Rosa, I., Sabino, J., Savarino, E., Spinelli, A., Stassen, L., Uzzan, M., Vavricka, S., Verstockt, B., Warusavitarne, J., Zmora, O., Fiorino, G., on behalf of the European Crohn's and Colitis Organisation [ECCO]: ECCO guidelines on therapeutics in Crohn's disease: medical treatment. *J. Crohns Colitis* **14**, 4–22 (2020). <https://doi.org/10.1093/ecco-jcc/jjz180>
- Raine, T., Bonovas, S., Burisch, J., Kucharzik, T., Adamina, M., Annese, V., Bachmann, O., Bettenworth, D., Chaparro, M., Czuber-Dochan, W., Eder, P., Ellul, P., Fidalgo, C., Fiorino, G., Gionchetti, P., Gisbert, J.P., Gordon, H., Hedin, C., Holubar, S., Iacucci, M., Karmiris, K., Katsanos, K., Kopylov, U., Lakatos, P.L., Lytras, T., Lyutakov, I., Noor, N., Pellino, G., Piovani, D., Savarino, E., Selvaggi, F., Verstockt, B., Spinelli, A., Panis, Y., Doherty, G.: ECCO guidelines on therapeutics in ulcerative colitis: medical treatment. *J. Crohns Colitis* **16**, 2–17 (2022). <https://doi.org/10.1093/ecco-jcc/jjab178>
- Verstockt, B., Ferrante, M., Vermeire, S., Van Assche, G.: New treatment options for inflammatory bowel diseases. *J. Gastroenterol.* **53**, 585–590 (2018). <https://doi.org/10.1007/s00535-018-1449-z>
- Briggs, A., Sculpher, M., Claxton, K.: *Decision Modelling for Health Economic Evaluation*. OUP, Oxford (2006)
- Yang, F., Devlin, N., Luo, N.: Cost-utility analysis using EQ-5D-5L data: does how the utilities are derived matter? *Value Health* **22**, 45–49 (2019). <https://doi.org/10.1016/j.jval.2018.05.008>
- Knowles, S.R., Keefer, L., Wilding, H., Hewitt, C., Graff, L.A., Mikocka-Walus, A.: Quality of life in inflammatory bowel disease: a systematic review and meta-analyses—part II. *Inflamm. Bowel Dis.* **24**, 966–976 (2018). <https://doi.org/10.1093/ibd/izy015>
- Kolotkin, R.L., Andersen, J.R.: A systematic review of reviews: exploring the relationship between obesity, weight loss and health-related quality of life. *Clin. Obes.* **7**, 273–289 (2017). <https://doi.org/10.1111/cob.12203>
- Cooper, V., Clatworthy, J., Harding, R., Whetham, J., Brown, A., Leon, A., Gonzalez, E., Garcia, D., Lockhart, D., Marent, B., West, B., Caceres, C., Fatz, D., Wallitt, E., Beck, E., Gomez, E., Teofilo, E., Garcia, F., Henwood, F., Rodrigues, G., Whetham, J., Begovac, J., Block, K., Apers, L., Pereira, L., Borges, M.,

- Darking, M., Chausa, P., Zekan, S., Bremner, S., Hoornaert, S., Mandalia, S.: Emerge Consortium: Measuring quality of life among people living with HIV: a systematic review of reviews. *Health Qual. Life Outcomes* **15**, 220 (2017). <https://doi.org/10.1186/s12955-017-0778-6>
12. Chen, X.-L., Zhong, L., Wen, Y., Liu, T.-W., Li, X.-Y., Hou, Z.-K., Hu, Y., Mo, C., Liu, F.-B.: Inflammatory bowel disease-specific health-related quality of life instruments: a systematic review of measurement properties. *Health Qual. Life Outcomes* **15**, 177 (2017). <https://doi.org/10.1186/s12955-017-0753-2>
 13. Buxton, M.J., Lacey, L.A., Feagan, B.G., Niecko, T., Miller, D.W., Townsend, R.J.: Mapping from disease-specific measures to utility: an analysis of the relationships between the inflammatory bowel disease questionnaire and Crohn's disease activity index in Crohn's disease and measures of utility. *Value Health* **10**, 214–220 (2007). <https://doi.org/10.1111/j.1524-4733.2007.00171.x>
 14. Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., Gray, A.: Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. *Int. J. Technol. Assess. Health Care* **31**, 230–235 (2015). <https://doi.org/10.1017/S0266462315000379>
 15. Irvine, E.J., Zhou, Q., Thompson, A.K.: The short inflammatory bowel disease questionnaire: a quality of life instrument for community physicians managing inflammatory bowel disease. *Am. J. Gastroenterol.* **91**, 1571–1578 (1996)
 16. Irvine, E.J.: Development and subsequent refinement of the inflammatory bowel disease questionnaire: a quality-of-life instrument for adult patients with inflammatory bowel disease. *J. Pediatr. Gastroenterol. Nutr.* **28**, S23 (1999)
 17. Williet, N., Sarter, H., Gower-Rousseau, C., Adrianjafy, C., Olympie, A., Buisson, A., Beaugerie, L., Peyrin-Biroulet, L.: Patient-reported outcomes in a French Nationwide Survey of inflammatory bowel disease patients. *J. Crohns Colitis* **11**, 165–174 (2017). <https://doi.org/10.1093/ecco-jcc/jjw145>
 18. Roseira, J., Sousa, H.T., Marreiros, A., Contente, L.F., Magro, F.: Short Inflammatory Bowel Disease Questionnaire: translation and validation to the Portuguese language. *Health Qual. Life Outcomes* **19**, 59 (2021). <https://doi.org/10.1186/s12955-021-01698-9>
 19. Sun, D., Chi, L., Liu, J., Liang, J., Guo, S., Li, S.: Psychometric validation of the Chinese version of the Short Inflammatory Bowel Disease Questionnaire and evaluation of its measurement invariance across sex. *Health Qual. Life Outcomes* **19**, 253 (2021). <https://doi.org/10.1186/s12955-021-01890-x>
 20. Luo, N., Wang, Y., How, C.H., Tay, E.G., Thumboo, J., Herdman, M.: Interpretation and use of the 5-level EQ-5D response labels varied with survey language among Asians in Singapore. *J. Clin. Epidemiol.* **68**, 1195–1204 (2015). <https://doi.org/10.1016/j.jclinepi.2015.04.011>
 21. Ludwig, K., Graf von der Schulenburg, J.-M., Greiner, W.: German value set for the EQ-5D-5L. *Pharmacoeconomics* **36**, 663–674 (2018). <https://doi.org/10.1007/s40273-018-0615-8>
 22. Feng, Y., Devlin, N., Bateman, A., Zamora, B., Parkin, D.: Distribution of the EQ-5D-5L profiles and values in three patient groups. *Value Health* **22**, 355–361 (2019). <https://doi.org/10.1016/j.jval.2018.08.012>
 23. Feng, Y.-S., Kohlmann, T., Janssen, M.F., Buchholz, I.: Psychometric properties of the EQ-5D-5L: a systematic review of the literature. *Qual. Life Res.* **30**, 647–673 (2021). <https://doi.org/10.1007/s11136-020-02688-y>
 24. Wailoo, A.J., Hernandez-Alava, M., Manca, A., Mejia, A., Ray, J., Crawford, B., Botteman, M., Busschbach, J.: Mapping to estimate health-state utility from non-preference-based outcome measures: an ISPOR good practices for outcomes research task force report. *Value Health* **20**, 18–27 (2017). <https://doi.org/10.1016/j.jval.2016.11.006>
 25. Grochtdreis, T., Dams, J., König, H.-H., Konnopka, A.: Health-related quality of life measured with the EQ-5D-5L: estimation of normative index values based on a representative German population sample and value set. *Eur. J. Health Econ.* **20**, 933–944 (2019). <https://doi.org/10.1007/s10198-019-01054-1>
 26. Herhaus, B., Kersting, A., Brähler, E., Petrowski, K.: Depression, anxiety and health status across different BMI classes: a representative study in Germany. *J. Affect. Disord.* **276**, 45–52 (2020). <https://doi.org/10.1016/j.jad.2020.07.020>
 27. Rozich, J.J., Holmer, A., Singh, S.: Effect of lifestyle factors on outcomes in patients with inflammatory bowel diseases. *Am. J. Gastroenterol.* **115**, 832–840 (2020). <https://doi.org/10.14309/ajg.0000000000000608>
 28. Dakin, H., Abel, L., Burns, R., Yang, Y.: Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health Qual. Life Outcomes* **16**, 31 (2018). <https://doi.org/10.1186/s12955-018-0857-3>
 29. Mukuria, C., Rowen, D., Harnan, S., Rawdin, A., Wong, R., Ara, R., Brazier, J.: 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* **17**, 295–313 (2019). <https://doi.org/10.1007/s40258-019-00467-6>
 30. Tobin, J.: Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24–36 (1958). <https://doi.org/10.2307/1907382>
 31. Hernández Alava, M., Wailoo, A.J., Ara, R.: Tails from the Peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value Health* **15**, 550–561 (2012). <https://doi.org/10.1016/j.jval.2011.12.014>
 32. Oliveira Gonçalves, A.S., Werdin, S., Kurth, T., Panteli, D.: Mapping studies to estimate health-state utilities from nonpreference-based outcome measures: a systematic review on how repeated measurements are taken into account. *Value Health* **26**, 589–597 (2023)
 33. Hajem, A., Bellavance, F., Larocque, D.: Mixed-effects random forest for clustered data. *J. Stat. Comput. Simul.* **84**, 1313–1328 (2014)
 34. Krennmair, P., Schmid, T.: Flexible domain prediction using mixed effects random forests. *J. R. Stat. Soc. Ser. C Appl. Stat.* (2022). <https://doi.org/10.48550/arXiv.2201.10933>
 35. Athey, S., Tibshirani, J., Wager, S.: Generalized random forests. *Ann. Stat.* **47**, 1148–1178 (2019). <https://doi.org/10.1214/18-AOS1709>
 36. Tibshirani, R.: Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **58**, 267–288 (1996). <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
 37. Oliveira Gonçalves, A.S., Panteli, D., Neeb, L., Kurth, T., Aigner, A.: HIT-6 and EQ-5D-5L in patients with migraine: assessment of common latent constructs and development of a mapping algorithm. *Eur. J. Health Econ.* **23**, 47–57 (2022). <https://doi.org/10.1007/s10198-021-01342-9>
 38. Kay, S., Tolley, K., Colayco, D., Khalaf, K., Anderson, P., Globe, D.: Mapping EQ-5D utility scores from the incontinence quality of life questionnaire among patients with neurogenic and idiopathic overactive bladder. *Value Health* **16**, 394–402 (2013). <https://doi.org/10.1016/j.jval.2012.12.005>
 39. Xu, R.H., Wong, E.L.Y., Jin, J., Dou, Y., Dong, D.: Mapping of the EORTC QLQ-C30 to EQ-5D-5L index in patients with lymphomas. *Eur. J. Health Econ.* **21**, 1363–1373 (2020). <https://doi.org/10.1007/s10198-020-01220-w>
 40. Thankappan, K., Patel, T., Ajithkumar, K.K., Balasubramanian, D., Raj, M., Subramanian, S., Iyer, S.: Mapping of head and neck cancer patient concerns inventory scores on to Euroqol-Five Dimensions-Five Levels (EQ-5D-5L) health utility scores. *Eur.*

- J. Health Econ. **23**, 225–235 (2022). <https://doi.org/10.1007/s10198-021-01369-y>
41. Aghdaee, M., Parkinson, B., Sinha, K., Gu, Y., Sharma, R., Olin, E., Cutler, H.: An examination of machine learning to map non-preference based patient reported outcome measures to health state utility values. *Health Econ.* **31**, 1525–1557 (2022). <https://doi.org/10.1002/hec.4503>
 42. Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B.: Shiny: Web application framework for R. <https://CRAN.R-project.org/package=shiny> (2021)
 43. Bokemeyer, B., Hardt, J., Hüppe, D., Prenzler, A., Conrad, S., Düffelmeyer, M., Hartmann, P., Hoffstadt, M., Klugmann, T., Schmidt, C., Weismüller, J., Mittendorf, T., Raspe, H.: Clinical status, psychosocial impairments, medical treatment and health care costs for patients with inflammatory bowel disease (IBD) in Germany: an online IBD registry. *J. Crohns Colitis* **7**, 355–368 (2013)
 44. Hein, R., Köster, I., Bollschweiler, E., Schubert, I.: Prevalence of inflammatory bowel disease: estimates for 2010 and trends in Germany from a large insurance-based regional cohort. *Scand. J. Gastroenterol.* **49**, 1325–1335 (2014)
 45. Statistisches Bundesamt: Mikrozensus - Fragen zur Gesundheit - Körpermaße der Bevölkerung, 2017. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Gesundheitszustand-und-Relevantes-Verhalten/_inhalt.html (2018)
 46. Severs, M., van Erp, S.J.H., van der Valk, M.E., Mangen, M.J.J., Fidder, H.H., van der Have, M., van Bodegraven, A.A., de Jong, D.J., van der Woude, C.J., Romberg-Camps, M.J.L., Clemens, C.H.M., Jansen, J.M., van de Meeberg, P.C., Mahmmoud, N., Ponsoen, C.Y., Bolwerk, C., Vermeijden, J.R., Pierik, M.J., Siersema, P.D., Leenders, M., van der Meulen-de Jong, A.E., Dijkstra, G., Oldenburg, B., The Dutch Initiative on Crohn and Colitis: Smoking is associated with extra-intestinal manifestations in inflammatory bowel disease. *J. Crohns Colitis.* **10**, 455–461 (2016). <https://doi.org/10.1093/ecco-jcc/jjv238>
 47. Rencz, F., Lakatos, P.L., Gulácsi, L., Brodszky, V., Kürti, Z., Lovas, S., Banai, J., Herszényi, L., Cserni, T., Molnár, T., Péntek, M., Palatka, K.: Validity of the EQ-5D-5L and EQ-5D-3L in patients with Crohn's disease. *Qual. Life Res.* **28**, 141–152 (2019). <https://doi.org/10.1007/s11136-018-2003-4>
 48. Ruiz-Casas, L., Evans, J., Rose, A., Pedra, G.G., Lobo, A., Finnegan, A., Hayee, B., Peyrin-Biroulet, L., Sturm, A., Burisch, J., Terry, H., Avedano, L., Tucknott, S., Fiorino, G., Limdi, J.: The LUCID study: living with ulcerative colitis; identifying the socio-economic burden in Europe. *BMC Gastroenterol.* **21**, 1–13 (2021). <https://doi.org/10.1186/s12876-021-02028-5>
 49. Van Assche, G., Peyrin-Biroulet, L., Sturm, A., Gisbert, J.P., Gaya, D.R., Bokemeyer, B., Mantzaris, G.J., Armuzzi, A., Sebastian, S., Lara, N., Lynam, M., Rojas-Farreras, S., Fan, T., Ding, Q., Black, C.M., Kachroo, S.: Burden of disease and patient-reported outcomes in patients with moderate to severe ulcerative colitis in the last 12 months—multicenter European cohort study. *Dig. Liver Dis.* **48**, 592–600 (2016). <https://doi.org/10.1016/j.dld.2016.01.011>
 50. Hagelund, L.M., Elkjær Stallknecht, S., Jensen, H.H.: Quality of life and patient preferences among Danish patients with ulcerative colitis—results from a survey study. *Curr. Med. Res. Opin.* **36**, 771–779 (2020). <https://doi.org/10.1080/03007995.2020.1716704>
 51. Slonim-Nevo, V., Sarid, O., Friger, M., Schwartz, D., Sergienko, R., Pereg, A., Vardi, H., Singer, T., Chernin, E., Greenberg, D., Odes, S.: Israeli IBD research nucleus (IIRN): effect of social support on psychological distress and disease activity in inflammatory bowel disease patients. *Inflamm. Bowel Dis.* **24**, 1389–1400 (2018). <https://doi.org/10.1093/ibd/izy041>
 52. Rolston, V.S., Boroujerdi, L., Long, M.D., McGovern, D.P.B., Chen, W., Martin, C.F., Sandler, R.S., Carmichael, J.D., Dubinsky, M., Melmed, G.Y.: The influence of hormonal fluctuation on inflammatory bowel disease symptom severity—a cross-sectional cohort study. *Inflamm. Bowel Dis.* **24**, 387–393 (2018). <https://doi.org/10.1093/ibd/izx004>
 53. Perera, L.P., Bhandari, S., Liu, R., Guilday, C., Zadvornova, Y., Saeian, K., Eastwood, D.: Advanced age does not negatively impact health-related quality of life in inflammatory bowel disease. *Dig. Dis. Sci.* **63**, 1787–1793 (2018). <https://doi.org/10.1007/s10620-018-5076-6>
 54. Lokkerbol, J., Wijnen, B.F.M., Chatterji, S., Kessler, R.C., Chisholm, D.: Mapping of the World Health Organization's disability assessment schedule 2.0 to disability weights using the multi-country survey study on health and responsiveness. *Int. J. Methods Psychiatr. Res.* **30**, e1886 (2021). <https://doi.org/10.1002/mp.1886>
 55. Gao, L., Luo, W., Tonmukayakul, U., Moodie, M., Chen, G.: Mapping MacNew Heart Disease Quality of Life Questionnaire onto country-specific EQ-5D-5L utility scores: a comparison of traditional regression models with a machine learning technique. *Eur. J. Health Econ.* **22**, 341–350 (2021). <https://doi.org/10.1007/s10198-020-01259-9>
 56. Round, J., Hawton, A.: Statistical alchemy: conceptual validity and mapping to generate health state utility values. *Pharmacoeconomics Open.* **1**, 233–239 (2017). <https://doi.org/10.1007/s41669-017-0027-2>
 57. Xu, Y., Goodacre, R.: On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *J. Anal. Test.* **2**, 249–262 (2018)

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