Quality of Life Research (2022) 31:2153–2165
https://doi.org/10.1007/s11136-022-03082-6

Can EQ-5D-3L utility values of low back pain patients be validly predicted by the Oswestry Disability Index for use in cost-effectiveness analyses?

Sylvia Pellekooren[1,2] · Ângela J. Ben[3] · Judith E. Bosmans[3] · Raymond W. J. G. Ostelo[1,4] · Maurits W. van Tulder[2] · Esther T. Maas[1] · Frank J. P. M. Huygen[5,6] · Teddy Oosterhuis[7,8] · Adri T. Apeldoorn[9] · Miranda L. van Hooff[10,11] · Johanna M. van Dongen[1,3]

Accepted: 7 January 2022 / Published online: 17 January 2022

Abstract

Purpose To assess whether regression modeling can be used to predict EQ-5D-3L utility values from the Oswestry Disability Index (ODI) in low back pain (LBP) patients for use in cost-effectiveness analysis.

Methods EQ-5D-3L utility values of LBP patients were estimated using their ODI scores as independent variables using regression analyses, while adjusting for case-mix variables. Six different models were estimated: (1) Ordinary Least Squares (OLS) regression, with total ODI score, (2) OLS, with ODI item scores as continuous variables, (3) OLS, with ODI item scores as ordinal variables, (4) Tobit model, with total ODI score, (5) Tobit model, with ODI item scores as continuous variables, and (6) Tobit model, with ODI item scores as ordinal variables. The models' performance was assessed using explained variance (R^2) and root mean squared error (RMSE). The potential impact of using predicted instead of observed EQ-5D-3L utility values on cost-effectiveness outcomes was evaluated in two empirical cost-effectiveness analysis.

Results Complete individual patient data of 18,692 low back pain patients were analyzed. All models had a more or less similar R^2 (range 45–52%) and RMSE (range 0.21–0.22). The two best performing models produced similar probabilities of cost-effectiveness for a range of willingness-to-pay (WTP) values compared to those based on the observed EQ-5D-3L values. For example, the difference in probabilities ranged from 2 to 5% at a WTP of 50,000 €/QALY gained.

Conclusion Results suggest that the ODI can be validly used to predict low back pain patients' EQ-5D-3L utility values and QALYs for use in cost-effectiveness analyses.

Keywords Low back pain · Utility scores · Oswestry Disability Index · Ordinary least squares · Tobit · EQ-5D

✉ Sylvia Pellekooren
s.pellekooren@vu.nl

[1] Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences Research Institute, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

[2] Department Human Movement Sciences, Faculty of Behavioral & Movement Sciences, Vrije Universiteit, Amsterdam, The Netherlands

[3] Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

[4] Department of Epidemiology and Data Science, Amsterdam UMC, Location VUmc, Amsterdam Movement Sciences Research Institute, Amsterdam, The Netherlands

[5] Center of Pain Medicine Erasmusmc, Rotterdam, The Netherlands

[6] Center of Pain Medicine UMCU, Utrecht, The Netherlands

[7] Netherlands Society of Occupational Medicine, Centre of Excellence, Utrecht, the Netherlands

[8] Coronel Institute of Occupational Health, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

[9] Rehabilitation Departement, Noordwest Ziekenhuisgroep, Alkmaar, Netherlands, Breederode Hogeschool, Rotterdam, Netherlands

[10] Departement Research, Sint Maartenskliniek, Nijmegen, The Netherlands

[11] Department of Orthopedic Surgery, Radboud University Medical Center, Nijmegen, The Netherlands

Plain English summary

Quality-Adjusted Life Years (QALY) are an essential outcome in economic evaluations that assess whether a new intervention is cost-effective (i.e., provides good value for money) compared to an alternative intervention. However, not all economic evaluations among low back pain patients measure quality of life using a preference-based measure which is necessary to calculate QALY. If a preference-based quality-of-life measure is missing, utility values may be predicted using other measurement instruments, such as a condition-specific questionnaire on low back pain complaints, such as the frequently used Oswestry Disability Index. However, it is unclear whether this results in valid estimates of the utility values. Therefore, we developed six different models that predicted utility values based on the Oswestry Disability Index and assessed their predictive ability. Additionally, we assessed the extent to which cost-effectiveness outcomes differed between the predicted and actual utility values for the EQ-5D-3L. Results suggest that the ODI can be validly used to predict low back pain patients' QALYs in an economic evaluation when preference-based quality-of-life data are missing.

Introduction

Low back pain (LBP) has an estimated incidence of 250 million people worldwide and is characterized by a high burden of disease [1]. Patients with LBP typically experience difficulties in different aspects of health-related quality of life, such as their daily functioning, social participation [2, 3], and working ability [4, 5]. These difficulties may affect patients' health-related quality of life considerably [3, 6] and have significant impact on healthcare and societal costs [7, 8]. As limited (healthcare) resources are available, decision-makers are not only interested in the effectiveness of LBP treatments recommended in international guidelines, but also in their cost-effectiveness compared to alternative treatments.

Cost-effectiveness analysis provide insight into relative cost-effectiveness of treatments by comparing their incremental costs to their incremental effects [9]. These effects are often expressed in Quality-Adjusted Life Years (QALYs), which combine both the quality and quantity of life into a single outcome [10]. For estimating QALYs, health-related quality of life is typically measured using preference-based quality-of-life measures. Health states obtained from these measures can be converted into utility values, which represent the preferences of the general population of a country for given health states [11]. In many countries, it is recommended to estimate utility values using the EuroQol five-dimension questionnaire (EQ-5D) and national tariffs to account for the fact that health state preferences differ across countries [12–14]. Unfortunately, EQ-5D data are not always available in clinical trials [15], as higher priority is sometimes given to condition-specific measures that assess more clinically relevant outcomes [16].

When utility values are missing, QALYs cannot be calculated. However, information about incremental cost per QALY gained is typically required by healthcare decision-makers, particularly at the national level [12, 13]. In the absence of the EQ-5D or another generic preference-based quality-of-life measure, a condition-specific measure might be used to predict utility values [17]. In LBP, one of the most frequently used condition-specific measures is the Oswestry Disability Index (ODI) [18]. The ODI measures limitations of a patient's performance [19] and is recommended in the core outcome set for clinical trials in nonspecific LBP [20] and management of LBP [21].

A previous study assessed the predictive ability of the ODI in estimating utility values from the EQ-5D-3L by using data from 14,544 patients with lumbar degenerative pathology treated in a tertiary spine center [22]. Linear regression analysis was performed to predict the patients' EQ-5D utility values based on their ODI total or individual item scores and patients reported severity of back and leg pain. Based on a root mean square error (RMSE) of 0.14, authors concluded that it is not possible to estimate EQ-5D-3L utility values based on the ODI. However, given the bounded nature of EQ-5D data as well as the possible existence of other contextual factors that influence health-related quality of life in LBP, it is likely that the models' performance might be improved by using a Tobit model to account for possible ceiling effects. The model's performance might also be improved by including a wider variety of LBP patients treated in various settings, while adjusting for more case-mix variables. Moreover, the authors only based their conclusions on the models' RMSE without assessing the impact of using predicted utility scores in cost-effectiveness. Therefore, this study aimed to assess the feasibility of using different regression models to predict EQ-5D-3L utility values in LBP patients based on the ODI in cost-effectiveness analyses while adjusting for a broad range of case-mix characteristics.

Method

Source of data

Individual patient data included in this study originated from four previously conducted prospective studies; i.e., the minimal interventional treatments (MINT) study, the rehabilitation after lumbar disk surgery (REALISE) study, the Nijmegen Decision Tool study, and a study evaluating a treatment-based classification system [23–32]. These studies were conducted among sub-acute and chronic LBP patients treated in primary care, secondary care, and/or tertiary care. For all patients, various sociodemographic variables were assessed at baseline, and both the ODI and EQ-5D-3L utility values were assessed at baseline and at one or more follow-up moments. In total, 21,500 patients were included in these studies. For developing the models, only baseline data were used in the present study, because the proportion of participants with missing data was low at baseline (i.e., < 5%), thereby preventing the need for imputation of missing values. To assess the final models' performance in a trial-based cost-effectiveness analysis setting, baseline as well as follow-up data were used of the MINT study [23–25], and the treatment-based classification system study [29, 30].

The MINT study [23–25], the REALISE study [31, 32], and the treatment-based classification system study [29, 30] obtained ethical approval from the Medical Ethics Committee of the Erasmus Medical Centre Rotterdam or Medical Ethics Committee of the VU University Medical Centre in Amsterdam. For the Nijmegen Decision Tool study [26–28], ethical approval was not required, because the "*Dutch Act on Medical Research involving Human Subjects*" does not apply to screening questionnaires that are part of routine practice. More detailed information on the design and study population of the different studies is provided in Supplementary Appendix A.

Utility values

Utility values were based on the EQ-5D-3L, which is a generic preference-based measure that asks participants to describe their health state on five health dimensions (i.e., mobility, self-care, usual activities, pain/ discomfort, and anxiety/depression) using three severity levels (i.e., no problems, moderate problems, and severe problems)[33]. The participants' EQ-5D-3L health states were converted into utility values using the Dutch tariff[34]. Utility values are presented on a continuous scale that is anchored at 1 (indicating full health) to 0 (indicating a state as bad as being dead). Negative values may also occur, which represent health states that are regarded as worse than a state that is as bad as being dead [10]. Dutch EQ-5D-3L utility values can range between − 0.33 and 1.

Oswestry Disability Index

The ODI measures the limitations of a patient's performance compared with that of a fit person, and consists of ten items assessing various aspects of daily living (e.g., lifting, walking, and traveling). Each item is scored on a six-point scale, ranging from 0 to 5. The overall ODI score was estimated by summing the values of all individual items, subsequently dividing this score by the total possible score, and multiplying this score by 100. The total score ranges from 0 to 100%, with higher scores indicate higher level of disability [19, 35]. For this study, the "sex life" (item 8) was not included, as this item is frequently omitted in applied studies as well [36–38]. Including this item would have hampered the generalization of the results to a large number of LBP studies. The cross cultural adapted Dutch language version of the ODI version 2.1a was used in all studies included [39].

Predictors

The following case-mix variables were included; age (years), gender (male/female), education level (low/moderate/high), living together with a partner (yes/no), type of LBP (subacute/chronic), setting (primary care/secondary care/ tertiary care), and back pain (Numeric Rating Scale (NRS: 0–10) Pain score: low 0–3, moderate 4–6, and severe 7–10) [40]. Given error proneness of overly detailed models and benefits of ease of use, NRS scores were categorized using cut-off points from an earlier conducted study, which categorized NRS pain scores based on pain-related interference with functioning in patients with chronic musculoskeletal pain [41]. These variables were included, because they were expected to increase the predictive value of the models [42–47] and to be measured in most applied studies, thereby increasing applicability of the models.

Statistical analysis

Baseline characteristics were described using frequencies and percentages for categorical variables and means and standard deviations for continuous variables. Prior to the development of the models, linearity and additivity assumptions (i.e., normally distributed residuals, homoscedasticity, influential cases and outliers) were assessed using diagnostic plots (i.e., scatterplot, density plot, and boxplots), and diagnostic tests (e.g., Grubbs test). Pearson's correlation coefficient was used to assess the strength of the linear relationship between the patients' EQ-5D-3L based utility values and ODI total scores. To assess the agreement between the EQ-5D-3L and the ODI the Intra Class Correlation (ICC) was calculated using a two-way random effects model.

Model development and variable selection

Models were developed using two regression techniques; i.e., Ordinary Least Squares (OLS) regression and Tobit regression (i.e., censored or truncated regression). OLS regression was included, because it is still one of the most frequently used linear modeling techniques. OLS regression is used to estimate the strength of the association between a continuous outcome variable and one or more independent variables [48]. OLS, however, does not take into account the bounded nature of utility values which can be accounted for in a Tobit regression [49]. This model can estimate linear relationships between variables, where the range of the dependent variable is constrained. This is done using a so-called latent variable that accounts for the fact that the true independent variable is—in our case—bounded at 1. Hereby, biased and inconsistent estimates, that may occur when using OLS regression, may be prevented [50].

For both the OLS and Tobit model, three different regression models were developed: (1) including the overall ODI score as independent variable, (2) using all nine ODI items scores as independent variables and assuming them to be continuous, and (3) using all nine ODI items scores as independent variables and assuming them to be ordered. This resulted in six different models: (1) OLS, with the total ODI score, (2) OLS, with the ODI item scores as continuous variables, (3) OLS, with the ODI item scores as ordinal variables, (4) Tobit model, with the total ODI score, (5) Tobit model, with the ODI item scores as continuous variables, (6) Tobit model, with the ODI item scores as ordinal variables. To assess which variables increased the predictive value of the models, a bi-directional stepwise selection procedure [51], using Akaike Information Criterion (i.e., the trade-off between the goodness of fit of the model and the simplicity of the model) [52], with a 5% significance level was used. Stepwise selection combines the elements of forward and backward selection by sequentially adding variables, based on the most contributing predictors, and omitting variables that no longer provide an improvement in the model fit after adding a new variable to the model. Final models only included case-mix variables that increased the predictive value.

Model performance and internal validation

The original dataset was split into a training sample (70%), and a validation sample (30%) using the 'create Data Partition' function in R. This function creates a balanced split of the data by performing a stratified random split of the data based on the mean of the dependent variable, which leads to a comparable mean EQ-5D-3L utility value in both the training and validation dataset. After developing the models in the training sample, their performance was assessed in the validation sample using the RMSE (i.e., the absolute fit of the model) and the adjusted R^2 (i.e., the relative fit of the model). The minimal important difference (MID) of the EQ-5D-3L was used to determine an acceptable RSME, which was set at a cut of point of 0.03 [53]. A correlation of 0.5 or higher (i.e., a relatively moderate correlation as the R squared indicates that about half of the variance of the utility values is explained by the ODI) was considered sufficient for performing regression analysis. Recommended models were selected based on parsimony, which is the trade-off between simplicity of the model (i.e., low AIC) and explanatory predictive power (i.e., high R^2). To assess agreement between the actual and estimated EQ-5D-3L based utility values a Bland Altman analysis was performed for all models.

Sensitivity analyses

In addition to the main analysis, three sensitivity analyses (SA) were performed. In the first sensitivity analysis (SA1) the variable mental health status was added to the case-mix variables (SA1). SA1 was only performed on a sub-set of the data, as only one of the four datasets (i.e., the MINT study [23–25]) assessed mental health using the Four Dimensional Symptom Questionnaire (4DSQ) [53], and only part of the sample ($n = 4123$) completed this questionnaire. The 4DSQ assesses four different aspects of mental health (i.e., distress, depression, anxiety, and somatisation), all of which were included in the models as a separate variable. In SA2, the variable living with a partner was omitted. In SA3 the patients' EQ-5D-3L utility values were converted to EQ-5D-5L utility values using the reverse crosswalk (SA3) [55]. Reversed cross walk values make it possible to link EQ-5D-3L responses to EQ-5D-5L value sets, and can be used when 5L values are wanted, but only 3L data are available [55, 56]. The 5-level EQ-5D version is an adapted version of the EQ-5D-3L, which is known to be more sensitive and has less ceiling effects, including through changing the number of levels of perceived problems per dimension from 3 to 5[57].

Cost-effectiveness analysis

To assess the models' impact on cost-effectiveness outcomes, complete cases from two randomized controlled trials were used, i.e., empirical dataset 1 ($n = 68$; Apeldoorn et al. [29, 30]) and empirical dataset 2 ($n = 424$; Maas et al. [23–25]). In both studies, QALYs were estimated based on both the actual EQ-5D-3L scores (i.e., actual QALY values) and based on the patients' ODI scores (i.e., predicted QALY values). Agreement between the actual and estimated EQ-5D-3L based utility values was assessed by performing a Bland Altman analysis for each of the empirical datasets.

Then, full trial-based cost-effectiveness analysis were conducted for each of the six models as well as the patients' actual QALY values (i.e., QALYs based on the measured EQ-5D-3L scores). For each trial-based cost-effectiveness analysis, mean differences in costs and QALYs between treatment groups were estimated using seemingly unrelated regression analyses. Incremental Cost-Effectiveness Ratios (ICERs) were calculated by dividing the difference in costs by the difference in effects. Uncertainty around cost and QALY differences was estimated using bootstrapping. The percentage of bootstrapped cost-effect pairs was reported per quadrant of the Cost-Effectiveness Plane (i.e., north east, south east, north west, and south west). Subsequently, Cost-Acceptability Curves (CEACs) were plotted. CEACs indicate an intervention's probability of cost-effectiveness compared to control for a range of willingness-to-pay (WTP) values (i.e., thresholds of 0, 30,000 euro and 50,000). These probabilities were assessed on their decision sensitivity (i.e., how sensitive is the conclusion of a cost-effectiveness analysis is to using a particular statistical method) [58] . Analyses were performed in R software, version 3.4.0.

Results

Participants

Out of the individual patient data that included 21,500 patients, 18,692 complete cases were included for analysis. These patients had sub-acute ($n = 3248$) or chronic LBP ($n = 15{,}444$). The mean age of the patients was 53.9 years (SD = 14.7, range 18.1–91.9) and 61% of the sample was female. The patients' mean ODI score at baseline was 41.23 (SD = 15.4, range 0–100) and their mean baseline EQ-5D-3L based utility value was 0.46 (SD = 0.29, range -0.3290–1.00). More details on the patients' characteristics are shown in Table 1.

Table 1 Baseline characteristics of patients included

Characteristic	$n = 18{,}692$
Age (mean (SD), range)	53.9 (14.7), 18.1–91.9
Gender; female (*n*, %)	11,345 (60.7)
Education (*n*, %)	
Low (no education, primary level education, lower vocational and lower secondary education)	5,398 (28.9)
Moderate (higher secondary education or undergraduate)	9,078 (48.6)
High (tertiary, university level, postgraduate)	4,216 (22.6)
Living with a partner (*n*, %)	14,085 (75.4)
Type of LBP (*n*, %)	
Sub-acute (< 3 months)	3,248 (17.4)
Chronic (> 3 months)	15,444 (82.6)
Post-surgery (*n*, %)	1,587 (8.5)
Setting (*n*, %)	
Primary care (i.e., physiotherapy clinics)	150 (0.8)
Secondary care (i.e., pain clinics)	4,123 (22.1)
Tertiary care (i.e., hospital)	14,419 (77.1)
NRS pain (mean (SD))	6.99 (1.9)
Utility score (mean (SD), range)	0.467 (0.299), − 0.3290–1.00
ODI score[a] (mean (SD), range)	41.23 (15.4), 0–100
ODI 1 mean (SD)/median (IQR)	2.66 (0.93)/3 (2–4)
ODI 2 mean (SD)/median (IQR)	1.11 (1.04)/1 (0–2)
ODI 3 mean (SD)/median (IQR)	2.78 (1.32)/3 (2–4)
ODI 4 mean (SD)/median (IQR)	1.44 (1.22)/1 (0–2)
ODI 5 mean (SD)/median (IQR)	2.11 (1.09)/2 (1–3)
ODI 6 mean (SD)/median (IQR)	2.85 (1.29)/3 (2–4)
ODI 7 mean (SD)/median (IQR)	1.49 (1.09)/1 (0–2)
ODI 9 mean (SD)/median (IQR)	2.14 (1.20)/2 (1–3)
ODI 10 mean (SD)/median (IQR)	1.98 (1.32)/2 (1–3)

[a] Excluding item 8 sex life

LBP low back pain, *NRS* numeric rating scale (range 0–10), utility (range − 0.33 to 1), *ODI* oswestry disability scale (range 0–100), ODI individual item (range 0–5), *SD* standard deviation, *IQR* inter quartile range

Variables included and model performance

The diagnostic plots showed a linear relationship between EQ-5D-3L based utility values and the ODI, and homogeneity of variance of the residuals. Even though the patients' baseline EQ-5D-3L based utility values followed a bimodal distribution, the corresponding residuals were normally distributed. Hence, the normality of residuals assumption of linear regression was met. No outliers or influential cases were identified. Pearson's correlation coefficient between the patients' baseline EQ-5D-3L utility values and ODI total score was 0.63. The ICC showed an agreement of 0.23 between individual ODI items and EQ-5D-3L items.

An overview of the independent variables that were included in the final models, as well as their respective regression coefficients, can be found in Supplementary Appendix B. The case-mix variables age, gender, education, partner, and NRS were included in all models, whereas type of LBP was not included in any of the models. The variable setting was included in all models except for model 1 (i.e., OLS with ODI total scores). In the models using Tobit regression, 74 of the 13,087 observations in the training set were right censored.

The performance of the different models was more or less the same, with explained variances ranging from 45 to 51% and RMSEs ranging from 0.21 to 0.22. Based on parsimony of the models, model 2 and 5 seem most appropriate to use. More details on the performance of the different models are shown in Table 2.

The mean difference between estimated and actual utility values for model 2 was -0.068 (95%CI -0.495, 0.359), and for model 5 -0.086 (95%CI -0.512, 0.341). Bland Altman plots of models 2 and 5 are shown in Fig. 1. The plots for other all models are presented in Supplementary Appendix C.

Sensitivity analysis

Adding mental health variable(s) to the models resulted in an increase of the explained variance of 2–4%, whereas the RMSE remained similar. Omission of the variable 'living with a partner' (SA2) did not change the models'

Table 2 Performance measures in the training set

	Performance in the training set ($n = 13{,}087$)		AIC	Performance in validation set ($n = 5605$)		AIC
	R^2	RMSE		R^2	RMSE	
Model 1: OLS with ODI total scores	0.45	0.22	– 2326.48	0.46	0.22	– 1083.26
Model 2: OLS with ODI individual item total scores continuous	0.50	0.21	– 3423.24	0.50	0.21	– 1513.73
Model 3: OLS with ODI individual item total scores ordered	0.51	0.21	– 3769.51	0.52	0.21	– 1638.09
Model 4: Tobit with ODI total scores	0.45	0.22	– 2061.91	0.46	0.22	– 951.61
Model 5: Tobit with ODI individual item total scores continuous	0.50	0.21	– 3164.37	0.50	0.21	– 1385.32
Model 6 Tobit with individual item total scores ordered	0.51	0.21	– 3474.88	0.52	0.21	– 1494.06

OLS ordinary least squares regression, *ODI* oswestery disability index, R^2 proportion of variance for the dependent variable, *RMSE* root mean squared error, *AIC* akaike information criteria

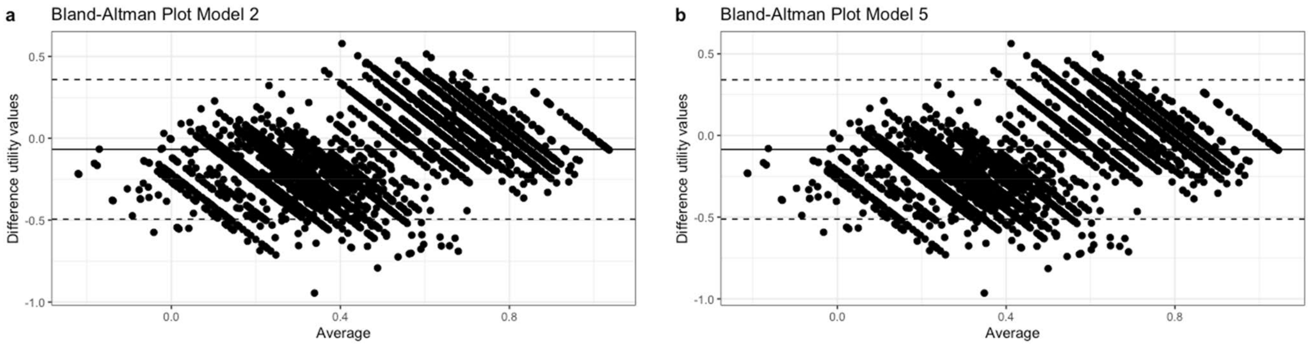

Fig. 1 Bland Altman plot model 2 and 5. *X*-axis: average measurement of the estimated and actual utility values, *Y*-axis: difference in measurements between the two instruments. Solid line: Average difference in measurements between the estimated and actual utility values, Dashed lines: 95% confidence interval limits for the average difference

performance. Using the patients' reversed cross-walked EQ-5D-5L utility values (SA3) improved the models' explained variance by 3–4%, and the RMSE reduced with 0.06–0.07. More details on the results of the sensitivity analyses are provided in Supplementary Appendix D.

Results cost-effectiveness analysis

The mean difference between estimated and actual utility values for empirical dataset 1 model 2 was -0.039 (95%CI -0.075, -0.002), and for model 5 -0.057 (95%CI -0.097, -0.018). The mean difference between estimated and actual utility values for empirical dataset 2 model 2 was 0.295 (95%CI 0.246, 0.344), and for model 5 the mean difference was 0.294 (95%CI 0.248, 0.341). Bland Altman plots of models 2 and 5 for both empirical datasets are shown in Fig. 2. The plots for other all models are presented in Supplementary Appendix E.

In both empirical datasets, the difference between the predicted and actual differences in QALYs was small for the two most parsimonious models (i.e., models 2 and 5:$\Delta \leq 0.004$) and the distributions of cost-effect pairs across the four quadrants of the cost-effectiveness plane were comparable. The cost-effectiveness acceptability curves based on both predicted and actual QALY values were also similar. The predicted probability of an intervention being cost effective at a willingness to pay of 50,000 was slightly higher in both models than the actual probabilities (i.e., 2–5% in model 2, and 3–5% in model 5). More details on the cost-effectiveness outcomes for all models in both empirical studies are shown in Table 3 and Fig. 3.

Discussion

Main findings

There were no large differences in the models' performance between OLS and Tobit regression, nor between using the patients' total ODI scores and ODI individual item scores. The explained variance of the developed models ranged from 45 to 51%, and the RMSE ranged from 0.21 to 0.22. Models 2 and 5 are recommended based on the best fit and parsimony. The models' relatively low absolute fit (RMSE) indicates that they are not suitable for estimating utility values for individual patients. Nonetheless, they can be used to predict differences in LBP patients' EQ-5D-3L utility values and QALY's, as the systematic bias in mean scores does not affect the differences between the groups. Cost-effectiveness outcomes of models 2 and 5 based on predicted and actual values were similar. These findings enable researchers to

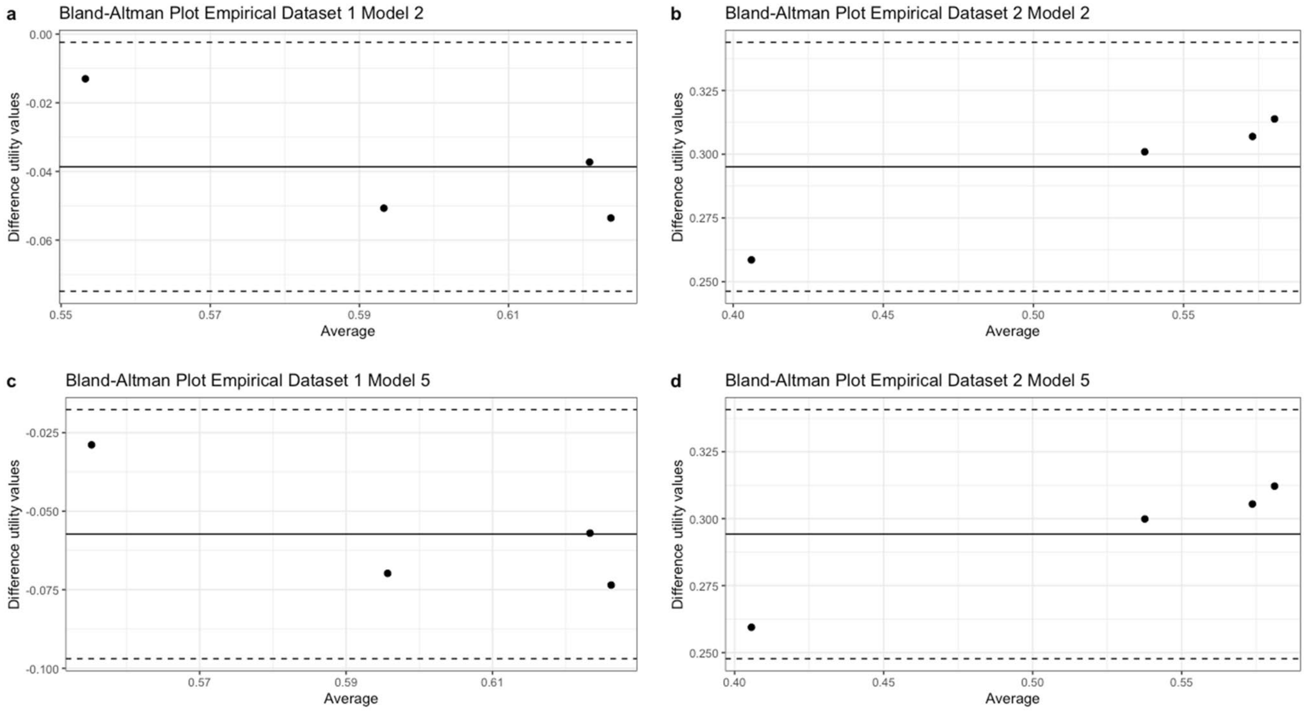

Fig. 2 Bland Altman plot model 2 and 5 empirical datasets. *X*-axis: average measurement of the estimated and actual utility values, *Y*-axis: difference in measurements between the two instruments. Solid line: average difference in measurements between the estimated and actual utility values, Dashed lines: 95% confidence interval limits for the average difference

Table 3 Cost-effectiveness outcomes for an intervention in comparison with usual care by predictive models

Predictive models	ΔE (95% CI)	ΔC (95% CI)	ICER	Cost-effectiveness plane				Cost-effectiveness acceptability curve			
				NE (%)	SE (%)	SW (%)	NW (%)	P_{CE} (0)	P_{CE} (10,000)	P_{CE} (30,000)	P_{CE} (50,000)
Empirical dataset 1 [28, 29] $N=86$											
Actual values	– 0.041 (– 0.091; 0.009)	– 110 (– 1761; 1283)	2697	2	4	51	42	0.55	0.36	0.16	0.11
Model 1	– 0.035 (– 0.094; 0.021)	– 110 (– 1761; 1283)	3091	1	10	45	44	0.55	0.39	0.25	0.20
Model 2	**– 0.043 (– 0.106; 0.015)**	**– 110 (– 1761; 1283)**	**2559**	**1**	**7**	**48**	**44**	**0.55**	**0.36**	**0.21**	**0.16**
Model 3	– 0.027 (– 0.081; 0.018)	– 110 (– 1761; 1283)	4068	1	13	43	43	0.55	0.42	0.30	0.24
Model 4	– 0.036 (– 0.095; 0.021)	– 110 (– 1761; 1283)	3058	1	10	45	44	0.55	0.39	0.25	0.20
Model 5	**– 0.044 (– 0.107; 0.015)**	**– 110 (– 1761; 1283)**	**2514**	**1**	**7**	**48**	**44**	**0.55**	**0.36**	**0.21**	**0.16**
Model 6	– 0.027 (– 0.080; 0.021)	– 110 (– 1761; 1283)	4084	2	13	42	43	0.55	0.42	0.30	0.25
Empirical dataset 2 [22–24] $N=424$											
Actual values	– 0.004 (– 0.034; 0.027)	1576 (596; 2575)	– 371,566	38	0	0	62	0.001	0.002	0.017	0.048
Model 1	– 0.007 (– 0.037; 0.023)	1576 (596; 2575)	– 226,441	32	0	0	68	0.001	0.002	0.014	0.037
Model 2	**0.0002 (– 0.030; 0.029)**	**1576 (596; 2575)**	**6,670,132**	**51**	**0**	**0**	**49**	**0.001**	**0.003**	**0.025**	**0.070**
Model 3	– 0.001 (– 0.026; 0.024)	1576 (596; 2575)	– 2,099,247	48	0	0	52	0.001	0.002	0.015	0.028
Model 4	– 0.007 (– 0.037; 0.024)	1576 (596; 2575)	– 224,080	32	0	0	67	0.001	0.002	0.014	0.038
Model 5	**0.0003 (– 0.030; 0.030)**	**1576 (596; 2575)**	**5,105,447**	**51**	**0**	**0**	**49**	**0.001**	**0.003**	**0.025**	**0.073**
Model 6	– 0.001 (– 0.027; 0.026)	1576 (596; 2575)	– 2,417,793	48	0	0	51	0.001	0.002	0.018	0.053

Recommended models are presented as bold text

N number of observations in the analysis, *ΔC* difference in costs, *95% CI* 95% confidence interval. *ΔE* difference in effects, *ICER* incremental cost-effectiveness ratio, *NE* north east, *SE* south east, *SW* south west, *NW* north west, P_{CE} *(0)* probability that the intervention is cost-effective as compared to usual care with a threshold of 0, P_{CE} *( )* probability that the intervention is cost-effective as compared to usual care with willingness-to-pay thresholds of 0, 10,000, 30,000, and 50,000 Euros

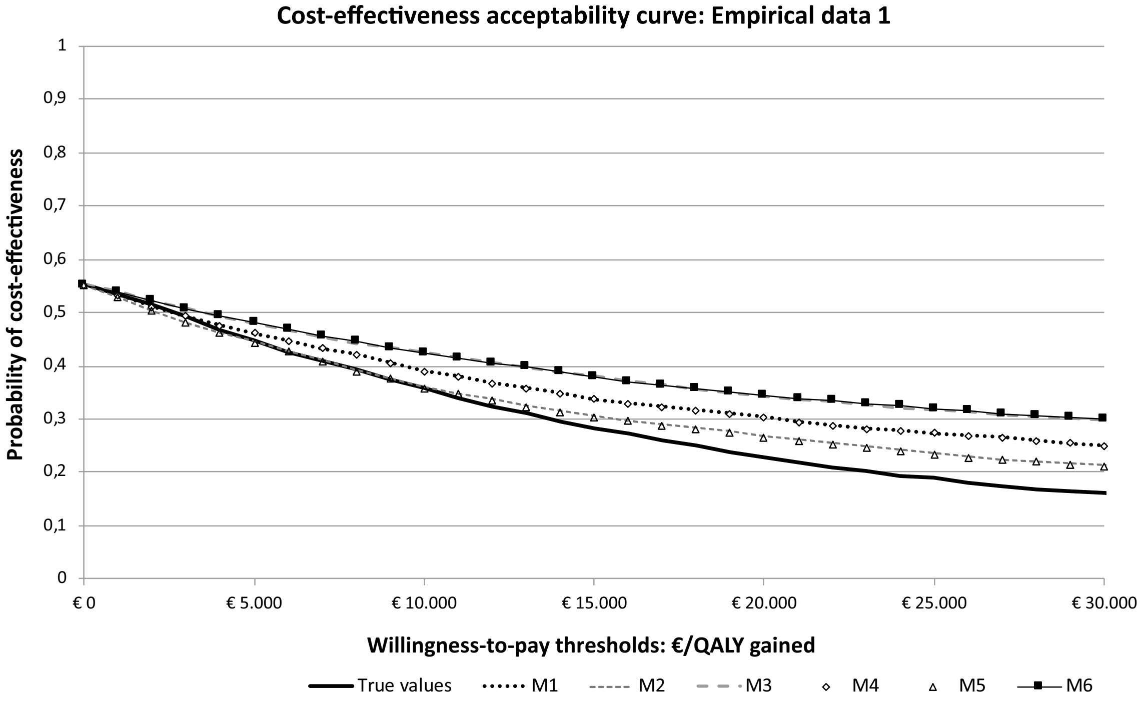

Fig. 3 Cost-effectiveness acceptability curves empirical dataset 1. *M1* model 1, *M2* model 2, *M3* model 3, *M4* model 4, *M5* model 5, *and M6* model 6

perform a cost-effectiveness analysis with QALYs as the outcome measure, even if EQ-5D-3L data are missing.

Comparison with literature

Our findings regarding the performance measures are more or less in line with the previous study by Carreon et al. [20], who aimed to predict individual LBP patients' EQ-5D-3L utility values based on their ODI scores. Their model performed slightly better in terms of its explained variance (i.e., R^2 was 61%) and its absolute fit (i.e., RMSE is 0.149), which is probably the result of a more homogenous study population, and therefore may indicate an overfitting of their model. Based on the RMSE, Carreon et al. [20] concluded that individual patients' EQ-5D-3L utility values could not validly be predicted from their ODI scores. Although we agree with this conclusion, we would like to stress that a low RMSE does not necessarily mean that the models cannot be used in the context of an cost-effectiveness analysis. This is true, when the bias surrounding the predicted utility values does not translate into relevant differences in incremental QALYs and the probability of the intervention being cost-effective compared to the control group (i.e., decision-based validity) [58]. This may be explained by the fact that the bias is likely to be similar in the intervention and control groups, thereby not affecting incremental QALYs and CEACs [59].

Strengths and limitations

To develop the models, a large sample of LBP patients from various settings (i.e., primary, secondary, and tertiary care) and with various complaint durations (i.e., subacute and chronic LBP) was used, which increases both the reliability and generalisability of the models. Moreover, next to OLS models, Tobit models were used to account for the constrained range of utility values [49, 50]. Although the added value of the Tobit model in this LBP population turned out to be rather limited, this might be different for LBP populations with milder symptoms, in which a larger share of patients is expected to report full health (i.e., a utility value of 1).

Our study also had some limitations. First, part of the sample was derived from two RCTs. Although RCT data may have limited generalisability, we chose to add these RCTs to our sample to create a more diverse sample and provide a better representation of the LPB population. Second, during the analysis, balanced data splitting was used to create the training and validation set. Although this balanced split provides better distribution of data then a random split, it might have been more appropriate to use K-fold cross validation[60]. Unfortunately, running the Tobit model using *k*-fold cross validation was not feasible as the R package for the Tobit model was not compatible with the *K*-fold package. In a post hoc analysis we developed and validated the

OLS models with *k*-fold cross validation and this produced similar results as our main analysis (data not shown). We also expect this to be the case for the Tobit models. Third, EQ-5D-3L utilities were used instead of EQ-5D-5L utilities. This is a limitation because EQ-5D-5L is known to be more sensitive and therefore recommended in pharmacoeconomic guidelines. Nonetheless, some countries still use the EQ-5D-3L. Therefore, we preferred to use the current relatively large dataset with EQ-5D-3L utility values of nearly 20,000 patients for developing and validating the models, instead of using a relatively small dataset with EQ-5D-5L. As the performance measures in the sensitivity analysis using the EQ-5D-5L reversed cross walk were comparable with those of the EQ-5D-3L version, we expect that EQ-5D-5L values can also be validly estimated using ODI scores. Fourth, the models were based on Dutch utility values. Previous research has shown that there are differences in utilities, QALYs, ICERs, and CEACs between countries due to the use of different value sets per country [14]. Therefore, we added the regression coefficients of models 2 and 5 for different countries in Supplementary Appendix F. These regression coefficients are based on the available value sets (tariffs) for different countries, and can be used to calculate utility values and QALYs. Fifth, some data that were used were to assess the performance of the developed models in a trial-based cost-effectiveness analysis setting were also part of the training set. However, as this was only a small percentage of the total training set (3.1%), we do not expect it to have influenced the validity of our finding that the difference between the estimated and true QALYs is small. Last, for assessing the performance of the developed models in a trial-based cost-effectiveness analysis setting, we only used data of two clinical trials, both of which found the intervention far from being cost-effective. That is, the probability of the interventions being cost-effective was low regardless of the willingness-to-pay threshold. In datasets where the interventions' cost-effectiveness is less conclusive, even small differences in the probability of an intervention being cost-effective might impact the overall conclusion of a study. Further research in the form of a simulation study, using simulated data to examine the generalisability beyond the datasets, is needed to assess the performance of the developed models in a wide range of trial-based cost-effectiveness analysis settings.

Implications for research and practice

Our findings suggest that predictive modeling can be used to estimate utility values from disease-specific measures, such as the ODI among LBP patients, when assessing incremental costs per QALY gained (as part of a cost-effectiveness analysis) or differences in utilities between groups. This is helpful for assessing cost-effectiveness in trials that did not directly measure utilities. Given the relatively large RMSE (i.e., low absolute fit of the models) and the relatively low r-square value (i.e., low relative fit) it is strongly discouraged to use the developed models to estimate the utility values of individual patients. Further research is needed to validate the models in order to (1) assess whether these models yield comparable results in other empirical datasets on LBP interventions, especially in analysis on interventions that are expected not to be conclusive in their cost-effectiveness, and (2) to improve their generalisability among different LBP patients by external validation in another sample. This study focused on assessing the validity of predictive regression modeling in estimating EQ-5D-3L utility values from the ODI and the impact of these estimated utility values on cost-effectiveness analysis. Results show that this is feasible for estimating QALYs and ICERs, but not for estimating individual utility scores. Further research is needed to explore whether other mapping methods, such as response mapping approaches like non-parametric and multinomial logistic regression [16, 54, 55], result in better predictive accuracy in estimating individual utility values of preference-based measures, such as the EQ-5D. This is important because studies suggest these mapping methods might be better at preventing regression to the mean [61]. Additional research might not only result in more accurate estimated utility values, but would also provide insight into the relative performance of different methods to estimate these values.

In the meantime, researchers can use the developed models in their cost-effectiveness analysis when utility values are lacking. Of them, the OLS model (i.e., model 2) is recommended in samples in which only a small number of patients has a utility value of 1 at baseline or follow-up measurement, whereas the Tobit model (i.e., model 5) is recommended in samples in which a substantial part of the sample has a utility score at baseline or at follow-up measurement. Although it seems possible to estimate utility values from disease-specific measures it is important to stress that it is still preferred to use preference-based quality-of-life measurements when setting up new studies.

Conclusion

Results of this study suggest that the ODI can be used to predict LBP patients' EQ-5D-3L utility values when the aim is to perform an cost-effectiveness analysis for QALYs, if utility values are missing, meaning in order to compare difference between groups of patients. The models are not suitable for estimating utility values for individual patients. Further research is needed to validate the models in order to assess whether these models yield comparable results in other empirical datasets on LBP interventions, to improve generalisability of the estimated models, and to compare the performance of predictive modeling compared to a mapping approach for estimating utility values. In the meantime, researchers can use the developed models in their cost-effectiveness analysis when utility values are lacking.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11136-022-03082-6.

Author contributions **SP:** methodology, data curation, formal analysis, writing—original draft, and project administration. **ÂJB:** methodology, data curation, formal analysis, and writing—review & editing. **JEB:** conceptualization, methodology, writing—review & editing, supervision, and funding acquisition. **RWJGO**: conceptualization, methodology, and writing—review & editing. **MWT**: resources, funding acquisition, and writing—review & editing. **ETM:** resources, and writing—review & editing. **FJPMH:** resources and writing—review & editing. **TO:** resources and writing—review & editing. **ATA:** Resources and writing—review & editing. **MLH:** resources and writing—review & editing. **JMD**: conceptualization, methodology, and writing—review & editing supervision and funding acquisition. All authors have agreed to submission of the final paper.

Funding Financial support was provided through a personal grant (receiver: JM van Dongen) issued by the Faculty of Sciences of the Vrije Universiteit Amsterdam.

Data availability Data are available through the corresponding author upon reasonable request.

Code availability Codes are available through the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet (London, England), 392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7
2. MacNeela, P., Doyle, C., O'Gorman, D., Ruane, N., & McGuire, B. E. (2015). Experiences of chronic low back pain: A meta-ethnography of qualitative research. *Health psychology review, 9*(1), 63–82. https://doi.org/10.1080/17437199.2013.840951
3. Froud, R., Patterson, S., Eldridge, S., Seale, C., Pincus, T., Rajendran, D., Fossum, C., & Underwood, M. (2014). A systematic review and meta-synthesis of the impact of low back pain on people's lives. *BMC Musculoskeletal Disorders, 15*, 50. https://doi.org/10.1186/1471-2474-15-50
4. Ihlebaek, C., Hansson, T. H., Laerum, E., Brage, S., Eriksen, H. R., Holm, S. H., Svendsrød, R., & Indahl, A. (2006). Prevalence of low back pain and sickness absence: A "borderline" study in Norway and Sweden. *Scandinavian Journal of Public Health, 34*(5), 555–558. https://doi.org/10.1080/14034940600552051
5. Steenstra, I. A., Munhall, C., Irvin, E., Oranye, N., Passmore, S., Van Eerd, D., Mahood, Q., & Hogg-Johnson, S. (2017). Systematic Review of Prognostic Factors for Return to Work in Workers with Sub Acute and Chronic Low Back Pain. *Journal of Occupational Rehabilitation, 27*(3), 369–381. https://doi.org/10.1007/s10926-016-9666-x
6. Hush, J. M., Refshauge, K., Sullivan, G., De Souza, L., Maher, C. G., & McAuley, J. H. (2009). Recovery: What does this mean to patients with low back pain? *Arthritis and Rheumatism, 61*(1), 124–131. https://doi.org/10.1002/art.24162
7. Maniadakis, N., & Gray, A. (2000). The economic burden of back pain in the UK. *Pain, 84*(1), 95–103. https://doi.org/10.1016/S0304-3959(99)00187-6
8. Lambeek, L. C., van Tulder, M. W., Swinkels, I. C., Koppes, L. L., Anema, J. R., & van Mechelen, W. (2011). The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. *Spine, 36*(13), 1050–1058. https://doi.org/10.1097/BRS.0b013e3181e70488
9. Drummond, M. F., Sculpher, M. J., Torrance, G. J., O'Brien, B. J., & Stoddart, G. L. (2005). *Methods for the economic evaluation of health care programmes*. Oxford University Press.
10. Brazier, J., Ratcliffe, J., Saloman, J., & Tsuchiya, A. (2016). *Measuring and valuing health benefits for economic evaluation*. Oxford University Press.
11. Froberg, D. G., & Kane, R. L. (1989). Methodology for measuring health-state preferences—I: Measurement strategies. *Journal of Clinical Epidemiology, 42*(4), 345–354. https://doi.org/10.1016/0895-4356(89)90039-5
12. Hakkaart-van Roijen L., van der Linden N., Bouwmans C. A. M., Kanters T. A., & Tan S. S. (2015) *Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg*. Zorginstituut Nederland. Retrieved August 30, 2021, from https://www.zorginstituutnederland.nl/over-ons/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg
13. National Institute for Health and Care Excellence. (2013). *Guide to the methods of technology appraisal*. Retrieved August 30, 2021, from https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l
14. van Dongen, J. M., Jornada Ben, Â., Finch, A. P., Rossenaar, M., Biesheuvel-Leliefeld, K., Apeldoorn, A. T., Ostelo, R., van Tulder, M. W., van Marwijk, H., & Bosmans, J. E. (2021). Assessing the impact of EQ-5D country-specific value sets on cost-utility outcomes. *Medical care, 59*(1), 82–90. https://doi.org/10.1097/MLR.0000000000001417
15. Gianola, S., Frigerio, P., Agostini, M., Bolotta, R., Castellini, G., Corbetta, D., Gasparini, M., Gozzer, P., Guariento, E., Li, L. C., Pecoraro, V., Sirtori, V., Turolla, A., Andreano, A., & Moja, L. (2016). Completeness of outcomes description reported in low back pain rehabilitation interventions: A Survey of 185 randomized trials. *Physiotherapy Canada, 68*(3), 267–274. https://doi.org/10.3138/ptc.2015-30.PMID:27909376;PMCID:PMC5125456

16. Fayers, P. M., & Hays, R. D. (2014). Should linking replace regression when mapping from profile-based measures to preference-based measures? *Value in Health, 17*(2), 261–265. https://doi.org/10.1016/j.jval.2013.12.002
17. Longworth, L., & Rowen, D. (2013). Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value in Health, 16*(1), 202–210. https://doi.org/10.1016/j.jval.2012.10.010
18. Chapman, J. R., Norvell, D. C., Hermsmeyer, J. T., Bransford, R. J., DeVine, J., McGirt, M. J., & Lee, M. J. (2011). Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine, 36*(21 Suppl), S54–S68. https://doi.org/10.1097/BRS.0b013e31822ef74d
19. Fairbank, J. C., Couper, J., Davies, J. B., & O'Brien, J. P. (1980). The Oswestry low back pain disability questionnaire. *Physiotherapy, 66*(8), 271–273.
20. Chiarotto, A., Boers, M., Deyo, R. A., Buchbinder, R., Corbin, T. P., Costa, L., Foster, N. E., Grotle, M., Koes, B. W., Kovacs, F. M., Lin, C. C., Maher, C. G., Pearson, A. M., Peul, W. C., Schoene, M. L., Turk, D. C., van Tulder, M. W., Terwee, C. B., & Ostelo, R. W. (2018). Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain, 159*(3), 481–495. https://doi.org/10.1097/j.pain.0000000000001117
21. Clement, R. C., Welander, A., Stowell, C., Cha, T. D., Chen, J. L., Davies, M., Fairbank, J. C., Foley, K. T., Gehrchen, M., Hagg, O., Jacobs, W. C., Kahler, R., Khan, S. N., Lieberman, I. H., Morisson, B., Ohnmeiss, D. D., Peul, W. C., Shonnard, N. H., Smuck, M. W., et al. (2015). A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthopaedica, 86*(5), 523–533. https://doi.org/10.3109/17453674.2015.1036696
22. Carreon, L. Y., Bratcher, K. R., Das, N., Nienhuis, J. B., & Glassman, S. D. (2014). Estimating EQ-5D values from the Oswestry Disability Index and numeric rating scales for back and leg pain. *Spine, 39*(8), 678–682. https://doi.org/10.1097/BRS.0000000000000220
23. Mutubuki, E. N., Luitjens, M. A., Maas, E. T., Huygen, F., Ostelo, R., van Tulder, M. W., & van Dongen, J. M. (2020). Predictive factors of high societal costs among chronic low back pain patients. *European Journal of Pain (London, England), 24*(2), 325–337. https://doi.org/10.1002/ejp.1488
24. Maas, E. T., Juch, J. N., Groeneweg, J. G., Ostelo, R. W., Koes, B. W., Verhagen, A. P., van Raamt, M., Wille, F., Huygen, F. J., & van Tulder, M. W. (2012). Cost-effectiveness of minimal interventional procedures for chronic mechanical low back pain: Design of four randomised controlled trials with an economic evaluation. *BMC Musculoskeletal Disorders, 13*, 260. https://doi.org/10.1186/1471-2474-13-260
25. Juch, J., Maas, E. T., Ostelo, R., Groeneweg, J. G., Kallewaard, J. W., Koes, B. W., Verhagen, A. P., van Dongen, J. M., Huygen, F., & van Tulder, M. W. (2017). Effect of radiofrequency denervation on pain intensity among patients with chronic low back pain: The mint randomized clinical trials. *JAMA, 318*(1), 68–81. https://doi.org/10.1001/jama.2017.7918
26. van Hooff, M. L., van Loon, J., van Limbeek, J., & de Kleuver, M. (2014). The Nijmegen decision tool for chronic low back pain. Development of a clinical decision tool for secondary or tertiary spine care specialists. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0104226
27. van Dongen, J. M., van Hooff, M. L., Spruit, M., de Kleuver, M., & Ostelo, R. (2017). Which patient-reported factors predict referral to spinal surgery? A cohort study among 4987 chronic low back pain patients. *European Spine Journal, 26*(11), 2782–2788. https://doi.org/10.1007/s00586-017-5201-9
28. van Hooff, M. L., van Dongen, J. M., Coupé, V. M., Spruit, M., Ostelo, R., & de Kleuver, M. (2018). Can patient-reported profiles avoid unnecessary referral to a spine surgeon? An observational study to further develop the Nijmegen Decision Tool for Chronic Low Back Pain. *PLoS ONE, 13*(9), e0203518. https://doi.org/10.1371/journal.pone.0203518
29. Apeldoorn, A. T., Ostelo, R. W., van Helvoirt, H., Fritz, J. M., de Vet, H. C., & van Tulder, M. W. (2010). The cost-effectiveness of a treatment-based classification system for low back pain: Design of a randomised controlled trial and economic evaluation. *BMC Musculoskeletal Disorders, 11*, 58. https://doi.org/10.1186/1471-2474-11-58
30. Apeldoorn, A. T., Ostelo, R. W., van Helvoirt, H., Fritz, J. M., Knol, D. L., van Tulder, M. W., & de Vet, H. C. (2012). A randomized controlled trial on the effectiveness of a classification-based system for subacute and chronic low back pain. *Spine, 37*(16), 1347–1356. https://doi.org/10.1097/BRS.0b013e31824d9f2
31. Oosterhuis, T., van Tulder, M., Peul, W., Bosmans, J., Vleggeert-Lankamp, C., Smakman, L., Arts, M., & Ostelo, R. (2013). Effectiveness and cost-effectiveness of rehabilitation after lumbar disc surgery (REALISE): Design of a randomised controlled trial. *BMC Musculoskeletal Disorders, 14*, 124. https://doi.org/10.1186/1471-2474-14-124
32. Oosterhuis, T., Ostelo, R. W., van Dongen, J. M., Peul, W. C., de Boer, M. R., Bosmans, J. E., Vleggeert-Lankamp, C. L., Arts, M. P., & van Tulder, M. W. (2017). Early rehabilitation after lumbar disc surgery is not effective or cost-effective compared to no referral: A randomised trial and economic evaluation. *Journal of Physiotherapy, 63*(3), 144–153. https://doi.org/10.1016/j.jphys.2017.05.016
33. EuroQol Group. (1990). EuroQol–a new facility for the measurement of health-related quality of life. *Health Policy (Amsterdam, Netherlands), 16*(3), 199–208. https://doi.org/10.1016/0168-8510(90)90421-9
34. Lamers, L. M., Stalmeier, P. F., McDonnell, J., Krabbe, P. F., & van Busschbach, J. J. (2005). Kwaliteit van leven meten in economische evaluaties: Het Nederlands EQ-5D-tarief [Measuring the quality of life in economic evaluations: The Dutch EQ-5D tariff]. *Nederlands Tijdschrift Voor Geneeskunde, 149*(28), 1574–1578.
35. Fairbank, J. C., & Pynsent, P. B. (2000). The Oswestry Disability Index. *Spine, 25*(22), 2940–2952. https://doi.org/10.1097/00007632-200011150-00017
36. Hudson-Cook, N., Tomes-Nicholson, K., & Breen, A. A. (1989). Revised Oswestry disability questionnaire. In M. Roland & J. R. Jenner (Eds.), *Back pain: New approaches to rehabilitation and education* (pp. 187–204). Manchester University Press.
37. Yeomans, S. G., & Liebenson, C. (1997). Applying outcomes management to clinical practice. *Journal of the Neuromusculoskeletal System, 5*(1), 1–14.
38. Shearer H. M. (2007). Rehabilitation of the Spine—A practitioner's manual, 2nd Ed. *The Journal of the Canadian Chiropractic Association, 51*(1), 62.
39. van Hooff, M. L., Spruit, M., Fairbank, J. C., van Limbeek, J., & Jacobs, W. C. (2015). The Oswestry Disability Index (version 2.1a): validation of a Dutch language version. *Spine, 40*(2), E83–E90. https://doi.org/10.1097/BRS.0000000000000683
40. Downie, W. W., Leatham, P. A., Rhind, V. M., Wright, V., Branco, J. A., & Anderson, J. A. (1978). Studies with pain rating scales. *Annals of the Rheumatic Diseases, 37*(4), 378–381. https://doi.org/10.1136/ard.37.4.378
41. Boonstra, A. M., Stewart, R. E., Köke, A. J., Oosterwijk, R. F., Swaan, J. L., Schreurs, K. M., & Schiphorst Preuper, H. R. (2016). Cut-off points for mild, moderate, and severe pain on the numeric rating scale for pain in patients with chronic musculoskeletal pain: Variability and influence of sex and catastrophizing. *Frontiers in Psychology, 7*, 1466. https://doi.org/10.3389/fpsyg.2016.01466

42. Husky, M. M., Ferdous Farin, F., Compagnone, P., Fermanian, C., & Kovess-Masfety, V. (2018). Chronic back pain and its association with quality of life in a large French population survey. *Health and Quality of Life Outcomes, 16*(1), 195. https://doi.org/10.1186/s12955-018-1018-4
43. Horng, Y. S., Hwang, Y. H., Wu, H. C., Liang, H. W., Mhe, Y. J., Twu, F. C., & Wang, J. D. (2005). Predicting health-related quality of life in patients with low back pain. *Spine, 30*(5), 551–555. https://doi.org/10.1097/01.brs.0000154623.20778.f0
44. Kovacs, F. M., Abraira, V., Zamora, J., Fernández, C., & Spanish Back Pain Research Network. (2005). The transition from acute to subacute and chronic low back pain: A study based on determinants of quality of life and prediction of chronic disability. *Spine, 30*(15), 1786–1792. https://doi.org/10.1097/01.brs.0000172159.47152.dc
45. Lamé, I. E., Peters, M. L., Vlaeyen, J. W., Kleef, M., & Patijn, J. (2005). Quality of life in chronic pain is more associated with beliefs about pain, than with pain intensity. *European Journal of Pain (London, England), 9*(1), 15–24. https://doi.org/10.1016/j.ejpain.2004.02.006
46. Bentsen, S. B., Wahl, A. K., Strand, L. I., & Hanestad, B. R. (2007). Relationships between demographic, clinical and pain variables and health-related quality of life in patients with chronic low back pain treated with instrumented fusion. *Scandinavian Journal of Caring Sciences, 21*(1), 134–143. https://doi.org/10.1111/j.1471-6712.2007.00440.x
47. Coste, J., Lefrançois, G., Guillemin, F., Pouchot, J., & French Study Group for Quality of Life in Rheumatology. (2004). Prognosis and quality of life in patients with acute low back pain: Insights from a comprehensive inception cohort study. *Arthritis and Rheumatism, 51*(2), 168–176. https://doi.org/10.1002/art.20235
48. Hutcheson, G. D. (1999). *The multivariate social scientist*. SAGE Publications, Ltd.
49. Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrics, 26*(1), 24–36.
50. Austin, P. C., Escobar, M., & Kopec, J. A. (2000). The use of the Tobit model for analyzing measures of health status. *Quality of Life Research, 9*(8), 901–910. https://doi.org/10.1023/a:1008938326604
51. Smith, G. (2018). Step away from stepwise. *Journal of Big Data*. https://doi.org/10.1186/s40537-018-0143-6
52. Chowdhury, M., & Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health*. https://doi.org/10.1136/fmch-2019-000262
53. Soer, R., Reneman, M. F., Speijer, B. L., Coppes, M. H., & Vroomen, P. C. (2012). Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. *The Spine Journal, 12*(11), 1035–1039. https://doi.org/10.1016/j.spinee.2012.10.030
54. Terluin, B., van Marwijk, H. W., Adèr, H. J., de Vet, H. C., Penninx, B. W., Hermens, M. L., van Boeijen, C. A., van Balkom, A. J., van der Klink, J. J., & Stalman, W. A. (2006). The Four-Dimensional Symptom Questionnaire (4DSQ): A validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. *BMC Psychiatry, 6*, 34. https://doi.org/10.1186/1471-244X-6-34
55. van Hout, B. A., & Shaw, J. W. (2021). Mapping EQ-5D-3L to EQ-5D-5L. *Value in Health, 24*(9), 1285–1293. https://doi.org/10.1016/j.jval.2021.03.009
56. Euroqol. *Cross-walk and reverse cross-walk*. Retrieved August 30, 2021, from https://euroqol.org/support/tools/analysis-tools/cross-walk-reverse-cross-walk/
57. Euroqol. *EQ-5D-5L version*. Retrieved August 30, 2021, from https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/
58. Griffin, S. (2010). *Dealing with uncertainty in the economic evaluation of health care technologies*. University of York.
59. Ben, Â., Finch, A. P., van Dongen, J. M., de Wit, M., van Dijk, S., Snoek, F. J., Adriaanse, M. C., van Tulder, M. W., & Bosmans, J. E. (2020). Comparing the EQ-5D-5L crosswalks and value sets for England, the Netherlands and Spain: Exploring their impact on cost-utility results. *Health Economics, 29*(5), 640–651. https://doi.org/10.1002/hec.4008
60. Blum, A., Kalai, A., & Langford, J. (1999). Beating the hold-out: Bounds for k-fold and progressive cross- validation. In Association for Computing Machinery, *Proceedings of the twelfth annual conference on Computational learning theory (COLT '99)*. (pp. 203–208).
61. Thompson, N. R., Lapin, B. R., & Katzan, I. L. (2017). Mapping PROMIS global health items to EuroQol (EQ-5D) utility scores using linear and equipercentile equating. *PharmacoEconomics, 35*(11), 1167–1176. https://doi.org/10.1007/s40273-017-0541-1