Ayala, A., Forjaz, M. J., Ramallo-Fariña, Y., Martín-Fernández, J., García-Pérez, L., & Bilbao, A. (2021). Response Mapping Methods to Estimate the EQ-5D-5L From the Western Ontario McMaster Universities Osteoarthritis in Patients With Hip or Knee Osteoarthritis. *Value in Health,* 24 (6), pp. 874-883.

DOI: 10.1016/j.jval.2021.01.003

Response Mapping Methods to Estimate the EQ-5D-5L From the Western Ontario McMaster Universities Osteoarthritis in Patients With Hip or Knee Osteoarthritis

Alba Ayala, MSc, Maria João Forjaz, PhD, Yolanda Ramallo-Fariña, MSc, Jesús Martín-Fernández, MD, PhD, Lidia García-Pérez, MSc, Amaia Bilbao, PhD

A B S T R A C T

Objectives: The mapping technique can estimate generic preference-based measure scores through a specific measure that cannot be used in economic evaluations. This study compared 2 response mapping methods to estimate EQ-5D-5L scores using the Western Ontario McMaster Universities Osteoarthritis (WOMAC).

Methods: The sample consisted of 758 patients with the hip or knee osteoarthritis recruited in baseline. Bayesian networks (BN) and multinomial logistic regression (ML) were used as response mapping models. Predictions were obtained using the 6-month follow-up as a validation sample. The mean absolute error, mean squared error, deviation from the root mean squared error and intraclass correlation coefficient were calculated as precision measures.

Results: There was 5.5% of missing data, which was removed. The mean age was 69.6 years (standard deviation = 10.5), with 61.6% of women. The BN model presented lower mean absolute error, mean squared error, root mean squared error and higher intraclass correlation coefficient than the ML model. Only the WOMAC items pain and physical function items were related with the EQ-5D-5L dimensions.

Conclusion: BN response mapping models are more robust methods, with better prediction results, than ML models. The BN model also provided a graphic representation of the dependency relationships between the EQ-5D-5L dimensions and the different WOMAC items that could be useful in the clinical investigation of patients with hip or knee osteoarthritis.

Keywords: Bayesian networks, EQ-5D-5L, osteoarthritis, response mapping, WOMAC.

VALUE HEALTH. 2021; 24(6):874–883

Introduction

Knee and hip osteoarthritis (OA) is the most common and prevalent joint degenerative disease in older population and the main cause of functional disability and chronic pain.[1] These consequences impair patient quality of life and increase hospitalization rates, which is why the economic burden and indirect costs of OA are especially high.[2] Therefore, the costs and quality-of-life impact of OA must be analyzed by economic evaluation to ensure efficient use of limited healthcare resources.[3]

In clinical practice, economic evaluation techniques help to prioritize and decide which intervention is appropriate when resources are scarce.[4] Investigators try to evaluate interventions in terms of costs and quality-adjusted life years (QALYs). To facilitate the calculation of a QALY measure, the EuroQol group developed the EQ-5D, a health-related quality of life (HRQoL) instrument. The EQ-5D comprises 5 health questions or dimensions, which can be converted into a single index as a utility measure. The EQ-5D utility index provides a simple, generic measure that can be used in clinical and economic evaluation.[5] Apart from being one of the tools most commonly used in this type of evaluation,[6] its use is recommended in all economic evaluations to ensure comparability between studies.[7] The current version is EQ-5D-5L,[8] which compared with the previous 3-level measure (EQ-5D-3L),[5] presents a series of advantages such as better reliability and sensitivity, maintaining the viability and reducing the ceiling effect.[9]

In both versions, EQ-5D is a valid and reliable measure in patients with OA of the hip or knee.[10-12] However, in clinical trials of patients with OA, it is usual to rely on specific HRQoL measures, such as the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC).[13,14] This WOMAC measure has been chosen in many clinical trials because it is highly sensitive to change and to the detection of a minimally important difference in patients with OA.[15] One of the disadvantages of WOMAC is that its scores do not provide a preference-based measure, making it unsuitable for economic evaluations. The best way to address these studies would be to use the utility measure directly. However, mapping or cross-walking techniques can be a solution for considering nonpreference-based measures, such as the WOMAC as QALYs and thus suitable for its use in economic evaluations.

Two types of strategies are frequently used in mapping studies with the EQ-5D. The direct methods used most often consist in predicting the EQ-5D index through nonpreference-based measures. Indirect methods or response mapping predict the response categories of the EQ-5D dimensions through the other measure items, and subsequently the index is calculated. Provided that the sample size is large enough, response mapping can be a useful technique because it does not assume, a priori, that the responses follow a normal distribution.[16,17]

Previous mapping studies have been conducted between the EQ-5D-3L and the WOMAC, all with direct methods.[18-23] In addition, the EQ-5D update from 3 to 5 levels requires additional analysis to verify the viability of using the WOMAC to estimate the scores of the EQ-5D-5L. A recent article developed functions to map the WOMAC onto the EQ-5D-5L in patients with hip or knee OA using direct methods, although indirect methods (response mapping) have not been evaluated yet.[24] Therefore, this study aims to use 2 mapping response algorithms previously developed, and to evaluate which technique is most appropriate for estimating the EQ-5D-5L scores through the different WOMAC items in patients with hip or knee OA.

Methods

Sample

The sample used data from a cohort of 758 patients recruited in different hospitals and primary care (PC) centers of 3 autonomous regions in Spain: Basque Country, Canary Islands, and Madrid. Patients diagnosed with OA of the hip or knee according to the American Rheumatism Association criteria,[25,26] who visited traumatology and rheumatology departments, or had a PC appointment, were included in the study as a convenience sample. Any patients who refused to participate in the study or who could not complete the measuring instruments, patients with organic or psychiatric pathologies, and any who could not read or did not understand Spanish, were excluded. The study was approved by the institutional review boards of each institution. All participants signed informed consent forms and took part in the study on a voluntary basis. More information of the sample can be found in another mapping article.[24] Because our models only work with complete cases, individuals who did not answer the questionnaires in baseline for the 2 measures used (EQ-5D-5L and WOMAC) were eliminated in all statistical analyses and represented a total 5.5% of missing values. The baseline sample (training data) was used to develop the mapping models and the 6-month follow-up data to validate these models.

Measures

The EQ-5D-5L is a descriptive system comprising 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.[8] Each dimension is answered in 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The EQ-5D-5L descriptive system presents 3125 profiles that are combined to calculate a single index, where the best health status is "11111" and the worst is "55555." For this study, the index was calculated using the set of specific values of the Spanish version of the EQ-5D-5L.[27] The index has a value of 1 for the best health status and negative values for those states of health valued as worse than death, ranging from −0.416 to 1 for Spanish population.[27]

In this study, we used the Spanish version of WOMAC.[14] It consists of 24 items grouped in 3 components: pain (5 items), stiffness (2 items), and physical function (17 items). Each of the items has 5 Likert-type response options, ranging from 0 (no problem) to 4 (severe problems). The total score was obtained from the standardized sum, ranging from 0 to 100, of each of its 3 components, where a high value means a worse quality of life. The domains scores were calculated in a similar way.

Data Analysis

Exploratory analysis

The frequencies and percentages of each response category of the items of both questionnaires were calculated along with the mean and standard deviation (SD) for the continuous variables. The distributions of the EQ-5D-5L index and WOMAC dimensions were graphically observed through boxplots and the relationship between them through scatterplots. Spearman correlations between the EQ-5D-5L index and the WOMAC dimensions also were calculated, considering a moderate correlation between 0.5 and 0.7 and a high correlation for a value greater than 0.7.[28]

Predictive Models

The technique applied in this study, response mapping, consisted in estimating the scores of each EQ-5D-5L dimension (outcome variables) through the 24 WOMAC items (predictors), using a model for each EQ-5D-5L dimension. Therefore, Bayesian networks (BN)[29] and multinomial logistic regression (ML) were applied because these models use EQ-5D-5L dimensions as dependent variables. Other direct mapping methods to compare with response mapping were performed in another article using the same database.[24]

BN have been previously used as mapping algorithms in other studies.[30,31] BN models can be represented graphically, showing dependency relationships between variables clearly and intuitively, and handling uncertainty through the theory of probability.[32] In our analysis, the algorithm used for learning the BN structure was the Bayesian classifier Tree-Augmented Naive Bayes.[33,34] This classifier provides a tree structure that starts by relating the predictor variables (WOMAC items) to each other and then connects these predictors with the outcome variable (each of the EQ-5D-5L dimensions). The WOMAC items were introduced into BN model as categorical variables.

ML regression was also used as a response mapping method since the outcome variable is polytomous.[16] The ML regression uses a linear model to predict the probabilities of the different possible dependent variable categories, given a set of independent variables; the model's coefficients are interpreted by calculating the odds ratio (OR), and the statistical significance of the independents variables is evaluated through the Wald statistic. The WOMAC items were introduced into ML model as categorical variables.

Variable Selection

Strategies were developed for each model to eliminate any variables that could generate noise in the predictions, in order to obtain the simplest possible model according to the parsimony principle. In each BN, the WOMAC items that showed dependency relationships with the outcome variable (EQ-5D-5L dimensions) were selected, and the relationships were evaluated through the χ^2 test (significance level of 5%), a method used in other mapping studies with BN.[30,31] In the models developed with ML, the WOMAC items were selected by calculating the Akaike information criterion (AIC) for all possible models and choosing the model with the lowest AIC. The analyses were also performed using the

Table 1. EQ-5D-5L and WOMAC distributions in frequencies and percentages (baseline sample, n = 716).

EQ-5D-5L		Categories									
		No problems		Slight problems		Moderate problems		Severe problems		Extreme problems	
		1		2		3		4		5	
		n	%	N	%	n	%	n	%	n	%
	Mobility	83	11.6	143	20.0	306	42.7	170	23.7	14	2.0
	Self-care	208	29.1	198	27.7	225	31.4	80	11.2	5	0.7
	Usual activities	119	16.6	178	24.9	266	37.2	117	16.3	36	5.0
	Pain/discomfort	46	6.4	156	21.8	260	36.3	220	30.7	34	4.7
	Anxiety/depression	314	43.9	168	23.5	135	18.9	77	10.8	22	3.1
WOMAC (mean, 48.1; SD, 20.2; range, 0-100)		0		1		2		3		4	
		n	%	N	%	n	%	n	%	n	%
Pain dimension: Rate your pain when											
W1a	Walking on an even floor	49	6.8	238	33.2	267	37.3	115	16.1	47	6.6
W1b	Going up or down stairs	30	4.2	94	13.1	221	30.9	220	30.7	151	21.1
W1c	In bed at night	140	19.6	238	33.2	186	26.0	97	13.5	55	7.7
W1d	Sitting or lying down	133	18.6	321	44.8	169	23.6	69	9.6	24	3.4
W1e	Standing	51	7.1	190	26.5	248	34.6	155	21.6	72	10.1
Stiffness dimension											
W2a	Stiffness after getting up in the morning	83	11.6	195	27.2	231	32.3	135	18.9	72	10.1
W2b	Stiffness the rest of the day after sitting, lying or resting	59	8.2	192	26.8	271	37.8	139	19.4	55	7.7
Physical function dimension											
W3a	Going down stairs	37	5.2	122	17.0	245	34.2	191	26.7	121	16.9
W3b	Going up stairs	28	3.9	117	16.3	246	34.4	200	27.9	125	17.5
W3c	Standing up after sitting	25	3.5	142	19.8	251	35.1	192	26.8	106	14.8
W3d	Standing	53	7.4	171	23.9	260	36.3	169	23.6	63	8.8
W3e	Bending and picking something up	34	4.7	123	17.2	198	27.7	169	23.6	192	26.8
W3f	Walking on an even floor	76	10.6	241	33.7	258	36.0	109	15.2	32	4.5
W3g	Getting in and out of a car	29	4.1	116	16.2	235	32.8	181	25.3	155	21.6
W3h	Going shopping	55	7.7	137	19.1	236	33.0	181	25.3	107	14.9
W3i	Putting on socks or stockings	54	7.5	137	19.1	199	27.8	141	19.7	185	25.8
W3j	Getting up from bed	72	10.1	194	27.1	233	32.5	150	20.9	67	9.4
W3k	Taking off stockings or socks	64	8.9	157	21.9	200	27.9	130	18.2	165	23.0
W3l	Lying in bed	161	22.5	287	40.1	160	22.3	77	10.8	31	4.3
W3m	Getting in/out of shower/bath	118	16.5	212	29.6	206	28.8	108	15.1	72	10.1
W3n	Sitting	143	20.0	314	43.9	169	23.6	72	10.1	18	2.5
W3o	Getting on/off toilet	87	12.2	201	28.1	212	29.6	145	20.3	71	9.9
W3p	Doing heavy domestic duties	23	3.2	72	10.1	197	27.5	209	29.2	215	30.0
W3q	Doing light domestic duties	83	11.6	258	36.0	230	32.1	121	16.9	24	3.4

SD indicates standard deviation; WOMAC, Western Ontario McMaster Universities Osteoarthritis.

Bayesian Information Criterion. However, with this criterion only 1 or 2 WOMAC items were related to the EQ-5D-5L dimensions in the ML model. Also, the predicted errors were lower than the ones reported in this study (>0.3). Therefore, the AIC method was used as it provided the best prediction results for ML models.

Predictions

The predictions and the base case result of the 2 developed models were calculated using both the most-likely probability and the expected value methods. The most-likely probability method consists of predicting the response categories of each of the EQ-5D-5L dimensions and subsequently calculating the index of the Spanish version of the EQ-5D-5L.[27] The expected value method[30,35] calculates the predicted utility EQ-5D-5L index through the following algebraic equation, using the value set of the Spanish population:

$$
\begin{aligned}
&\text{Utility EQ}-5\text{D}-5\text{L index} = \\
&1 - 0.084 \times P(MO=2) + 0.099 \times P(MO=3) + 0.250 \times P(MO=4) \\
&\quad + 0.337 \times P(MO=5) + 0.050 \times P(SC=2) + 0.053 \times P(SC=3) \\
&\quad + 0.164 \times P(SC=4) + 0.196 \times P(SC=5) + 0.044 \times P(UA=2) \\
&\quad + 0.049 \times P(UA=3) + 0.135 \times P(UA=4) + 0.153 \times P(UA=5) \\
&\quad + 0.078 \times P(PD=2) + 0.101 \times P(PD=3) + 0.245 \times P(PD=4) \\
&\quad + 0.382 \times P(PD=5) + 0.081 \times P(AD=2) + 0.128 \times P(AD=3) \\
&\quad + 0.270 \times P(AD=4) + 0.348 \times P(AD=5)
\end{aligned}
$$

where Pij represents the probability of j response on dimension i estimated with BN and ML models.

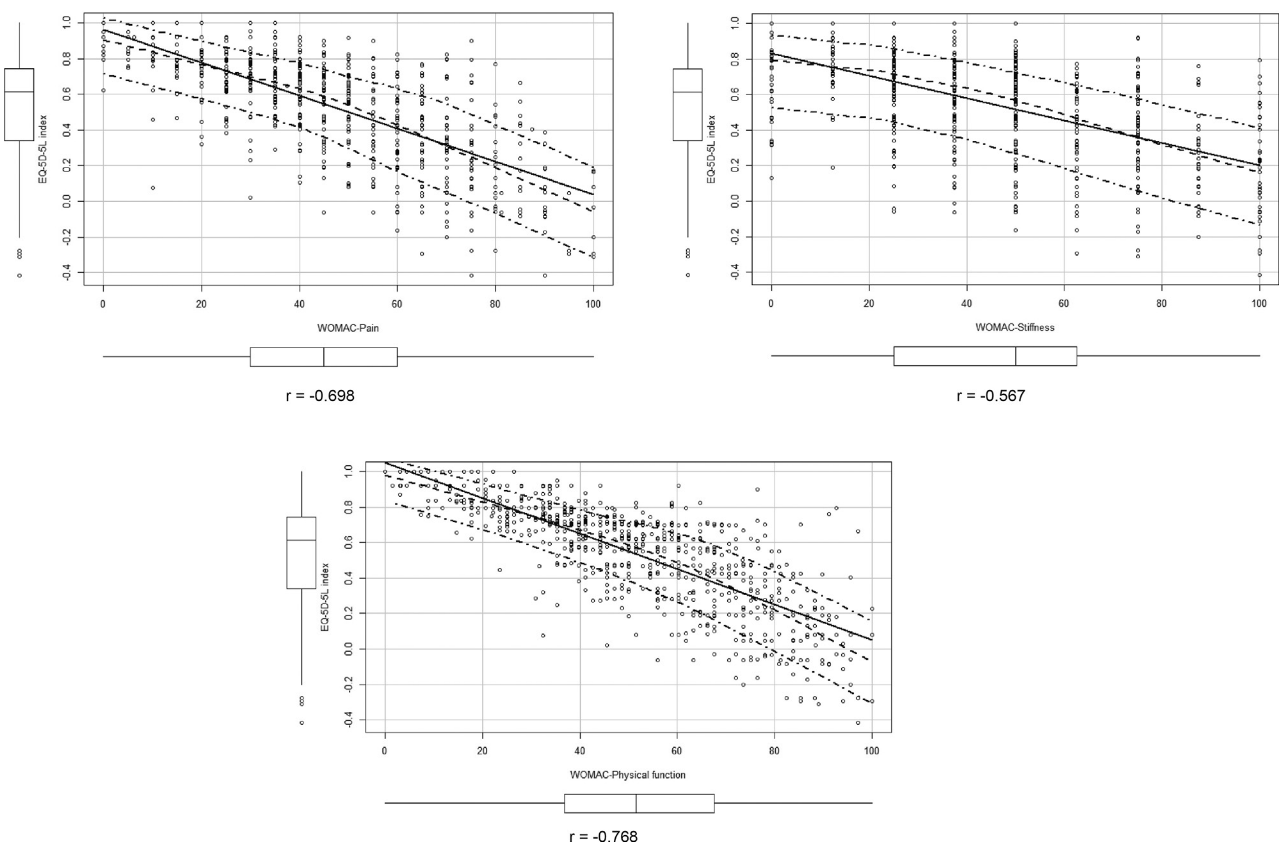
The baseline sample (training data) was used to develop the mapping models (BN and ML) and the 6-month follow-up (validation) data to validate these models.

Predictive Accuracy

The BN and ML models were compared using the EQ-5D-5L index prediction errors: mean absolute error (MAE), mean squared error (MSE) and deviation from the root mean squared error (RMSE). The 95% probability interval of these measures was calculated using the bootstrap method. To this end, the MAE, MSE, and RMSE were estimated using the predicted and the observed value of the validation sample in each replication of 1000 bootstrapped samples. The probability interval lower and upper bounds were calculated as the 2.5th and 97.5th percentile of the bootstrapped value distribution. In addition, we calculated the proportion of agreement between the prediction and observed values of EQ-5D index measured through the intraclass correlation coefficient (ICC). Predictive accuracy was evaluated according to intervals of observed EQ-5D-5L index and WOMAC score. The prediction distributions were also compared to the observed values of the index through graphs.

Data were analyzed with the program R, version 3.1.3, using the commands "tree.bayes" and "predict" from the bnlearn package[36,37] to develop the BN, the "multinom" command from de *nnet* package for the ML model and the hydroGOF package to compare

Figure 1. Scatterplots and boxplots between the EQ-5D-5L and Western Ontario McMaster Universities Osteoarthritis dimensions. r: Spearman correlation coefficient. The straight line in the scatterplots corresponds to the linear regression formed by both variables. The smooth line and the 95% confidence intervals correspond to a nonparametric estimate of the mean function of the vertical axis variable given the horizontal axis. In the box plot, the minimum, first quartile, median, third quartile, and maximum are drawn.



the predictions between models (calculation of MAE, MSE and RMSE).

Results

The average age of the baseline sample of 716 patients with OA of the hip or knee was 69.6 years (SD = 10.5), and 61.6% were women. The EQ-5D-5L index presented an average of 0.53 (SD = 0.29), while the means of the WOMAC dimensions were: pain, 46.43 (SD = 21.60); stiffness, 47.50 (SD = 25.71); and physical function, 51.72 (SD = 21.97). The follow-up sample included 633 patients, 1.7% of missing data were excluded from this sample.

The highest frequency found in the EQ-5D-5L dimensions was in the moderate-problem category for all dimensions in baseline (Table 1) and in follow-up sample (Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.003), except in anxiety/depression, where patients most frequently responded that they did not have any problem (43.9%). Because they were patients with OA, severe problems also predominated in the mobility (23.7%) and pain/discomfort (30.7%) dimensions.

In the WOMAC, items W3e (bending and picking something up), W3i (putting on socks), and W3p (doing heavy domestic duties) had a ceiling effect (26.8, 25.8 and 30.0%, respectively). A nonlinear relationship was observed between the EQ-5D-5L index and the WOMAC dimensions (Fig. 1). The EQ-5D-5L index presented a moderate, negative correlation with 2 of the WOMAC dimensions, pain ($r = -0.698$) and stiffness ($r = -0.567$), and a high negative correlation with the physical function dimension ($r = -0.768$), with $P < .05$ in all cases.

The predictive accuracy was higher using the expected value method than using the most-likely probability method (Table 2). For the expected value method, the BN model presented a lower MAE, MSE, RMSE and higher ICC values than the ML model. The predictive accuracy according to observed EQ-5D-5L index and WOMAC scores can be observed in Table 3 for the expected value method. The results showed that the MAE and the RMSE of both models were higher for negative health states (worse than death) than for positive health states. Also, the prediction errors of both models vary with the WOMAC score, with worse values for higher WOMAC scores. This pattern was also observed with the most-likely probability method to predict the EQ-5D-5L index (Table S2 Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.003). Figure 2 presents the distribution of predicted EQ-5D-5L index for both models compared to observed index.

The graphic structure of the BN models can be seen in Fig. 3. The direction of the arrows indicates a causal relationship between the constructs measured by EQ-5D-5L dimensions and WOMAC items. The EQ-5D-5L mobility and daily activities dimensions are associated to most WOMAC items. Two items of the WOMAC pain dimension are present in all the BN models, the W1a item (pain when walking on an even floor) is related to all the EQ-5D-5L dimensions except for the personal care dimension, and item W1b (pain when going up or down stairs) is observed in all dimensions except in anxiety/depression. The other WOMAC items shown in the BN models belong to the physical function dimension. No WOMAC stiffness dimension item was related to the EQ-5D-5L dimensions. The conditional probabilities for each combination of WOMAC items can be found in a .csv file (https://www.dropbox.com/s/i858bfjtqohn2vk/conditional_probabilities_BN_Supplementary_material.xlsx?dl = 0), and the description in Appendix Tables 3 and 4 (in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.003) shows the final ML model parameters or coefficients. These coefficients can be used to estimate the probabilities of each EQ-5D-5L dimension. More

Table 2. Descriptive statistics of the predictions of the EQ-5D-5L index and precision measurements for each of the models (follow-up validation sample).

	BN*	BN†	ML*	ML†	Observation EQ-5D-5L
Predictive accuracy					
MAE (95% PI)	0.132	0.121	0.149	0.134	-
	(0.122-0.142)	(0.113-0.130)	(0.138-0.160)	(0124-0143)	
MSE (95% PI)	0.033	0.027	0.042	0.035	-
	(0.029-0.039)	(0.023-0.032)	(0.036-0.048)	(0.029-0.038)	
RMSE (95% PI)	0.182	0.165	0.205	0.183	-
	(0.169-0.196)	(0.153-0.178)	(0.191-0.220)	(0.169-0.195)	
ICC	0.784	0.806	0.756	0.775	-
(95% PI)	(0.751-0.813)	(0.775-0.832)	(0.719-0.788)	(0.741-0.805)	
EQ-5D-5L index					
Mean	0.642	0.599	0.608	0.587	0.598
SD	0.265	0.240	0.299	0.257	0.288
Maximum	1.000	0.852	1.000	0.975	1.000
Minimum	−0.186	−0.079	−0.339	−0.273	−0.398

BN indicates Bayesian network; ICC, intraclass correlation coefficient; SD, standard deviation; MAE, mean absolute error; ML, multinomial logistic regression; MSE, mean squared error; PI, probability interval; RMSE, root mean squared error.
*Predictions with the most-likely probability method.
†Predictions with expected-utility method.

Table 3. Predictive accuracy according to intervals of the observed EQ-5D-5L utility index and WOMAC total score in the validation sample (follow-up; expected-utility method).

EQ-5D-5L utility index	n	BN		ML	
		MAE	RMSE	MAE	RMSE
EQ-5D-5L ≤ −0.3	2	0.404	0.421	0.300	0.332
−0.3 < EQ-5D-5L ≤ −0.2	11	0.360	0.386	0.328	0.377
−0.2 < EQ-5D-5L ≤ −0.1	10	0.250	0.286	0.230	0.271
−0.1 < EQ-5D-5L ≤0	16	0.308	0.347	0.337	0.383
0 < EQ-5D-5L ≤0.1	15	0.201	0.254	0.237	0.258
0.1 < EQ-5D-5L ≤0.2	21	0.180	0.224	0.209	0.239
0.2 < EQ-5D-5L ≤0.3	27	0.102	0.128	0.133	0.169
0.3 < EQ-5D-5L ≤0.4	20	0.153	0.183	0.134	0.170
0.4 < EQ-5D-5L ≤0.5	49	0.155	0.183	0.149	0.190
0.5 < EQ-5D-5L ≤0.6	57	0.135	0.167	0.153	0.192
0.6 < EQ-5D-5L ≤0.7	100	0.103	0.143	0.112	0.159
0.7 < EQ-5D-5L ≤0.8	174	0.093	0.126	0.117	0.162
0.8 < EQ-5D-5L ≤0.9	63	0.078	0.106	0.086	0.122
EQ-5D-5L >0.9	68	0.078	0.103	0.082	0.118
WOMAC total score					
0 ≤ WOMAC ≤10	57	0.064	0.082	0.076	0.113
10 < WOMAC ≤20	65	0.077	0.100	0.130	0.169
20 < WOMAC ≤30	74	0.083	0.105	0.107	0.143
30 < WOMAC ≤40	112	0.090	0.127	0.097	0.141
40 < WOMAC ≤50	94	0.119	0.161	0.149	0.202
50 < WOMAC ≤60	68	0.164	0.197	0.180	0.232
60 < WOMAC ≤70	67	0.194	0.240	0.247	0.294
70 < WOMAC ≤80	42	0.203	0.253	0.256	0.313
80 < WOMAC ≤90	39	0.204	0.242	0.234	0.293
90 < WOMAC ≤100	14	0.153	0.175	0.139	0.185

MAE indicates mean absolute Error; RMSE, root mean squared error; WOMAC, Western Ontario McMaster Universities Osteoarthritis.

detailed information can be found in this Appendix. The ML model variance-covariance matrix is available at the following link: https://www.dropbox.com/s/q6ztb0fo9h274e0/var-covar-matrix_ML.xlsx?dl=0.

Discussion

In this study, response mapping methods have been developed to estimate the EQ-5D-5L scores through the different WOMAC items in a sample of patients with hip or knee OA, following the recommendations of "the MAPS Statement for mapping studies."[38-40] Also, according to ISPOR good practices,[41] the model coefficients have been reported to 3 decimal places to permit accurate estimation and the variance-covariance matrix was included in this study to allow a probabilistic sensitivity analysis. The results suggest that BN model offers better prediction accuracy than applying the ML model.

It is usual in mapping studies for the EQ-5D that items show a floor effect,[42] although the index has a ceiling effect. The use of the 5-level EQ-5D avoided the floor effect problems of the items in our study, and only 2.7% patients presented a value of 1 in the EQ-5D-5L index in the baseline sample. Unlike another mapping study of the EQ-5D-3L through the WOMAC, using a sample of patients with hip and knee OA in Spain,[21] no bimodal distribution was observed in our study, possibly because we used the EQ-5D-5L.

In our study, the prediction errors across EQ-5D-5L health states revealed that the ML model overestimated the values of severe health states like other studies which estimated the EQ-5D-3L through the WOMAC.[18,22] This systematic bias is probably due to the use of the 5-level EQ-5D and to the fact that most of the previous works use linear regression to calculate the index,[43] a method that is considered inappropriate for mapping studies.[44] Compared with other traditional response mapping methods, such as multinomial regression, Bayesian methods can solve these overestimation problems,[45,46] provide a robust method and have better prediction accuracy.[30,31] However, our study also indicates that the BN model does not predict the low values of the index well.

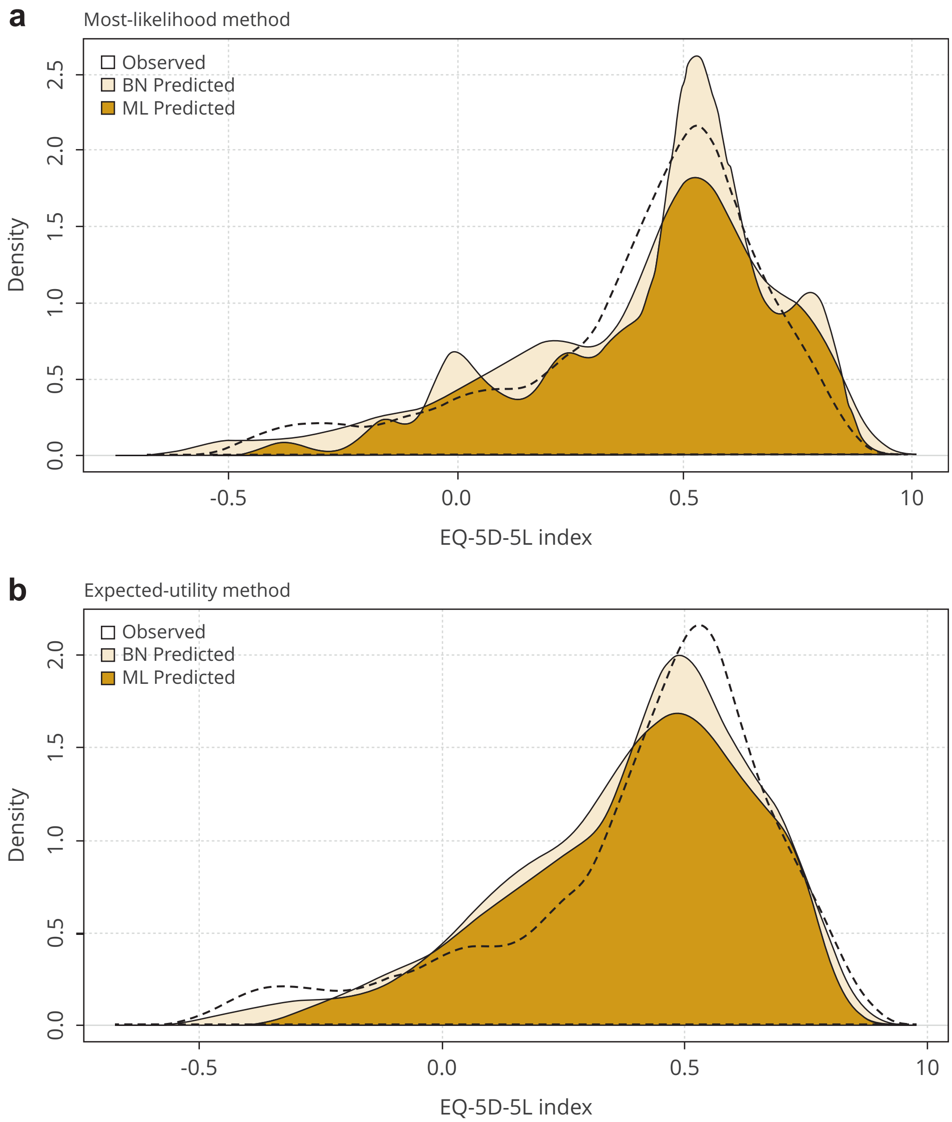
In terms of prediction accuracy, the MSE has not been calculated in any of the EQ-5D mapping studies using the WOMAC as a predictor. However, its use is recommended as a measure of precision in mapping studies.[39] The MSE values obtained are similar to those reported previously.[16] The RMSE values of 3 of the studies between the EQ-5D and the WOMAC evaluated in an external sample were higher than those found in this study, while the MAEs were lower.[22] Also, the predictive accuracy measured through MAE, RMSE and ICC in our study with response mapping was higher than the one obtained with the direct method mapping using EQ-5D-5L.[24]

No WOMAC stiffness dimension item forms part of the final model, while certain pain and physical function dimension items are strongly related to the EQ-5D-5L dimensions. This result shows that the WOMAC stiffness dimension is not included in the EQ-5D-5L questionnaire and is consistent with other mapping studies between the 3-level EQ-5D and the WOMAC.[18,19,21] In the relationship graphic provided by the BN model, it was observed that the 2 WOMAC pain component items (pain when going up or down stairs, and pain sitting or lying down) were found in BN model of the EQ-5D-5L anxiety/depression dimension, which confirms the evidence of the relationship between pain and depression.[47,48]

In addition, 2 WOMAC physical function component items (lying in bed and going shopping) that were not associated with any other dimension of the EQ-5D-5L, were found in BN models of the EQ-5D-5L anxiety/depression dimension. This suggests that these WOMAC items can be used to estimate the anxiety/depression dimension scores, unlike what happens in other studies that also use BN models.[30] Such studies indicate the need to include items that measure emotional problems, such as anxiety or depression, to improve the accuracy of predictions in mapping models.[30] On the other hand, as expected, the 2 WOMAC domestic duties items (doing heavy and light duties) were only related to the daily activities dimension of the EQ-5D but not to the rest of the dimensions.

The distribution of the BN predictions showed an overestimation of higher values of the observed EQ-5D-EL index. In addition, the BN model with most-likely method overestimated the mean value of the EQ-5D-5L index and showed a 0.044 absolute difference between the predictions and the observations, while the difference in the ML model was smaller (0.010). However, the absolute differences between predicted and observed values decreased in BN model when the expected value method was used (0.001). Also, the BN model prediction presented a greater index range than that obtained with the ML model. This also occurs in other response mapping studies that compare both models, where the difference between the observed value and the

Figure 2. Distribution of the observed and predicted EQ-5D-5L (index in validation sample).



BN indicates Bayesian network; ML, multinomial logistic regression.

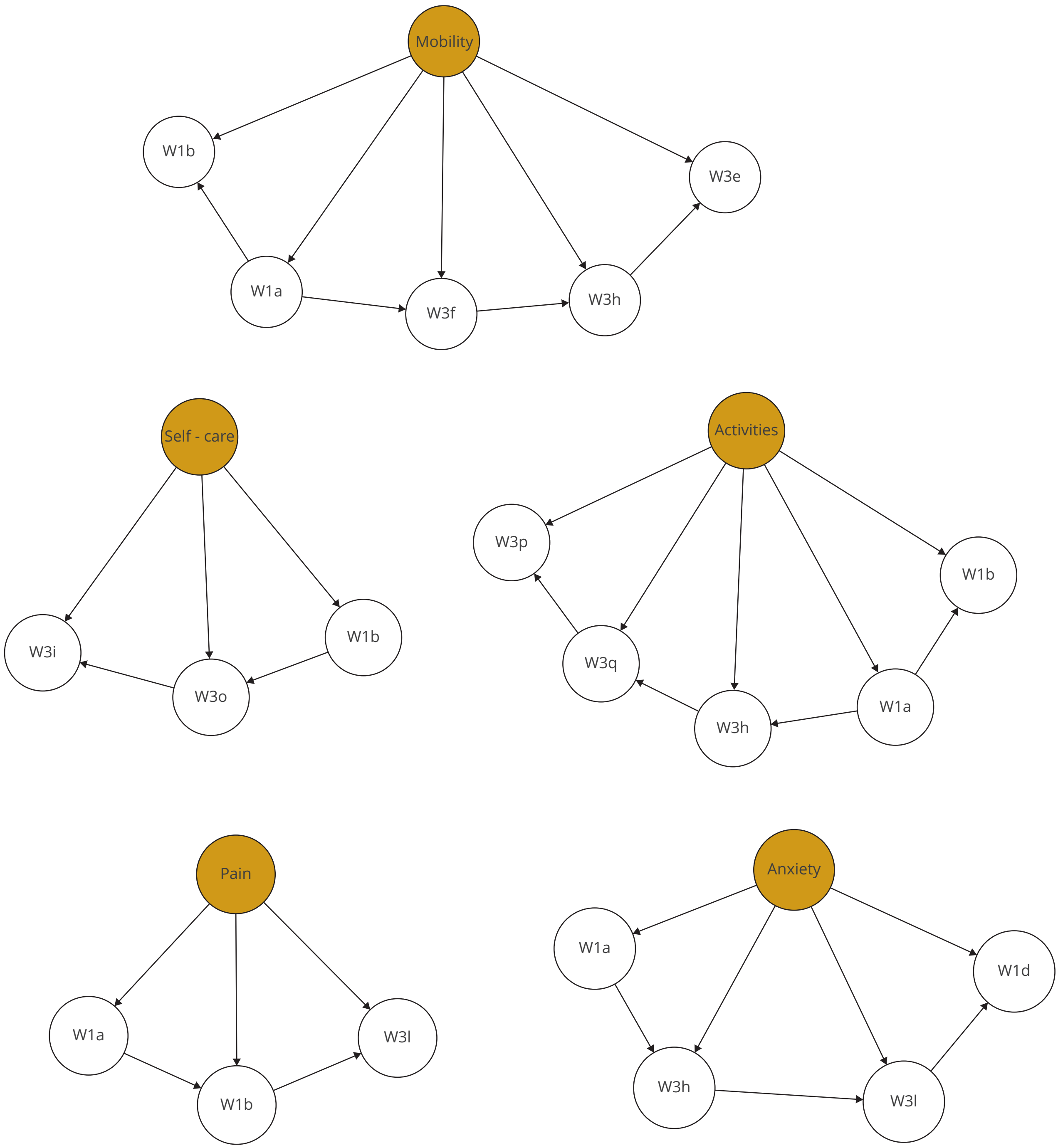
prediction is lower in the ML model than in the BN model.[30] Finally, our results also show that the absolute difference between the mean of the observations and the mean of the predicted values reflects the bias in the most likely method. This indicates that it is more appropriate to use the expected value to response mapping studies using the EQ-5D-5L.

Certain limitations of this study must be considered. First, the models were only built with the complete data in the 2 questionnaires, and 5.5% of missing data were eliminated from the total sample. Data imputation techniques used in other mapping studies, such as the multiple imputation algorithm with Expectation-Maximization,[49] could improve the obtained EQ-5D-5L estimates. Second, our results provide low predictive accuracy in extreme levels of the utility index and the models present convergence issues, possibly due to the few cases in this range. The sample of people with severe problems should be increased in future studies. Third, the sample used to develop the models is not representative of the Spanish population with hip or knee OA. However, this sample was collected from the different geographic regions and the patients present a wide range of disease severity. Finally, the predictions have been calculated using the same patients in the follow-up. This a common practice in mapping studies and can reduce estimation errors. However, the training and validation sample are formed by the same patients and therefore a lack of external validity may be observed, and the variances may be underestimated. External validation with different patients is recommended for future research.[41]

Conclusion

Bayesian networks are appropriate for estimating EQ-5D-5L scores through WOMAC as they show nonlinear dependency relationships between the variables and provide robust estimates and better prediction accuracy than ML. This algorithm

Figure 3. Bayesian network for each of the EQ-5D-5L dimensions with the Western Ontario McMaster Universities Osteoarthritis items as predictors.



might be used in all health states except in OA patients with severe health problems (ie, with negative values in the EQ-5D-5L index or WOMAC values of ≥70). The visual representation obtained through Bayesian networks allows researchers to interpret the dependency relationships between the preference-based measure and the specific measure of HRQOL in hip or knee OA, which may be useful in future research. These results provide 2 response mapping algorithms which will allow researchers to predict EQ-5D-5L values from WOMAC scores, to use the utility index in economic evaluations.

Author Affiliations: University Carlos III of Madrid, Madrid, Spain (Ayala); National Epidemiology Centre, Institute of Health Carlos III, Madrid, Spain (Forjaz); Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC), Santa Cruz de Tenerife, Tenerife, Spain (Ramallo-Fariña, García-Pérez); Oeste Multiprofessional Teaching Unit of Primary and Community Care, Primary Healthcare Management, Madrid Health Service, Madrid, Spain (Martín-Fernández); Osakidetza Basque Health Service, Basurto University Hospital, Research Unit, Bilbao, Spain (Bilbao); Health Service Research Network on Chronic Diseases (REDISSEC) (Ayala, Forjaz, Ramallo-Fariña, Martín-Fernández, García-Pérez, Bilbao); Kronikgune Institute for Health Services Research, Barakaldo, Spain (Bilbao).

Correspondence: Alba Ayala, MSc, University Carlos III of Madrid, School of Law and Social Sciences. Profesor Asociado, Department of Statistics, C/ Madrid, 126, 28903 Getafe (Madrid) Spain. Email: aayala@est-econ.uc3m.es

Author Contributions: *Concept and design*: Ayala, Forjaz, Ramallo-Fariña, Martín-Fernández, García-Pérez, Bilbao
Acquisition of data: Martín-Fernández, García-Pérez
Analysis and interpretation of data: Ayala, Ramallo-Fariña, García-Pérez, Bilbao
Drafting of the manuscript: Ayala, Ramallo-Fariña, Bilbao
Critical revision of paper for important intellectual content: Forjaz, Ramallo-Fariña, Martín-Fernández, García-Pérez, Bilbao
Statistical analysis: Ayala
Provision of study materials or patients: Martín-Fernández, García-Pérez
Obtaining funding: Martín-Fernández, García-Pérez, Bilbao
Administrative, technical, or logistic support: García-Pérez
Supervision: Forjaz, Bilbao

Conflict of Interest Disclosures: The authors reported no conflicts of interest.

Funding/Support: This study was supported by grants from the Carlos III Health Institute (Refs: PI13/00560, PI13/00518 and PI13/ 00648) and the European Regional Development Fund.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. *Osteoarthritis Cartilage*. 2019;27(11):1578–1589.
2. Hiligsmann M, Reginster J. The economic weight of osteoarthritis in Europe. *Medicographia*. 2013;35:197–202.
3. Hiligsmann M, Cooper C, Arden N, et al. Health economics in the field of osteoarthritis: an expert's consensus paper from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). *Semin Arthritis Rheum*. 2013;43(3):303–313.
4. Sacristán JA, Ortun V, Rovira J, Prieto L, García-Alonso F. Economic assessment in medicine. *Med Clin (Barc)*. 2004;122(10):379–382.
5. EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy Amst Neth*. 1990;16(3):199–208.
6. Räsänen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynänen O-P, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: a systematic literature review. *Int J Technol Assess Health Care*. 2006;22(2):235–241.
7. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013*. National Institute for Health and Care Excellence (NICE); 2013. . Accessed May 23, 2017.
8. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
9. Oemar M, Janssen B. EQ-5D-5L user guide-basic information on how to use the EQ-5D-5L instrument. Rotterdam EuroQol Group. Published online 2013. https://www.unmc.edu/centric/_documents/EQ-5D-5L.pdf. Version 2.0. Accessed October 2013.
10. Xie F, Li S-C, Luo N, et al. Comparison of the EuroQol and short form 6D in Singapore multiethnic Asian knee osteoarthritis patients scheduled for total knee replacement. *Arthritis Rheum*. 2007;57(6):1043–1049.
11. Conner-Spady BL, Marshall DA, Bohm E, et al. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Qual Life Res*. 2015;24(7):1775–1784.
12. Bilbao A, García-Pérez L, Arenaza JC, et al. Psychometric properties of the EQ-5D-5L in patients with hip or knee osteoarthritis: reliability, validity and responsiveness. *Qual Life Res*. 2018;27(11):2897–2908.
13. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15(12):1833–1840.
14. Escobar A, Quintana JM, Bilbao A, Azkárate J, Güenaga JI. Validation of the Spanish version of the WOMAC questionnaire for patients with hip or knee osteoarthritis. Western Ontario and McMaster Universities Osteoarthritis Index. *Clin Rheumatol*. 2002;21(6):466–471.
15. Bellamy N. WOMAC: a 20-year experiential review of a patient-centered self-reported health status questionnaire. *J Rheumatol*. 2002;29(12):2473–2476.
16. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health Qual Life Outcomes*. 2013;11:151.
17. 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*. 2018;16(1):31.
18. Grootendorst P, Marshall D, Pericak D, Bellamy N, Feeny D, Torrance GW. A model to estimate health utilities index mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the knee. *J Rheumatol*. 2007;34(3):534–542.
19. Barton GR, Sach TH, Jenkinson C, Avery AJ, Doherty M, Muir KR. Do estimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? *Health Qual Life Outcomes*. 2008;6:51.
20. Xie F, Pullenayegum EM, Li S-C, Hopkins R, Thumboo J, Lo N-N. Use of a disease-specific instrument in economic evaluations: mapping WOMAC onto the EQ-5D utility index. *Value Health J*. 2010;13(8):873–878.
21. Wailoo A, Hernandez Alava M, Escobar Martinez A. Modelling the relationship between the WOMAC Osteoarthritis Index and EQ-5D. *Health Qual Life Outcomes*. 2014;12:37.
22. Kiadaliri AA, Englund M. Assessing the external validity of algorithms to estimate EQ-5D-3L from the WOMAC. *Health Qual Life Outcomes*. 2016;14(1):141.
23. Price AJ, Smith J, Dakin H, et al. The Arthroplasty Candidacy Help Engine tool to select candidates for hip and knee replacement surgery: development and economic modelling. *Health Technol Assess*. 2019;23(32).
24. Bilbao A, Martín-Fernández J, García-Pérez L, et al. Mapping WOMAC onto the EQ-5D-5L utility index in patients with hip or knee osteoarthritis. *Value Health*. 2020;23(3):379–387.
25. Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum*. 1986;29(8):1039–1049.
26. Altman RD. Criteria for classification of clinical osteoarthritis. *J Rheumatol Suppl*. 1991;27:10–12.
27. Ramos-Goñi JM, Craig BM, Oppe M, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value Health*. 2018;21(5):596–604.
28. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med*. 2012;24(3):69–71.
29. Pearl J. *Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning*. University of California (Los Angeles). Computer Science Department; 1985.
30. Le QA, Doctor JN. Probabilistic mapping of descriptive health status responses onto health state utilities using Bayesian networks: an empirical analysis converting SF-12 into EQ-5D utility index in a national US sample. *Med Care*. 2011;49(5):451–460.
31. Le QA. Probabilistic mapping of the health status measure SF-12 onto the health utility measure EQ-5D using the US-population-based scoring models. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2014;23(2):459–466.
32. Margaritis D. Learning Bayesian network model structure from data. http://www.cs.cmu.edu/afs/cs/Web/People/dmarg/Papers/PhD-Thesis-Margaritis.pdf. Accessed May 24, 2017.
33. Borgelt C, Steinbrecher M, Kruse RR. *Graphical Models: Representations for Learning, Reasoning and Data Mining*. Pittsburgh, PA: John Wiley & Sons; 2009.
34. Friedman N, Goldszmidt M. Building classifiers using Bayesian networks. In: *Proceedings of the National Conference on Artificial Intelligence*. 1996:1277–1284.
35. Ramos-Goni JM, Rivero-Arias O, Dakin H, others. Response mapping to translate health outcomes into the generic health-related quality-of-life instrument EQ-5D: introducing the mrs2eq and oks2eq commands. *Stata J*. 2011;13:474–491.
36. Scutari M. bnlearn: Bayesian network structure learning, parameter learning and inference. *R Package Version*. 2012;3. . Accessed May 24, 2017.
37. Nagarajan R, Scutari M, Lèbre S. *Bayesian networks in R*. Springer; 2013:125–127.
38. Petrou S, Rivero-Arias O, Dakin H, et al. Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. *PharmacoEconomics*. 2015;33(10):985–991.
39. Petrou S, Rivero-Arias O, Dakin H, et al. The MAPS reporting statement for studies mapping onto generic preference-based outcome measures: explanation and elaboration. *PharmacoEconomics*. 2015;33(10):993–1011.

40. Petrou S, Rivero-Arias O, Dakin H, et al. Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. *Qual Life Res*. 2016;25(2):275–281.
41. Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to estimate health-state utility from non-preference-based outcome measures: an ISPOR good practices for outcomes research task force report. *Value Health*. 2017;20(1):18–27.
42. Longworth L, Rowen D. *NICE DSU Technical Support Document 10: The Use of Mapping Methods to Estimate Health State Utility Values*. London: National Institute for Health and Care Excellence (NICE); 2011.
43. Mukuria C, Rowen D, Harnan S, et al. An updated systematic review of studies mapping (or cross-walking) measures of health-related quality of life to generic preference-based measures to generate utility values. *Appl Health Econ Health Policy*. 2019;17(3):295–313.
44. Hernández Alava M, Wailoo A, Wolfe F, Michaud K. A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes. *Med Decis Mak*. 2014;34(7):919–930.
45. Kharroubi SA, Edlin R, Meads D, Browne C, Brown J, McCabe C. Use of Bayesian Markov chain Monte Carlo methods to estimate EQ-5D utility scores from EORTC QLQ data in myeloma for use in cost-effectiveness analysis. *Med Decis Mak*. 2015;35(3):351–360.
46. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) nonpreference-based measures of health to generic preference-based measures. *Eur J Health Econ*. 2010;11(2):215–225.
47. Costa J, Pinto-Gouveia J, Marôco J. Chronic pain experience on depression and physical disability: the importance of acceptance and mindfulness-based processes in a sample with rheumatoid arthritis. *J Health Psychol*. 2019;24(2):153–165.
48. Kleiber B, Jain S, Trivedi MH. Depression and pain. *Psychiatry Edgmont*. 2005;2(5):12–18.
49. Young MK, Ng S-K, Mellick G, Scuffham PA. Mapping of the PDQ-39 to EQ-5D scores in patients with Parkinson's disease. *Qual Life Res*. 2013;22(5):1065–1072.