



Research paper

Properties of the EQ-5D-5L when prospective longitudinal data from 28,902 total hip arthroplasty procedures are applied to different European EQ-5D-5L value sets

Anders Joelson^{a,b,*}, Peter Wildeman^{a,b}, Freyr Gauti Sigmundsson^{a,b}, Ola Rolfson^{c,d,e}, Jan Karlsson^f

^a Orebro University School of Medical Sciences, Orebro, Sweden

^b Department of Orthopedics, Orebro University Hospital, Orebro, Sweden

^c Department of Orthopedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

^d Department of Orthopedics, Sahlgrenska University Hospital, Gothenburg, Sweden

^e Swedish Hip Arthroplasty Register, Gothenburg, Sweden

^f University Health Care Research Center, Faculty of Medicine and Health, Orebro University, Orebro, Sweden

ARTICLE INFO

Article History:

Received 2 May 2021

Revised 4 June 2021

Accepted 7 June 2021

Available online 14 July 2021

ABSTRACT

Background: The purpose of this study was to evaluate the impact of using different country-specific value sets in EQ-5D-5L based outcome analyses.

Methods: We obtained data on patients surgically treated with total hip arthroplasty (THA) between 2017 and 2019 from the national Swedish Hip Arthroplasty Register. Preoperative and one-year postoperative data on a total of 28,902 procedures were available for analysis. The EQ-5D-5L health states were coded to the EQ-5D-5L preference indices using 13 European value sets. The EQ-5D-5L index distributions were then estimated with kernel density estimation. The change in EQ-5D-5L index before and one year after treatment was evaluated with the standardized response mean (SRM). The lifetime gain in quality-adjusted life years (QALYs) was estimated with a 3.5% annual QALY discount rate.

Findings: There was a marked variability in means and shapes of the resulting EQ-5D-5L index distributions. There were also considerable differences in the EQ-5D-5L index distribution shape before and after the treatment using the same value set. The effect sizes of one-year change (SRM) were similar for all value sets. However, the differences in estimated QALY gains were substantial.

Interpretation: The EQ-5D-5L index distributions varied considerably when a single large data set was applied to different European EQ-5D-5L value sets. The most pronounced differences were between the value sets based on experience-based valuation and the value sets based on hypothetical valuation. This illustrates that experience-based and hypothetical value sets are inherently different and also that QALY gains derived with different value sets are not comparable. Our findings are of importance in study planning since the results and conclusions of a study depend on the choice of value set.

Funding: None.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

The EuroQol 5-dimensional instrument (EQ-5D) [1] is a multilevel preference-based measure for assessment of general health. The instrument is primarily used for calculation of quality-adjusted life years (QALYs) in economic evaluation of health interventions. The

original version of the instrument (EQ-5D-3L) has 3 response levels for the 5 dimensions. In 2011, the 5-level version of the instrument (EQ-5D-5L) was introduced with the intention to reduce ceiling effects and to improve the instrument's ability to measure small changes in health [2]. Both versions use specific national value sets to adjust for national differences in experience of health-related quality of life (HRQoL).

There are several published and ongoing valuation studies for different EQ-5D-5L value sets [3]. For studies conducted in

* Corresponding author at: Department of Orthopedics, Orebro University Hospital, Orebro, Sweden.

E-mail address: anders@joelson.se (A. Joelson).

Research in context

Evidence before this study

Since its introduction in 1990, the EQ-5D instrument has been used worldwide in clinical trials, population studies and in clinical settings for assessment of health-related quality of life and QALY estimation. The most recent version, EQ-5D-5L, was launched in 2011 and the translation process is ongoing for many languages. We searched PubMed on March 30, 2021, for prospective studies with pre- and post-treatment data attributing national differences in EQ-5D-5L index distributions. The search terms used were: (EQ-5D-5L) AND (comparison OR difference* OR distribution OR density OR cross-cultural) with no limits applied to publication dates. Our search yielded 571 studies. We could not identify any prospective studies with pre- and post-treatment data on national differences in EQ-5D-5L index distributions. Cross-sectional data, however, have suggested that EQ-5D-5L data based on different national value sets are not comparable. The current study aimed to address the research gap from a prospective perspective.

Added value of this study

To our knowledge, this is the first study to apply prospectively collected pre- and post-treatment data to several national EQ-5D-5L value sets. We found a marked variability in EQ-5D-5L index distributions between the different national value sets. The most pronounced differences were between the value sets based on experience-based valuation and the value sets based on hypothetical valuation.

Implications of all the available evidence

The available evidence suggests that EQ-5D-5L data from studies conducted in different countries using different national value sets are not comparable. This observation is of particular importance in study planning since the results and conclusions of a study depend on the choice of value set. Our work should heighten awareness among researchers, health-care professionals and policy makers about the limited comparability of EQ-5D-5L data derived from different value sets.

2.2. EQ-5D-5L

EQ-5D-5L is a multilevel preference-based measure for assessment of HRQoL [2]. The dimensions are: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each item has 5 response options, coded on an ordinal scale from 1 to 5 (1=no problems, 2=slight problems, 3=moderate problems, 4=severe problems, and 5=extreme problems). The answers are assembled to a 5-digit health state reflecting the score on each dimension (in total $5^5=3,125$ states, 11111 being the best and 55555 the worst).

Each health state can be coded to a summary index using a value set. The summary index range from less than 0 (where 0 is the value of a health state equivalent to death while negative values represent values that are worse than death) to 1 (the value of full health) [3]. There are several national value sets for coding health states to summary indices. The value sets are derived using different standardized valuation protocols including valuation protocols such as of time trade-off (TTO), visual analogue scale (VAS) and a discrete choice experiment (DCE) [8]. We wanted to explore differences in EQ-5D-5L index distributions for neighboring countries. At the time of our study (Spring 2021), the valuations of 13 European value sets were completed. We used these 13 value sets for our investigation: Denmark [9], England [10], France [11], Germany (2 value sets) [12,13], Hungary [14], Ireland [15], the Netherlands [16], Poland [17], Portugal [18], Spain [19,20], Sweden Time Trade-off (TTO) and Visual Analogue Scale (VAS) [21]. One of the German value sets and the Swedish value sets are experience-based (respondents value their own current health state) whereas the other value sets are derived using hypothetical valuation (members of the general public are asked to value hypothetical health states). For all value sets, the index is calculated as decrements from full health based on the health state. The Swedish TTO model has an additional term (N5) adding a decrement if any dimension score is on level 5 (extreme problems). The Swedish VAS model has 3 additional terms (N2, N3, and N4) adding decrements if any dimension score is on level 2 to 5. The German experience-based value set has 4 additional terms (MPL2, MPL3, MPL4 and MPL5, maximum problem level) adding a decrement based on the maximum score of all dimensions. Value set characteristics are summarized in Table 1. Supplementary Fig. S1 provides illustrations of the decrements in preference weights for each of the EQ-5D-5L dimensions for the 13 value sets.

2.3. The Swedish hip arthroplasty register (SHAR)

The SHAR was launched in 1979, the national coverage is 97-99% of all primary THA procedures in Sweden. The one-year follow-up rate is 82% [7]. The SHAR started to collect EQ-5D-5L data in 2017.

2.4. Patient selection

From SHAR, we obtained data on all patients operated with primary total hip arthroplasty due to osteoarthritis between 2017 and 2019 (45,966 THA procedures in 42,685 patients). Data on both operations were included for individuals with bilateral observations during the study period. Preoperative or one-year postoperative EQ-5D-5L data were incomplete for 17064 (37%) of the procedures which gave 28,902 procedures eligible for analysis. The baseline characteristics of the patients are shown in Table 2.

2.5. Data transformation

EQ-5D-5L data was coded to EQ-5D-5L preference indices using the 13 European value sets. The conversion from health state to summary index was made using R (R Foundation for Statistical Computing, Vienna, Austria, 2017) using the models given in the references for the 13 value sets. The EQ-5D-5L index distributions (preop, one-

countries or regions without a country-specific value set the authors may use the Western Preference Pattern [4], or a value set from a neighboring or culturally similar county. Alternatively, EQ-5D-5L health states can be converted to EQ-5D-3L health states by using crosswalk methods [5,6]. In the current study, we posed three research questions with regard to value set selection: (1) What are the characteristics of the distributions of EQ-5D-5L index scores calculated with different European value sets in a total hip arthroplasty (THA) population? (2) What value set features explain the patterns of distribution? (3) How does the choice of value set influence effect size and QALY calculations? For our evaluation, we used data from the national Swedish Hip Arthroplasty Register (SHAR) [7].

2. Patients and methods

2.1. Study design

The present study was a register study with prospectively collected longitudinal THA data from the SHAR.

Table 1
Overview of value sets.

Country	Reference	N	Type	Valuation method	Regression model	Range	Other characteristics
Denmark	Elgaard Jensen et al. (2021) (hybrid model)	1,014	Hypothetical	TTO + DCE	hybrid	1.76 (-0.76 to 1)	
England	Devlin et al. (2017)	912	Hypothetical	TTO + DCE	hybrid	1.28 (-0.28 to 1)	
France	Andrade et al.(2020) (model 4)	1,048	Hypothetical	TTO + DCE	hybrid	1.52 (-0.52 to 1)	
Germany	Ludwig et al. (2018) (model 3b)	1,158	Hypothetical	TTO + DCE	hybrid	1.66 (-0.66 to 1)	
Hungary	Rencz et al. (2020) (model 5)	1,000	Hypothetical	TTO + DCE	tobit	1.85 (-0.85 to 1)	
Ireland	Hobbins et al. (2018)	1,160	Hypothetical	TTO + DCE	hybrid	1.97 (-0.97 to 1)	
Netherlands	Versteegh et al. (2016) (model 3)	979	Hypothetical	TTO + DCE	tobit	1.40 (-0.45 to 0.95)	
Poland	Golicki et al. (2019) (final model)	1,252	Hypothetical	TTO + DCE	MCMC	1.59 (-0.59 to 1)	
Portugal	Ferreira et al. (2019) (hybrid model)	1,451	Hypothetical	TTO + DCE	hybrid	1.60 (-0.60 to 1)	
Spain	Ramos-Goni et al. (2017, 2018) (model 3)	973	Hypothetical	TTO + DCE	hybrid	1.42 (-0.42 to 1)	
Germany	Leidl et al. (2017) (model 3)	8,114	Experience-based	VAS	ML	0.82 (0.10 to 0.92)	MPL2-5 terms
Sweden	Burström et al. (2020) (model 5 VAS)	23,899	Experience-based	VAS	OLS	0.87 (0.02 to 0.89)	N2-4 terms
Sweden	Burström et al. (2020) (model 5 TTO)	13,381	Experience-based	TTO	OLS	0.74 (0.24 to 0.98)	N5 term

TTO=Time Trade-off, DCE=Discrete Choice Experiment, VAS=Visual Analogue Scale, ML=Maximum Likelihood, MCMC=Markov Chain Monte Carlo simulation, OLS=Ordinary Least Squares, MPL=Maximum Problem Level

Table 2
Characteristics of the study population.

Parameter	Value
n	28,902
Age, Mean (SE)	69 (0.057)
Life expectancy, Mean (SE)	16 (0.056)
BMI, Mean (SE)	27 (0.025)
Women %	58

year postop, and difference between one-year postop and preop) were then estimated with kernel density estimation.

2.6. Quality-adjusted life years

To determine the gain in QALYs after THA surgery, we used preoperative baseline data and one-year follow-up data. We assumed that the degradation in health was equal for both baseline and postoperative data. Consequently, we applied a 3.5% annual discount rate for both baseline and postoperative data [22]. We then calculated the accumulated QALY gain for the remaining life expectancy. The life expectancy for men and women was determined using publicly available data from Statistics Sweden [23]. The number of remaining years at age 65 in Sweden 2018 was 21 years for women and 19 years for men. The remaining life expectancy was calculated as the remaining years to age 86 and 84 years respectively. We did not adjust the remaining life expectancy calculation for differences in expected life expectancy based on age for every individual; however, we used the number of remaining years at age 65 for all patients.

2.7. Statistics

Data are presented as mean and standard error of the mean (SE) and/or median and interquartile range (IQR). Standardized response mean (SRM) for paired data, i.e. the difference in means divided by the standard deviation of the difference, was used to evaluate responsiveness. Controversy still exists regarding the definition of responsiveness and on how responsiveness should be quantified. A comprehensive review of methods for the quantification of responsiveness is given by Husted et al. [24] who argue in favor of the standardized response mean (SRM) for the assessment of responsiveness. Also, Fayers et al. [25] recommend SRM when assessing responsiveness. The SRM was interpreted as follows: <0.2 no effect, 0.2 to 0.4 small effect, 0.5 to 0.7 moderate effect, >0.7 large effect [25]. The improvement/deterioration in EQ-5D-5L index was evaluated with percent change from baseline (%CFB) i.e. the difference in means divided by the baseline mean. We used kernel density estimation with Gaussian kernels to estimate the

EQ-5D-5L index distributions (R functions `geom_density` and `geom_stat`, R Foundation for Statistical Computing, Vienna, Austria, 2017). We used the default method for bandwidth selection based on the sample standard deviation and interquartile range [26]. Ceiling and floor effects, defined as 15% or more of the respondents achieving the highest/lowest possible score, were calculated [27].

2.8. Role of the funding source

There was no funding source for this study.

3. Results

The EQ-5D-5L response histograms distribution showed highest values for the pain/discomfort and mobility dimensions (Fig. 1). The response histograms shifted towards lower values postoperatively. The best possible health state 11111 was reported for 0.062% of the procedures preoperatively and for 28.6% of the procedures postoperatively. The worst possible health state 55555 was not reported preoperatively or postoperatively. For all health states eligible for analysis ($2 \times 28,902$, i.e. preop and postop), 2,963 of the 3,125 health states (94.8%) were used by less than 0.1% of the health scorings, and 1,914 of the 3,125 health states (61.2%) were unused. The EQ-5D-5L state distribution is illustrated in supplementary Fig. S2.

The comparison of the different value sets showed a marked variability in EQ-5D-5L index distributions preoperatively (Fig. 2). There was a tendency towards unimodal distributions for the experience-based value sets (Germany, Sweden) while the hypothetical value sets had bimodal distributions. Postoperatively, the EQ-5D-5L index distributions were skewed towards higher values. All density functions shifted towards higher values after treatment with peaks near 1.0 (full health), whereas no density function had peaks at full health before treatment. Also for the difference distributions, the variability was substantial (Fig. 2).

Table 3 summarizes EQ-5D-5L index data for the different national EQ-5D-5L value sets. The mean and median values were similar for a given national value set. There were, however, substantial differences between the national value sets. The effect sizes of change (SRM) were similar for all value sets (supplementary Fig. S3).

Estimated QALY gains based on a remaining life expectancy of 16 years are presented in Table 3. There were considerable variations in QALY gains for the different value sets. The Spearman rank correlation between QALY gain and value set range was 0.76 (supplementary Fig. S4).

4. Discussion

We found a marked variability in EQ-5D-5L index distributions when a large THA data set was applied to different European EQ-5D-

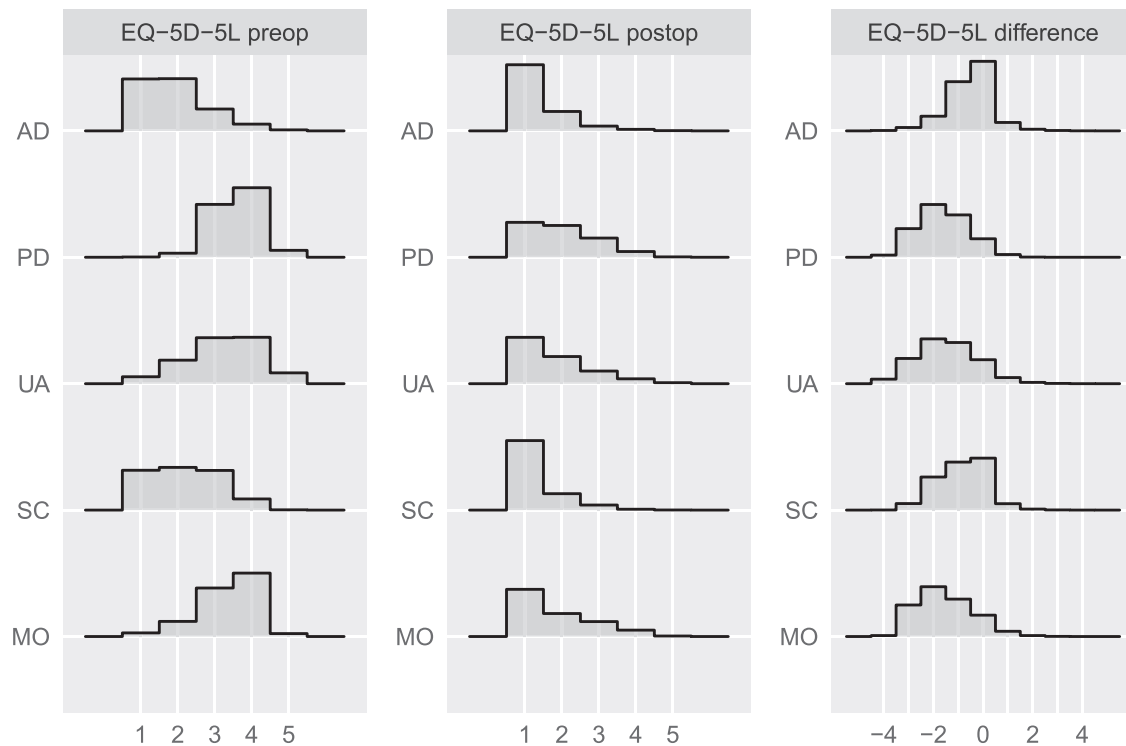


Fig. 1. Response histograms for the 5 dimensions of EQ-5D-5L before and one year after total hip arthroplasty ($n=28,902$). MO = mobility, SC = self-care, UA = usual activities, PD = pain/discomfort, AD = anxiety/depression. 1 = no problems, 2 = slight problems, 3 = moderate problems, 4 = severe problems, 5 = extreme problems.

5L value sets (Fig. 2). The most pronounced differences were between the value sets based on experience-based valuation (Germany and Sweden) and the value sets based on hypothetical valuation. Preoperatively, the index distributions based on the experience-based value sets were unimodal, while the distributions based on hypothetical value sets were in all cases bimodal. Postoperatively, the differences were less pronounced. The valuation method also affected the difference distributions (Fig. 2). The difference distributions based on hypothetical value sets were more irregular than the corresponding distributions based on experience-based value sets. Consequently, not only the preferences of the population of a country/region but also the valuation method have a considerable impact on the index distribution.

One factor that might contribute to the difference in shapes between experience-based and the hypothetical data is the smaller width of the experience-based value sets of our study (Table 1). There might be some loss of information when the EQ-5D-5L states are compressed to a smaller range of values. This hypothesis is supported by the fact that the differences in distributions are less pronounced for the postoperative data that are limited in range because of ceiling effects.

For EQ-5D-3L, previous studies have reported differences between value sets based on experienced-based valuation and hypothetical valuation [28]. Kiadaliri et al. [29] used EQ-5D-3L data from the Swedish national diabetes register to compare experience-based valuation with hypothetical valuation and concluded that the choice of valuation method might have important impact on economic evaluation and funding decisions. Our data confirms that experience-based and hypothetical valuation results in different index distributions also for EQ-5D-5L (Fig. 2). Moreover, for experience-based valuation, the Swedish TTO model produces higher means but less improvement than the Swedish and German VAS models. A possible confounder is that the Swedish TTO cohort, in contrast to the Swedish and German VAS cohorts, had an upper age limit of 69 years.

Nevertheless, we agree with Burström et al. [21] that value sets based on VAS and TTO may yield different health outcomes in practical evaluations.

There were differences in EQ-5D-5L index distributions before and after surgery using the same value set. This means that it is not only the mean/median EQ-5D-5L index that may change after a medical intervention, the entire shape of the distributions may be different after an intervention. Moreover, the difference distributions were non-normally distributed for several value sets. This finding has consequences for the statistical inference on paired data when the EQ-5D-5L index before and after a medical intervention is evaluated. Assumptions on normality and/or variance equality are violated, and parametric methods might not be applicable.

EQ-5D-5L was introduced with the intention to reduce ceiling effects and to improve the instrument's ability to measure small changes in health [2]. The skewness of the postoperative distributions (Fig. 2) is most probably explained by the marked ceiling effects in the postoperative data. Our results confirm the findings of previous studies that EQ-5D-5L is limited by ceiling effects, especially for general population data [30]. For data limited by ceiling or floor effects, subdividing the patients into two or more subgroups might provide a better understanding of the benefits of a given treatment [31].

The preoperative bimodality of the hypothetical value sets is not easily explained. Since index distributions of the experience-based value sets are all unimodal, there is a possibility that the bimodality of the hypothetical value sets is an artifact of the value set scoring system rather than a true separate grouping of the patients [32].

The effect sizes of one-year change (SRM), i.e. the change in terms of standard deviations, were large for all value sets. The SRM is often used to evaluate responsiveness to changes in psychometric evaluations of HRQoL instruments. Consequently, EQ-5D-5L is responsive to change for THA surgery irrespective of choice of value set. The SRMs of our study are similar to the SRM reported by Bilbao et al. [33] for a Spanish cohort of patients who underwent hip or knee arthroplasty

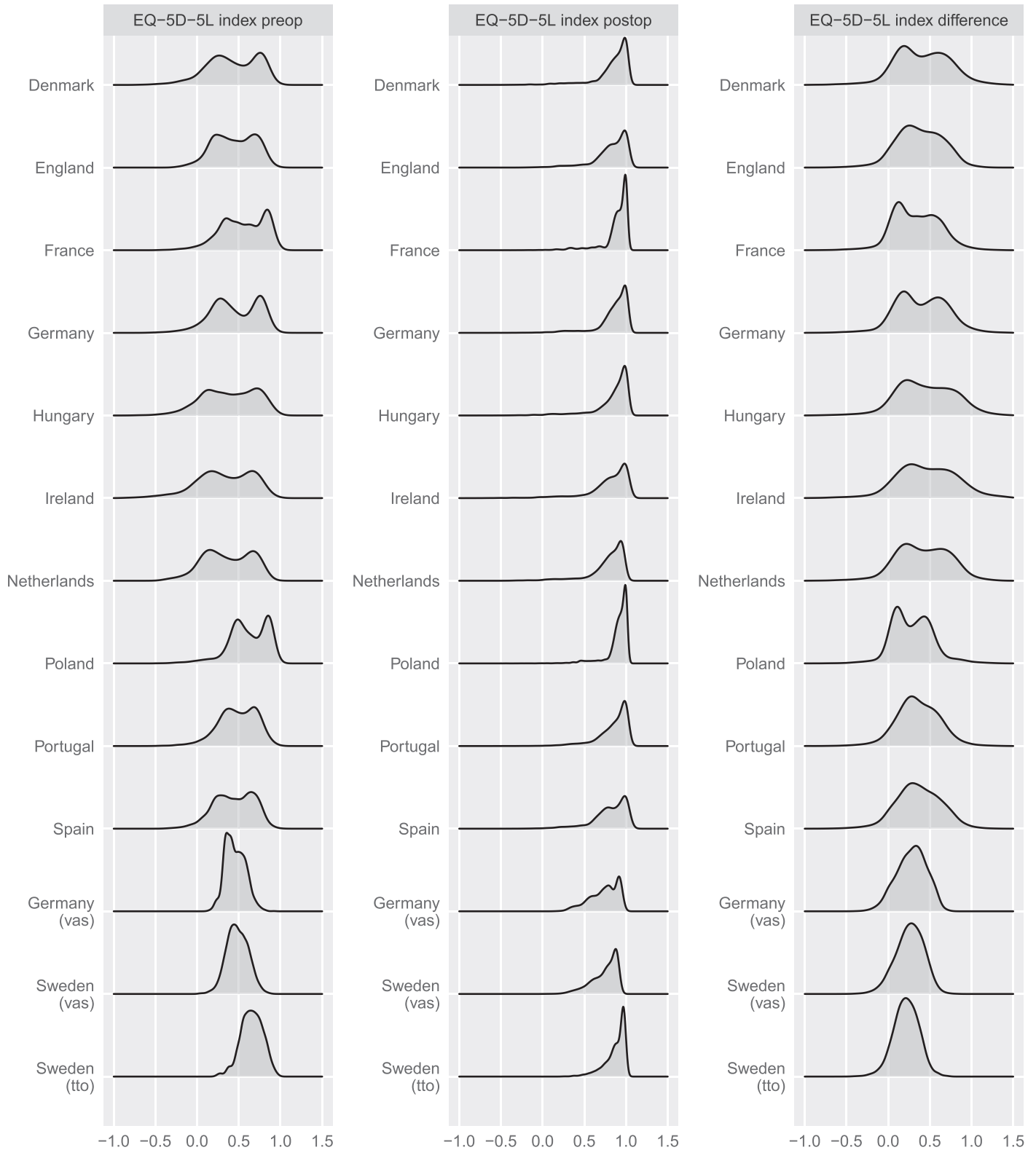


Fig. 2. Kernel estimates of the EQ-5D-5L index distributions for different European EQ-5D-5L value sets for total hip arthroplasty (n=28,902) before surgery (left), one-year after surgery (middle) and difference between one-year after and before surgery (right). The bottom 3 value sets are experience-based.

surgery. The definition of the SRM is based on differences and standard deviations of paired data. Fig. 2 suggest that the difference between postop and preop data is not necessarily normally distributed. This means that SRM based on EQ-5D-5L index has to be interpreted with caution.

The QALY gains after THA have been reported previously. Appleby et al.[22] reported a lifetime QALY gain of 2.77 for THA based on data from a large UK national database. Jenkins et al.[34] reported a lifetime QALY gain of 6.35 based on 348 THA procedures in the UK. Fawcitt et al.[35] reported a lifetime QALY gain of 7.99 based on data of

Table 3

EQ-5D-5L index data preoperatively and one year postoperatively for 28,902 THR procedures for different national value sets. The bottom 3 value sets are experience-based.

	Preop		One year postop		Difference		SRM	QALY
	Mean (SE)	Median (IQR)	Mean (SE)	Median (IQR)	Mean (SE)	%CFB	Mean (SE)	Mean (SE)
Denmark	0.43 (0.0018)	0.42 (0.20-0.70)	0.84 (0.0013)	0.91 (0.80-1)	0.41 (0.0019)	97	1.27 (0.0079)	5.12 (0.030)
England	0.45 (0.0014)	0.45 (0.25-0.66)	0.82 (0.0012)	0.88 (0.75-1)	0.38 (0.0016)	84	1.37 (0.0082)	4.68 (0.026)
France	0.55 (0.0016)	0.56 (0.35-0.81)	0.89 (0.001)	0.94 (0.87-1)	0.34 (0.0016)	62	1.22 (0.0078)	4.20 (0.025)
Germany	0.45 (0.0017)	0.42 (0.25-0.72)	0.85 (0.0012)	0.91 (0.80-1)	0.40 (0.0018)	88	1.31 (0.0080)	4.94 (0.028)
Hungary	0.39 (0.0019)	0.40 (0.14-0.67)	0.84 (0.0014)	0.92 (0.80-1)	0.45 (0.0020)	116	1.31 (0.0080)	5.58 (0.032)
Ireland	0.33 (0.0020)	0.33 (0.09-0.62)	0.79 (0.0015)	0.87 (0.73-1)	0.47 (0.0021)	142	1.29 (0.0080)	5.79 (0.034)
Netherlands	0.36 (0.0017)	0.34 (0.13-0.61)	0.78 (0.0013)	0.85 (0.72-0.95)	0.42 (0.0018)	115	1.34 (0.0081)	5.18 (0.029)
Poland	0.61 (0.0014)	0.62 (0.47-0.83)	0.90 (0.0008)	0.94 (0.88-1)	0.29 (0.0014)	48	1.19 (0.0077)	3.61 (0.022)
Portugal	0.48 (0.0013)	0.50 (0.33-0.67)	0.84 (0.0011)	0.91 (0.77-1)	0.36 (0.0015)	74	1.40 (0.0083)	4.46 (0.024)
Spain	0.44 (0.0014)	0.46 (0.24-0.62)	0.81 (0.0012)	0.84 (0.71-1)	0.37 (0.0016)	83	1.37 (0.0082)	4.55 (0.026)
Germany (VAS)	0.46 (0.0007)	0.44 (0.36-0.54)	0.74 (0.0010)	0.77 (0.62-0.92)	0.28 (0.0011)	61	1.56 (0.0088)	3.48 (0.017)
Sweden (VAS)	0.48 (0.0008)	0.48 (0.39-0.57)	0.74 (0.0009)	0.79 (0.65-0.89)	0.26 (0.0010)	53	1.52 (0.0086)	3.20 (0.016)
Sweden (TTO)	0.66 (0.0008)	0.66 (0.57-0.76)	0.87 (0.0007)	0.91 (0.82-0.98)	0.22 (0.0009)	33	1.41 (0.0083)	2.68 (0.015)

%CFB = percent change from baseline.

the UK and Swedish hip joint registers. Several factors may explain the differences in the QALY gain, e.g. remaining life expectancy, QALY annual discount rate, HRQoL instrument, and choice of value set. In our study, the QALYs for the different value sets ranged between 2.68 and 5.79. This means that we report lower QALY gains compared with the UK studies [22,34,35]. In contrast to the UK studies, our data are based on EQ-5D-5L. One explanation for limited QALY gain when using Swedish value set might be that the anxiety/depression dimension has the highest impact on the Swedish EQ-5D-5L index values (cf. Fig. S1) [21]. THA, in contrast, primarily addresses pain and mobility. The anxiety/depression dimension, however, has the highest impact also on the Danish and Irish value set, and the QALY gains of these value sets are substantial. This illustrates the complexity of multi-dimensional preference-based measurements.

The EQ-5D-5L indices based on the experience-based value sets (German and Swedish) generated the lowest QALY gains of the value sets studied. Consequently, not only cultural similarities, but also the value set valuation method, has to be taken in consideration when selecting value set. Furthermore, our correlation analysis suggests an association between QALY gain and the value set range. The smaller ranges, and consequently the smaller possible QALY gains, of the experience-based value sets are expected since experience-based value sets are based on ratings on the 0–1 scale [12] (Sweden 0-100) [21] not allowing for values less than zero. This illustrates that experience-based and hypothetical value sets are inherently different and that QALY gains derived with different value sets are not comparable. For EQ-5D-3L and for 5L to 3L crosswalk, previous studies have found that applying different national EQ-5D value sets to the same data may result in substantially different incremental QALY estimates [36,37]. The current study extends these results to EQ-5D-5L.

Our study evaluated differences and similarities between different value sets by using graphical exploration, responsiveness to THA, and QALY estimation. There are different approaches to evaluate differences between value sets. Craig et al.[38] used correlation analysis and found that the US value set was highly correlated with the value set of Canada and England. The correlation with the value set from the Netherlands was weaker, although the same valuation method was used for all value sets. The authors concluded that similarity in language (English) may be more important than valuation protocol. Henry et al.[39] used simulations to estimate the minimally important difference (MID) for eight value sets and found that mean MID ranged from 0.072 to 0.101 and that MID was correlated with value set range.

The findings of our study should be evaluated in the light of several limitations. First, the data were limited to THA patients, i.e. persons with problems mainly related to the musculoskeletal system

which limits external validity. Second, the conversion of the 3,125 EQ-5D-5L health states to the EQ-5D-5L index represents a non-linear multivariate transformation on discrete, sometime clustered, data. The analysis of such model is mathematically challenging. Instead of more complex mathematical methods, we used descriptive statistics and graphical representations to explore our data. Third, we restricted our analysis to European countries. Selecting a different set of countries would have given different EQ-5D-5L index distributions. Our main purpose, however, was to explore differences and with our data set we found considerable differences when equal response patterns were applied to the value sets. Fourth, data were incomplete for 37% of the procedures. Fifth, we used the number of remaining years at age 65 for all patients in the QALY calculations, i.e. we did not adjust the remaining life expectancy calculation for differences in expected life expectancy based on age for every individual. This means that the QALY calculation may be biased. The bias, however, is evenly applied to the scales so there should be no impact in terms of comparisons.

5. Conclusion

We found a marked variability in EQ-5D-5L index distributions when a single large THA data set was applied to different European EQ-5D-5L value sets. The most pronounced differences were between the value sets based on experience-based valuation and the value sets based on hypothetical valuation. This illustrates that experience-based and hypothetical value sets are inherently different and also that QALY gains derived with different value sets are not comparable. Our findings are of importance in study planning since the results and conclusions of a study depend on the choice of value set.

Authors' contribution

All authors designed the study. AJ did the literature review and analyzed the data. AJ and PW accessed and verified the data. All authors interpreted the data. AJ wrote the manuscript with contributions from PW, FGS, OR, and JK. All authors approved the final version of the manuscript.

Ethics approval

The study was approved by the Swedish Ethical Review Authority (registration number: 2020-06299).

Data sharing

Data are available from the SHAR after approval by the Swedish Ethical Review Authority, and according to regulations in the General Data Protection Regulation and the Swedish Patient Data Act.

Declaration of Interests

We declare no competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.lanepe.2021.100165](https://doi.org/10.1016/j.lanepe.2021.100165).

References

- [1] EuroQolGroup. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- [2] Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–36.
- [3] EuroQol Research Foundation. EQ-5D-5L user guide: basic information on how to use the EQ-5D-5L instrument, version 3.0. Rotterdam: EuroQol Research Foundation; 2019.
- [4] Olsen JA, Lamu AN, Cairns J. In search of a common currency: a comparison of seven EQ-5D-5L value sets. *Health Econ* 2018;27:9–49.
- [5] van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;15:708–15.
- [6] Ben A, Finch AP, van Dongen JM, de Wit M, van Dijk SEM, Snoek FJ, Adriaanse MC, van Tulder MW, Bosmans JE. Comparing the EQ-5D-5L crosswalks and value sets for England, the Netherlands and Spain: exploring their impact on cost-utility results. *Health Econ* 2020;29:640–51.
- [7] Kärrholm J, Rogmark C, Nauclér E, Nätman J, Vinblad J, Mohaddes M, Rolfson O. Swedish hip arthroplasty register: annual report 2019. Gothenburg: Swedish Hip Arthroplasty Register; 2019.
- [8] Stolk E, Ludwig K, Rand K, van Hout B, Ramos-Goni JM. Overview, update, and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. *Value Health* 2019;22:23–30.
- [9] Elgaard Jensen C, Sorensen SS, Gudex C, Jensen MB, Pedersen KM, Ehlers LH. The Danish EQ-5D-5L value set: a hybrid model using cTTO and DCE data. *Appl Health Econ Health Policy* 2021 [Epub ahead of print].
- [10] Devlin NJ, Brooks R. EQ-5D and the EuroQol group: past, present and future. *Appl Health Econ Health Policy* 2017;15:127–37.
- [11] Andrade LF, Ludwig K, Goni JMR, Oppe M, de Pouvourville G. A French value set for the EQ-5D-5L. *Pharmacoeconomics* 2020;38:413–25.
- [12] Leidl R, Reitmeir P. An experience-based value set for the EQ-5D-5L in Germany. *Value Health* 2017;20:1150–6.
- [13] Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. *Pharmacoeconomics* 2018;36:663–74.
- [14] Rencz F, Brodsky V, Gulacsi L, Golicki D, Ruzsa G, Pickard AS, Law EH, Pentek M. Parallel valuation of the EQ-5D-3L and EQ-5D-5L by time trade-off in Hungary. *Value Health* 2020;23:1235–45.
- [15] Hobbins A, Barry L, Kelleher D, Shah K, Devlin N, Goni JMR, O'Neill C. Utility values for health states in Ireland: a value set for the EQ-5D-5L. *Pharmacoeconomics* 2018;36:1345–53.
- [16] Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value Health* 2016;19:343–52.
- [17] Golicki D, Jakubczyk M, Graczyk K, Niewada M. Valuation of EQ-5D-5L health states in Poland: the First EQ-VT-based study in central and eastern Europe. *Pharmacoeconomics* 2019;37:1165–76.
- [18] Ferreira PL, Antunes P, Ferreira LN, Pereira LN, Ramos-Goni JM. A hybrid modeling approach for eliciting health state preferences: the Portuguese EQ-5D-5L value set. *Qual Life Res* 2019;28:3163–75.
- [19] Ramos-Goni JM, Pinto-Prades JL, Oppe M, Cabases JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Med Care* 2017;55:e51–8.
- [20] Ramos-Goni JM, Craig BM, Oppe M, Ramallo-Farina Y, Pinto-Prades JL, Luo N, Rivero-Arias O. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value Health* 2018;21:596–604.
- [21] Burström K, Teni FS, Gerdtham UG, Leidl R, Helgesson G, Rolfson O, Henriksson M. Experience-based Swedish TTO and VAS value sets for EQ-5D-5L health states. *Pharmacoeconomics* 2020;38:839–56.
- [22] Appleby J, Poteliakhoff E, Shah K, Devlin N. Using patient-reported outcome measures to estimate cost-effectiveness of hip replacements in English hospitals. *J R Soc Med* 2013;106:323–31.
- [23] Statistics Sweden. Forecast on deaths and life expectancy in 2020. Stockholm: Statistics Sweden; 2020.
- [24] Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459–68.
- [25] Fayers PM, Machin D. Quality of life: the assessment, analysis and reporting of patient-reported outcomes. 3rd ed. Chichester: John Wiley and Sons Ltd; 2016.
- [26] Rizzo ML. Statistical computing with R. 2nd ed. Boca Raton: CRC Press Taylor and Francis Group; 2019.
- [27] Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- [28] Rand-Hendriksen K, Augestad LA, Kristiansen IS, Stavem K. Comparison of hypothetical and experienced EQ-5D valuations: relative weights of the five dimensions. *Qual Life Res* 2012;21:1005–12.
- [29] Kiadaliri AA, Eliasson B, Gerdtham UG. Does the choice of EQ-5D tariff matter? A comparison of the Swedish EQ-5D-3L index score with UK, US, Germany and Denmark among type 2 diabetes patients. *Health Qual Life Outcomes* 2015;13:145.
- [30] Thompson AJ, Turner AJ. A comparison of the EQ-5D-3L and EQ-5D-5L. *Pharmacoeconomics* 2020;38:575–91.
- [31] Greene ME, Rolfson O, Garellick G, Gordon M, Nemes S. Improved statistical analysis of pre- and post-treatment patient-reported outcome measures (PROMs): the applicability of piecewise linear regression splines. *Qual Life Res* 2015;24:567–73.
- [32] Rolfson O, Kärrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish hip arthroplasty register: results of a nationwide prospective observational study. *J Bone Joint Surg Br* 2011;93:867–75.
- [33] Bilbao A, Garcia-Perez L, Arenaza JC, Garcia I, Ariza-Cardiel G, Trujillo-Martin E, Forjaz MJ, Martin-Fernandez J. Psychometric properties of the EQ-5D-5L in patients with hip or knee osteoarthritis: reliability, validity and responsiveness. *Qual Life Res* 2018;27:2897–908.
- [34] Jenkins PJ, Clement ND, Hamilton DF, Gaston P, Patton JT, Howie CR. Predicting the cost-effectiveness of total hip and knee replacement: a health economic analysis. *Bone Joint J* 2013;95:115–21.
- [35] Fawsitt CG, Thom HHZ, Hunt LP, Nemes S, Blom AW, Welton NJ, Hollingworth W, Lopez-Lopez JA, Beswick AD, Burston A, Rolfson O, Garellick G, Marques EMR. Choice of prosthetic implant combinations in total hip replacement: cost-effectiveness analysis using UK and Swedish hip joint registries data. *Value Health* 2019;22:303–12.
- [36] Karlsson JA, Nilsson JA, Neovius M, Kristensen LE, Gülfe A, Saxne T, et al. National EQ-5D tariffs and quality-adjusted life-year estimation: comparison of UK, US and Danish utilities in south Swedish rheumatoid arthritis patients. *Ann Rheum Dis* 2011;70:2163–6.
- [37] van Dongen JM, Jornada Ben A, Finch AP, Rossenaar MMM, Biesheuvel-Leliefeld KEM, Apeldoorn AT, Ostelo RWJG, van Tulder MW, van Marwijk HWJ, Bosmans JE. Assessing the impact of EQ-5D country-specific value sets on cost-utility outcomes. *Med Care* 2021;59:82–90.
- [38] Craig BM, Rand K. Choice defines QALYs: a US valuation of the EQ-5D-5L. *Med Care* 2018;56:529–36.
- [39] Henry EB, Barry LE, Hobbins AP, McClure NS, O'Neill C. Estimation of an instrument-defined minimally important difference in EQ-5D-5L index scores based on scoring algorithms derived using the EQ-VT version 2 valuation protocols. *Value Health* 2020;23:936–44.