Comparing the EQ-5D-5L crosswalks and value sets for England, the Netherlands and Spain

Jornada Ben, Angela; Finch, Aureliano Paolo; van Dongen, Johanna M; de Wit, Maartje; van Dijk, Susan E M; Snoek, Frank J; Adriaanse, Marcel C; van Tulder, Maurits W; Bosmans, Judith E

published in
Health Economics
2020

DOI (link to publisher)
10.1002/hec.4008

document version
Publisher's PDF, also known as Version of record

Received: 4 January 2019 | Revised: 14 December 2019 | Accepted: 21 January 2020

DOI: 10.1002/hec.4008

RESEARCH ARTICLE

Health Economics WILEY

Comparing the EQ-5D-5L crosswalks and value sets for England, the Netherlands and Spain: Exploring their impact on cost-utility results

Ângela Ben[1] | Aureliano Paolo Finch[1] | Johanna M. van Dongen[1] | Maartje de Wit[2] | Susan E.M. van Dijk[1] | Frank J. Snoek[2] | Marcel C. Adriaanse[1] | Maurits W. van Tulder[1] | Judith E. Bosmans[1]

[1]Health Technology Assessment Section, Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2]Department of Medical Psychology, Amsterdam University Medical Centers – VUmc, Amsterdam, The Netherlands

Correspondence
Aureliano Paolo Finch, PhD, Department of Health Sciences, Health Technology Assessment Section, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 MV Amsterdam, North Holland, The Netherlands.
Email: a.p.finch@vu.nl

Abstract

This study compares the five-level EuroQol five-dimension questionnaire (EQ-5D-5L) crosswalks and the 5L value sets for England, the Netherlands, and Spain and explores the implication of using one or the other for the results of cost–utility analyses. Data from two randomized controlled trials in depression and diabetes were used. Utility value distributions were compared, and mean differences in utility values between the EQ-5D-5L crosswalk and the 5L value set were described by country. Quality-adjusted life years (QALYs) were calculated using the area-under-the-curve method. Incremental cost-effectiveness ratios (ICERs) were calculated, and uncertainty around ICERs was estimated using bootstrapping and graphically shown in cost-effectiveness acceptability curves. For all countries investigated, utility value distributions differed between the EQ-5D-5L crosswalk and 5L value set. In both case studies, mean utility values were lower for the EQ-5D-5L crosswalk compared with the 5L value set in England and Spain, but higher in the Netherlands. However, these differences in utility values did not translate into relevant differences across utility estimation methods in incremental QALYs and the interventions' probability of cost-effectiveness. Thus, our results suggest that EQ-5D-5L crosswalks and 5L value sets can be used interchangeably in patients affected by mild or moderate conditions. Further research is needed to establish whether these findings are generalizable to economic evaluations among severely ill patients.

KEYWORDS

EuroQol, utility, EQ-5D, EQ-5D-5L, crosswalk, value set, quality of life

Ângela J. Ben, PhD, and Aureliano P. Finch, PhD, contributed equally to this work.

1 | INTRODUCTION

Generic preference-based measures of health are pre-scored utility measures that can be used to calculate quality-adjusted life years (QALYs) for economic evaluations (Finch, Brazier, Mukuria, & Bjorner, 2017). The EuroQol five-dimension questionnaire (EQ-5D) is the most frequently used generic preference-based measures of health, and it is recommended by health technology assessment agencies, such as the National Institute for Health and Care Excellence (National Institute for Health and Care Excellence, 2018) and the Dutch National Health Care Institute (Zorginstituut Nederland, 2016).

The EQ-5D comprises a standardized descriptive system through which health is described and a value set that reflects the strength of preferences of the general public for the health states is described. The original descriptive system of the EQ-5D (commonly referred to as three-level [3L] EQ-5D [EQ-5D-3L] or 3L version) comprised five health dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each of which had three levels of response, that is, no problems, some problems, and extreme problems (EuroQol Group, 1990). Although theoretically the EQ-5D-3L is applicable across all disease areas and conditions, evidence suggests that it lacks responsiveness in specific populations, such as patients with mental health issues, visual disorders, or neoplasms (Finch, Brazier, & Mukuria, 2017). One possible explanation for this could be the absence of an appropriate number of response levels to capture relevant changes in the patients' health-related quality of life. This has led to the development of a new EQ-5D version; the five-level (5L) EQ-5D (EQ-5D-5L; commonly referred to as EQ-5D-5L or 5L version). This measure describes health using the same dimensions of the EQ-5D-3L, but it uses five response levels, that is, no problems, slight problems, moderate problems, severe problems, and extreme problems (Herdman et al., 2011; Janssen et al., 2013; Janssen, Bonsel, & Luo, 2018).

The uptake of the EQ-5D-5L amongst researchers has increased significantly the last years, and numerous economic evaluations in the areas of musculoskeletal diseases, cardiovascular diseases, and mental health are now using this measure (Jensen, Riis, Pedersen, Jensen, & Petersen, 2014; Ruo et al., 2003; van Dijk et al., 2013). Given the importance of employing country-specific value sets because of sociocultural differences among populations (Xie, Gaebel, Perampaladas, Doble, & Pullenayegum, 2013), valuation studies of the EQ-5D-5L are being undertaken in many countries. Nonetheless, 5L value sets are not available yet for many countries, while the 3L values sets are. For this reason, many economic evaluations used and plan to use a country-specific interim crosswalk (van Hout et al., 2012), which estimates the EQ-5D-5L utility values from the EQ-5D-3L tariffs (Jensen et al., 2014; Krog et al., 2017).

The predictive ability of the crosswalk may be affected by the protocols used for the valuation of the EQ-5D-3L and EQ-5D-5L, that is, the measurement and valuation of health for the EQ-5D-3L and the EuroQol valuation technology for the EQ-5D-5L (Oppe, Devlin, van Hout, Krabbe, & de Charro, 2014; Oppe, Rand-Hendriksen, Shah, Ramos-Goñi, & Luo, 2016). Differences in protocols include, among the others, the use of different elicitation methods, for example, visual analogue scale or discrete choice experiments, modes of administration, for example, face to face or computer-assisted personal interviews and models specifications, for example, N3 or main effect models. Moreover, there are also differences in the design of the valuation studies for the same EQ-5D version between countries that might affect the precision of the crosswalk. For example, in the English and Spanish valuation studies, 43 health states were directly valued, and 12 health states were administered per responder, whereas Dutch valuation study directly valued only 17 health states, with each responder valuing all of them (Badia, Roset, Herdman, & Kind, 2001; Dolan, 1997; Lamers, Stalmeier, McDonnell, Krabbe, & J.van Busschbach J., 2005).

Two studies have shown that the EQ-5D-5L crosswalk and 5L value set for England differ in terms of the decrements associated with the response levels and utility range/distribution (Camacho et al., 2018; Mulhern et al., 2018). With the release of 5L value sets for the Netherlands and Spain, it became possible to investigate whether such differences also exist between the crosswalk and value set for other countries. Moreover, the impact of using one utility estimation method over the other on the statistical uncertainty surrounding cost–utility estimates is still unclear. The latter is particularly important as it provides insight into the validity of reimbursement decisions based on economic evaluations that relied on interim crosswalk QALY estimates (Brazier, Briggs, & Bryan, 2018).

This study builds on the work of Mulhern et al. (2018) and Camacho et al. (2018) by assessing whether the differences found between the EQ-5D-5L crosswalk and the 5L value set of England, also apply to the Dutch and Spanish crosswalks and value sets. Additionally, this study explores whether the different utility estimation methods have an impact on the outcomes of a cost–utility analysis, such as incremental QALYs and an intervention's probability of being cost-effective. For this purpose, data from two pragmatic randomized clinical trials in patients affected by depression

and diabetes were used. These conditions are relevant for the comparison of the crosswalks and the value sets as they are associated with significant impairments in health-related quality of life (Evans Kreider, Pereira, & Padilla, 2017; Ruo et al., 2003).

2 | METHODS

2.1 | Data

This study used data from two pragmatic cluster-randomized controlled trials performed in the Netherlands. Full details of the trials are described elsewhere (van Dijk et al., 2013; Wit, Rondags, Tulder, Snoek, & Bosmans, 2018). In brief, the first trial (referred to as depression study in this paper) assessed the cost-effectiveness of a program consisting of four sequential treatment steps to prevent major depression in comparison with usual care among patients with subthreshold depression symptoms. In the second trial (referred to as diabetes study in this paper), the intervention consisted of group sessions aiming at improving symptom recognition and management of hypoglycemia by patients with diabetes. This intervention was compared with current practice.

Both studies used the Dutch EQ-5D-5L value set for estimating utilities and found that the intervention under study was not cost-effective compared with usual care. In the depression study, the EQ-5D-5L was administered at baseline and every 3 months until 12-month follow-up. In the diabetes study, it was administered at baseline and at 2, 4, and 6-month follow-up. Costs were measured from a societal perspective and included costs of the study interventions, health care utilization, medication, and lost productivity (absenteeism and presenteeism; van Dijk et al., 2013; Wit et al., 2018).

2.2 | The EQ-5D-5L utility estimation methods

For both case studies, utility values were estimated using the EQ-5D-5L crosswalk approach (van Hout et al., 2012) and the 5L value sets of England (Devlin, Shah, Feng, Mulhern, & van Hout, 2018), the Netherlands (M. Versteegh et al., 2016), and Spain (Ramos-Goñi et al., 2017). In the crosswalk method, utility values for the 5L version were predicted from the 3L utilities of the United Kingdom (U.K. crosswalk; Dolan, 1997), the Netherlands (Dutch crosswalk; Lamers et al., 2005), and Spain (Spanish crosswalk; Badia et al., 2001).

2.3 | Analyses

2.3.1 | Hypothetical health states

First, both utility estimation methods were compared in terms of possible changes in utility values between hypothetical adjacent health states. For this purpose, the definition of adjacent health states published by Mulhern et al. (2018) was used, that is, "as having one dimension with one level difference." For instance, in the mobility dimension, the utility differences between health states 11111 and 21111; 21111 and 31111; 31111 and 41111; 41111 and 51111 were calculated. This approach was also applied to the other four dimensions, resulting in 25 adjacent hypothetical health states. Subsequently, the utility values of the 25 hypothetical health states were calculated based on both methods and compared. This analysis provides an overview of the potential magnitude of the differences due to the utility estimation method used. We extend the analysis performed by Mulhern et al. by also including the Dutch and Spanish sets.

2.3.2 | Case studies—Comparison of utility values

Hereafter, the pooled utility value distributions of all measurement points obtained from the samples included in the case studies were compared using Kernel density histograms. Using these data, the mean utility values, standard deviations of the mean utility values, ranges and 95% confidence intervals were estimated. Mean differences in utility

values between the EQ-5D-5L crosswalk and the 5L value set, including their 95% confidence intervals, were described by country. Also, the utility values generated by both methods were used to calculate QALYs. QALYs were estimated by the area-under-the-curve method with linear interpolation between time points.

2.3.3 | Case studies—Comparison of cost–utility outcomes

Subsequently, two cost–utility analyses were conducted per country, one using the EQ-5D-5L crosswalk and one using the 5L value set. For both case studies, missing data were imputed using multiple imputation by chained equations (van Buuren & Groothuis-Oudshoorn, 2011). The EQ-5D-5L descriptive system was imputed, rather than the corresponding utility values, to allow for the calculation of utility values based on different value sets using the imputed EQ-5D-5L responses. Predictive mean matching was used to deal with the skewed distribution of the costs and the categorical nature of the EQ-5D-5L descriptive system data (White, Royston, & Wood, 2011). All potential variables associated with the "missingness" of data, cost and effect outcomes, and possible confounders were included in the multiple imputation model. Datasets were imputed in order for the loss of efficiency to be less than 5% (White et al., 2011). Ten datasets were analyzed separately, and estimates were pooled using Rubin's rules (Rubin, 2004).

Incremental costs and incremental QALYs were estimated using seemingly unrelated regression analyses (Willan, Briggs, & Hoch, 2004), in which two separate regression models for costs and QALYs are estimated and the correlation between costs and QALYs is accounted for through correlated error terms. Bias-corrected and accelerated bootstrapping with 5,000 replications was used to estimate confidence intervals surrounding incremental costs (Chaudhary & Stearns, 1996). Incremental cost-effectiveness ratios (ICERs) were calculated by dividing incremental costs by incremental QALYs. Uncertainty surrounding the ICERs was estimated using bootstrapping (Black, 1990; Thompson & Barber, 2000) and by plotting cost-effectiveness acceptability curves (CEACs; Fenwick, O'Brien, & Briggs, 2004). CEACs indicate the probability of an intervention being cost-effective compared with a control for a range of willingness-to-pay thresholds (Fenwick et al., 2004). All data analyses were performed in Stata statistical software 14th version.

3 | RESULTS

3.1 | Hypothetical health states

Table 1 compares the differences between the EQ-5D-5L crosswalks and the 5L value sets for the 25 adjacent hypothetical health states. The United Kingdom, the Dutch, and the Spanish crosswalks resulted in largest utility decrements between response Levels 5 "extreme" and 4 "severe," whereas the largest decrements in their respective 5L value sets occurred between levels 4 "severe" and 3 "moderate."

In the EQ-5D-5L crosswalk for the United Kingdom, the utility decrement between Levels 2 "slight" and 1 "no problems" was larger than in the 5L value set for England in all five dimensions. In the Dutch crosswalk, the utility decrement between levels "slight" and "no problems" was very similar and slightly higher as compared with the Dutch 5L value set in all dimensions, except in the usual activities dimension. In the Spanish crosswalk, the utility decrement between levels "slight" and "no problems" was larger (but of small magnitude as compared with the U.K. crosswalk) than in the Spanish 5L value set, except in the anxiety/depression dimension.

3.2 | Case studies—Comparison of utility values

Table S1 presents the background characteristics of the two samples at baseline. Figure 1 shows the kernel density histograms for the EQ-5D-5L crosswalks and the 5L value sets by case study. Both utility estimation methods result in a left skewed distribution of utility values. In the depression study, the EQ-5D-5L crosswalks and the 5L value sets resulted in a unimodal distribution, whereas in the diabetes study, they resulted in a bimodal distribution. In both case studies, peaks for the U.K. crosswalk were different than for the England 5L value set. However, there were only slight differences in the utility value distributions for the Dutch and Spanish crosswalks and their 5L value sets, and the peaks for the two methods were similar.

TABLE 1 Comparing the utility decrement between adjacent health states by the EQ-5D-5L crosswalk and the 5L value sets of England, the Netherlands, and Spain

	England				The Netherlands				Spain			
Health states	U. K. Crosswalk	Diff	5L value set	Diff	Crosswalk	Diff	5L value set	Diff	Crosswalk	Diff	5L value set	Diff
11111	1.000		1.000		1.000		1.000		1.000		1.000	
21111	0.877	0.123	0.942	0.058	0.912	0.088	0.918	0.082	0.893	0.107	0.916	0.084
31111	0.850	0.027	0.924	0.018	0.893	0.019	0.896	0.022	0.870	0.023	0.901	0.015
41111	0.813	0.037	0.793	0.131	0.867	0.026	0.787	0.109	0.826	0.044	0.750	0.150
51111	0.336	0.477	0.726	0.067	0.534	0.333	0.750	0.037	0.255	0.571	0.663	0.088
11111	1.000		1.000		1.000		1.000		1.000		1.000	
12111	0.846	0.154	0.950	0.050	0.873	0.127	0.915	0.085	0.868	0.132	0.950	0.050
13111	0.815	0.031	0.920	0.030	0.847	0.026	0.892	0.023	0.842	0.026	0.947	0.003
14111	0.723	0.092	0.836	0.084	0.773	0.074	0.785	0.107	0.729	0.113	0.836	0.111
15111	0.436	0.287	0.797	0.039	0.543	0.230	0.785	0.000	0.376	0.353	0.804	0.032
11111	1.000		1.000		1.000		1.000		1.000		1.000	
11211	0.906	0.094	0.950	0.050	0.917	0.083	0.914	0.086	0.924	0.076	0.956	0.044
11311	0.883	0.023	0.937	0.013	0.897	0.020	0.866	0.048	0.905	0.019	0.951	0.005
14111	0.776	0.107	0.838	0.099	0.812	0.085	0.761	0.105	0.769	0.136	0.865	0.086
15111	0.556	0.220	0.816	0.022	0.638	0.174	0.761	0.000	0.490	0.279	0.847	0.018
11111	1.000		1.000		1.000		1.000		1.000		1.000	
11121	0.837	0.163	0.937	0.063	0.874	0.126	0.887	0.113	0.909	0.091	0.922	0.078
11131	0.796	0.041	0.916	0.021	0.843	0.031	0.861	0.026	0.887	0.022	0.899	0.023
11141	0.583	0.213	0.724	0.192	0.652	0.191	0.593	0.268	0.702	0.185	0.755	0.144
11151	0.264	0.319	0.665	0.059	0.366	0.286	0.538	0.055	0.424	0.278	0.618	0.136
11111	1.000		1.000		1.000		1.000		1.000		1.000	
11112	0.879	0.121	0.922	0.078	0.845	0.155	0.883	0.117	0.932	0.068	0.919	0.081
11113	0.848	0.031	0.896	0.026	0.805	0.040	0.808	0.075	0.914	0.018	0.872	0.047
11114	0.635	0.213	0.715	0.181	0.592	0.213	0.597	0.211	0.731	0.183	0.730	0.143
11115	0.414	0.221	0.711	0.004	0.370	0.222	0.532	0.065	0.541	0.190	0.652	0.077
55555	−0.594		−0.285		−0.329		−0.446		−0.654		−0.416	

Note. Diff: utility decrement between adjacent health state.
Gray shadow: In the U.K. crosswalk, the utility decrement between Levels 1 and 2 was larger than in the England 5L value set in all dimensions. In the Dutch crosswalk, the utility decrement between Levels 2 and 1 was very similar and slightly higher as compared with the Dutch 5L value set in all dimensions, except in the usual activities dimension (where the utility decrement is slightly small in the Dutch crosswalk than the Dutch 5L value set). In the Spanish crosswalk the utility decrement between Levels 2 and 1 was larger than in the Spanish 5L value set, except in the anxiety/depression dimension (where the utility decrement is slightly smaller in the Spanish crosswalk than the Spanish 5L value set).
Abbreviations: 5LEQ-5D-5L, five-level EuroQol five-dimension questionnaire

Table 2 reports the minimum, maximum, and mean utility values estimated by the EQ-5D-5L crosswalks and the 5L value sets in the two case studies. The maximum utility value was 1.000 for both utility estimation methods in both case studies and for all countries. Differences in minimum utility values were found between the EQ-5D-5L crosswalk and the 5L value set in both case studies, and those differences were most pronounced for the Netherlands. Mean utility values were lower when generated using the EQ-5D-5L crosswalk compared with the 5L value set in both case studies for England and Spain, whereas the opposite was true for the Netherlands.

3.3 | Case studies—Comparison of cost–utility outcomes

Table 3 shows the results of the economic evaluations using both utility estimation methods in each case study per country. In both case studies and for all countries, relatively small differences in incremental QALYs were found between the EQ-5D-5L crosswalk and the 5L value set (≤0.0038). In all scenarios, the intervention was on average less costly and less effective compared with the control condition, resulting in negative ICERs. Thus, both interventions were dominated by the control group for all countries. ICER point estimates differed between the two utility

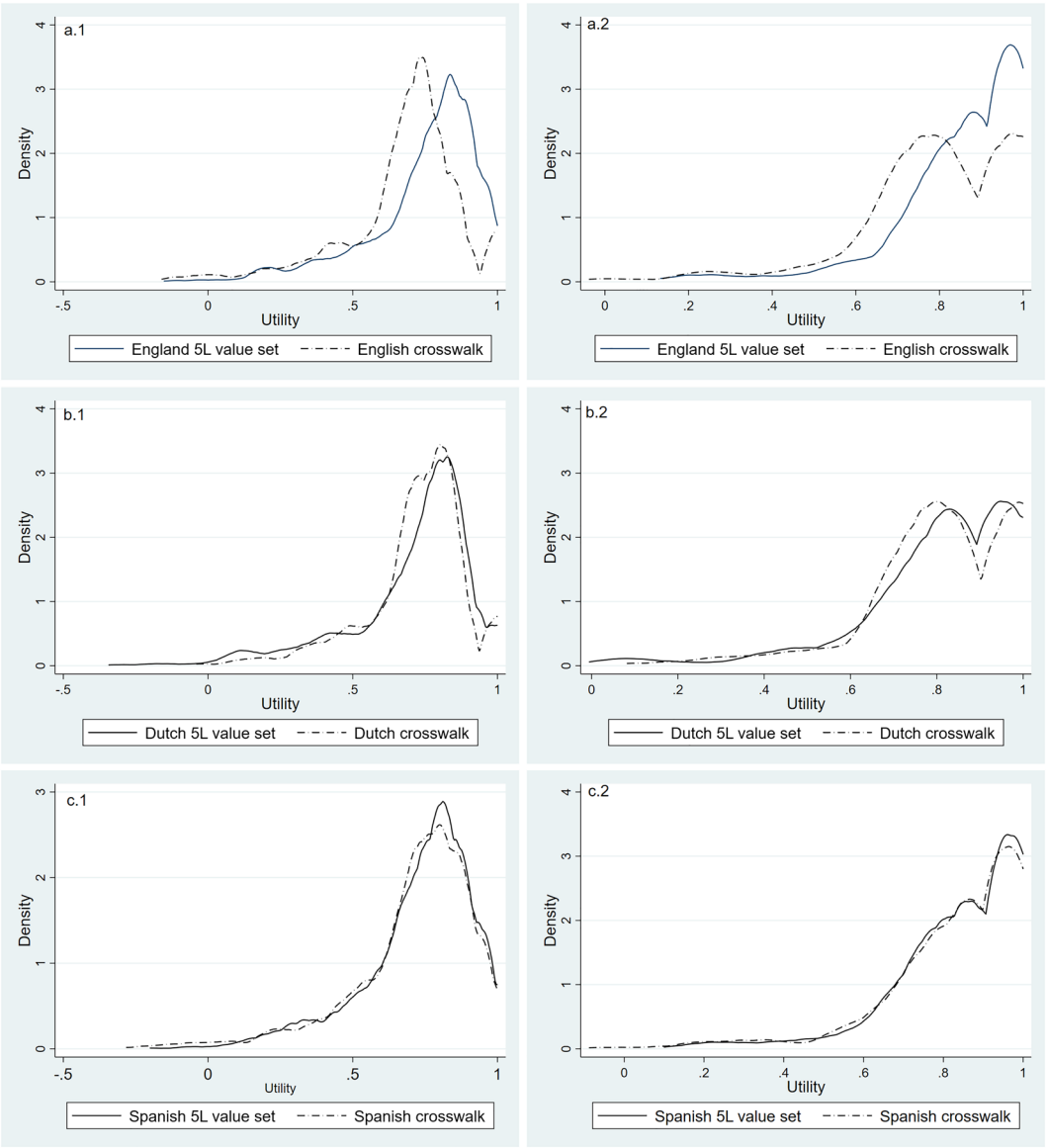

FIGURE 1 Kernel density histograms comparing the utility values distribution by the five-level (5L) EuroQol five-dimension questionnaire crosswalks and the 5L value sets by country. Depression study: a.1, b.1, and c.1. Diabetes study: a.2, b.2, and c.2 [Colour figure can be viewed at wileyonlinelibrary.com]

estimation methods for both case studies and all countries, with differences in the magnitude of the ICER point estimates ranging from €−35,998/QALY for Spain to €20,000/QALY for the Netherlands in the depression study and from €−4,063/QALY for England to €369/QALY for Spain in the diabetes study. Although the impact of using the EQ-5D-5L crosswalk or the 5L value set on the deterministic estimates of the ICER was large in both case studies and in all countries, this was mainly due to the small differences in QALYs between the treatment groups and not to the utility estimation methods.

Despite the observed differences in utility value distributions, mean utility values, and ICER point estimates, the distribution of bootstrapped cost–effect pairs on the cost-effectiveness plane was similar for the EQ-5D-5L crosswalks and 5L value sets in both case studies for all countries (Table 3).

Table 4 and Figure 2 show the differences in the probability of cost-effectiveness between both utility estimation methods by country and by case study. In both case studies, the probabilities of cost-effectiveness are relatively similar across methods in all three countries for all willingness-to-pay thresholds (Table 4). Consequently, the CEACs of both utility estimation methods look similar for all countries and case studies (Figure 2).

TABLE 2 Comparing utility values estimated by the EQ-5D-5L crosswalks and the 5L value sets of England, the Netherlands, and Spain

Method	Mean utility (*SD*)	Min	Max	95% CI
Depression study				
U.K. crosswalk	0.676 (0.19)	−0.160	1	0.664; 0.687
England 5L value set	0.752 (0.20)	−0.153	1	0.741; 0.763
Mean difference	−0.076 (0.02)			−0.073; −0.080
Dutch crosswalk	0.720 (0.17)	−0.041	1	0.710; 0.730
Dutch 5L value set	0.706 (0.22)	−0.344	1	0.694; 0.719
Mean difference	0.014 (0.07)			0.010; 0.018
Spanish crosswalk	0.712 (0.21)	−0.282	1	0.700; 0.725
Spanish 5L value set	0.730 (0.19)	−0.201	1	0.719; 0.741
Mean difference	−0.017 (0.07)			−0.013; −0.021
Diabetes study				
U.K. crosswalk	0.808 (0.19)	−0.038	1	0.792; 0.824
England 5L value set	0.862 (0.16)	0.132	1	0.849; 0.875
Mean difference	−0.054 (0.05)			−0.049; −0.059
Dutch crosswalk	0.8234 (0.17)	0.082	1	0.809; 0.837
Dutch 5L value set	0.8231 (0.19)	−0.006	1	0.808; 0.839
Mean difference	0.0003 (0.05)			0.003; 0.004
Spanish crosswalk	0.842 (0.18)	−0.088	1	0.827; 0.857
Spanish 5L value set	0.850 (0.16)	0.010	1	0.836; 0.864
Mean difference	−0.008 (0.05)			−0.004; −0.012

Abbreviations: 5L, five level; CI, confidence interval; EQ-5D-5L, five-level EuroQol five-dimension questionnaire; *SD*, standard deviation.

4 | DISCUSSION

The aim of the current study was to compare utility values generated using EQ-5D-5L crosswalks and 5L value sets for England, the Netherlands, and Spain and to assess whether the use of these different utility estimation methods affected cost–utility outcomes in two case studies. Results showed that differences exist in the utility value distribution, mean utility values, and ICER point estimates between the EQ-5D-5L crosswalks and the 5L value sets in all countries investigated. However, in our case studies, differences in mean utility values between both utility estimation methods were smaller than the minimum clinically important difference for EQ-5D utility values of 0.074 (Walters & Brazier, 2005). Moreover, the impact of using either one of those methods on the estimated utility values was similar in the intervention and control groups, thereby not translating into relevant differences in incremental QALYs and probabilities of the interventions being cost-effective.

1The observed differences in utility values between the EQ-5D-5L crosswalk and the 5L value set within and between countries could be due to differences in descriptive system between EQ-5D versions, the use of differing modeling techniques, and differences in sociodemographic characteristics of the study populations. It is noteworthy that mean utility values derived using the EQ-5D-5L crosswalk were lower than those derived using the 5L value set in England and Spain, but higher in the Netherlands. As EQ-5D-5L health states and predicted EQ-5D-3L health states were similar in all countries, this is likely due to differences across countries regarding the decrements members of the general public attribute to the severity levels of both versions of the EQ-5D. Moreover, although the observed differences did not impact the probability of cost-effectiveness in the uncertainty analyses, deterministic outcomes of economic evaluations were affected by these differences. Especially striking is the large difference in utility values between the U.K. EQ-5D-5L crosswalk and the England 5L value set, which is in contrast with the results for Spain and the Netherlands. This finding supports the recommendation of NICE to calculate utility values using the crosswalk from the 3L value set and not the EQ-5D-5L values (National Institute for Health and Care Excellence, 2019).

To the best of our knowledge, only three studies have been published comparing the EQ-5D-5L crosswalk and 5L value set (Camacho et al., 2018; Mulhern et al., 2018; Yang, Devlin, & Luo, 2019). The first study, conducted by Mulhern et al. (2018) observed that the England 5L value set produces utility values that are on average 0.085 points higher than the U.K. EQ-5D-5L crosswalk across 12 health conditions, including depression and diabetes. In the current study, we observed similar differences in utility values for England (e.g., a 0.076-point difference among depressed patients), but also that these differences were less pronounced for the Netherlands and Spain.

TABLE 3 Results of the cost–utility analysis

Method	Incremental costs (95% CI)	Incremental QALYs (95% CI)	ICER (€)	Distribution of the cost-effectiveness plane			
				Northeast[a] (%)	Southeast[b] (%)	Southwest[c] (%)	Northwest[d] (%)
Depression study							
U.K. crosswalk	€2,000 [€−935, €5,122]	−0.0085 [0.0553, 0.0387]	−236,282	28	8	3	61
England 5L value set	€2,000 [€−951, €5,172]	−0.0080 [−0.0526, 0.0365]	−248,856	27	7	3	63
Difference	—	0.0005	−12,574	1	1	0	2
Dutch crosswalk	€2,000 [€−957, €5,159]	−0.0067 [−0.0462, 0.0327]	−297,131	29	7	3	61
Dutch 5L value set	€2,000 [€−970, €5,150]	−0.0072 [−0.058, 0.043]	−277,071	31	7	3	59
Difference	—	−0.0005	20,060	2	0	0	2
Spanish crosswalk	€2,000 [€−935, €5,166]	−0.0084 [−0.0582, 0.0415]	−238,703	28	8	2	61
Spanish 5L value set	€2,000 [€−966, €5,183]	−0.0073 [−0.0530, 0.0385]	−274,700	30	7	3	60
Difference	—	0.0011	−35,997	2	1	1	1
Diabetes study							
U.K. crosswalk	€49 [€−1,205, €1,090]	−0.0090 [−0.0378, 0.0197]	−5,441	15	12	33	40
England 5L value set	€49 [€−1,206, €1,077]	−0.0052 [−0.0300, 0.0197]	−9,504	19	16	30	35
Difference	—	0.0038	−4,063	4	4	3	5
Dutch crosswalk	€49 [€−1,202, €1,097]	−0.0082 [−0.0342, 0.0179]	−6,014	15	12	34	39
Dutch 5L value set	€49 [€−1,205, €1,090]	−0.0066 [−0.363, 0.0231]	−7,424	18	15	31	36
Difference	—	0.0016	−1,410	3	3	3	3
Spanish crosswalk	€49 [€−1,204, €1,098]	−0.0071 [−0.0344, 0.0202]	−6,932	16	14	32	38
Spanish 5L value set	€49 [€−1,204, €1,098]	−0.0075 [−0.0332, 0.0182]	−6,563	16	13	33	38
Difference	—	−0.0004	369	0	1	1	0

Note. In both case studies, the new intervention was on average more costly and less effective and therefore was dominated by control.

Abbreviations: CI, confidence interval; ICER, cost-effectiveness ratio; QALY, quality-adjusted life-year.

[a]New intervention is more effective but more costly compared with control.

[b]New intervention is more effective and less costly compared with control.

[c]New intervention is less effective and less costly compared with control.

[d]New intervention is less effective and more costly compared with control.

TABLE 4 Comparing the probability of cost-effectiveness between the EQ-5D-5L crosswalk and the 5L value set by country and case study at different willingness-to-pay thresholds

Method	€0 /QALY gained	€20,000 /QALY gained	€34,000[a] /QALY gained
Depression study			
England 5L value set	0.1039	0.1226	0.1382
U.K. crosswalk	0.1041	0.1256	0.1425
Difference	−0.0002	−0.0030	−0.0043
Dutch 5L value set	0.1039	0.1281	0.1480
Dutch crosswalk	0.1025	0.1202	0.1346
Difference	0.0014	0.0079	0.0133
Spanish 5L value set	0.1041	0.1244	0.1413
Spanish crosswalk	0.1032	0.1292	0.1487
Difference	0.0009	−0.0048	−0.0073
Diabetes study			
England 5L value set	0.4668	0.4059	0.3789
U.K. crosswalk	0.4669	0.3644	0.3234
Difference	−0.0001	0.0415	0.0554
Dutch 5L value set	0.4669	0.3915	0.3621
Dutch crosswalk	0.4668	0.3703	0.3284
Difference	0.0001	0.0212	0.0337
Spanish 5L value set	0.4668	0.3783	0.3396
Spanish crosswalk	0.4669	0.3869	0.3536
Difference	−0.0001	−0.0086	−0.0140

Note. The probability of cost-effectiveness is determined as the proportion of the bootstrapped cost–effect pairs where the intervention is cost-effective (i.e., the proportion of the bootstrapped cost–effect pairs falling to the South–East quadrant of the CE plane) given a willingness-to-pay threshold (e.g., the probability of the intervention being cost-effective was.1382 using U.K. crosswalk and.1425 using 5L value set for England at willingness-to-pay of €34,000/QALY gained, a difference of.0043).

Abbreviations: 5L, five level; EQ-5D-5L, five-level EuroQol five-dimension questionnaire; QALY, quality-adjusted life year.

[a]U.K. upper commonly accepted willingness-to-pay threshold.

The second study, conducted by Camacho et al., found mean utility values estimated using the England 5L value set to be approximately 0.08 points higher than mean utility values estimated using the U.K. EQ-5D-5L crosswalk in people with mental health issues. Consequently, they found ICER point estimates to be higher for the U.K. EQ-5D-5L crosswalk than for the England 5L value set (Camacho et al., 2018). In our study, similar differences in mean utility values and ICER point estimates were found across utility estimation methods in both case studies for all countries investigated.

More recently, Yang et al. (2019) explored the impact of using EQ-5D-5L crosswalks or 5L value sets on incremental QALYs in a model-based economic evaluation for seven countries (including England and the Netherlands, but not Spain). For this, they used data from patients at end-stage renal disease. Similar to our results, they found that both utility estimation methods resulted in comparable incremental QALYs for the Netherlands (mean difference: 0.009). However, they found a relatively large difference in incremental QALYs between the U.K. EQ-5D-5L crosswalk and England 5L value set (mean difference: 0.098), which was in contrast with our results. This difference in results may be due to the fact that Yang et al. (2019) included severely ill patients, whereas moderately ill patients were included in our case studies.

The current study went beyond the aforementioned studies by assessing the impact of using either one of the utility estimation methods on utility values as well as cost–utility analysis outcomes, such as incremental QALYs, ICER point estimates, and CEACs. Our results indicated that the identified differences in the utility value distribution, mean utility values, and ICER point estimates between the EQ-5D-5L crosswalks and the 5L value sets for England, the Netherlands, and Spain did not translate into relevant differences in incremental QALYs and the probability of the interventions being cost-effective as compared with the control group. This is likely caused by the high level of uncertainty surrounding utility values and QALYs and, more importantly, by the fact that the impact of using either one of the utility estimation methods was similar in the intervention and control groups, thereby not affecting cost–utility outcomes, such as incremental QALYs and CEACs. However, it is unclear whether this also applies to populations with more severe health conditions.

This study has some limitations. First, both case studies included relatively few patients with more severe EQ-5D-5L health states and the interventions under study were primarily aimed at improving quality of life, instead of increasing life expectancy. Therefore, further research is needed to assess whether the current findings also apply to interventions aimed at more severely ill patients. This is particularly important because utility decrements between response Levels

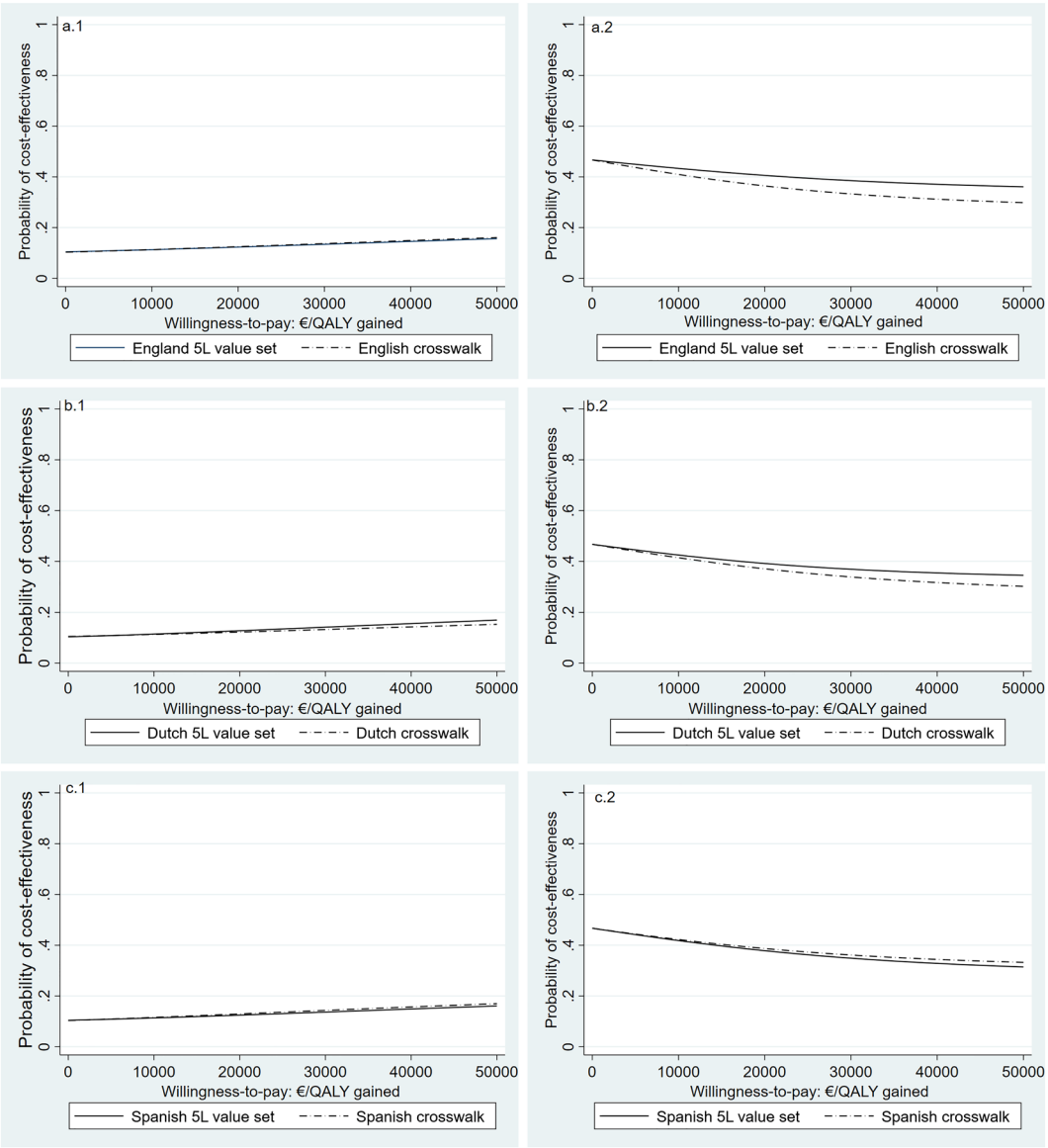

FIGURE 2 The five-level (5L) EuroQol five-dimension questionnaire crosswalks and the 5L value sets cost-effectiveness acceptability curves by country and case study. The probability of the intervention being cost-effective at different willingness-to-pay thresholds. QALY, quality-adjusted life year. Depression study: a.1, b.1, and c.1. Diabetes study: a.2, b.2, and c.2 [Colour figure can be viewed at wileyonlinelibrary.com]

5 and 4 are typically larger in the EQ-5D-5L crosswalks than in the 5L value sets. Another potential limitation is that the difference in QALYs between the intervention and control groups were relatively small and in favor of the control group in both case studies. Future research should indicate whether the probability of an intervention being cost-effective differs in economic evaluations where the difference in QALYs across treatment groups is larger and in favor of the intervention group.

Strengths of our study are that we did not only explore differences in utility values but also performed a full cost–utility analysis to assess the impact on cost–utility outcomes, that is, incremental QALYs and ICERs. Moreover, to the best of our knowledge, we are the first to assess the joint uncertainty around ICER in two empirical longitudinal datasets in relation to the use of these different utility estimation methods. Bearing in mind the importance of using country-specific value sets in order to account for sociocultural differences among populations (Xie et al., 2013); another strength is that we explored the impact of using the crosswalk or value set in three different countries.

5 | CONCLUSIONS

In line with previous research, the current study found differences to exist in utility value distributions, mean utility values and ICER point estimates between EQ-5D-5L crosswalks and 5L value sets for England, the Netherlands, and Spain. However, in our case studies, these differences did not translate into relevant differences in incremental QALYs and the probability of the intervention being cost-effective as compared with the control group at different willingness-to-pay thresholds. In spite of the identified differences between both utility valuation methods, this suggests that EQ-5D-5L crosswalks and 5L value sets can be used interchangeably in economic evaluations of patients affected with mild or moderate conditions. Further research is needed to establish whether these findings also apply in situations where the EQ-5D-5L is administered to severely ill patients.

CONFLICT OF INTEREST

All authors declare no conflict of interests.

ORCID

Ângela Ben https://orcid.org/0000-0003-4793-9026
Aureliano Paolo Finch https://orcid.org/0000-0003-4472-197X

REFERENCES

Badia, X., Roset, M., Herdman, M., & Kind, P. (2001). A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Medical Decision Making*, *21*, 7–16. https://doi.org/10.1177/0272989X0102100102

Black, W. C. (1990). The CE plane: A graphic representation of cost-effectiveness. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, *10*(3), 212–214. https://doi.org/10.1177/0272989X9001000308

Brazier, J., Briggs, A., & Bryan, S. (2018). EQ-5D-5L: Smaller steps but a major step change? *Health Economics*, *27*(1), 4–6. https://doi.org/10.1002/hec.3627

van Buuren S, Groothuis-Oudshoorn K. 2011. MICE: Multivariate imputation by chained equations in R. Journal of Statistical Software. Available from: https://www.jstatsoft.org/article/view/v045i03 [].

Camacho, E. M., Shields, G., Lovell, K., Coventry, P. A., Morrison, A. P., & Davies, L. M. (2018). A (five-) level playing field for mental health conditions?: Exploratory analysis of EQ-5D-5L-derived utility values. *Quality of Life Research*, *27*(3), 717–724. https://doi.org/10.1007/s11136-017-1768-1

Chaudhary, M., & Stearns, S. (1996). Estimating confidence intervals for cost-effectiveness ratios: An example from a randomized trial. *Statist Med*, *15*(13), 1447–1458.

Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & van Hout, B. (2018). Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Economics*, *27*(1), 7–22. https://doi.org/10.1002/hec.3564

van Dijk, S. E., Pols, A. D., Adriaanse, M. C., Bosmans, J. E., Elders, P. J., van Marwijk, H. W., & van Tulder, M. W. (2013). Cost-effectiveness of a stepped-care intervention to prevent major depression in patients with type 2 diabetes mellitus and/or coronary heart disease and subthreshold depression: Design of a cluster-randomized controlled trial. *BMC Psychiatry*, *13*, 2–9, 128.

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care*, *35*(11), 1095–1108.

EuroQol Group (1990). EuroQol-A new facility for the measurement of health-related quality of life. *Health Policy*, *16*(3), 199–208.

Evans Kreider, K., Pereira, K., & Padilla, B. I. (2017). Practical approaches to diagnosing, treating and preventing hypoglycemia in diabetes. *Diabetes Therapy*, *8*(6), 1427–1435. https://doi.org/10.1007/s13300-017-0325-9

Fenwick, E., O'Brien, B. J., & Briggs, A. (2004). Cost-effectiveness acceptability curves – Facts, fallacies and frequently asked questions. *Health Economics*, *13*(5), 405–415. https://doi.org/10.1002/hec.903

Finch, A. P., Brazier, J. E., & Mukuria, C. (2017). What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *The European Journal of Health Economics*, *19*(4), 1–14.

Finch, A. P., Brazier, J. E., Mukuria, C., & Bjorner, J. B. (2017). An exploratory study on using principal-component analysis and confirmatory factor analysis to identify bolt-on dimensions: The EQ-5D case study. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *20*(10), 1362–1375. https://doi.org/10.1016/j.jval.2017.06.002

Herdman, M., Gudex, C., Lloyd, A., Janssen, M. F., Kind, P., Parkin, D., ... Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*(10), 1727–1736. https://doi.org/10.1007/s11136-011-9903-x

van Hout, B., Janssen, M. F., Feng, Y.-S., Kohlmann, T., Busschbach, J., Golicki, D., ... Pickard, A. S. (2012). Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health*, *15*(5), 708–715. https://doi.org/10.1016/j.jval.2012.02.008

Janssen, M. F., Bonsel, G. J., & Luo, N. (2018). Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *PharmacoEconomics*, *36*(6), 675–697.

Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., ... Busschbach, J. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: A multi-country study. *Quality of Life Research*, *22*(7), 1717–1727. https://doi.org/10.1007/s11136-012-0322-4

Jensen, C. E., Riis, A., Pedersen, K. M., Jensen, M. B., & Petersen, K. D. (2014). Study protocol of an economic evaluation of an extended implementation strategy for the treatment of low back pain in general practice: A cluster randomised controlled trial. *Implementation Science*, *9*(1), 2–7, 140. https://doi.org/10.1186/s13012-014-0140-x

Krog, A. H., Sahba, M., Pettersen, E. M., Wisløff, T., Sundhagen, J. O., & Kazmi, S. S. (2017). Cost-utility analysis comparing laparoscopic vs open aortobifemoral bypass surgery. *Vascular Health and Risk Management*, *13*, 217–224. https://doi.org/10.2147/VHRM.S138516

Lamers, L. M., Stalmeier, P. F. M., McDonnell, J., Krabbe, P. F. M., & J.van Busschbach J. (2005). Kwaliteit van leven meten in economische evaluaties: Het Nederlands EQ-5D-tarief. *Ned Tijdschr Geneeskd*, *149*, 1574–1578.

Mulhern, B., Feng, Y., Shah, K., Janssen, M. F., Herdman, M., van Hout, B., & Devlin, N. (2018). Comparing the UK EQ-5D-3L and English EQ-5D-5L Value Sets. *PharmacoEconomics*, *36*(6), 699–713.

National Institute for Health and Care Excellence. 2018. NICE position statement on the EQ-5D-5L – EQ-5D. Available from: https://euroqol.org/nice-position-statement-on-the-eq-5d-5l/ [].

National Institute for Health and Care Excellence. 2019. Position statement on use of the EQ-5D-5L value set for England (updated October 2019)|NICE technology appraisal guidance|NICE guidance|Our programmes|What we do|About|NICE. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l [].

Nederland, Z. (2016). Zorginstituut Nederland - Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. In *Ministerie van Volksgezondheid*. Welzijn en: Sport. Available from:. https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg

Oppe, M., Devlin, N. J., van Hout, B., Krabbe, P. F. M., & de Charro, F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health*, *17*(4), 445–453. https://doi.org/10.1016/j.jval.2014.04.002

Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goñi, J. M., & Luo, N. (2016). EuroQol protocols for time trade-off valuation of health outcomes. *PharmacoEconomics*, *34*(10), 993–1004. https://doi.org/10.1007/s40273-016-0404-1

Ramos-Goñi, J. M., Pinto-Prades, J. L., Oppe, M., Cabasés, J. M., Serrano-Aguilar, P., & Rivero-Arias, O. (2017). Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Medical Care*, *55*(7), e51.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. n/a: John Wiley & Sons.

Ruo, B., Rumsfeld, J. S., Hlatky, M. A., Liu, H., Browner, W. S., & Whooley, M. A. (2003). Depressive symptoms and health-related quality of life: The heart and soul study. *JAMA*, *290*(2), 215–221. https://doi.org/10.1001/jama.290.2.215

Thompson, S. G., & Barber, J. A. (2000). How should cost data in pragmatic randomised trials be analysed? *BMJ: British Medical Journal*, *320*(7243), 1197–1200. https://doi.org/10.1136/bmj.320.7243.1197

Versteegh, M. M., Vermeulen, K. M., Evers, S. M. A. A., de Wit, G. A., Prenger, R., & Stolk, E. A. (2016). Dutch tariff for the five-level version of EQ-5D. *Value in Health*, *19*(4), 343–352.

Walters, S. J., & Brazier, J. E. (2005). Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research*, *14*(6), 1523–1532. https://doi.org/10.1007/s11136-004-7713-0

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399. https://doi.org/10.1002/sim.4067

Willan, A. R., Briggs, A. H., & Hoch, J. S. (2004). Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Economics*, *13*(5), 461–475.

Wit, M., Rondags, S. M. P. A., Tulder, M. W., Snoek, F. J., & Bosmans, J. E. (2018). Cost-effectiveness of the psycho-educational blended (group and online) intervention HypoAware compared with usual care for people with Type 1 and insulin-treated Type 2 diabetes with problematic hypoglycaemia: Analyses of a cluster-randomized controlled trial. *Diabetic Medicine*, *35*(2), 214–222. https://doi.org/10.1111/dme.13548

Xie, F., Gaebel, K., Perampaladas, K., Doble, B. M., & Pullenayegum, E. (2013). Comparing EQ-5D valuation studies: A systematic review and methodological reporting checklist. *Value in Health*, *16*(3), A44–A45.

Yang, F., Devlin, N., & Luo, N. (2019). Cost-utility analysis using EQ-5D-5L data: Does how the utilities are derived matter? *Value in Health*, *22*(1), 45–49. Available from:. https://www.valueinhealthjournal.com/article/S1098-3015(18)32210-1/fulltext

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ben Â, Finch AP, van Dongen JM, et al. Comparing the EQ-5D-5L crosswalks and value sets for England, the Netherlands and Spain: Exploring their impact on cost-utility results. *Health Economics*. 2020;29:640–651. https://doi.org/10.1002/hec.4008