Assessing the impact of EQ-5D country-specific value sets on cost-utility outcomes

Article  (Accepted Version)


This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/101934/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher’s version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
Abstract

Purpose: To assess the impact of EQ-5D country-specific value sets on cost-utility outcomes.

Methods: Data from two RCTs on low back pain (LBP) and depression were used. 3L value sets were identified from the EuroQol website. A nonparametric crosswalk was employed for each tariff to obtain the likely 5L values. Differences in QALYs between countries were tested using paired t-tests, with the UK as reference. Cost-utility outcomes were estimated for both studies and both EQ-5D versions, including differences in Quality Adjusted Life Years (QALYs) and Cost-Effectiveness Acceptability Curves.

Results: For the 3L, QALYs ranged between 0.650 (Taiwan) and 0.892 (United States) in the LBP study and between 0.619 (Taiwan) and 0.879 (United States) in the depression study. In both studies, most country-specific QALY estimates differed statistically significantly from that of the UK. ICERs ranged between €2,044/QALY (Taiwan) and €5,897/QALY (Zimbabwe) in the LBP study and between €38,287/QALY (Singapore) and €96,550/QALY (Japan) in the depression study. At the NICE threshold of €23,300/QALY (≈£20,000/QALY), the intervention’s probability of being cost-effective versus control ranged between 0.751 (Zimbabwe) and 0.952 (Taiwan) and between 0.230 (Canada) and 0.396 (Singapore) in the LBP study and depression study, respectively. Similar results were found for the 5L, with extensive differences in ICERs and moderate differences in the probability of cost-effectiveness.

Conclusions: This study indicates that the use of different EQ-5D country-specific value sets impacts on cost-utility outcomes. Therefore, to account for the fact that health state preferences are affected by sociocultural differences, relevant country-specific value sets should be used.

Keywords: cost-utility analysis, EQ-5D, value sets
Introduction

In economic evaluations of healthcare interventions, Quality-Adjusted Life Years (QALYs) are often used as a metric of health effects. QALYs combine the length of life and the health-related quality of life (HRQoL) into one metric. HRQoL is measured in terms of utility units. A utility is a weight that typically indicates the general public’s strength of preference or desirability for a given health state or condition. A utility of zero represents dead and a utility of one full health, while negative values are assigned to those states considered worse than death\(^1,2\).

Utility values are typically estimated using generic preference-based measures of health\(^3,4\), including the EuroQol five-dimension questionnaire (EQ-5D)\(^5\), the short form 36 health survey questionnaire (SF-6D)\(^6\), the Health Utilities Index Mark 3 (HUI3)\(^7\), the Assessment of Quality of Life instrument (AQoL)\(^8\), and the 15-dimensional measure of HRQoL (15D)\(^9\). These measures describe health in terms of dimensions and severity levels, and come with a value set or tariff that assigns a utility value to each of the health states described\(^2\). Such value sets have been traditionally estimated using cardinal response methods\(^10\), such as the Time Trade-Off (TTO), the Visual Analogue Scale (VAS) and the Standard Gamble (SG), and more recently using ordinal response methods, such as a Discrete Choice Experiment including duration (DCE\(_{TTO}\)) on a subset of all possible health states\(^11,12\). The derived data are modelled and result in a tariff through which the full set of health states can be obtained\(^2,13,14\).

The EQ-5D\(^5\) is, among the other generic preference-based measures, the most commonly used worldwide\(^2\). The original EQ-5D uses a classification system consisting of five health dimensions, i.e. mobility, self-care, usual activities, pain/discomfort and anxiety/depression, and three severity levels, i.e. no problems, some problems, severe problems (further referred to as the 3L)\(^5\). To increase the 3L’s sensitivity to changes in health and to reduce its commonly observed ceiling effects\(^15\), a 5-level version of the EQ-5D was developed (further referred to as the 5L)\(^16\). The 5L maintained the same health dimensions, but described health using five severity levels. Currently, a large number of value sets is available for estimating EQ-5D health states’ utility values. As evidence suggests that health state preferences are affected by sociocultural differences, such value sets are typically developed per country separately\(^17\).

Previous studies indicate that different country-specific EQ-5D value sets result in different utility values and QALY estimates\(^13,18,19\). However, this does not necessarily
mean that the identified differences also impact cost-utility outcomes, and thus, the health technology assessment (HTA) process. This is because the identified differences across countries may be equal in the intervention and control group, thereby not affecting cost-utility outcomes, such as incremental cost-effectiveness ratios and cost-effectiveness acceptability curves\textsuperscript{19}. To address this research gap, the present study aimed to explore the impact of country-specific EQ-5D value sets on cost-utility outcomes.
Methods

The impact of country-specific EQ-5D value sets on cost-utility outcomes was explored by comparing the results of two cost-utility analyses, employing different EQ-5D country-specific value sets, for both the 3L and 5L.

Data of two published randomized controlled trials (RCTs) performed in the Netherlands were used to estimate cost-utility outcomes\textsuperscript{20,21}. As the EQ-5D focuses on physical as well as mental aspects of health, one of these studies included patients with a physical disorder (i.e. low back pain [LBP]) and the other included patients with a mental disorder (i.e. depression).

The LBP study evaluated the cost-effectiveness of a treatment-based classification system for sub-acute and chronic LBP patients in comparison with usual care. The study had a 52-week follow-up. A total of 156 patients were included, of which 74 in the intervention group and 82 in the control group\textsuperscript{20}. The depression study evaluated a nurse-led self-help treatment in combination with usual care for recurrent depression patients in comparison with usual care alone. The study had a 65-week follow-up. For the purpose of this study, only data collected up until 52 weeks were used. A total of 248 patients were included, of which 124 in the intervention group and 124 in the control group\textsuperscript{21}.

In both studies, costs were measured from a societal perspective, including healthcare and lost productivity costs. All costs were expressed in Euros 2013. A detailed description of the example studies’ measurement and valuation of costs can be found elsewhere\textsuperscript{20-23}. Baseline characteristics of the example studies’ participants can be found in Appendix 1. A summary of the health states of the example studies’ participants can be found in Appendix 2.

Country-specific EQ-5D utility values

In the LBP study, the 3L was administered at baseline, 8, 26, and 52 weeks. In the depression study, it was administered at baseline, 13, 26, 39, and 52 weeks. Country-specific 3L value sets were identified from the EuroQol website (http://euroqol.org). In total, 33 3L value sets were available on the EuroQol website, of which 23 were developed using TTO and 10 were developed using VAS or a combination of both. Of them, we only included TTO-based value sets, because previous evidence
indicates that TTO values tend to be systematically higher than VAS values, irrespective of country. Herewith, we wanted to ensure that possible differences across countries were due to sociocultural differences, rather than the previously established differences between VAS and TTO values. On top of that, value sets were excluded if they were derived from populations other than the general population, and if they contained interaction terms, quadratic variables, and/or additional health states, such as death or unconscious. Eventually, 16 different country-specific value sets were included (Appendix 3). The participants’ EQ-5D health states were converted into 3L utility values using all of the identified value sets. Additionally, 5L utility values were estimated for all countries using the crosswalk approach. To do that, the observed 3L health states were converted into possible 5L health states using the nonparametric crosswalk developed by van Hout et al.

**Statistical analysis**

Missing EQ-5D descriptive system data and cost data were imputed using multivariate imputation by chained equations. Pooled estimates were calculated using Rubin’s rules. Imputation models were constructed per study, and included total costs, EQ-5D item responses at all measurement points, variables related to the “missingness” of data, variables differing between treatment groups at baseline, and variables that were related to the outcomes. Imputation models were stratified for study group.

Cost-utility outcomes were estimated for both studies, all countries, and both versions of the EQ-5D. QALYs were estimated by multiplying the participants’ utility value of a health state by the duration of time they spent in that health state using linear interpolation between measurement points. Subsequently, it was explored whether there were minimally clinically important differences between the country-specific QALY estimates and the UK QALY estimate. The UK QALY estimate was used as reference, as a quick Google Scholar search indicated the UK value set to be the most frequently referenced one. The minimally clinically important difference for QALYs was set at 0.057. Paired t-tests were used to test whether country-specific QALY estimates statistically significantly differed from that of the UK.

Incremental cost-effectiveness ratios (ICERs) were calculated by dividing the mean differences in costs by the mean differences in QALYs. Mean differences in costs and QALYs across groups were estimated using seemingly unrelated regression analyses.
For all country-specific ICERs, it was explored whether they were above or below the upper and lower bound of the Dutch willingness-to-pay (WTP) threshold (i.e. €10,000/QALY to €80,000/QALY) as well as the NICE threshold of £20,000/QALY (i.e. about €23,300/QALY). Bias-corrected and accelerated bootstrapping with 5,000 replications was used to estimate the uncertainty surrounding cost differences and ICERs. Bootstrapped cost-effect pairs were plotted on cost-effectiveness planes. To provide a summary measure of the joint uncertainty surrounding cost and QALY differences, cost-effectiveness acceptability curves (CEAC) were constructed\textsuperscript{52}. A CEAC provides insight into the probability of an intervention being cost-effective compared with a control for a range of ceiling ratios (i.e. the maximum amount of money decision-makers are willing to pay per additional unit of effect)\textsuperscript{52}. Analyses were performed in STATA v12 and statistical significance was set at p<0.05.

**Sensitivity analysis**

For countries in which an EQ-5D value set is already available for the 5L (See Appendix 1), alternative 5L utility values were estimated. In these cases, the transition probabilities of van Hout et al. (2012)\textsuperscript{47} were multiplied by the 5L utility values instead of the 3L utility values. Using the alternative 5L utility values, cost-utility outcomes were estimated for both studies and all applicable countries again.
Results

QALYs

In the LBP study, QALY estimates derived from the 3L ranged between 0.650 (Taiwan) and 0.892 (United States [US]) (Table 1). Except for the Japanese case, country-specific QALY estimates statistically significantly differed from that of the UK. For Germany, the United States, South Korea, Singapore, and Taiwan, this difference was also clinically relevant (Table 1). Similar results were found when calculating QALYs for the 5L using the nonparametric crosswalk, with QALY estimates ranging between 0.644 (Taiwan) and 0.884 (US); 14 out of 15 country-specific QALY estimates statistically significantly differing from that of the UK, and four out of 15 country-specific QALY estimates clinically relevantly differing from that of the UK.

In the depression study, QALY estimates derived from the 3L ranged between 0.619 (Taiwan) and 0.879 (US) (Table 1). Once again, except for the Japanese case, country-specific QALY estimates statistically significantly differed from that of the UK, and for Germany, the United States, South Korea, Singapore and Taiwan this difference was clinically relevant (Table 1). Similar results were found when calculating QALYs for the 5L using the nonparametric crosswalk, with these ranging between 0.612 (Taiwan) and 0.870 (US), 14 out of 15 country-specific QALY values statistically significantly differing from that of the UK, and four out of 15 country-specific QALY values clinically relevantly differing from that of the UK.

ICERs and CEACs

In the LBP study, ICERs derived from the 3L ranged between €2,044 per QALY (Taiwan) and €5,897 per QALY (Zimbabwe) (Table 2). All ICERs were below the Dutch as well as the NICE threshold. The corresponding probabilities of the intervention being cost-effective compared with the control differed moderately across countries; at the lower bound of the Dutch threshold (i.e. €10,000 per QALY), this probability ranged between 0.569 (Zimbabwe) and 0.810 (Taiwan) and at the NICE threshold it ranged between 0.751 (Zimbabwe) and 0.952 (Taiwan) (Table 4; Figure 1a). Similar results were found for the 5L, with extensive differences in ICERs and the probability of cost-effectiveness ranging between 0.565 (Zimbabwe) and 0.806 (Taiwan) at the lower bound of the Dutch threshold (Table 2-4).
In the depression study, ICERs derived from the 3L ranged between €38,287 per QALY (Singapore) and €96,550 per QALY (Japan) (Table 3). None of the ICERs was below the lower bound of the Dutch threshold and the NICE threshold, and 14 (88%) ICERs were below the upper bound of the Dutch threshold. Again, differences were found across countries in the probability of the intervention being cost-effective compared with the control. At the NICE threshold, this probability ranged between 0.230 (Canada) and 0.396 (Singapore) (Table 4; Figure 1b). Similar results were found for the 5L, with extensive differences in ICERs and the probability of cost-effectiveness ranging between 0.283 (Japan) and 0.394 (Singapore) at the NICE threshold (Table 2-4).

Please note that countries with the most similar QALY estimates (Table 1; e.g. UK and Japan) do not necessarily have the most similar ICERs (Table 2 & 3). This is due to the fact that QALYs were estimated for all participants of the example studies, whereas ICERs were based on the mean differences in QALYs across study groups.

**Sensitivity analysis**

Results of the sensitivity analysis were similar to those of the main analysis, with QALY estimates and ICERs differing extensively across countries and moderate differences across countries in the probability of the intervention being cost-effective compared with the control at all thresholds, in both studies (Table 2-4).
Discussion

This study explored the impact of using different country-specific EQ-5D value sets on cost-utility outcomes by employing 16 different EQ-5D country-specific value sets for both the 3L and 5L. Results showed that the use of different value sets may impact on cost-utility outcomes. To illustrate, for the 3L, ICERs ranged between €2,044/QALY (Taiwan) and €5,897/QALY (Zimbabwe) in the LBP study and between €38,287/QALY (Singapore) and €96,550/QALY (Japan) in the depression study. Moreover, at the lower bound of the Dutch QALY threshold, the probability of a treatment-based classification system for sub-acute and chronic LPB patients being cost-effective compared with usual care was found to range between 0.569 in Zimbabwe and 0.810 in Taiwan.

Previous studies on the impact of country-specific EQ-5D value sets were mostly based on the 3L. In line with the present findings, these studies found Danish, German, Dutch, United States, and Japanese utility values derived from the 3L to be generally higher than those of the UK. Also, Badia et al. (2001) and Kiadaliri et al (2015) found Spanish and Swedish 3L utility values to be either higher or lower than those of the UK, depending of the severity of a health state. Another study of van Lien et al. (2015) found that United States, Danish, French, German, Japanese, and Dutch ICERs derived from the 3L differed extensively with that of Canada, with relative differences ranging from -17% (Germany) to +16% (United Kingdom). Similar results have recently been found for the 5L, when using both crosswalks and available 5L value sets. To the best of our knowledge, the impact of country-specific EQ-5D value sets on the probability of an intervention being cost-effective compared with the control has never been explored, whereas this is one of the most important outcomes of a cost-utility analysis for decision-making purposes.

Strengths of this study include the fact that it was the first to systematically assess the impact of country-specific EQ-5D value sets on the probability of an intervention being cost-effective, its inclusion of a broad range of countries, and its use of empirical study data of patients suffering from both a physical and a mental health problem. Also, special efforts were made to ensure that the observed differences across countries could be ascribed to sociocultural differences instead of methodological ones. This was done by excluding value sets derived using elicitation methods other than the TTO, value sets that were derived from populations other than the general population, and value sets that contained interaction terms, quadratic variables, and/or additional health states,
such as death or unconscious. By using the crosswalk approach, we were able to assess the impact of country-specific EQ-5D value sets for both the 3L and the 5L. Even though evidence suggests that utility values derived using the nonparametric crosswalk approach cannot be used interchangeably with those derived using available 5L value sets, both have previously been found to be associated with similar differences across countries\textsuperscript{19,57}. This suggests that our use of crosswalks as a proxy of actual 5L utility values provides important preliminary evidence that using different country-specific EQ-5D value sets may impact on cost-utility outcomes. Further research using actual 5L utility values is needed to confirm this.

Several limitations are noteworthy as well. First, even though special efforts were made to ensure comparability across countries with regards to the methods they used to elicit population preferences, the impact of using different value sets could also depend on factors other than sociocultural differences, such as a) the applied modelling technique and b) the quality of the underlying data\textsuperscript{57,58}. The way in which these factors impact on cost-utility outcomes should, therefore, be explored in a future study. Second, in both example studies, some cost and EQ-5D descriptive system data were missing. Missing data were handled using multiple imputation under the assumption that they were only related to observed data and not to unobserved data (i.e. Missing At Random). As missing data are common in trial-based economic evaluations and multiple imputation is the recommended approach for dealing with missing data in such studies we do not expect the presence of missing data to have severely biased our results\textsuperscript{59,60}. Also, it is important to mention that the present cost-utility analyses were only intended to illustrate the possible implications of the use of different country-specific value sets and not to serve as a bona fide cost-utility analysis of the interventions under study. For the latter, we refer to previous publications\textsuperscript{22,23}. Third, both example studies included data from two relatively small Dutch RCTs with specific patient populations and relatively few patients presenting more severe EQ-5D health states. Therefore, our findings should be viewed as an important indication that the use of different EQ-5D country-specific value sets might result in different cost-utility outcomes and further research is needed to investigate whether these results are generalizable to other studies, countries, and/or patient groups. \textbf{Fourth, the non-parametric crosswalk of van Hout et al. (2012) is based on self-reported health status of 3L and 5L versions of the EQ-5D in responders of six European countries, which include Denmark, the UK, Italy, the Netherlands, Poland, and Scotland}\textsuperscript{47}. Therefore, it may not adequately capture
differences in 5L utility values that might exist for other countries. To assess the possible impact of this issue, a sensitivity analysis was performed, in which the transition probabilities of van Hout et al. (2012) were multiplied by 5L utility values, instead of 3L utility values, for countries that already have a 5L value set. As the results of this sensitivity analysis were in line with those of the main analysis, we do not expect our reliance on the nonparametric crosswalk to have severely biased our results.

This study indicates that the use of different EQ-5D country-specific value sets has an impact on cost-utility outcomes, for both the 3L and 5L. This indicates that cost-utility outcomes may not be directly transferable across countries. Transferring economic evaluation results across countries is sometimes necessary, because many jurisdictions request information on the cost-effectiveness of interventions, while it is not possible to provide this information for every jurisdiction separately. As country-specific value sets are thought to better reflect sociocultural differences, this does not mean that more generic value sets, such as the European one, are preferred. Instead, researchers are encouraged to conduct sensitivity analyses to assess the impact of using different country-specific value sets on their results and to develop methods for transferring cost-utility outcomes across countries. Healthcare decision-makers, on the other hand, are recommended to interpret cost-utility analysis results from other countries with caution.

Until now, various studies have been performed to assess the transferability of cost-utility outcomes across countries and/or to develop methods for improving their transferability. These studies indicate that various factors influence the transferability of cost-utility outcomes across countries, including not only differences in value sets, but also differences in resource use patterns, unit costs, and baseline risks. Up until now, the majority of research focused on the cost side of the equation. Recently, however, the “ISPOR Good Research Practices Task Force” emphasized the importance of the fact that utilities are not directly transferable across countries as well; something which was confirmed by the present study. Oddersshede et al. (2015) were the first to develop a method for converting Dutch and German 3L utility values into UK ones. This set of analyses, however, only included a limited number of countries and did not account for the severity level of a health state. Therefore, more research into this area is warranted, particularly because the existence of successful methods for transferring cost-utility outcomes across countries could reduce the
number of required studies and ensures that healthcare decisions can be made in a much more timely fashion\textsuperscript{64}.

In conclusion, this set of analyses indicates that the use of different EQ-5D country-specific value sets has an impact on cost-utility outcomes. Therefore, to account for the fact that health state preferences are affected by sociocultural differences, relevant country-specific value sets should be used in cost-utility analyses. Healthcare decision-makers are recommended to interpret cost-utility outcomes from other countries with caution and researchers are encouraged to develop methods for successfully transferring cost-utility outcomes from one country to another.
References


35. Lamers LM. The Transformation of Utilities for Health States Worse Than... : Medical Care. Medical Care. 2007 Mar;45(3):238–44.


63. Oddershede L, Petersen KD. Adjustment of foreign EQ-5D-3L utilities can increase their transferability. CEOR. 2015 Dec;629.
