Health State Utility instruments compared: Inquiring into non-linearity across EQ-5D-5L, SF-6D, HUI3, and 15D

Thor Gamst-Klaussen*, Gang Chen**, Admassu N. Lamu* and Jan Abel Olsen*

*Department of Community Medicine, University of Tromsø, Norway

** Flinders Centre for Innovation in Cancer and Flinders Health Economics Group, Flinders University, Australia

Corresponding author:
thor.klaussen@uit.no
Phone: +47 77 62 07 10

Abstract

Purpose Different health state utility (HSU) instruments produce different utilities for the same individuals, thereby compromising the intended comparability of economic evaluations of health care interventions. When developing crosswalks, previous studies have indicated non-linear relationships. This paper inquires into the degree of non-linearity across the four most widely used HSU-instruments, and proposes exchange rates that differ depending on the severity levels of the health state utility scale.

Method Overall, 7933 respondents from six countries; 1760 in a non-diagnosed *healthy group* and 6173 in seven *disease groups,* reported their health states using four different instruments: EQ-5D-5L, SF-6D, HUI-3 and 15D. Quantile regressions investigate the degree of non-linear relationships between these instruments. To compare the instruments across different disease severities, we split the health state utility scale into utility intervals with 0.2 successive decrements in utility starting from perfect health at 1.00. *Exchange Rates (ERs)* are calculated as the mean utility difference between two utility intervals on one HSU-instrument divided by the difference in mean utility on another HSU-instrument.

Result Quantile regressions reveal significant non-linear relationships across all four HSU-instruments. The degrees of non-linearities differ, with a maximum degree of difference in the coefficients along the health state utility scale of 3.34 when SF-6D is regressed on EQ-5D. At the lower end of the health state utility scale, the exchange rate from SF-6D to EQ-5D is 2.19, whilst at the upper end it is 0.35.

Conclusion Comparisons at different utility levels illustrates the fallacy of using linear functions as crosswalks between HSU-instruments. The existence of non-linear relationships between different HSU-instruments suggests that level-specific exchange rates should be used when converting a change in utility on the instrument used, onto a corresponding utility change had another instrument

been used. Accounting for non-linearities will increase the validity of the comparison for decision makers when faced with a choice between interventions whose calculations of QALY-gains have been based on different HSU-instruments.

Keywords: Health state utility instruments, crosswalks, non-linearity, exchange rates, health-related quality of life, economic evaluation.

Introduction

The rationale for developing the QALY-metric (Quality Adjusted Life Year) was to make health care programmes comparable in terms of their effectiveness in producing health outcomes. The unique feature of the QALY as a measure of health outcomes is that it accounts for both quantity (increased life time) and quality (improved health state). Health states are valued on a scale where the value of being dead must be 0, whilst the upper end represents best imaginable health with a value of 1. To permit comparability and aggregation, the value scale is assumed to have interval scale properties such that a given interval increase at the lower end of the scale is equally valuable as the same interval change at the upper end of the scale [1].

The health state, or the Q in the QALY, is measured by a *generic preference-based* instrument, often referred to as a health state utility (HSU) (or multi-attribute utility) instrument. HSU-instruments consist of a descriptive system, and a scoring algorithm that assign a value – or utility – to each possible combination in the descriptive system. These ‘utilities’ are intended to reflect peoples’ preferences over the relative importance of each health dimension, as well as the trade-offs they are prepared to make between quality of life gains *vs* life-time gains.

The problem is that the health state utility inferred from a respondent depends on which instrument has been applied. Such differences in utilities are caused by i) different dimensions included in the descriptive systems, and; ii) different methods used for assigning preferences to health states (e.g. standard gamble, time-trade-off, visual analogue scale, discrete choice experiment), including the mathematical model used for computing the tariff (e.g. additive, multiplicative) [2, 3]. Hence, the intended comparability across health care programmes is compromised when their QALY calculations have been based on different HSU-instruments.

In the current paper we compare the four most widely used instruments, covering more than 90% of applied studies. A comprehensive review showed that the EQ-5D is the dominant instrument (63.2% of the studies), followed by HUI (2-3) (14.4%), SF-6D (8.8%) and 15D (6.9%)[4]. A recent review confirms the dominant position of the EQ-5D (70%), followed by SF-6D (10%), HUI (5%) and 15D (4%)[5].

Most comparisons of HSU-instruments have included only two instruments[4]. Head-to-head comparisons were identified in 392 studies, suggesting low level of agreement, and hence a need for developing algorithms to compare utility scores across the various instruments used. Crosswalks or 'mappings' between a disease specific instrument and a HSU-instruments are commonly used to enable QALY estimations[6, 7]. However, while such crosswalks between HSU-instruments are rare[8-11], previous studies have indicated non-linear relationships[8, 9, 11]. By the use of item response analysis, Fryback and colleagues investigated the interrelationship of five instruments (EQ-5D-3L, HUI2, HUI3, QWB-SA, SF-6D) in a sample of US adults (N=3844). They concluded that 'simple linear functions may serve as crosswalks among these indexes only for lower health states albeit with low precision. Ceiling effects make crosswalks among most of the indexes ill specified above a certain level of health'[9]. Responses from two HSU-instruments, EQ-5D-3L and SF-6D (converted from SF-36) were compared in a sample of patients with coronary heart disease (N=1493) using quantile regression. An important conclusion was that the 'strength of association between the instruments is different across different parts of the health distribution' [11]. The presence of such non-linear relationships have important implications for health care decision making if crosswalks between instruments are based on linear models, and the health state increments under consideration happen to be at the part of the scales where the real association between the increments is very different from the linear estimation. Insights into such relationships are important when comparing results from studies that have been based on different HSU-instruments.

The current paper aims to: i) investigate into the degree of non-linear relationships across the four most widely used HSU-instruments, and; ii) provide exchange rates (coefficients) that differ depending on which intervals of the scales that are considered. The analysis is based on a unique data set from the Multi Instrument Comparison (MIC) project[12]. Given the dominant role of the EQ-5D in this literature, an important contribution is to use the new 5-level descriptive system (EQ-5D-

5L), and its most recent tariff based on preferences expressed using this descriptive system[13], rather than the crosswalk tariff [14] which was based on preferences expressed through the 3-level descriptive system[15].

Methods

Data

Data is based on an online survey administered in six countries (Australia, Canada, Germany, Norway, UK , US) by a global panel company, CINT Australia Pty Ltd [12]. Respondents were initially asked to rate their overall health on a [0 – 100] visual analogue scale (VAS) (where 0 represents the least desirable health and 100 represents the best possible physical, mental and social health) and indicate if they had any chronic disease. Respondents qualified for the healthy group if they reported no chronic disease and a rating of overall health of at least 70 on the [0 – 100] scale. In each country we sought a demographically representative sample of 300 according to age, gender and education. For each of the seven disease groups (arthritis, asthma, cancer, depression, diabetes, hearing loss, heart diseases), we had a target of 150 respondents. A total of 7933 subjects are included. For further details on respondent recruitment, see Richardson et al., 2012 [12].

Measures

EQ-5D-5L consists of five items/dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, each described along five levels, giving a total of 3125 possible health states. We adopted the new EQ-5D-5L value set for England based on both time-trade-off (TTO) and discrete choice experiment (DCE)[13], which will be referred to as the *new England tariff*. In one descriptive analysis, the interim UK cross-walk tariff was also included[14, 15].

SF-6D (derived from SF-36) is based on 11 items measuring six dimensions: physical functioning, role limitations, social functioning, pain, mental health, and vitality. Each item has between 4-6 levels, giving 18,000 possible health states. Scores are based on standard gamble[16].

HUI-3 includes eight items/dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. Each dimension has five or six levels giving 972,000 possible health states. Utility weights are based on a representative sample of adult Canadians who used the VAS in their valuation exercise, as well as standard gamble[17].

15D consist of 15 items/dimensions: mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort/symptoms, depression, distress, vitality, and sexual activity. Each item has five levels giving more than 30 billion possible health states. The valuation was based on five random samples of the Finnish general population using a variant of the VAS[18].

Analysis

First, descriptive analyses report means, standard deviation (SD), and range. Spearman's rank-order correlations and intra-class correlation coefficients (ICCs) are computed across the four HSU-instruments. Second, *quantile regression models* are used to study the relationship between instruments *i*, and the conditional quantiles of instrument *j* to formally test the existence of non-linearity between two HSU-instruments. Different from the classical ordinary least squares (OLS) regression which focuses on the conditional mean of the dependent variable, quantile regression models the relationship between an independent variable and the conditional quantiles of the dependent variable, which could further reveal the complicated relationship between two HSU-instruments in more detail. (For a theoretical background on quantile regression, see[19]). Wald F-statistics were used to test for equality of coefficients across the quantile regression results. The degree of non-linearity between HSU-instruments is given by dividing the highest coefficient in each estimation by the lowest coefficient, which will be referred to as the *maximum degree of differences in coefficients* (MDDC). F-tests and MDDC are shown for each subsample (disease group) to inquire into variations in the degree of non-linearity across disease groups. Considering the extensive use of the EQ-5D and the growing use of the SF-6D[4, 5], they have the role as the two alternative target instruments on which each of the other instruments' utilities are being converted. Conversions to the HUI-3 and 15D scales are presented in the Appendix.

Lastly, we split the HSU scales into six utility intervals with 0.2 successive decrements in utility starting from perfect health at 1.00 {<0.2, [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), [0.8, 1), 1}, except for SF-6D and 15D which did not include the bottom interval since their minimum utilities were above 0.2. The rationale for the choice of these intervals was to account for distinctly different severity levels: from those expressing perfect health (1.00); mild symptoms [0.8, 1); moderate health problems on the middle of the scale [0.4, 0.6), and; those with extreme problems expressing utilities less than 0.2. In addition, there are the two remaining utility intervals between mild and moderate symptoms [0.6, 0.8) and between moderate and extreme problems [0.2, 0.4).

Scale dependent *exchange rates (ERs)* were developed, and defined as the change in utilities measured using HSU-instrument *i* (ΔU_i) divided by the change in utilities measured using HSU-instrument j (ΔU_j):

$$ER_{ij} = \Delta U_i / \Delta U_j \quad (1)$$

Differences in exchange rates along the scale indicate non-linear associations across instruments. The 95% confidence intervals on each *ER* were calculated using a bootstrap method (i.e. randomly draw 60% sample from the full sample to calculate the *ER* and replicate the procedure 1000 times).

The tool (i.e. the *ER*) presented here differs from the traditional mapping algorithm produced which is usually derived using regression techniques (e.g. OLS) and based on individual-level data (e.g. Chen et al.[8]). The *ER* in this paper is based on a simple calculation relying on aggregate data. The results could be used to help policy-makers and researchers understand the extent of comparability at various severity levels on the health state utility scale.

All statistical analyses were performed in Stata version 13.1 (StataCorp LP), except Figures 2 and 3 which were constructed using SPSS version 21.0 (IBM Corp).

Results

Sample characteristics on age, sex, education and disease groups are provided in the Appendix (Tables A1 and A2). Summary statistics on the four HSU-instruments are reported in Table 1. As for the EQ-5D-5L, the interim crosswalk tariff is included in Table 1 to compare with the new England tariff.

The *mean* utilities in the HSU-instruments differed from 0.71 (SF-6D and HUI-3) to 0.85 (15D), whilst the *median* ranged from 0.70 (SF-6D) to 0.88 (15D). The ceiling effect (i.e. HSU = 1.00) was most evident in EQ-5D (19.3%), followed by HUI-3 (7.2%) and 15D (7.0%). The EQ-5D and HUI-3 scales allow for utilities below zero, and have larger proportions at the bottom end of the scales. Because of the different scale lengths, the differences between health state utilities varied depending on what range we are comparing. The large differences at the 10th percentile imply that the potential utility gain involved from a cure would differ a lot depending on the instrument used.

—Table 1 about here—

The Spearman's rank correlations across HSU-instruments vary from 0.724 (between SF-6D and HUI-3) to 0.841 (between EQ-5D and 15D). The level of absolute agreement measured by the intra-class correlation ICC was the lowest between SF-6D and HUI-3 (ICC=0.591), whilst the highest ICC was between SF-6D and 15D (ICC=0.780).

Figure 1 shows the quantile regression estimates with 95% confidence intervals. If the relationship between each pair of HSU-instruments were linear, quantile regression result would show a constant coefficient across all quantiles (i.e. horizontal lines). However, the non-linear relationship is evident. F-tests rejected the null hypothesis of the equality of coefficients across quantiles. The 95% confidence intervals tell whether the estimated coefficients are statistically significant at 5% criteria. Additionally, OLS estimates were included to illustrate the difference when using the two approaches. Note, however, that the scale on the vertical axis is different across illustrations in the Figure 1.

As an example, take regressing 15D utility onto EQ-5D utility (i.e. top left of Figure 1): One unit change on 15D leads to a 1.8 unit change on EQ-5D at the 0.1 quantile, whilst at the 0.9 quantile, one unit change on 15D was associated with 0.7 unit change on EQ-5D ($p<0.01$). This pattern is similar when the two other HSU-instruments are regressed on EQ-5D, with largest impact at the 0.1 quantile, then decreasing. When SF-6D is the dependent variable (right side of Figure 1), the impact of the independent variable is associated with the smallest change at the 0.1 quantile while the largest impact is indicated at the 0.6 and 0.7 quantiles. The slight decline beyond the 0.7th quantile in Figure 1 is due to the fact that SF-6D generally has lower values at the upper end of the scale, and a much lower ceiling effect (1.4% in Table 1) as compared to the other instruments. Corresponding result when using HUI-3 and 15D as dependent variables are provided in the Appendix (Figure A1).

--Figure 1 about here--

The non-linearity was also investigated by Wald F-test across quantiles with EQ-5D and SF-6D as dependent variable within each disease group (Table 2). All tests indicate significant difference ($p<0.01$) between coefficients, except when regressing on SF-6D in the depression group with HUI-3 and 15D as independent variables.

To investigate the degree of non-linearity across full sample and subsamples, we calculated the ratio between the highest and the lowest coefficients, referred to as the *maximum degree of difference in the coefficients* (MDDC) at each estimation. For example, as illustrated in Figure 1, when regressing SF-6D onto EQ-5D, the highest coefficient is at the 0.1 quantile and the lowest at the 0.9 quantile (MDDC = 1.37/0.41 = 3.34). On the contrary, when regressing EQ-5D onto SF-6D the coefficient is lowest at the 0.1 quantile and highest on the 0.7 quantile (MDDC = 0.60/0.46 = 1.26). For the full sample (first column in Table 2), the degree of non-linearity is smaller when regressing onto SF-6D (MDDC; 1.26-1.32) than on EQ-5D (MDDC; 2.49-3.34). Similar results are found within the disease groups.

—Table 2 about here—

Figure 2 illustrates the relationships across HSU-instruments. An important difference between Figure 2 and Figure 1 (the quantile regression results) is that in Figure 2 respondents were divided into five or six intervals on the health state utility scale depending on the utility scores within each instrument. In contrast, quantile regression coefficients depicted in Figure 1a and b demonstrate the relationships across HSU-instruments at nine different quantiles (as represented on the horizontal axes) of EQ-5D-5L and SF-6D, respectively. In Figure 2, the horizontal axis is changed to the HSU *utility intervals*. For example: the horizontal axis in Figure 2 (a) represents the five 15D utility intervals (the bottom interval starts at < 0.4 since there were no 15D utilities lower than 0.25). Then, the mean utilities for EQ-5D and SF-6D respectively were plotted accordingly. Note that the 15D produced the highest scores (except for the lowest interval). SF-6D produced the lowest scores at the top three 15D intervals, whilst EQ-5D produces the lowest scores for the two lowest 15D intervals. At the top end of the scale the 15D and the EQ-5D run parallel, indicating an exchange rate of 1.00. The slope differences along the scale are larger between 15D vs SF-6D, than between 15D vs EQ-5D. Figure 2 (b) shows the *six* HUI-3 intervals (now the bottom interval starts at < 0.2 since HUI-3 utilities drops below zero) against the group means of the other HSU-instruments. Figure 2 (c and d) shows the five SF-6D and six EQ-5D utility intervals, respectively.

—Figure 2 about here—

Table 3-6 provides the same information as contained in Figure 2, in that the exchange rates are illustrated by the slopes of the curves at different intervals. The first HSU-instrument given in the table (15D in Table 3) represents the *source instrument*, while the subsequent SF-6D and/or EQ-5D represent the *target instrument*(s) for which corresponding mean utilities are computed within each interval, and to which exchange rates are derived. Given the dominant role of the EQ-5D, and the increasing use of the SF-6D [3,4], we consider these two instruments to act as the currencies onto which other instruments' utilities are being converted. In Table 3, for example, the mean utilities (U) within the five 15D utility intervals are computed, together with the corresponding EQ-5D, as well as SF-6D, based utilities. The *ER*s between each utility interval on the 15D scale tells what a 15D increment has to be multiplied with to get the corresponding change in utility had the EQ-5D been applied (or the SF-6D). As an example, take the *ER*s between 15D and EQ-5D: If 15D had been applied on a study where the patient group at baseline are found in the [0.4 – 0.6] interval with a mean utility of 0.53, and after treatment they are in the [0.6 – 0.8] interval with a mean utility of 0.72, it represents an increase of ΔU_{15D} = 0.19 on the 15D utility scale. The corresponding increment on the EQ-5D scale would be from 0.38 to 0.67 ($\Delta U_{EQ\text{-}5D}$ =0.29). Hence, the 15D increment has to be multiplied with an *exchange rate* = 1.51 ($\Delta U_{EQ\text{-}5D} / \Delta U_{15D}$) to make the utility increment comparable to a situation when the study under consideration had applied EQ-5D rather than 15D. Table 4 provides similar result if HUI-3 had been the source instrument and were to be converted to corresponding changes on the EQ-5D or SF-6D scales. Lastly, Table 5 presents exchange rates when converting SF-6D utilities onto the EQ-5D scale, and vice versa in Table 6.

—Table 3-6 about here—

Discussion

This paper confirms earlier findings that the health state utility a patient reports depends on which instrument has been applied. Our comprehensive data set of 8,000 subjects from six countries, who expressed their health state utilities along several instruments, have revealed large differences across the instruments. We have shown that the degree of such differences in health state utilities is *scale dependent*, i.e. they depend on which interval on the health state utility scale that is considered. These non-linear associations are important to take into account when comparing health care programmes whose effectiveness have been measured by different HSU-instruments.

For the most widely used instrument, the EQ-5D, quantile regressions reveal strong non-linear relationships with any of the other HSU-instruments. This finding is consistent across the seven disease groups. As for SF-6D, the degree of non-linearity is less than for EQ-5D, still the non-linearity are replicated across all disease groups with only one exception (Figure 1 and Table 2).

These findings corroborate with other studies that have indicated non-linear relationships across HSU-instruments [8, 9, 11]. Scaling effect could be key factor to explain the observed non-linear relationship [3, 20]. We confirm earlier finding by Seymour et al. [11] that the effect of EQ-5D differ at different parts of the SF-6D distribution. However, we extend the analysis to reveal that the associations differ between the four most widely used HSU-instruments across different parts of the distribution of each instrument.

As for the calculated exchange rates, the result indicate that transformation of changes in utilities across HSU-instruments can vary considerably depending on what interval of the scale that are considered (Table 3-6). These results further reveal the non-linear relationships and the consequences of applying OLS regression when estimating crosswalks. For instance, comparing the quantile regression estimates to the OLS estimates reveals that the use of OLS will produce poor estimates by over-predicting the EQ-5D utility and under-predicting the SF-6D utility for respondents in poor health. The opposite are observed for respondents in moderate to good health by under-predicting the EQ-5D utility and over-predicting the SF-6D utility.

Decision makers would have a problem comparing QALY gains from studies using different instruments. Therefore, the exchange rates could be used to increase comparability. Consider the following study based on HUI-3 [21], in which the disease state 'Liver transplantation, first year' was associated with a utility of 0.5, while 'Liver transplantation, subsequent years' was assigned a utility of 0.7. In Table 4, we see that the exchange rate at this part of the scale when converting from HUI-3 onto EQ-5D is 0.59. Hence, the increased HUI-3 utility from 0.5 to 0.7 can be converted to the EQ-5D and SF-6D scales by applying their respective exchange rates at this part of the scale; for EQ-5D the *ER* is 0.59, which multiplied by the ΔU_{HUI-3} of 0.2 gives a EQ-5D gain = 0.118. As for the SF-6D, the exchange rate from HUI-3 at this part of the scale is 0.31, which implies a corresponding SF-6D gain 0.31*0.2 = 0.062. Hence, for institutional bodies that apply the EQ-5D and/or the SF-6D as reference cases (e.g. NICE in the UK), our scale dependent exchange rates could be applied when comparing the cost-effectiveness of competing programmes whose evaluations have been based on different HSU-instruments, and whose utility gains happen to take place at different parts of the health state utility scales.

While our extensive data set is unique, some study limitations should be acknowledged. First, there is a general concern regarding the quality of online data. However, a series of data cleaning procedures were performed to guarantee the quality (i.e. repetition of items to check consistency in responses throughout the survey) [14]. Second, as for possible lack of representativeness this should not be a problem in the current context, since studying the relationship between different instruments does not require a representative sample of the population. Third, it would be ideal to have a longitudinal dataset and further study the exchange rates on incremental utility using individual level data. However, it is out of the scope with the current data set. A strength of the current study is that it includes dissimilar patient groups on which sub-group analyses could be performed. Fourth, the EQ-5D-5L utilities are based on a UK value set that has yet to be published in a peer-reviewed journal.

To conclude, when comparing health state utilities produced by different instruments, we observe clear non-linear relationships, implying that the same exchange rate should not be applied across all levels of the scale. This paper suggests level-specific exchange rates that can be used to convert a change in utility on a given instrument onto a corresponding utility change on another instrument. When health care decision makers compare alternative programmes whose QALY-calculations have been based on different HSU-instruments, accounting for the non-linear relationships will increase the validity of these comparisons.

References

1. Weinstein, M. C., Torrance, G., & McGuire, A. (2009). QALYs: The Basics. *Value in Health*, *12*, S5-S9, doi:10.1111/j.1524-4733.2009.00515.x.

2. Brazier, J. (2007). *Measuring and Valuing Health Benefits for Economic Evaluation* (Vol.). New York: OUP Oxford.

3. Richardson, J., Iezzi, A., & Khan, M. (2015). Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility' effects. *Quality of Life Research*, 1-9, doi:10.1007/s11136-015-0926-6.

4. Richardson, J., McKie, J., & Bariola, E. Multi attribute utility instruments and their use In C. AJ (Ed.), *Encyclopedia of health economics* (pp. 341-357,Vol. San Diego: Elsevier Science.

5. Wisloff, T., Hagen, G., Hamidi, V., Movik, E., Klemp, M., & Olsen, J. A. (2014). Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. *Pharmacoeconomics*, *32*(4), 367-375, doi:10.1007/s40273-014-0136-z.

6. Brazier, J. E., Yang, Y., Tsuchiya, A., & Rowen, D. L. (2010). A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*, *11*(2), 215-225, doi:10.1007/s10198-009-0168-z.

7. Mortimer, D., & Segal, L. (2008). Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. *Med Decis Making*, *28*(1), 66-89, doi:10.1177/0272989x07309642.

8. Chen, G., Khan, M. A., Iezzi, A., Ratcliffe, J., & Richardson, J. (2015). Mapping between 6 Multiattribute Utility Instruments. *Med Decis Making*, doi:10.1177/0272989x15578127.

9. Fryback, D. G., Palta, M., Cherepanov, D., Bolt, D., & Kim, J. S. (2010). Comparison of 5 health-related quality-of-life indexes using item response theory analysis. *Med Decis Making*, *30*(1), 5-15, doi:10.1177/0272989x09347016.

10. Rowen, D., Brazier, J., Tsuchiya, A., & Alava, M. H. (2012). Valuing states from multiple measures on the same visual analogue sale: a feasibility study. *Health Econ*, *21*(6), 715-729, doi:10.1002/hec.1740.

11. Seymour, J., McNamee, P., Scott, A., & Tinelli, M. (2010). Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Econ*, *19*(6), 683-696, doi:10.1002/hec.1505.

12. Richardson, J., Kahn, M., Lezzi, A., & Maxwell, A. (2012). Cross-national comparison of twelve quality of life instruments: MIC Paper 1: Background, questions, instruments. Research paper 76. http://www.aqol.com.au/papers/researchpaper76.pdf. Accessed 14 May 2014.

13. Devlin, N., & van Hout, B. (2014). An EQ-5D-5L value set for England. Office of Health Economics (OHE), London. http://www.slideshare.net/ScHARR-HEDS/ohe-seminar-5-l-value-set-oct-30-2014-final-version-031114-1. Accessed 20 Jan 2015.

14. van Hout, B., Janssen, M. F., Feng, Y. S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P., & Pickard, A. S. (2012). Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*, *15*(5), 708-715, doi:http://10.1016/j.jval.2012.02.008.

15. Dolan, P. (1997). Modeling valuations for EuroQol health states. *Med Care*, *35*(11), 1095-1108.

16. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, *21*(2), 271-292, doi:http://dx.doi.org/10.1016/S0167-6296(01)00130-8.

17. Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., Denton, M., & Boyle, M. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care*, *40*(2), 113-128.

18. Sintonen, H., & Pekurinen, M. (1993). A fifteen-dimensional measure of health-related quality of life (15D) and its applications. In S. Walker & R. Rosser (Eds.), *Quality of Life Assessment: Key Issues in the 1990s* (pp. 185-195,Vol.: Springer Netherlands.

19. Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *Journal of Economic Perspectives*, *15*(4), 143-156, doi:doi: 10.1257/jep.15.4.143.

20. Chen, G., Flynn, T., Stevens, K., Brazier, J., Huynh, E., Sawyer, M., Roberts, R., & Ratcliffe, J. (2015). Assessing the Health-Related Quality of Life of Australian Adolescents: An Empirical Comparison of the Child Health Utility 9D and EQ-5D-Y Instruments. *Value Health*, *18*(4), 432-438, doi:10.1016/j.jval.2015.02.014.

21. Gheorghe, L., & Baculea, S. (2010). Cost-effectiveness of peginterferon alpha-2a and peginterferon alpha-2b combination regimens in genotype-1 naive patients with chronic hepatitis C. *Hepatogastroenterology*, *57*(101), 939-944.

Table 1 Summary statistics (N=7933)

	EQ-5D		SF-6D	HUI3	15D
	NET	CWT			
Mean	0.82	0.74	0.71	0.71	0.85
SD	0.19	0.22	0.14	0.26	0.13
Min	-0.21	-0.51	0.30	-0.34	0.25
10^{th} percentile	0.55	0.43	0.54	0.31	0.67
25^{th} percentile	0.76	0.66	0.61	0.58	0.78
Median	0.87	0.77	0.70	0.79	0.88
75^{th} percentile	0.94	0.88	0.83	0.92	0.95
90^{th} percentile	1.00	1.00	0.89	0.97	0.99
Max	1.00	1.00	1.00	1.00	1.00
Ceiling [HSU = 1 (in %)]	19.3	19.3	1.4	7.2	7.0
Bottom [HSUU < 0.40 (in %)]	4.8	8.7	1.2	13.7	0.3

Note: NET, New England tariff; CWT, Crosswalk tariff; and SD, standard deviation.

Table 2 Coefficient difference test (Wald test) across quantile regression results (quantiles; .10, .20, .30, .40, .50, .60, .70, .80, .90) for EQ-5D-5L and SF-6D entered one-by-one as independent variables.

	FS $F_{(8, 7930)}$	MDDC	Depr $F_{(8, 915)}$	MDDC	Diabetes $F_{(8, 922)}$	MDDC	Arthritis $F_{(8, 927)}$	MDDC	Asthma $F_{(8, 854)}$	MDDC	Cancer $F_{(8, 770)}$	MDDC	Hearing $F_{(8, 829)}$	MDDC	Heart $F_{(8, 941)}$	MDDC
EQ-5D-5L																
SF-6D	344.40**	3.34	32.23**	2.07	48.43**	3.73	57.09**	2.42	31.25**	3.34	41.84**	3.02	28.19**	3.23	30.91**	3.39
HUI3	586.46**	2.90	29.29**	2.53	60.39**	2.40	150.07**	2.53	18.51**	2.58	51.42**	2.87	33.11**	3.53	57.97**	2.44
15D	282.36**	2.49	25.77**	2.05	77.40**	2.52	107.82**	2.51	45.94**	2.25	43.03**	2.49	11.96**	2.49	30.00**	2.35
SF-6D																
EQ-5D	35.34**	1.26	2.31*	1.21	7.13**	1.30	3.54**	1.17	2.82*	1.24	11.74**	1.35	5.80**	1.24	4.19**	1.23
HUI3	89.17**	1.32	1.79	1.22	10.98**	1.28	30.21**	1.40	9.85**	1.40	19.96**	1.39	2.79*	1.23	17.96**	1.36
15D	77.81**	1.31	1.75	1.21	13.05**	1.27	7.27**	1.48	18.40**	1.48	13.08**	1.60	8.52**	1.20	5.12**	1.30

*p<0.05. **p<0.01.

Note: FS, full sample (N=7933); Depr, Depression; and MDDC, the maximum degree of difference in the coefficients (i.e. the ratio of the highest to the lowest coefficients).

Table 3 Exchange rates at different utility ranges for transformations of (mean) utility changes measured using 15D onto (mean) utility changes on EQ-5D-5L or SF-6D

Utility range (15D)	N	**15D**	**EQ-5D-5L**		**SF-6D**	
		U [95% CI]	U [95% CI]	ER [95% CI]	U [95% CI]	ER [95% CI]
1.00	555	1.00 [1.00, 1.00]	0.99 [0.99, 0.99]		0.87 [0.86, 0.88]	
				1.06 [1.05, 1.06]		**1.25 [1.25, 1.25]**
[0.8, 1.0)	5106	0.91 [0.91, 0.91]	0.90 [0.89, 0.90]		0.76 [0.75, 0.76]	
				1.21 [1.21, 1.21]		**0.89 [0.89, 0.89]**
[0.6, 0.8)	1876	0.72 [0.72, 0.72]	0.67 [0.66, 0.67]		0.59 [0.59, 0.59]	
				1.51 [1.51, 1.51]		**0.56 [0.56, 0.56]**
[0.4, 0.6)	372	0.53 [0.53, 0.54]	0.38 [0.36, 0.41]		0.48 [0.48, 0.49]	
				1.41 [1.39, 1.42]		**0.29 [0.28, 0.30]**
[0.2, 0.4)	21	0.35 [0.34, 0.37]	0.13 [0.01, 0.25]		0.43 [0.38, 0.48]	
OLS coef.	7930			***1.26***		***0.85***

Note: U, mean utility corresponding to each utility range; CI, Confidence intervals; ER, Exchange rates; and OLS, Ordinary least square coefficients. ER is calculated as change in mean utilities measured using instrument i (EQ-5D-5L or SF-6D) divided by change in mean utilities measured using instrument j (15D); i.e. $\Delta U_{EQ-5D} / \Delta U_{15D},\ and\ \Delta U_{SF-6D} / \Delta U_{15D}$, respectively. CIs for the ERs were calculated using a bootstrap method (with 1000 replications).

Table 4 Exchange rates at different utility ranges for transformations of (mean) utility changes measured using HUI-3 onto (mean) utility changes on EQ-5D-5L or SF-6D

Utility range (HUI-3)	N	**HUI-3**	**EQ-5D-5L**		**SF-6D**	
		U [95% CI]	U [95% CI]	ER [95% CI]	U [95% CI]	ER [95% CI]
1.00	572	1.00 [1.00, 1.00]	0.98 [0.98, 0.99]		0.85 [0.84, 0.85]	
				0.61 [0.61, 0.61]		**0.60 [0.59, 0.60]**
[0.8, 1.0)	3320	0.90 [0.90, 0.90]	0.92 [0.92, 0.93]		0.79 [0.78, 0.79]	
				0.49 [0.49, 0.49]		**0.53 [0.53, 0.53]**
[0.6, 0.8)	1931	0.71 [0.71, 0.72]	0.83 [0.82, 0.83]		0.69 [0.68, 0.69]	
				0.59 [0.59, 0.59]		**0.31 [0.31, 0.31]**
[0.4, 0.6)	1021	0.51 [0.51, 0.51]	0.71 [0.70, 0.72]		0.62 [0.62, 0.63]	
				0.57 [0.56, 0.57]		**0.29 [0.29, 0.29]**
[0.2, 0.4)	571	0.31 [0.31, 0.32]	0.60 [0.59, 0.62]		0.57 [0.56, 0.57]	
				0.72 [0.72, 0.72]		**0.28 [0.28, 0.28]**
<0.2	515	0.05 [0.04, 0.06]	0.41 [0.39, 0.43]		0.49 [0.49, 0.50]	
OLS coef.	7930			***0.58***		***0.37***

Note: U, mean utility corresponding to each utility range; CI, Confidence intervals; ER, Exchange rates; and OLS, Ordinary least square coefficients. ER is calculated as change in mean utilities measured using instrument i (SF-6D) divided by change in mean utilities measured using instrument j (HUI-3); i.e. $\Delta U_{SF-6D} / \Delta U_{HUI-3}$. CIs for the ERs were calculated using a bootstrap method (with 1000 replications).

Table 5 Exchange rates at different utility ranges for transformations of (mean) utility changes measured using SF-6D into onto (mean) utility changes on EQ-5D-5L

Utility range (SF-6D)	N	**SF-6D**	**EQ-5D-5L**	
		U [95% CI]	U [95% CI]	ER [95% CI]
1.00	107	1.00 [1.00, 1.00]	1.00 [1.00, 1.00]	
				0.35 [0.35, 0.35]
[0.8, 1.0)	2326	0.87 [0.87, 0.87]	0.95 [0.95, 0.95]	
				0.58 [0.58, 0.58]
[0.6, 0.8)	3755	0.69 [0.69, 0.69]	0.85 [0.85, 0.85]	
				1.62 [1.61, 1.62]
[0.4, 0.6)	1643	0.54 [0.54, 0.54]	0.60 [0.59, 0.61]	
				2.19 [2.18, 2.20]
[0.2, 0.4)	99	0.36 [0.36, 0.37]	0.22 [0.18, 0.27]	
OLS coef.	7930			***1.04***

Note: U, mean utility corresponding to each utility range; CI, Confidence intervals; ER, Exchange rates; and OLS, Ordinary least square coefficients. ER is calculated as change in mean utilities measured using instrument i (EQ-5D) divided by change in mean utilities measured using instrument j (SF-6D); i.e. $\Delta U_{EQ-5D} / \Delta U_{SF-6D}$. CIs for the ERs were calculated using a bootstrap method (with 1000 replications).

Table 6 Exchange rates at different utility ranges for transformations of (mean) utility changes measured using EQ-5D-5L onto (mean) utility changes on SF-6D

Utility range (EQ-5D-5L)	N	**EQ-5D-5L**	**SF-6D**	
		U [95% CI]	U [95% CI]	ER [95% CI]
1.00	1529	1.00 [1.00, 1.00]	0.85 [0.85, 0.86]	
				1.02 [1.02, 1.02]
[0.8, 1.0)	3968	0.89 [0.89, 0.89]	0.74 [0.74, 0.74]	
				0.72 [0.72, 0.72]
[0.6, 0.8)	1488	0.72 [0.72, 0.72]	0.62 [0.61, 0.62]	
				0.34 [0.34, 0.35]
[0.4, 0.6)	564	0.52 [0.51, 0.52]	0.55 [0.54, 0.55]	
				0.26 [0.26, 0.26]
[0.2, 0.4)	270	0.31 [0.30, 0.31]	0.49 [0.48, 0.50]	
				0.29 [0.29, 0.29]
<0.2	111	0.07 [0.05, 0.09]	0.42 [0.41, 0.44]	
OLS coef.	7930			***0.54***

Note. U, mean utility corresponding to each utility range; CI, Confidence intervals; ER, Exchange rates; and OLS, Ordinary least square coefficients. ER is calculated as change in mean utilities measured using instrument i (SF-6D) divided by change in mean utilities measured using instrument j (EQ-5D); i.e. $\Delta U_{SF-6D} / \Delta U_{EQ-5D}$. CIs for the ERs were calculated using a bootstrap method (with 1000 replications).

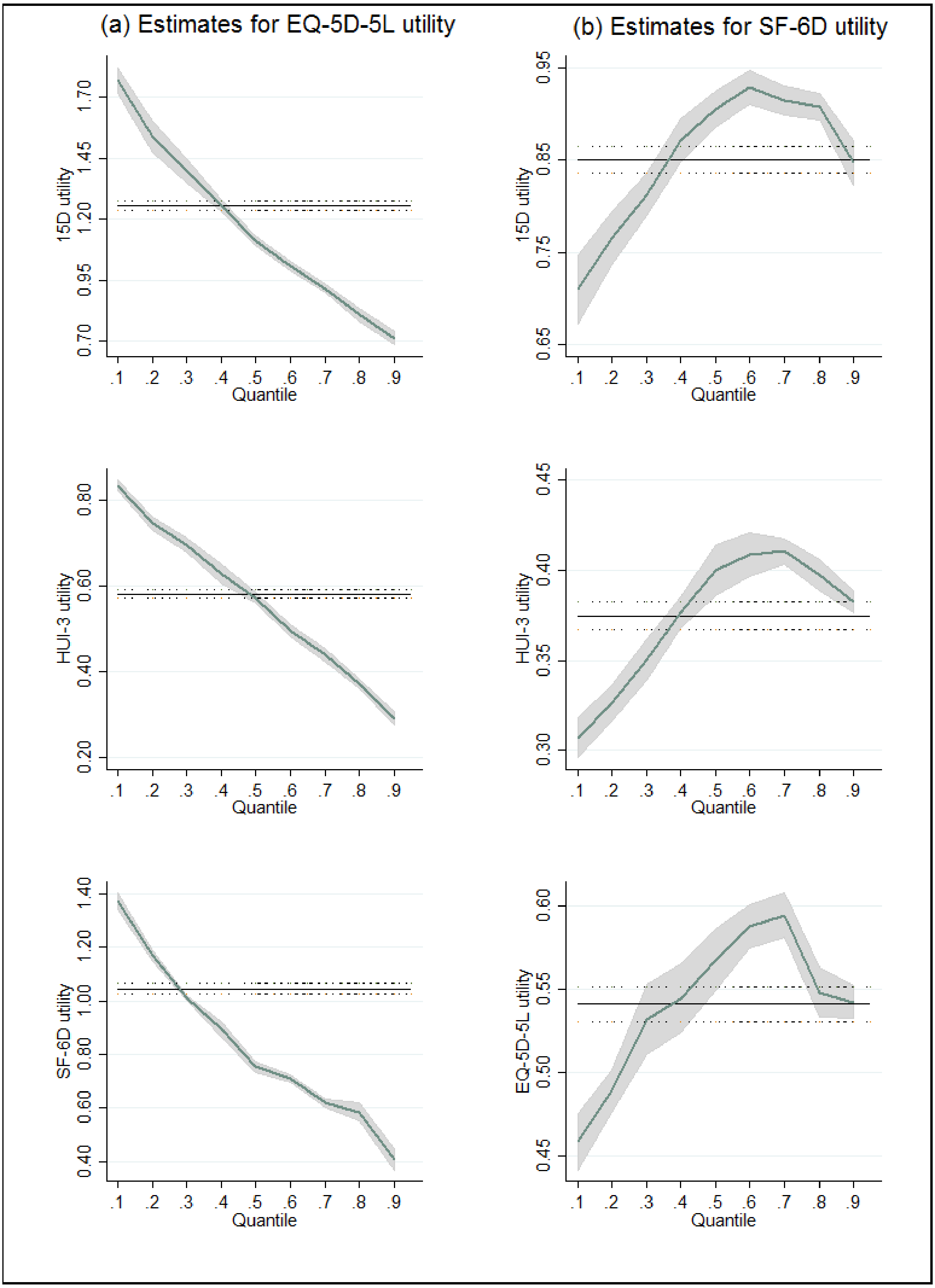

Fig 1 Quantile regression estimates for EQ-5D-5L and SF-6D (vertical axes represent estimates of named independent variable; horizontal axes show the quantiles of the dependent variable, EQ-5D-5L for the first column and SF-6D for the second column). Quantile regression estimates with 95% confidence intervals shaded. OLS regression estimates indicated by solid horizontal lines with 95% confidence intervals indicated by stippled lines. Note that the scales on the vertical axes differ.

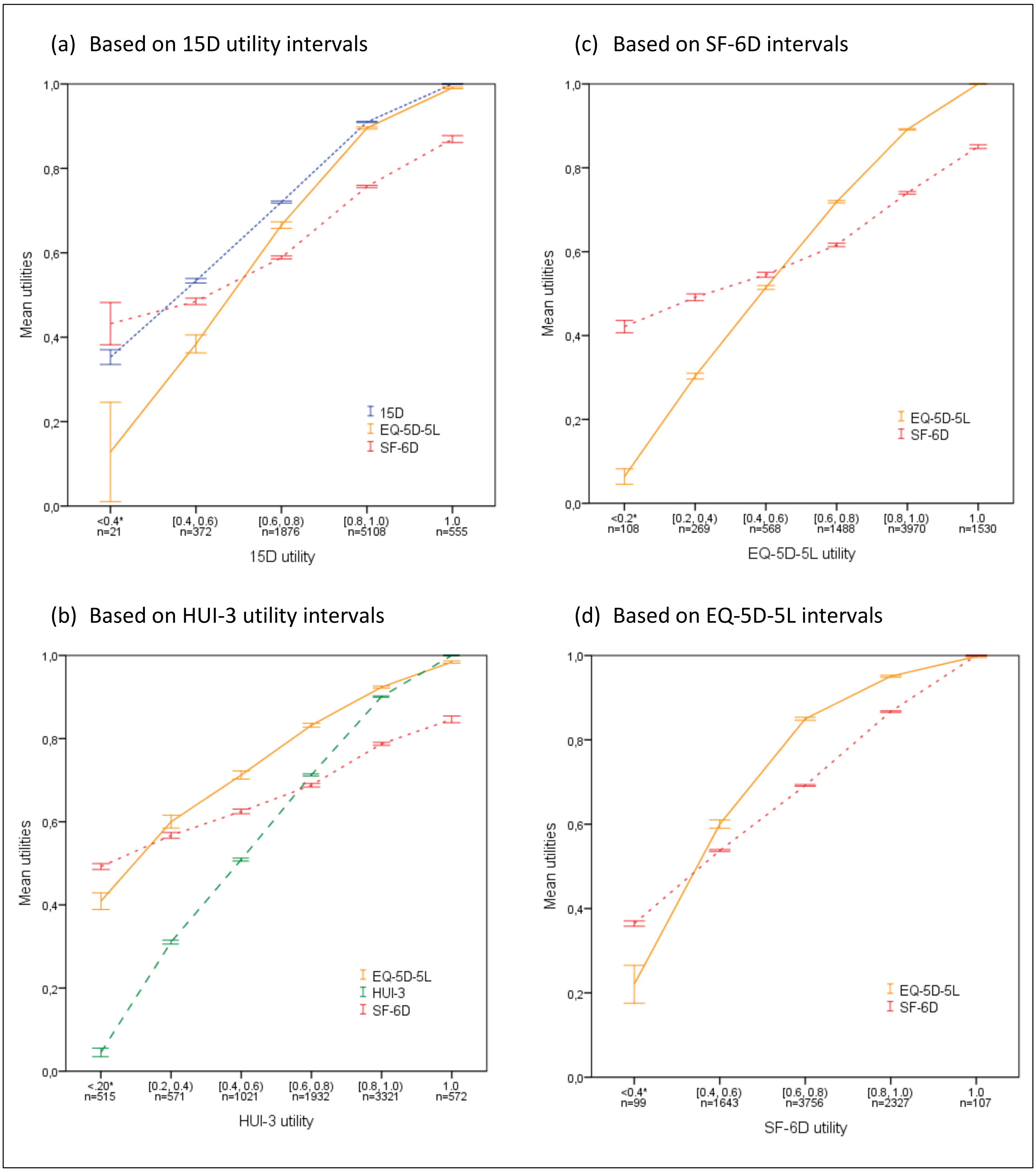

Fig 2 Comparison of mean HSU instruments at different utility intervals.
Note: Solid (orange) line is EQ-5D-5L; dashed (red) line is SF-6D; dotted (blue) line is 15D; and long-dash (green) line is HUI-3. The error bars depict 95% confidence intervals.