

EQ5D-3L Data Quality issues

Monica Hernandez, Steve Pudney, Allan Wailoo

9th Jan 2019

1. Data Quality

Sample size and coverage

3L had a larger sample size than 5L (3,395 vs 1,000). Valuations from TTO covered 42 states (17.7%) of the 243 (3^5) health states. For 5L the TTO covers 2.75% of the 5L health states. The DC experiments (pairwise choices) cover 12.54% of the health states (less than 0.01% of logically possible pairwise choices).

For 3L, the 42 health states were valued 856 times on average (median 751) and the minimum number of times was 723. For 5L, the average was 116 (median 99) and the minimum number of times was 79 in the TTO data.

Data Quality

Dolan et al (1994) report some aspects of data quality. They define inconsistency according to the same “strong” version that is used in the EEPRU report (see page 37 of Dolan et al 1994). Reported values are as follows: for the ranking exercise, 56% recorded no inconsistencies (implying 44% had one or more inconsistencies). For the VAS valuation exercise 57.4% had no inconsistency (implying 42.6% had one or more). Based on a sample size of 3347, which comprises the original sample less 48 respondents whose TTO data were excluded for this analysis, Dolan et al report that for the TTO, 28.2% of respondents had no inconsistency at all (implying 71.8% had one or more inconsistency). The equivalent figure for the 5L dataset, after sample deletions, is 92%. Dolan also reports that 5% had an inconsistency rate greater than or equal to 20%. 47% of respondents had inconsistencies in more than 20% of tasks in the 5L dataset.

We calculated data issues for the 3L using the same criteria we applied for the 5L. These are shown in Table 1 below. It should be noted that our calculations do not align with those reported by Dolan et al. Our analysis relates to the 2997 respondents whose data were used in the estimation of the UK Value Set, as reported in Dolan (1997). The table shows that the proportion of respondents that make one or more inconsistent choices from the set that they are asked to value using TTO is extremely high and similar to the 5L data. However, the proportion of respondents making multiple inconsistent choices is substantially lower. Half the proportion of 3L respondents rank 20% or more of the health states inconsistently compared to 5L respondents (24.4% vs 47.4%). This is also the case when considering the weaker version of inconsistency favoured by Devlin et al (4.5% vs 9%).

We found that there were fewer problematic responses of all other types in the 3L sample than for the 5L.

Table 1: Anomalous responses EQ5D-5L and EQ5D-3L, TTO data.

Anomalous type	outcome	Number of Individuals 5L (3L)	% of sampled individuals			
			Original TTO data ($n = 1,000$)	After sample deletions ¹ ($n = 912$)	After deletions and special treatment ² ($n = 604$)	3L ($n=2997$)
(1)	All individual's TTO trials result in same value	23 (13)	2.30%	0%	0%	0.38%
(2)	Individual reports at least 1 non-55555/33333 trial with same or lower value than trial of 55555/33333	668 (1458)	66.80%	63.80%	49.80%	48.65%
(3)	Individual reports at least 1 non-55555/33333 trial with strictly lower value than trial of 55555/33333	289 (780)	28.90%	26.10%	28.60%	26.03%
(4)	Individual reports fewer than 5 distinct values	309 (195)	30.90%	26.30%	20.90%	6.51%
(5)	Individual reports mild trial (1-point difference from 11111) with same or lower value as trial of 55555/33333	84 (58)	8.40%	0%	0%	1.94%
(6)	Individual reports values $T = 0, 10$ or 20 (-39, 0 or 10) in every trial	41 (22)	4.10%	2.70%	0%	0.73%
(7)	Individual reports all ten/twelve trial values T as multiple of 5 years	63 (24)	6.30%	4.20%	0.50%	0.80%
(8)	Individual gives only integer values for T	362 (63)	36.20%	35.10%	31.10%	2.10%
(9)	'Seam' outcome of $T = 10$ (0) in at least two trials with no outcome below 10 (0)	164 (54)	16.40%	16.40%	0%	1.80%
(10a)	Individual with any inconsistencies between the logical ordering of health states and the TTO valuation	922 (2,832)	92.20%	91.50%	88.40%	94.49%
(10b)	Individual with inconsistencies in more than 20% of tasks between the logical ordering of health states and TTO valuation	518 (732)	51.80%	47.40%	39.10%	24.42%

Individual displays any of anomalies (1), (3), (4) or (5)	520 (945)	52.00%	47.60%	44.20%	31.53%
Individual displays any of anomalies (1), (3), (4), (5), (7), (8) or (9)	711 (1019)	71.10%	68.40%	60.90%	34.00%
Individual displays any of anomalies (1), (3), (4), (5), (7), (8), (9) or (10b)	769 (1295)	76.90%	74.80%	67.40%	43.21%
Individual displays any of anomalies (1), (3), (4), (5), (7), (8), (9) or (10a)	940 (2834)	94.00%	93.40%	91.10%	94.56%

Comparisons between the quality of data obtained from respondents to the 3L and the 5L is further complicated by the fact that they were asked to complete different numbers and types of TTO tasks. 3L respondents valued 12 health states each, which allows 66 possible pairwise comparisons. 5L respondents valued 10 health states, allowing 45 pairwise comparisons. Furthermore, the design of the 3L valuation tasks included many more states that had a logical ordering than the 5L design did. The median number of pairwise comparisons of health states with a logical ordering each respondent faced was 34 (range 26 to 50) for the 3L (see Appendix Table 1). The median number for the 5L TTO tasks was 18 (range 16 to 26) (see Appendix Table 2). Therefore, 3L participants had the opportunity to make many more inconsistent choices than 5L respondents.

The proportion of inconsistent responses is lower for respondents in the 3L TTO experiments than the 5L TTO experiments, as shown in Table 2.

Table 2: Proportion of inconsistent responses made by respondents to the 3L and 5L TTO tasks

	3L	5L
Mean	0.15	0.26
Median	0.11	0.2
25 th Percentile	0.06	0.1
75 th Percentile	0.19	0.38

2. Summary

The 3L value set is based on direct observation of a greater proportion of the health states that comprise the descriptive system than 5L.

Those states that were valued were done so using TTO alone rather a mix of TTO and DCE methods. TTO provides more information than DCE because the latter is only a pairwise comparison for each task.

The sample size for the 3L was more than 3 times larger than the 5L. Each health state was, on average, valued over 7 times more often in the 3L valuation study than in the 5L valuation study.

Consequently, the 3L value set relies on far less extrapolation of values for those health states that were not valued than is the case for the 5L value set.

The design of the 3L TTO tasks meant that respondents valued more health states and more of those health states had a logical ordering. The mean proportion of inconsistent responses is lower in the 3L TTO data than the 5L (0.15 vs 0.26).

There are very high proportions of respondents that make at least 1 logical inconsistency across both the 3L and 5L data (the proportion is slightly higher for 3L compared to 5L). There is a much lower proportion of respondents that make multiple inconsistent valuations of health states in the 3L data than in the 5L data.

The proportions of respondents generating potentially problematic responses of all other types is lower for 3L than for 5L.

Refs:

Dolan P, Gudex C, Kind P, Williams A. (1994) "The measurement and valuation of health. First Report on the main survey", available at <https://www.york.ac.uk/media/che/documents/reports/MVH%20First%20Report.pdf> (last accessed 10th Dec 2018).

Dolan P, Gudex C, Kind P, Williams A. (1995) "A Social Tariff for EuroQoL: Results from a UK Population survey", York CHE Discussion paper 138 available at: <https://www.york.ac.uk/che/pdf/DP138.pdf> (last accessed 10th December 2018)

Dolan P. (1993) "Modeling Valuations for EuroQoL Health States", *Medical Care*, Vol. 35:1095-1108.

Appendix Tables:

Appendix Table 1: Number of pairwise comparisons per block of 12 EQ5D-3L health states with a logical ordering

Total number of comparisons with a logical ordering	Freq.	Percent	Cum.
26	7	0.23	0.23
27	26	0.87	1.1
28	52	1.74	2.84
29	86	2.87	5.71
30	132	4.4	10.11
31	221	7.37	17.48
32	257	8.58	26.06
33	266	8.88	34.93
34	274	9.14	44.08
35	304	10.14	54.22
36	289	9.64	63.86
37	262	8.74	72.61
38	227	7.57	80.18
39	164	5.47	85.65
40	136	4.54	90.19
41	101	3.37	93.56
42	67	2.24	95.8
43	43	1.43	97.23
44	30	1	98.23
45	20	0.67	98.9
46	12	0.4	99.3
47	7	0.23	99.53
48	9	0.3	99.83
49	2	0.07	99.9
50	3	0.1	100
Total	2,997	100.00	

Appendix Table 2: Number of pairwise comparisons per block of 10 EQ5D-5L health states with a logical ordering

Total number of comparisons with a logical ordering	Freq.	Percent	Cum.
16	97	9.7	9.7
18	209	20.9	30.6
19	200	20	50.6
20	218	21.8	72.4
23	79	7.9	80.3
25	99	9.9	90.2
26	98	9.8	100
Total	1,000	100.00	