ISSN 1178-2293 (Online)

University of Otago
Economics Discussion Papers
No. 1704

MARCH 2017

The true significance of 'high' correlations between EQ-5D value sets

Franz Ombler,[a*] Michael Albert,[a] Paul Hansen[b]

[a] *Department of Computer Science, University of Otago*

[b] *Department of Economics, University of Otago*

[*] Corresponding author

Email addresses:

franz@1000minds.com; michael.albert@otago.ac.nz; paul.hansen@otago.ac.nz

The true significance of ‘high’ correlations between EQ-5D value sets

Abstract

High correlation coefficients for EQ-5D value sets derived from different samples, e.g. across countries, are conventionally interpreted as evidence that the people in the respective samples have similar health-related quality of life preferences. However, EQ-5D value sets contain many inherent rankings of health state values *by design*. By calculating coefficients for value sets created from random data, we demonstrate that ‘high’ correlation coefficients are an artefact of these inherent rankings; e.g. median Pearson’s $r = 0.783$ for the EQ-5D-3L and 0.850 for the EQ-5D-5L instead of zero. Therefore, high correlation coefficients do not necessarily constitute evidence of ‘true’ associations. After calculating significance levels based on our simulations – available as a resource for other researchers – we find that many high coefficients are not as significant as conventionally interpreted, whereas other coefficients are not significant. These ‘high’ but insignificant correlations are in fact spurious.

JEL classification: C12, I18

Keywords: Correlation; statistical significance; EQ-5D; health-related quality of life

Highlights:

- Correlation coefficients from EQ-5D value sets are near-universally high.
- Tables of significance levels for EQ-5D correlation coefficients are provided.
- Correlations from the literature and derived from country studies are reassessed.
- Apparently ‘high’ but insignificant correlations are in fact spurious.

Acknowledgements: Thank you to Bas Janssen and Juan Manuel Ramos Goñi from the EuroQol Group for kindly sending us the EQ-5D value sets.

1. Introduction

The calculation of Quality-Adjusted Life Years, as used for cost-utility analysis, depends on the availability of value sets representing people's preferences with respect to health-related quality of life (HRQoL). A value set consists of HRQoL index values for each health state representable by the particular descriptive system used. Values are anchored at unity for 'full health' and zero for 'dead', with negative values for states worse than dead. Well-known descriptive systems include the HUI (Health Utilities Index), SF-6D (Short Form, 6 Dimensions), 15D (15 Dimensions), AQoL (Assessment of Quality of Life) and the EQ-5D (EuroQoL, 5 Dimensions). The EQ-5D is by far the most widely used system (Richardson et al., 2014), and so it is the focus of this paper; nonetheless, the main ideas in this paper apply analogously to other systems too.

The EQ-5D represents HRQoL on five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Two versions of the EQ-5D are available, differentiated by the number of levels on each dimension: (1) EQ-5D-3L, with three levels (Brooks, 1996); and (2) EQ-5D-5L, with five levels (Herdman et al., 2011). Each state is denoted by a five-digit number relating to the relevant levels for each dimension listed in the order above (e.g. 11111 = no problems on any dimension). These two EQ-5D systems – hereinafter referred to simply as 3L and 5L – are capable of representing 243 (3^5) and 3125 (5^5) health states respectively (in addition to 'dead').

Value sets for the 3L have been created for at least 22 countries since 1995, based on visual analogue scale (VAS) or time trade-off (TTO) techniques for eliciting people's HRQoL valuations (Janssen, Szende and Ramos-Goñi, 2014). For the 5L, a "standard protocol" for eliciting valuations based on discrete choice methods and a modified form of TTO was recently developed and piloted for five countries, and more national studies are underway (Oppe et al., 2014). "Ultimately, [these studies will] create a unique opportunity for international comparisons of values for the EQ-5D-5L" (Devlin and Krabbe, 2013, p. S2).

It is common practice for researchers to compare EQ-5D value sets derived from different samples – across countries and/or employing different valuation techniques – by calculating correlation coefficients: Pearson's r for linear correlation and Spearman's ρ for rank correlation (Kendall's τ is also possible). Correlation coefficients reported in the literature are presented in columns 3 and 4 of Table 1 (column 5 is explained later). As can be seen, high or very high correlation coefficients – e.g. >0.8 – are the near-universal finding, leading researchers to report strong correlations, which are usually interpreted as evidence in support of the hypothesis that the people in the respective samples have similar HRQoL preferences.

For example, according to Wittrup-Jensen et al. (2008, p. 32) "Not surprisingly both the Pearson and Spearman correlation matrices showed high and significant correlations, which indicates that [3L] tariffs are similar across different countries [Denmark, UK, Japan, Spain] and that it does not matter whether, for example, English or Danish tariffs are applied in Danish economic evaluation (e.g. cost-utility analysis)." The authors also analyse mean absolute differences (MADs) in values for the same states across value sets and find that "some differences exist", with the "implication … that caution should be taken in concluding that it does not matter which national tariffs are applied."

This apparent contradiction between strong correlations and relatively large MADs is a common finding. A good example is supplied by Lee et al. (2009, p. 1192): "In comparison with other studies, the [3L] value set obtained from our final model is highly correlated with the official value set in Japan (ρ = 0.969, P < 0.001), USA (ρ = 0.908, P < 0.001), and UK (ρ = 0.855, P < 0.001), respectively. The MAD between our Korean study and Japan is 0.056, with USA it is 0.105 and with the UK it is 0.322." This leads the authors to conclude (p. 1193): "When considering the correlation coefficients and MADs … the estimates here are closer to values in the Japanese study than those in the USA and UK. This observation could represent the cultural similarity between Korea and Japan which was also observed in the previous Korean study (Jo et al., 2008)."

According to Jo et al. (2008, p. 1188): "The rank correlation coefficient of estimated values between this study and foreign studies showed a strong positive correlation (the UK: 0.759, the USA: 0.747, Japan: 0.721)." And yet, again based on MAD analyses, Jo concludes: "Substantial differences in the EQ-5D value sets among countries were found. Special caution is needed when a value set from one country is applied to another with a different culture."

Similarly, with respect to the 5L, (Oppe et al., 2014, p. 450) concludes: "Predictions for the complete set of 3125 EQ-5D-5L questionnaire states were quite similar for the four countries. Correlations between the countries were high: from 0.88 (The Netherlands vs. United States) through 0.97 (Canada vs. England)."

The objective of the present paper is to investigate the extent to which high correlation coefficients are in fact artefacts of the many inherent rankings of health state values contained in EQ-5D value sets *by design*, rather than arising from the people in the respective samples having similar health preferences. Most obviously, state 11111 is always valued above all other states, and 33333 for the 3L and 55555 for the 5L are always valued lowest. Likewise, for example, 22123 is always valued above 23123 which is always valued above 23133, and so on.[1] This inherent ranking property of value sets is variously referred to as 'logical' (Devlin et al., 2003), 'primary' (Dolan and Kind, 1996) and 'internal' (Badia et al., 1999) consistency. Although respondents to valuation surveys are seldom perfectly consistent in their individual valuations (Devlin et al., 2003; Lamers et al., 2006), by excluding respondents judged to be 'excessively' inconsistent and imposing restrictions on the estimation method used, the resulting value sets are guaranteed to be consistent.[2]

In general, standard statistical tests for correlation based on the size of correlation coefficients are only valid when there are no inherent rankings – in other words, when it is *a priori* possible for the objects being compared (health state values in the present context) to be ranked in any order. Clearly, this requirement is not satisfied for EQ-5D value sets. Their inherent rankings almost invariably result in correlation coefficients that are commonly interpreted as revealing strong associations. To investigate the extent of this phenomenon, we calculate correlation coefficients for 3L and 5L value sets created from random data – in the process, demonstrating that 'high' correlation coefficients are an artefact of inherent rankings. This simulation-based approach allows us to discover the distribution of correlation coefficients possible from pairs of value sets with no 'true' association in terms of people's HRQoL preferences.

In the next section, we explain our method for generating the random-data 3L and 5L value sets and summarise the resulting correlation coefficients. We also report significance levels for evaluating the statistical significance of EQ-5D correlation coefficients in general – available as a resource for other researchers. In section 3, we present the results of applying these significance levels to the coefficients reported in Table 1 and others derived from published country value sets supplied by the EuroQol Group.

---

[1] Less obviously, due to the additive nature of how values are calculated and by transitivity, if 32133 is valued above 31233 and 13233 is valued above 23133, then 12333 must be valued above 21333; see Hansen and Ombler (2008, p. 90) for an analogous application of the 'joint-factor' independence property of additive models (Krantz, 1972).

[2] If value sets were not consistent, their use in cost-utility analysis might lead to inconsistent results.

Table 1
Correlation coefficients reported in the literature, and calculated significance levels based on simulations.

[1] Article	[2] Value sets compared	[3] Pearson's *r*	[4] Spearman's *ρ*	[5] Significance level (α)
EQ-5D-3L				
Augustovski et al. (2009)	Argentina VAS v. TTO		0.943	0.01
	Argentina TTO v. US TTO		0.963	0.00
Bansback et al. (2012)	Canada TTO v. US TTO	0.964		0.00
	Canada TTO v. UK TTO	0.963		0.00
Cleemput (2010)	Flanders VAS v. Europe VAS		0.974	0.00
	Flanders VAS v. NZ VAS		0.996	0.00
	Flanders VAS v. UK VAS		0.979	0.00
	Flanders VAS v. Spain VAS		0.961	0.00
	Flanders VAS v. Germany VAS		0.943	0.01
	Flanders VAS v. Denmark VAS		0.905	0.05
	Flanders VAS v. Finland VAS		0.869	0.14
	Flanders VAS v. Slovenia VAS		0.851	0.19
Craig et al. (2009)	8-country TTO v. VAS	0.972	0.970	0.00
	8-country TTO v. rank-based method 1	0.971	0.965	0.00
	8-country VAS v. rank-based method 1	0.992	0.989	0.00
	8-country TTO v. rank-based method 2	0.969	0.963	0.00
	8-country VAS v. rank-based method 2	0.992	0.990	0.00
Devlin et al. (2003)	New Zealand VAS sub- v. full sample		0.98	0.00
Golicki et al. (2010)	Poland TTO v. UK TTO	0.95		0.01
	Poland TTO v. German TTO	0.90		0.07
	Poland TTO v. European VAS	0.86		0.19
	Poland TTO v. Slovenia VAS	0.85		0.23
Jo et al. (2008)	South Korea TTO v. UK TTO		0.759	0.55
	South Korea TTO v. US TTO		0.747	0.60
	South Korea TTO v. Japan TTO		0.721	0.69
Lee et al. (2009)	South Korea TTO v. UK TTO		0.855	0.18
	South Korea TTO v. US TTO		0.908	0.05
	South Korea TTO v. Japan TTO		0.969	0.00
Lee et al. (2013)	Taiwan TTO v. UK TTO		0.924	0.02
	Taiwan TTO v. Japan TTO		0.879	0.11
	Taiwan TTO v. South Korea TTO		0.811	0.34
Liu et al. (2014)	China TTO v. Zimbabwe TTO	0.965		0.00
	China TTO v. UK TTO	0.930		0.02
	China TTO v. US TTO	0.929		0.02

	China TTO v. South Korea TTO (2007)	0.909		0.06
	China TTO v. South Korea TTO (2005)	0.904		0.07
	China TTO v. Argentina TTO	0.887		0.11
	China TTO v. Japan TTO	0.866		0.17
Luo et al. (2007)	US TTO v. UK TTO	0.99		0.00
Rand-Hendriksen et al. (2012)	UK TTO v. 'collapsed' TTO	0.999		0.00
Shaw et al. (2010)	US TTO D1 model v. MM-OC model		0.99	0.00
Viney et al. (2011)	Australia TTO 4 model specifications	≥0.96		≤0.003
			≥0.97	≤0.001
Wittrup-Jensen et al. (2008)	Denmark TTO v. UK TTO	0.974	0.971	0.00
	Denmark TTO v. Japan TTO	0.886		0.11
			0.897	0.07
	Denmark TTO v. Spain TTO	0.905		0.06
			0.905	0.05
	UK TTO v. Japan TTO	0.856		0.20
			0.875	0.14
	UK TTO v. Spain TTO	0.950		0.01
			0.920	0.03
	Spain TTO v. Japan TTO	0.911		0.05
			0.941	0.01
	Denmark TTO v. Denmark VAS	0.855		0.21
			0.847	0.24
EQ-5D-5L				
Krabbe et al. (2014)	Canada DCE v. England DCE	0.985		0.00
	Canada DCE v. The Netherlands DCE	0.971		0.00
	Canada DCE v. US DCE	0.981		0.00
	England DCE v. The Netherlands DCE	0.978		0.00
	England DCE v. US DCE	0.966		0.00
	The Netherlands DCE v. US DCE	0.937		0.02
Oppe et al. (2014)	Netherlands DC v. US DCE		0.88	0.10
	Canada DC v. England DCE		0.97	0.00
Shiroiwa et al. (2016)	DCE model 2 v. DCE model 3		0.98	0.00

Note: If both Pearson *r* and Spearman *ρ* are reported, they are in the same row if the significance level is the same; otherwise they are in different rows.

2. Correlation coefficients for value sets from random data

To investigate the extent to which high correlation coefficients in the literature are artefacts of the many inherent rankings of health state values, we used random data to create one million pairs of value sets for the 3L and 5L respectively, from which Pearson's r, Spearman's ρ and Kendall's τ were calculated for each pair.[3,4]

A fundamentally important property of most 'real' value sets estimated from survey data is that they allow for, and include, non-linearities between levels, usually via 'main effects' models (Xie et al., 2014).[5] We modelled such non-linearities for each simulated value set by randomly drawing numbers for each of the five EQ-5D dimensions in the range 0-1 and normalising and rounding the sums across all dimensions to the range 0.000 to 1.000 (corresponding to either 33333 = 0.000 or 55555 = 0.000 and 11111 = 1.000).[6] For each simulated 3L value set, three numbers were randomly drawn for each dimension (corresponding to three levels); and for each 5L value set, five numbers were drawn (five levels). Although negative values for states worse than 'dead' are not modelled here, in principle they could be by rescaling so that 33333 or 55555 has a negative value (and other severe states too); however, this is unnecessary as the rankings of the states are unaffected.

The relative frequency distributions of the correlation coefficients – i.e. based on one million observations each – are displayed in Fig. 1, and summary statistics are in Table 2. For comparison purposes, Fig. 1 also shows frequency distributions of the correlation coefficients for random datasets of 243 and 3125 values respectively that do not have any inherent rankings (as discussed in the introduction, a fundamental assumption of standard statistical tests for correlation is that there are no inherent rankings).

As can be seen in the figure and table, high correlation coefficients do not necessarily constitute evidence of a 'true' association. For example, for the 3L there is a higher than 50% chance of Pearson's $r > 0.783$, Spearman's $\rho > 0.772$ and Kendall's $\tau > 0.584$. For the 5L there

[3] We used Kendall's Tau-a statistic but, arguably, Tau-b would be more appropriate as this statistic makes adjustments for tied rankings in the value sets, whereas Tau-a does not.

[4] The simulations and analysis were performed in a custom software module based on 1000Minds decision-making software (www.1000minds.com), which implements the PAPRIKA method (Hansen and Ombler, 2008).

[5] Many estimated EQ-5D-3L models recognise an additional effect arising from any dimensions being at level 3 (extreme problems), as implemented by the so-called 'N3' variable.

[6] The appropriate granularity or possible spread of values for the simulated value sets is moot. Three significant digits were used for the simulations as this is the precision most commonly reported in the literature.

is a higher than 50% chance of $r > 0.850$, $\rho > 0.840$ and $\tau > 0.649$. Without inherent rankings, these statistics' median values are 0.0.

Table 2
Correlation coefficients, summary statistics – one million simulations.

	EQ-5D-3L (243 values)			**EQ-5D-5L** (3125 values)		
	Pearson's r	Spearman's ρ	Kendall's τ	Pearson's r	Spearman's ρ	Kendall's τ
median	0.783	0.772	0.584	0.850	0.840	0.649
mode[7]	0.81	0.78	0.58	0.86	0.85	0.65
mean	0.772	0.763	0.583	0.844	0.834	0.647
std. dev.	0.098	0.097	0.094	0.054	0.056	0.063
minimum	0.171	0.231	0.164	0.455	0.440	0.308
maximum	0.997	0.996	0.952	0.985	0.983	0.889

[7] The mode was calculated after rounding to two decimal places.

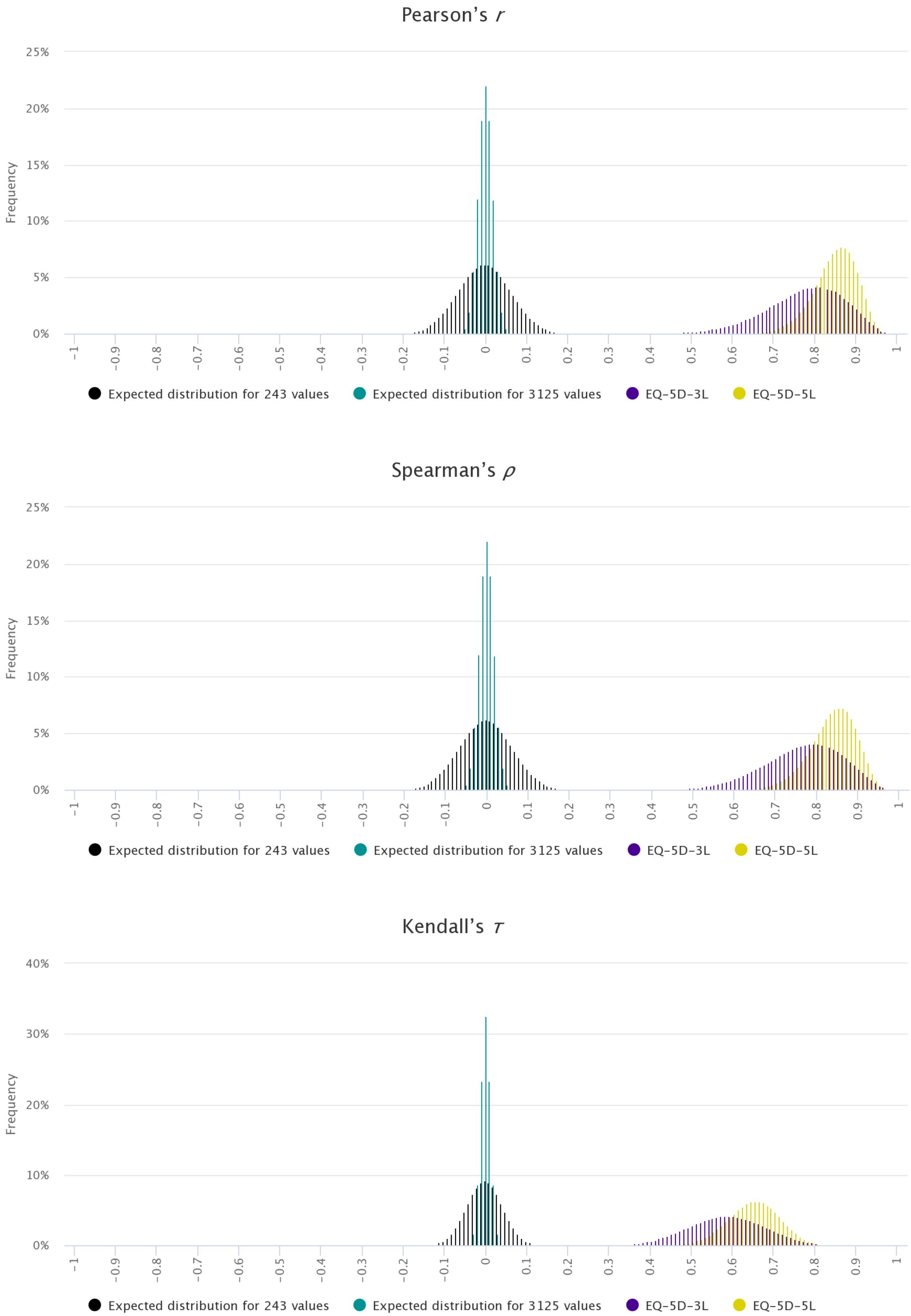

Fig. 1. Frequency distribution of correlation coefficients – one million simulations, 0.01 bins.

We also calculated significance levels (α) from the distributions for the purpose of evaluating the true statistical significance of correlation coefficients, i.e. to test the strength of evidence against the null hypothesis that two value sets are unrelated. A range of significance levels is reported in Table 3, and a tool for other researchers to look-up significance levels for correlation coefficients is available online at www.1000minds.com/sectors/health/eq5d.

As can be seen in the table, to achieve $\alpha < 0.05$ for the 3L: Pearson's $r > 0.91$, Spearman's $\rho > 0.91$ and Kendall's $\tau > 0.74$; and for the 5L: $r > 0.92$, $\rho > 0.92$ and $\tau > 0.75$. Note, however, that because alternative specifications of the model for generating the random data are possible, these significance levels should be regarded as being indicative rather than definitive in terms of their accuracy.

Table 3
Significance levels for Pearson's r, Spearman's ρ and Kendall's τ for pairs of value sets.[8]

Significance level (α)	Pearson's r	Spearman's ρ	Kendall's τ
EQ-5D-3L			
0.01	0.95	0.94	0.80
0.05	0.91	0.91	0.74
0.10	0.89	0.88	0.71
0.15	0.87	0.86	0.68
0.20	0.86	0.85	0.67
0.25	0.84	0.84	0.65
0.30	0.83	0.82	0.64
0.35	0.82	0.81	0.62
0.40	0.81	0.80	0.61
0.45	0.80	0.79	0.60
0.50	0.78	0.77	0.58
EQ-5D-5L			
0.01	0.94	0.94	0.79
0.05	0.92	0.92	0.75
0.10	0.91	0.90	0.73
0.15	0.9	0.89	0.71
0.20	0.89	0.88	0.70
0.25	0.88	0.88	0.69
0.30	0.88	0.87	0.68
0.35	0.87	0.86	0.67
0.40	0.86	0.85	0.67
0.45	0.86	0.85	0.66
0.50	0.85	0.84	0.65

[8] The one million simulations performed provide confidence in reporting values to two significant digits (Cuddington and Navidi, 2011, p. 726).

3. Significance of correlation coefficients in the EQ-5D literature

Applying the previous section's results, significance levels for the correlation coefficients reported in the literature are presented in column 5 of Table 1. As can be seen, many coefficients are statistically significant, though not as significant as conventionally interpreted, whereas other coefficients are not significant. For example, at the 5% level ($\alpha = 0.05$), for Pearson's r, 23 correlation coefficients are statistically significant and 11 are not; and for Spearman's ρ, 25 correlation coefficients are statistically significant and 12 are not.

These results enable particular findings in the literature to be re-evaluated, including the four studies discussed in the introduction. With respect to Wittrup-Jensen et al. (2008)'s finding for Denmark, UK, Japan, Spain that "Pearson and Spearman correlation matrices showed high and significant correlations", $\alpha > 0.05$ for five of these 12 correlation coefficients – i.e. contradicting the authors with respect to there being strong correlations for Denmark v. Japan, Denmark v. Spain and UK v. Japan (see Table 1). For Lee et al. (2009), Korea's value set is strongly correlated with both Japan ($\alpha = 0.00$) and – though less so – the US ($\alpha = 0.05$) but, in contrast to the authors' finding, not the UK ($\alpha = 0.18$). The seemingly contradictory conclusion in Jo et al. (2008) of substantial differences between the South Korea value set and those of the UK, US and Japan despite apparently "strong positive correlations" is validated with α ranging from 0.55 to 0.69. Finally, Oppe et al.'s (2014) conclusion that value sets were "… quite similar for the four countries. Correlations … were high: from 0.88 (The Netherlands vs. United States) through 0.97 (Canada vs. England)" is, arguably, only half correct – $\alpha = 0.00$ for Canada vs England, whereas $\alpha = 0.10$ for The Netherlands vs. the US.

In addition, we calculated correlation coefficients for country value sets supplied by the EuroQol Group (2016): 19 3L value sets (10 based on TTO and nine on VAS) and eight 5L value sets (all based on TTO). Their correlation coefficients and significance levels are reported in Tables 4a-4c for the 3L and Tables 5a-5c for the 5L. As can be seen in the tables, significance levels range from 0.00 to 0.69.

Table 4a
Pearson's *r* for EQ-5D-3L value sets in the upper triangular; significance levels in the lower triangular.

		TTO										VAS								
		DK	FR	DE	JP	NL	ES	TH	UK	US	ZW	BE	DK	EU	FI	DE	NZ	SI	ES	UK
TTO	DK	-	.893	.892	.883	.933	.896	.899	.955	.957	.943	.880	.878	.874	.858	.880	.891	.893	.840	.897
	FR	.09	-	.910	.873	.875	.965	.959	.938	.934	.933	.932	.970	.961	.915	.909	.930	.941	.968	.965
	DE	.10	.05	-	.862	.884	.949	.941	.968	.963	.898	.898	.837	.930	.787	.898	.913	.875	.906	.947
	JP	.12	.15	.19	-	.732	.907	.952	.854	.888	.818	.758	.859	.841	.891	.802	.775	.934	.806	.845
	NL	.02	.14	.12	.69	-	.859	.815	.964	.928	.924	.957	.841	.914	.818	.906	.960	.806	.871	.930
	ES	.09	.00	.01	.06	.19	-	.981	.952	.969	.917	.922	.922	.964	.881	.942	.919	.913	.970	.970
	TH	.08	.00	.01	.01	.37	.00	-	.926	.951	.894	.873	.928	.929	.886	.894	.884	.952	.931	.937
	UK	.01	.02	.00	.21	.00	.01	.03	-	.986	.957	.955	.890	.959	.858	.939	.961	.887	.929	.975
	US	.00	.02	.00	.11	.03	.00	.01	.00	-	.948	.929	.890	.943	.855	.952	.935	.890	.926	.959
VAS	ZW	.01	.02	.08	.36	.03	.04	.09	.00	.01	-	.903	.918	.908	.857	.885	.904	.880	.902	.925
	BE	.13	.02	.08	.60	.00	.04	.15	.01	.02	.07	-	.907	.974	.867	.943	.996	.853	.961	.979
	DK	.13	.00	.28	.20	.26	.04	.03	.10	.10	.04	.06	-	.924	.947	.859	.910	.960	.936	.925
	EU	.14	.00	.02	.26	.05	.00	.02	.00	.01	.06	.00	.03	-	.909	.946	.971	.892	.977	.996
	FI	.20	.05	.48	.10	.36	.13	.11	.20	.21	.20	.17	.01	.06	-	.828	.870	.927	.868	.893
	DE	.13	.06	.08	.42	.06	.01	.09	.02	.01	.11	.01	.20	.01	.31	-	.934	.817	.941	.953
	NZ	.10	.02	.05	.53	.00	.04	.12	.00	.02	.07	.00	.06	.00	.16	.02	-	.871	.952	.978
	SI	.09	.01	.14	.02	.40	.05	.01	.11	.10	.13	.22	.00	.10	.03	.36	.15	-	.880	.899
	ES	.27	.00	.06	.41	.16	.00	.02	.03	.03	.07	.00	.02	.00	.17	.01	.01	.13	-	.979
	UK	.08	.00	.01	.24	.02	.00	.02	.00	.00	.03	.00	.03	.00	.09	.01	.00	.08	.00	-

DK: Denmark, FR: France, DE: Germany, JP: Japan, NL: Netherlands, ES: Spain, TH: Thailand, UK: United Kingdom, US: United States, ZW: Zimbabwe, BE: Belgium, EU: European Union, FI: Finland, NZ: New Zealand, SI: Slovenia.

Table 4b
Spearman’s ρ for EQ-5D-3L value sets in the upper triangular; significance levels in the lower triangular.

		TTO										**VAS**								
		DK	FR	DE	JP	NL	ES	TH	UK	US	ZW	BE	DK	EU	FI	DE	NZ	SI	ES	UK
TTO	DK	-	.894	.897	.887	.940	.903	.893	.975	.974	.940	.911	.877	.902	.855	.967	.920	.883	.850	.930
	FR	.07	-	.888	.885	.834	.963	.949	.923	.920	.926	.929	.977	.969	.913	.923	.920	.939	.977	.976
	DE	.07	.09	-	.901	.818	.921	.939	.947	.943	.897	.826	.834	.892	.784	.869	.843	.899	.844	.916
	JP	.09	.10	.06	-	.731	.942	.960	.879	.915	.824	.784	.863	.887	.889	.889	.794	.940	.833	.894
	NL	.01	.25	.32	.66	-	.785	.756	.946	.889	.925	.942	.823	.872	.802	.886	.949	.780	.802	.895
	ES	.06	.00	.03	.01	.45	-	.986	.921	.957	.916	.874	.941	.949	.896	.951	.868	.935	.955	.957
	TH	.08	.01	.01	.00	.57	.00	-	.905	.944	.874	.841	.919	.924	.870	.919	.846	.953	.921	.935
	UK	.00	.03	.01	.11	.01	.03	.05	-	.977	.973	.932	.894	.942	.862	.948	.939	.894	.886	.965
	US	.00	.03	.01	.04	.08	.00	.01	.00	-	.957	.896	.895	.922	.857	.977	.900	.898	.894	.946
VAS	ZW	.01	.02	.07	.29	.02	.04	.12	.00	.00	-	.929	.909	.930	.841	.954	.920	.856	.918	.953
	BE	.04	.02	.28	.46	.01	.12	.23	.02	.07	.02	-	.944	.952	.901	.908	.996	.873	.929	.961
	DK	.11	.00	.25	.15	.29	.01	.03	.07	.07	.05	.01	-	.960	.941	.914	.934	.949	.976	.965
	EU	.06	.00	.08	.09	.13	.01	.03	.01	.03	.02	.00	.00	-	.948	.922	.947	.924	.958	.994
	FI	.18	.04	.45	.09	.38	.07	.13	.16	.17	.23	.06	.01	.01	-	.878	.897	.914	.891	.932
	DE	.00	.03	.14	.08	.09	.01	.03	.01	.00	.00	.05	.04	.03	.11	-	.902	.875	.913	.939
	NZ	.03	.03	.22	.41	.01	.14	.21	.01	.06	.03	.00	.02	.01	.07	.06	-	.884	.911	.958
	SI	.10	.01	.06	.01	.47	.01	.00	.07	.07	.17	.12	.01	.02	.04	.12	.10	-	.901	.938
	ES	.20	.00	.22	.26	.38	.00	.03	.09	.07	.03	.02	.00	.00	.08	.04	.04	.06	-	.961
	UK	.02	.00	.04	.07	.07	.00	.01	.00	.01	.00	.00	.00	.00	.02	.01	.00	.01	.00	-

DK: Denmark, FR: France, DE: Germany, JP: Japan, NL: Netherlands, ES: Spain, TH: Thailand, UK: United Kingdom, US: United States, ZW: Zimbabwe, BE: Belgium, EU: European Union, FI: Finland, NZ: New Zealand, SI: Slovenia.

Table 4c
Kendall's τ for EQ-5D-3L value sets in the upper triangular; significance levels in the lower triangular.

		TTO										VAS								
		DK	FR	DE	JP	NL	ES	TH	UK	US	ZW	BE	DK	EU	FI	DE	NZ	SI	ES	UK
TTO	DK	-	.732	.725	.717	.802	.742	.724	.866	.868	.802	.754	.714	.749	.677	.852	.767	.711	.698	.793
	FR	.06	-	.712	.706	.657	.837	.818	.767	.761	.776	.758	.868	.848	.742	.772	.753	.786	.871	.870
	DE	.07	.09	-	.726	.631	.763	.792	.803	.805	.728	.643	.645	.724	.590	.691	.656	.723	.665	.758
	JP	.08	.10	.07	-	.558	.796	.825	.710	.754	.660	.610	.689	.728	.719	.715	.618	.789	.660	.740
	NL	.01	.22	.31	.61	-	.607	.582	.814	.726	.768	.806	.639	.718	.621	.722	.821	.597	.626	.740
	ES	.05	.00	.03	.01	.40	-	.905	.764	.828	.757	.699	.792	.818	.723	.816	.693	.785	.842	.833
	TH	.07	.00	.01	.00	.51	.00	-	.744	.805	.715	.672	.757	.774	.686	.763	.676	.815	.779	.797
	UK	.00	.02	.01	.09	.01	.03	.04	-	.879	.859	.795	.717	.804	.677	.813	.804	.727	.715	.847
	US	.00	.03	.01	.03	.07	.00	.01	.00	-	.835	.734	.722	.773	.673	.871	.738	.729	.734	.810
VAS	ZW	.01	.02	.06	.21	.02	.03	.08	.00	.00	-	.776	.746	.783	.653	.811	.763	.675	.764	.822
	BE	.03	.03	.27	.40	.01	.11	.18	.01	.06	.02	-	.790	.819	.722	.747	.951	.694	.774	.835
	DK	.08	.00	.27	.13	.29	.01	.03	.08	.07	.04	.01	-	.826	.792	.749	.780	.804	.878	.845
	EU	.04	.00	.07	.06	.08	.00	.02	.01	.02	.01	.00	.00	-	.808	.775	.812	.767	.834	.940
	FI	.16	.05	.47	.08	.35	.07	.14	.17	.17	.24	.07	.01	.01	-	.697	.717	.749	.714	.776
	DE	.00	.02	.13	.08	.07	.00	.03	.01	.00	.01	.04	.04	.02	.12	-	.740	.706	.758	.802
	NZ	.02	.03	.23	.36	.00	.13	.17	.01	.05	.03	.00	.02	.01	.08	.05	-	.706	.748	.832
	SI	.09	.01	.07	.01	.45	.01	.01	.06	.06	.17	.12	.01	.02	.04	.10	.10	-	.728	.796
	ES	.12	.00	.20	.22	.33	.00	.02	.08	.06	.03	.02	.00	.00	.08	.03	.04	.06	-	.841
	UK	.01	.00	.03	.05	.05	.00	.01	.00	.01	.00	.00	.00	.00	.02	.01	.00	.01	.00	-

DK: Denmark, FR: France, DE: Germany, JP: Japan, NL: Netherlands, ES: Spain, TH: Thailand, UK: United Kingdom, US: United States, ZW: Zimbabwe, BE: Belgium, EU: European Union, FI: Finland, NZ: New Zealand, SI: Slovenia.

Table 5a
Pearson's r for EQ-5D-5L in the upper triangular; significance levels in the lower triangular.

	CA	CN	ES	KR	JP	NL	UK	UY
CA	-	.943	.962	.930	.927	.965	.981	.926
CN	.01	-	.955	.948	.986	.908	.960	.917
ES	.00	.00	-	.940	.955	.955	.984	.907
KR	.03	.01	.01	-	.949	.882	.943	.957
JP	.04	.00	.00	.01	-	.896	.953	.911
NL	.00	.11	.00	.26	.16	-	.972	.839
UK	.00	.00	.00	.01	.00	.00	-	.908
UY	.04	.06	.11	.00	.09	.59	.11	-

Table 5b
Spearman's ρ for EQ-5D-5L in the upper triangular; significance levels in the lower triangular.

	CA	CN	ES	KR	JP	NL	UK	UY
CA	-	.934	.958	.923	.917	.962	.980	.922
CN	.01	-	.951	.946	.985	.896	.955	.913
ES	.00	.00	-	.938	.950	.951	.983	.902
KR	.03	.01	.01	-	.947	.872	.938	.960
JP	.05	.00	.00	.00	-	.884	.947	.905
NL	.00	.13	.00	.28	.19	-	.970	.831
UK	.00	.00	.00	.01	.00	.00	-	.903
UY	.04	.06	.10	.00	.09	.57	.09	-

Table 5c
Kendall's τ for EQ-5D-5L in the upper triangular; significance levels in the lower triangular.

	CA	CN	ES	KR	JP	NL	UK	UY
CA	-	.780	.824	.764	.751	.831	.876	.760
CN	.01	-	.809	.797	.895	.726	.817	.744
ES	.00	.00	-	.784	.808	.808	.887	.728
KR	.03	.01	.01	-	.803	.691	.788	.832
JP	.05	.00	.00	.00	-	.713	.802	.734
NL	.00	.10	.00	.25	.15	-	.851	.644
UK	.00	.00	.00	.01	.00	.00	-	.732
UY	.03	.06	.10	.00	.08	.54	.09	-

CA: Canada, CN: China, ES: Spain, KR: South Korea, JP: Japan, NL: Netherlands, UK: United Kingdom, UY: Uruguay.

4. Conclusion

We show that high correlation coefficients are an artefact of the many inherent rankings contained in EQ-5D value sets by design. Therefore, high correlation coefficients do not necessarily constitute evidence of ‘true’ associations between value sets in terms of similar HRQoL preferences. The significance levels we calculated reveal that many of the high correlation coefficients reported in the literature are statistically significant, though not as significant as conventionally interpreted, whereas other coefficients are not significant. These apparently ‘high’ but insignificant correlations are in fact spurious. This finding serves to reconcile the seemingly contradictory results of both strong correlations and high mean absolute differences in values for the same states across value sets reported in the literature. These significance levels are available as a resource for other researchers using the EQ-5D. Future research applying a similar approach could be undertaken for other HRQoL descriptive systems such as the HUI, SF-6D, 15D and AQoL.

References

Augustovski, F.A., Irazola, V.E., Velazquez, A.P., Gibbons, L., Craig, B.M., 2009. Argentine valuation of the EQ-5D health states. *Value in Health* 12, 587-596.

Badia, X., Roset, M., Herdman, M., 1999. Inconsistent responses in three preference-elicitation methods for health states. *Social Science & Medicine* 49, 943-950.

Bansback, N., Tsuchiya, A., Brazier, J., Anis, A., 2012. Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. *PLoS ONE* 7, e31115.

Brooks, R., 1996. EuroQol: the current state of play. Health Policy 37, 53-72.

Cleemput, I., 2010. A social preference valuations set for EQ-5D health states in Flanders, Belgium. *The European Journal of Health Economics* 11, 205-213.

Craig, B.M., Busschbach, J.J.V., Salomon, J.A., 2009. Modeling ranking, time trade-off, and visual analog scale values for EQ-5D health states: a review and comparison of methods. *Medical Care* 47, 634-641.

Cuddington, J.T., Navidi, W., 2011. A critical assessment of simulated critical values. *Communications in Statistics – Simulation and Computation* 40, 719-727.

Devlin, N.J., Hansen, P., Kind, P., Williams, A., 2003. Logical inconsistencies in survey respondents' health state valuations - a methodological challenge for estimating social tariffs. *Health Economics* 12, 529-544.

Devlin, N.J., Krabbe, P.F.M., 2013. The development of new research methods for the valuation of EQ-5D-5L. *The European Journal of Health Economics* 14, 1-3.

Dolan, P., Kind, P., 1996. Inconsistency and health state valuations. *Social Science & Medicine* 42, 609-615.

EuroQol – EQ-5D-3L Value sets, n.d. www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-3l-value-sets.html (accessed 2 February 2017).

EuroQol – EQ-5D-5L Value sets, 2016. www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html (accessed 2 February 2017).

EuroQol – References, n.d. www.euroqol.org/about-eq-5d/valuation-of-eq-5d/references.html (accessed 2 February 2017).

Golicki, D., Jakubczyk, M., Niewada, M., Wrona, W., Busschbach, J.J.V., 2010. Valuation of EQ-5D health states in Poland: first TTO-based social value set in Central and Eastern Europe. *Value in Health* 13, 289-297.

Hansen, P., Ombler, F., 2008. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *Journal of Multi-Criteria Decision Analysis* 15, 87-107.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G., Badia, X., 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research* 20, 1727-1736.

Janssen, B., Szende, A., Ramos-Goñi, J.M., 2014. Data and Methods, in: Szende, A., Janssen, B., Cabases, J. (Eds.), *Self-Reported Population Health: An International Perspective Based on EQ-5D*. Springer Netherlands, Dordrecht, pp. 7-17.

Jo, M.-W., Yun, S.-C., Lee, S.-I., 2008. Estimating quality weights for EQ-5D health states with the time trade-off method in South Korea. *Value in Health* 11, 1186-1189.

Krabbe, P.F.M., Devlin, N.J., Stolk, E.A., Shah, K.K., Oppe, M., van Hout, B., Quik, E.H., Pickard, A.S., Xie, F., 2014. Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. *Medical Care* 52, 935-943.

Krantz, D.H., 1972. Measurement structures and psychological laws. *Science* 175, 1427-1435.

Lamers, L.M., 2006. Inconsistencies in TTO and VAS values for EQ-5D health states. *Medical Decision Making* 26, 173-181.

Lee, H.-Y., Hung, M.-C., Hu, F.-C., Chang, Y.-Y., Hsieh, C.-L., Wang, J.-D., 2013. Estimating quality weights for EQ-5D (EuroQol-5 dimensions) health states with the time trade-off method in Taiwan. *Journal of the Formosan Medical Association* 112, 699-706.

Lee, Y.-K., Nam, H.-S., Chuang, L.-H., Kim, K.-Y., Yang, H.-K., Kwon, I.-S., Kind, P., Kweon, S.-S., Kim, Y.-T., 2009. South Korean time trade-off values for EQ-5D health states: modeling with observed values for 101 health states. *Value in Health* 12, 1187-1193.

Liu, G.G., Wu, H., Li, M., Gao, C., Luo, N., 2014. Chinese time trade-off values for EQ-5D health states. *Value in Health* 17, 597-604.

Manly, B.F.J., 2007. *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 3rd ed., Texts in statistical science. Chapman & Hall/ CRC, Boca Raton, FL.

Nan Luo, Johnson, J.A., Shaw, J.W., Coons, S.J., 2007. A comparison of EQ-5D index scores derived from the US and UK population-based scoring functions. *Medical Decision Making* 27, 321-326.

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, 445-453.

Rand-Hendriksen, K., Augestad, L.A., Dahl, F.A., Kristiansen, I.S., Stavem, K., 2012. A shortcut to mean-based time tradeoff tariffs for the EQ-5D? *Medical Decision Making* 32, 569-577.

Richardson, J., McKie, J., Bariola, E., 2014. Multiattribute utility instruments and their use, in: *Encyclopedia of Health Economics*. Elsevier, pp. 341-357.

Shaw, J.W., Pickard, A.S., Yu, S., Chen, S., Iannacchione, V.G., Johnson, J.A., Coons, S.J., 2010. A median model for predicting United States population-based EQ-5D health state preferences. *Value in Health* 13, 278-288.

Shiroiwa, T., Ikeda, S., Noto, S., Igarashi, A., Fukuda, T., Saito, S., Shimozuma, K., 2016. Comparison of value set based on DCE and/or TTO data: scoring for EQ-5D-5L health states in Japan. *Value in Health* 19, 648-654.

Viney, R., Norman, R., King, M.T., Cronin, P., Street, D.J., Knox, S., Ratcliffe, J., 2011. Time trade-off derived EQ-5D weights for Australia. *Value in Health* 14, 928-936.

Wittrup-Jensen, K.U., Lauridsen, J., Pedersen, K.M., 2008. *Modelling Danish EuroQol (EQ-5D) Tariffs by Applying the Time Trade-Off Method*. Syddansk Universitet.

Xie, F., Gaebel, K., Perampaladas, K., Doble, B., Pullenayegum, E., 2014. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Medical Decision Making* 34, 8-20.