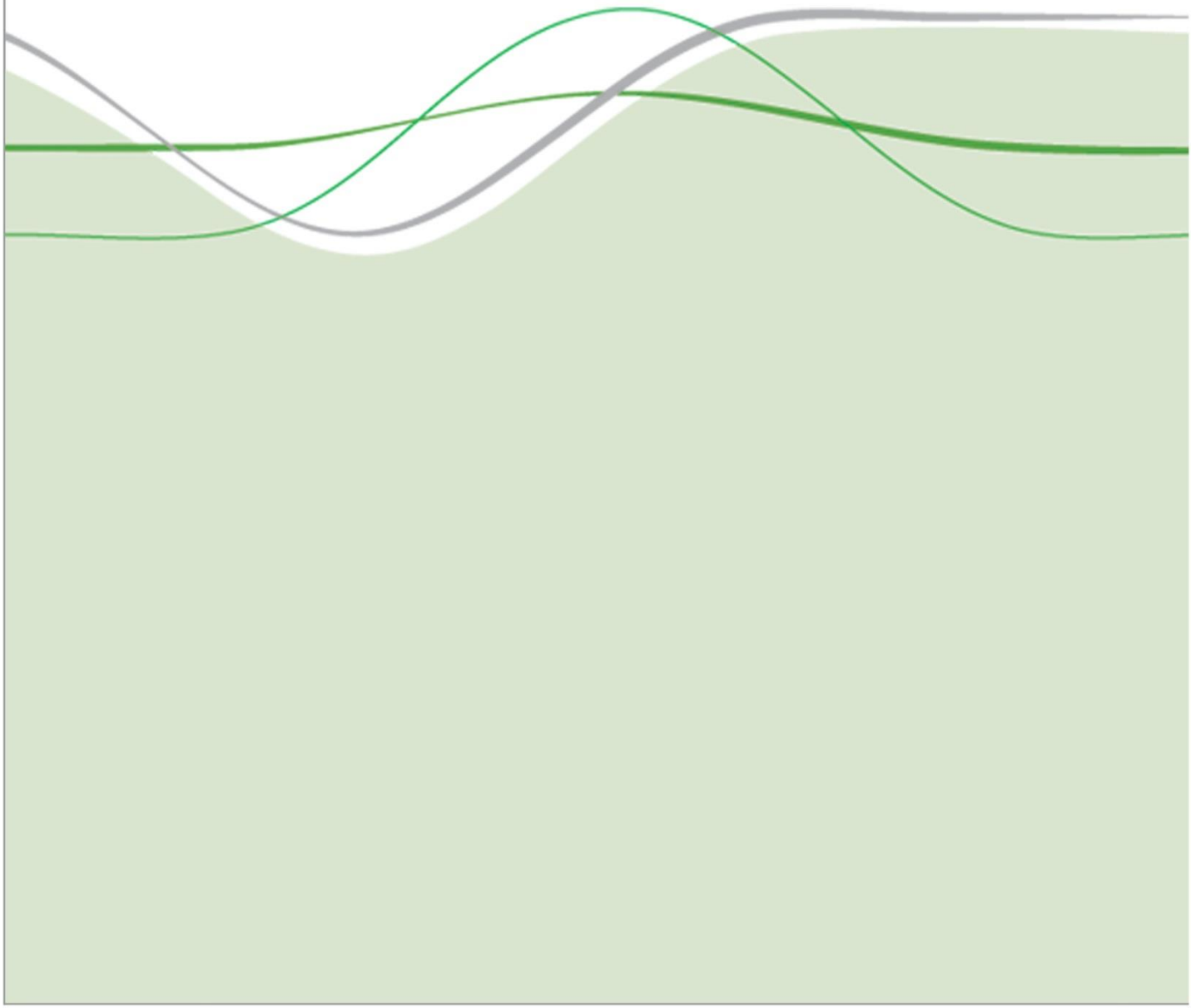
Office of Health Economics Research

Research Paper 16/04

The Distribution of the EQ-5D-5L Index in Patient Populations

October 2016

Yan Feng, Nancy Devlin, Andrew Bateman, Bernarda Zamora and David Parkin



The Distribution of the EQ-5D-5L Index in Patient Populations

Yan Feng [1], Nancy Devlin [1], Andrew Bateman [2,3], Bernarda Zamora [1] and David Parkin [4]

[1] Office of Health Economics

[2] Cambridgeshire Community Services NHS Trust

[3] University of Cambridge

[4] King's College London

October 2016

Research Paper 16/04

For further information please contact:

Yan Feng

yfeng@ohe.org

The Office of Health Economics (a company limited by guarantee of registered number 09848965)

Southside, 105 Victoria Street

London SW1E 6QT

United Kingdom

Tel: +44 207 747 8863

About OHE research papers

OHE Research Papers are intended to provide information on and encourage discussion about a topic in advance of formal publication.

Any views expressed are those of the authors and do not necessarily reflect the views or approval of OHE, its Editorial or Policy Board, or its sponsors.

Once a version of the Research Paper's content is published in a peer reviewed journal, that supersedes the Research Paper and readers are invited to cite the published version in preference to the original version.

Conflicts of interest

Nancy Devlin and David Parkin are members of the EuroQol Research Foundation, which developed and owns the EQ-5D.

Acknowledgements and disclaimers

This project was funded by a research grant from the EuroQol Research Foundation. The research was also supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East of England at Cambridgeshire and Peterborough NHS Foundation Trust. All views expressed are those of the authors, and not necessarily those of the EuroQol Research Foundation, the NHS or the NIHR.

The earlier versions of this paper have been presented at the 21st International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Annual International Meetings (Washington DC, 2016) and the Patient-Reported Outcome Measures (PROMs) Conference (Sheffield, 2016).

Abstract

Background: EQ-5D data are often summarised by an EQ-5D index, whose distribution for its original version, the EQ-5D-3L, often shows in patient populations two distinct groups, arising from both the distribution of ill health and how the index is constructed (Parkin *et al*., 2016). To date, there is little evidence about the distribution of the EQ-5D-5L index.

Aims: The aims of this study are: to explore whether or not the EQ-5D-5L index distribution also demonstrates clustering; to test the extent to which clustering of EQ-5D-5L profile data drives any observed clustering of the EQ-5D-5L index, and the extent to which clusters result from the value sets used to create the index; and to discuss the implications of our results for statistical analysis of EQ-5D-5L index data.

Data: Data from Cambridgeshire Community Services NHS's electronic patient records data warehouse were analysed. EQ-5D-5L profiles before treatment were obtained for 30,284 patients across three patient groups: community rehabilitation services (N=6,919); musculoskeletal therapy services (N=19,999); and nursing services (N=3,366).

Methods: The EQ-5D-5L index is calculated using both a 'mapped' (crosswalk) value set (MVS) and the English value set (EVS). We examined the distribution of 1,730 of the 3,125 profiles described by the EQ-5D-5L to check for clustering of the EQ-5D-5L index. The k-means cluster method and the Calinski–Harabasz pseudo-F index stopping rule were used to search for the clusters in the index. We examined the impact on the results of using different initial values in the clustering analysis.

Results: Clustering within the EQ-5D-5L index distribution is suggested by both clustering methods, for the three patient groups and all patients together. For the all patients' data, we found two robust clusters for the MVS-based index, compared to three robust clusters for the EVS-based index. The EQ-5D-5L profile data alone do not obviously drive the index clusters.

Conclusion: The results highlight the importance of undertaking careful exploratory data analysis for health related quality of life measures such as the EQ-5D, to ensure that statistical testing takes account of clustering and other features of the data distribution.

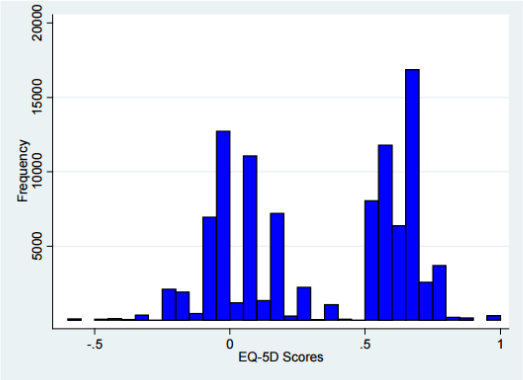
Key words: EQ-5D; EQ-5D-5L Index; EQ-5D-5L profile; EQ-5D-5L value sets; Clustering Analysis; Patient Reported Outcomes (PROs); NHS PROMs.

1. Introduction

The EQ-5D (Devlin and Brooks, 2016) has been extensively used worldwide in clinical trials, observational studies and clinical practice (Appleby *et al*, 2016). EQ-5D data are often summarised by a single number index, calculated by applying value sets to EQ-5D profiles (Szende *et al*, 2007). An EQ-5D index is anchored at 1, representing full health, and 0, meaning a health state as bad as being dead, a requirement for the use of these data to estimate quality-adjusted life years (QALYs) for purposes such as cost effectiveness analysis.

The EQ-5D describes health in terms of five dimensions. The original EQ-5D (now called the EQ-5D-3L) has three response options (no, some or extreme problems), and the subsequent EQ-5D-5L (Herdman *et al*, 2011) has five (no, slight, moderate, severe, or extreme/unable problems). It has been found that the distribution of the EQ-5D-3L index is often characterised by two distinct clusters or groups of observations - as shown in the example in Figure 1 - which could reflect the actual distribution of ill health, but might also be an artefact of how the index is constructed (Parkin *et al;* 2016).

Figure 1. Distribution of pre-surgery EQ-5D index scores for hip replacement patients, using the UK EQ-5D-3L value set.



Source: Parkin *et al* (2016)

Parkin *et al* (2016) show that this two-group distribution results from both how the EQ-5D-3L classification generates profiles and also the characteristics of EQ-5D-3L value sets. The causes include the low values assigned to profiles that include an extreme response in any dimension by, for example, the value set commonly used in the United Kingdom (Dolan, 1997).

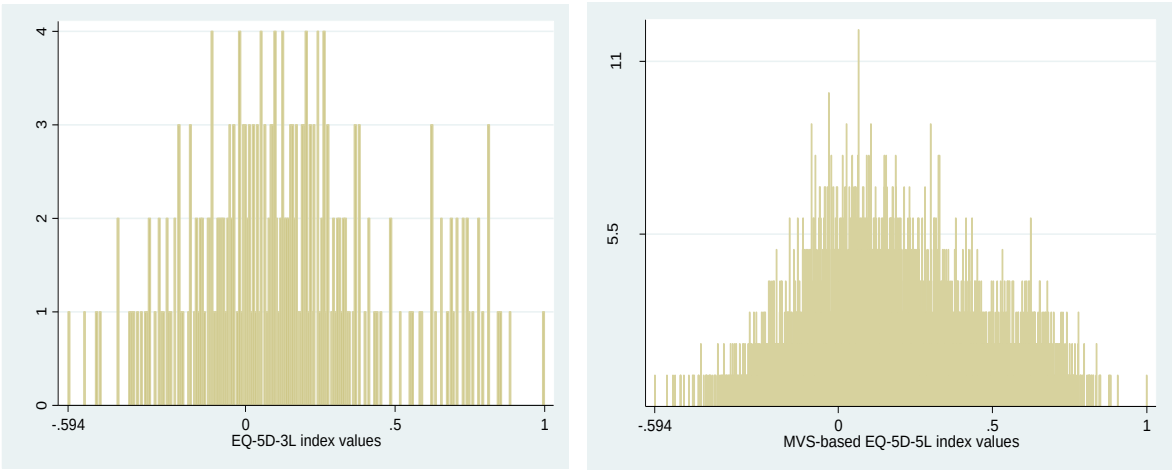
To date there has been little work to explore whether distributions of EQ-5D-5L data have similar characteristics. There are relatively few EQ-5D-5L data accessible to researchers and EQ-5D-5L value sets are only now becoming available (Devlin *et al*,

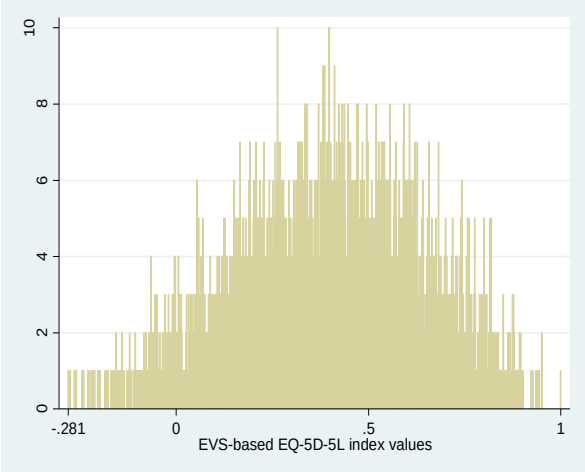
2016; Versteegh *et al*, 2016; Xie *et al*, 2015; Ramos-Goñi *et al*, 2015). Until recently, the EQ-5D-5L index could only be calculated using a mapping (sometimes called 'crosswalk') algorithm derived from the EQ-5D-3L descriptive system and value sets (van Hout *et al*, 2012). Increasing use of the EQ-5D-5L, for which requests for licenses now exceed those for the EQ-5D-3L (Devlin and Brooks, 2016), mean that understanding the characteristics of these distributions is timely.

There are good grounds for hypothesising that the EQ-5D-5L index data might *not* have the two-group distribution commonly observed for the EQ-5D-3L. First, a study comparing the distributions of EQ-5D-3L and EQ-5D-5L profile data from the general population sample in England found a wider spread of health states reported, including a larger proportion who reported severe problems (levels 4 and 5 in the 5L version; level 3 in the 3L version) and fewer who reported no problems (full health '11111' in both versions of the EQ-5D instrument) (Feng *et al*, 2015). Similar findings were reported by Craig *et al* (2014) using data from the general population sample in the US to compare the performance of 3L and 5L versions of the EQ-5D instrument. These results suggest that the EQ-5D-5L may generate no clustering, or a different type of clustering, in the distribution of index values.

Secondly, as Figure 2 shows, the distributions of EQ-5D-5L values over all possible profiles, both when based on the English value set (EVS) and on the mapped value set (MVS), do not have the two group shape of the EQ-5D-3L value set. This suggests that the value sets may not in themselves generate clusters, and even if clusters are generated by EQ-5D-5L patient profiles, this may not be reflected in the value set.

Figure 2. Frequency of values in the EQ-5D-3L and the MVS- and EVS-based EQ-5D-5L index over all possible profiles





This study aims to address the following:

1. Do EQ-5D-5L index datasets, summarised by the MVS, have a two-group or other sort of clustered distribution? If so, what explains that distribution?
2. Do EQ-5D-5L profile data, summarised by the EVS, have a two-group or other sort of clustered distribution? If so, what explains that distribution?
3. How does the distribution of EQ-5D-5L profiles compare with the distribution of EQ-5D-3L profiles? How do these affect the distribution of EQ-5D-5L index data?
4. To what extent do the characteristics of the EQ-5D-5L value set for England drive the distribution of the EQ-5D-5L index? Does this differ from the distribution generated by the mapped value set?
5. What are the implications for the analysis of EQ-5D-5L data?

This study also aims to develop further the methods reported in Parkin *et al* (2016) to investigate clustering – specifically, how the number of clusters identified in datasets is determined.

2. Data

The data were collected as part of routine clinical practice in Cambridgeshire Community Services, part of the English National Health Service (NHS). Clinicians provide a paper copy of the EQ-5D-5L to patients as part of their assessment and intervention planning. The individual responses are copied to the patient's electronic patient record. The Trust's data warehouse team extracts & collates the data along with basic demographic information. The clinical teams collecting the data provide services to distinctive patient groups. Rehabilitation services are provided by Occupational Therapists and Physiotherapists to people characterised broadly as frail elderly and people managing

conditions such as stroke. Home visits are geared toward improving participation in treatment through interventions such as providing mobility and self-care advice, assistive technology and housing adaptations. Specialist Nurses work in small disease-specific teams helping patients to manage their long term chronic conditions such as diabetes, Parkinson's disease, multiple sclerosis and chronic respiratory diseases. They see patients in clinic settings and in their own homes. In contrast the musculoskeletal (MSK) physiotherapy services are provided to service users across the life span with acute injuries and long term MSK chronic conditions covering rheumatology, orthopaedic and persistent pain. The team primarily comprises physiotherapists and clinical specialists whose focus is on spinal and peripheral joint pain and dysfunction. Patients attend outpatient clinics based in hospitals, medical centres and bespoke departments within primary care. Together the data collected represent a reasonably comprehensive survey of these service users over the months from January 2013 to March 2015.

Forty-six patients are excluded from the analysis because they were under 13 years old. There were therefore 30,284 patient observations across three patient groups: MSK physiotherapy services (N=19,999); specialist nursing services (N=3,366); and community rehabilitation services (N=6,919).

3. Methods

3.1. Distribution of the EQ-5D-5L profiles and EQ-5D-5L index

We first report the distribution of the EQ-5D-5L profiles and index for all patients and for each treatment group, with the aim of checking for clusters in each. We calculated two index values, one by applying to the profiles the MVS (van Hout *et al*, 2012) and the other the EVS (Devlin *et al*, 2016). We used a skewness and kurtosis test for normality of the resulting distribution.

3.2. Cluster analyses of the EQ-5D-5L profile data

We explored whether clusters of values can be generated by EQ-5D-5L profiles themselves in two ways, using the full EQ-5D-5L set of 3,125 profiles and both the MVS and the EVS. First, we divided profiles into two groups according to whether they had level 5 in any dimension or no level 5 in any dimension, and similarly for levels 4 and 5 and levels 3, 4 and 5. This follows the method used Parkin *et al* (2016), although it is more complicated because of the greater number of levels.

Second, we examined the differences between consecutive EQ-5D-5L values ordered by size, to see if we could identify any notable gaps in the distribution.

3.3. Cluster Analyses of the EQ-5D-5L index data

Cluster analysis was used to search for clusters within the EQ-5D-5L index distribution. We applied two search methods to determine the optimal number of clusters: the k-means cluster method and the Calinski–Harabasz pseudo-F index stopping rule. The criterion to select the optimal number of clusters is a combination of the k-means-derived within-sum-of-squares (WSS) statistic, the pseudo-F statistic, and the robustness of these two statistics when different initial values are applied.

The k-means cluster algorithm searches for the optimal partition in k clusters by minimising the within-sum-of-squares WSS(k) summed over all clusters (Makles, 2012; Dilts *et al*, 1995; McLachlan, 1992). We experimented with the number of clusters from k=1 to 20 and report four statistics for each: WSS(k); the natural log of WSS, log(WSS(k)); the proportional reduction of the WSS for each cluster solution k compared with the total sum of squares (TSS), $\eta^2(k)$; and the proportional reduction of the WSS for cluster solution k compared with the previous solution $k - 1$, (PRE(k)). We use graphs to search for kinks in the curve generated from the within sum of squares (WSS) or log(WSS), which correspond to a large proportional reduction according to $\eta^2(k)$ and PRE(k). As a sensitivity analysis, we check whether or not using the k-medians cluster method instead of the k-means cluster method affects the choice of the optimal number of clusters.

Many stopping rules are available to determine the optimal number of clusters. Milligan and Cooper (1985) evaluated 30, identifying the Calinski–Harabasz pseudo-F index (Calinski–Harabasz, 1974) and the Duda–Hart index (Duda–Hart, 2001) as the best. However, the Duda–Hart index works only for hierarchical cluster analysis, which does not apply to our dataset. We therefore calculated the pseudo-F statistic for each cluster, a larger value of which indicates more distinct clustering. The optimal cluster k is defined as the first for which the pseudo-F statistic decreases, that is F(k+1)<F(k).

Using both methods, the optimal number of clusters found may depend on the initial value specified for k in the clustering algorithm. To test this, we defined the k initial values using 50 random draws from the range of the EQ-5D-5L index distribution in our sample (MVS: -0.594 to 0.906; EVS: -0.281 to 0.951; both excluding index=1) and one initial value which assumed equal-sized partitions between clusters. The WSS and pseudo-F statistics for the optimal cluster k should be robust for different initial values in the clustering analyses. If more than two different solutions are generated among the 51 implementations of the algorithm, we do not regard the cluster to be a robust solution.

4. Results

4.1. Distributions of EQ-5D-5L profiles and index

1,730 profiles were observed in the all-patients data. Table 1 shows the number of patients in each level of each of the five dimensions for all patients and the three treatment groups. These distributions differ considerably across different dimensions and groups. For MSK, the dimension which had the largest proportion of no problems (level 1) was self-care (61.69%), while the smallest proportion was for pain/discomfort (3.63%). For specialist nursing, the largest proportion was again self-care (57.40%), but the smallest was mobility (33.10%). For community rehabilitation, the largest proportion was for anxiety/depression (45.70%) and the smallest usual activities (13.17%). The largest difference between the groups in the number reporting no problems was for pain/discomfort (33.77%) between MSK and specialist nursing, and the smallest was for anxiety/depression (10.46%) between MSK and community rehabilitation. The proportion of patients reporting level 5 is noticeably high among the community rehabilitation patients for the usual activities dimension (24.50%). It is confirmed by our clinical co-author that these patients are at that particular clinic because they cannot carry out their usual activities, which is an indicator of good face validity for the EQ-5D-5L. The characteristics of patients might be one of explanations for the multiple-modal distribution of the EQ-5D-5L index.

Table 1. Number of patients at each level of each EQ-5D dimension

All patients	All	MSK	Specialist nursing	Community rehabilitation
Mobility				
Level 1	9,463(31.25%)	7,413 (37.07%)	1,114 (33.10%)	936 (13.53%)
Level 2	7,735(25.54%)	5,392 (26.96%)	669 (19.88%)	1,674 (24.19%)
Level 3	8,321(27.48%)	4,887 (24.44%)	835 (24.81%)	2,599 (37.56%)
Level 4	4,064(13.42%)	2,165 (10.83%)	599 (17.80%)	1,300 (18.79%)
Level 5	701(2.31%)	142 (0.71%)	149 (4.43%)	410 (5.93%)
Self-care				
Level 1	16,517 (54.54%)	12,337 (61.69%)	1,932 (57.40%)	2,248 (32.49%)
Level 2	7,227 (23.86%)	4,599 (23.00%)	590 (17.53%)	2,038 (29.46%)
Level 3	4,631 (15.29%)	2,446 (12.23%)	482 (14.32%)	1,703 (24.61%)
Level 4	1,302 (4.30%)	527 (2.64%)	210 (6.24%)	565 (8.17%)
Level 5	607 (2.00%)	90 (0.45%)	152 (4.52%)	365 (5.28%)
Usual Activities				
Level 1	5,321 (17.57%)	3,209 (16.05%)	1,201 (35.68%)	911 (13.17%)
Level 2	8,733 (28.84%)	6,663 (33.32%)	741 (22.01%)	1,329 (19.21%)
Level 3	9,277 (30.63%)	6,562 (32.81%)	760 (22.58%)	1,955 (28.26%)
Level 4	4,025 (13.29%)	2,591 (12.96%)	405 (12.03%)	1,029 (14.87%)
Level 5	2,928 (9.67%)	974 (4.87%)	259 (7.69%)	1,695 (24.50%)
Pain/Discomfort				
Level 1	3,290 (10.86%)	725 (3.63%)	1,259 (37.40%)	1,306 (18.88%)
Level 2	8,159 (26.94%)	5,348 (26.74%)	961 (28.55%)	1,850 (26.74%)

Level 3	11,918 (39.35%)	8,739 (43.70%)	763 (22.67%)	2,416 (34.92%)
Level 4	5,736 (18.94%)	4,340 (21.70%)	312 (9.27%)	1,084 (15.67%)
Level 5	1,181 (3.90%)	847 (4.24%)	71 (2.11%)	263 (3.80%)
Anxiety/Depression				
Level 1	15,968 (52.73%)	11,231 (56.16%)	1,575 (46.79%)	3,162 (45.70%)
Level 2	7,783 (25.70%)	4,880 (24.40%)	964 (28.64%)	1,939 (28.02%)
Level 3	4,707 (15.54%)	2,802 (14.01%)	599 (17.80%)	1,306 (18.88%)
Level 4	1,272 (4.20%)	778 (3.89%)	156 (4.63%)	338 (4.89%)
Level 5	554 (1.83%)	308 (1.54%)	72 (2.14%)	174 (2.51%)
N	N=30,284	N=19,999	N=3,366	N=6,919

Figures 3 and 4 show the distributions of MVS- and EVS-based EQ-5D-5L index values for all patients and the three treatment groups, with imposed kernel estimates. Skewness and kurtosis tests reject the normality hypothesis for each of these distributions at the 1% level.

All of the distributions are negatively skewed. For the MVS, there is a noticeable gap between the values of 11111 and the next highest health state and, for community rehabilitation patients only, the kernel estimates suggest a bimodal distribution, similar to the two-group distributions found in EQ-5D-3L data, although there is no observable gap between these groups. There also appears to be a noticeable change in the shape of the distribution around the value 0.5. None of these observations is apparent for the EVS data. The modelling for the EVS but not the MVS took account of the right censored nature of the EQ-5D valuation data. This difference explains the much smaller gap between the value of 11111 and the two second best health states (12111 and 21111) in the EVS than the gap between the value of 11111 and the second best health state (11211) in the MVS. These observations suggest that the MVS values may inherit some of the characteristics of the EQ-5D-3L values on which they are based.

Figure 3. Distributions of MVS-based EQ-5D-5L values for all patients and three treatment groups

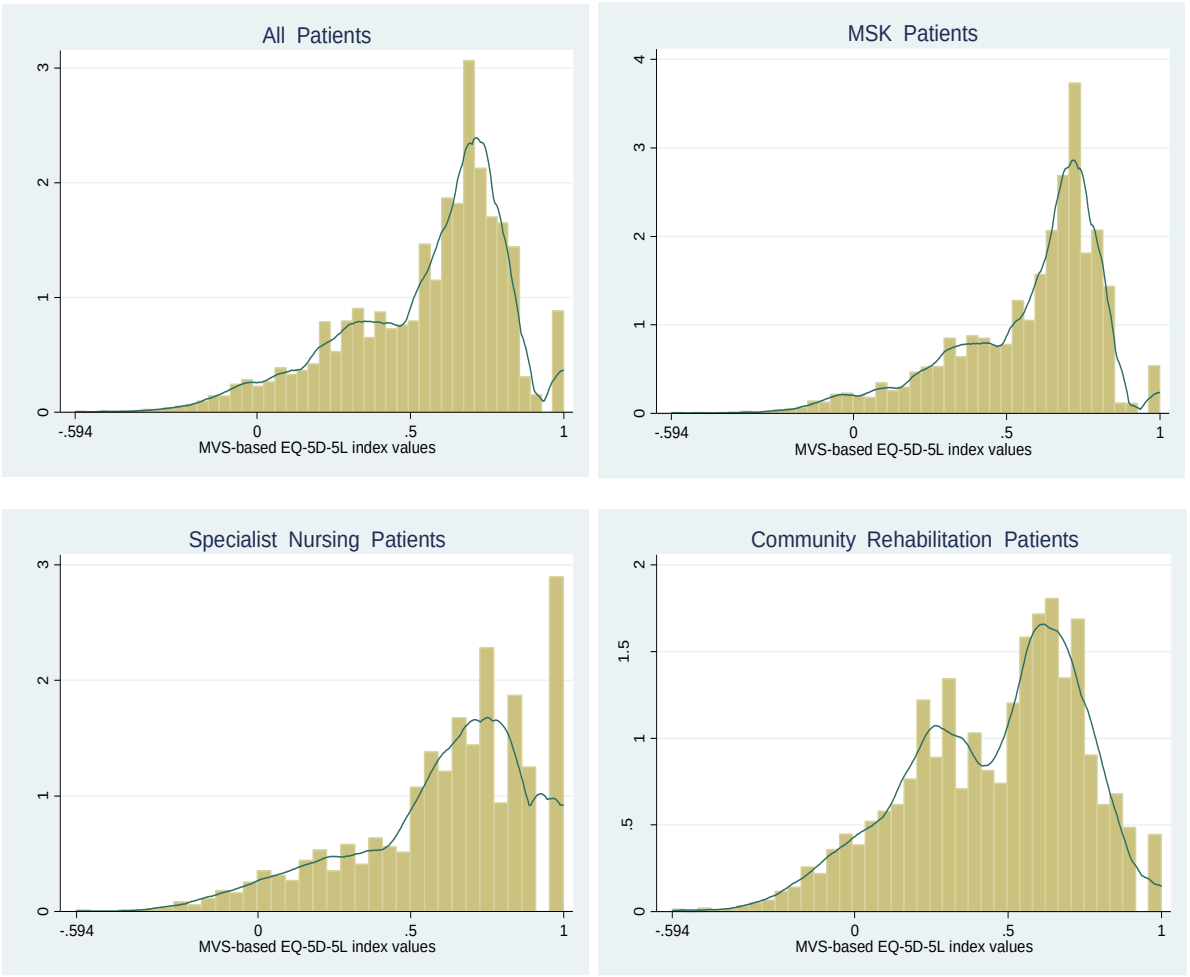
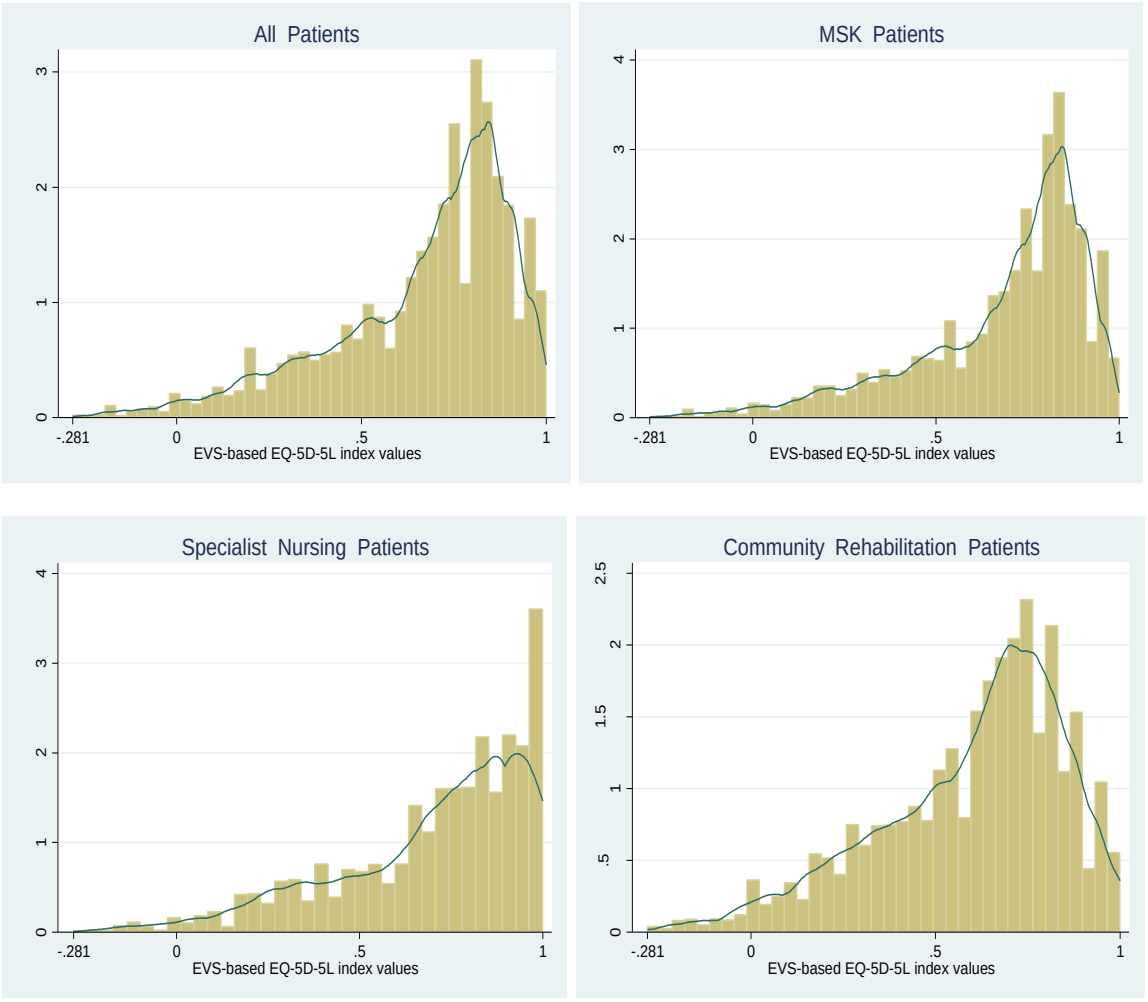


Figure 4. Distributions of EVS-based EQ-5D-5L values for all patients and three treatment groups



4.2. Searching for clusters of profiles

We used two ways to examine whether the composition of profiles and their values for the full set of EQ-5D-5L profiles is likely in itself to generate clusters of values within real datasets, for both the EVS and the MVS.

First, Table 2 shows how the presence of particular levels within dimensions affects the values that profiles have. Profiles were partitioned according to the presence or absence in any dimension of level 5, levels 4 and 5, and levels 3, 4 and 5. In each line of the table, we report in the third and fourth columns the lowest value taken by a profile that does not contain the level or levels and the highest value taken by a profile that does. The fifth column shows the number of profiles that take a value below the lowest for those that do not contain the level or levels, which by definition all do contain them. The eighth column similarly shows the number of profiles that take a value above the highest for those that contain the level or levels, all of which do not. Columns six and seven show the extent of overlap, the number of profiles whose values lie within these lowest

and highest values, for those that do and do not contain the level or levels. The final columns show the mean values for profiles according to whether they do or do not contain the level or levels.

The extent of the overlaps between profiles that do or do not contain worse levels does not suggest any obvious clustering. For profiles that do not have worse levels, many more have values that are within the range of values taken by profiles that do have them, rather than are above that range. Profiles that contain worse levels also form the large majority of those whose values lie in the range taken by profiles that do not.

Table 2. Impact on the distribution of health state values of the presence of worse levels

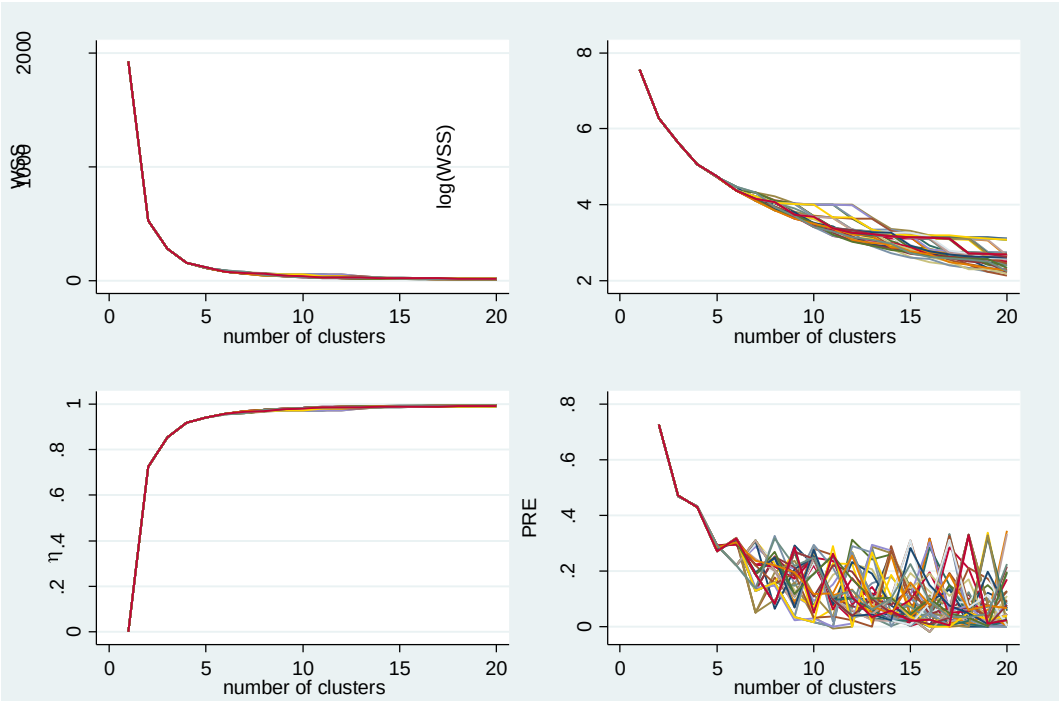
Levels	Value set	Lowest value without levels	Highest value with levels	Number ≤ lowest without	Number between lowest without and highest with		Number ≥ highest with	Mean value with levels	Mean value without levels
					with level	without level			
5	EVS	-0.102	0.816	47	2053	946	79	0.309	0.559
	MVS	0.036	0.556	1003	1099	654	369	0.037	0.484
4,5	EVS	0.628	0.832	2644	239	188	54	0.358	0.780
	MVS	0.516	0.813	2665	217	225	18	0.142	0.676
3,4,5	EVS	0.712	0.939	2870	223	26	6	0.386	0.856
	MVS	0.592	0.883	2820	272	30	3	0.178	0.762

Second, we ranked profiles by their values and calculated the difference between adjacent profiles, to see if there are any obvious 'gaps' in the distribution of values. Most differences are small; we used as an illustration a search for differences > 0.01. For the MVS, there are nine of these, the largest of which (0.094) is between health states 11111 (value=1) and 11211 (value = 0.906). Of the others, two are also close to the highest value, and six are close to the lowest value. For the EVS, there are six differences >0.01, the biggest of which (0.049) is between health states 11111 (value = 1) and 12111 (value = 0.951). Another is close to the highest index value, and four are close to the lowest value. Again, these data do not suggest any obvious clusters, just a slight spreading out of the distribution at the extremes.

4.3. Searching for clusters of index values

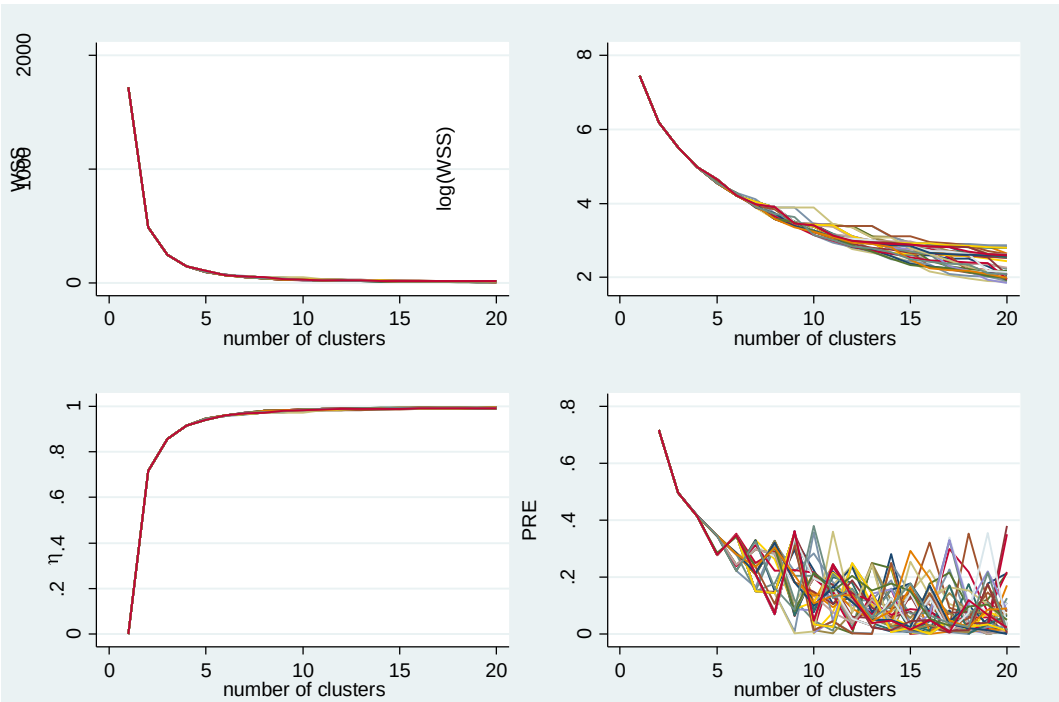
Figures 5 and 6 show the four key statistics from the k-means cluster analyses for all patients. The results are similar for the MVS indices and EVS indices (results for the three treatment groups are available from the authors). A summary of the general trends for four statistics between the two indices is shown as below.

Figure 5. K-means cluster analyses statistics for MVS-based EQ-5D-5L values for all patients



Kmean algorithm. 50 random draws of initial k values from sample, and 1 initial values from equal partition

Figure 6. K-means cluster analyses statistics for EVS-based EQ-5D-5L values for all patients



Kmean algorithm. 50 random draws of initial k values from sample, and 1 initial values from equal partition

WSS(k) decreases, as it must, with the increasing in the number of clusters, while $\eta^2(k)$ increases. These two statistics become stable when $k \geq 5$. Both log(WSS) and PRE

statistics show that WSS(k) is affected by the initial value assigned to the number of clusters, in particular when $k > 5$.

Tables 3 and 4 show the results of the sensitivity analyses using different initial values. The second row shows the stopping rule clusters k^* for the first decrease in pseudo-F. This stopping rule does not result in a unique optimal k^* but depends on the choice of initial values. Therefore, we present a range of partitions in k clusters for the MVS and the EVS, which varies from 5 to 19 and 6 to 18 respectively. This variability of stopping rule clusters is also observed for the three treatment groups.

Table 3. Calinski–Harabasz pseudo-F statistics from the MVS-based EQ-5D-5L indices

	All Patients	MSK	Specialist Nursing	Community Rehabilitation
Stopping rule clusters k*	6-15	5-18	6-18	8-19
Number of solution (k=2)	1	2 (1)	1	3
Number of solution (k=3)	3	3	2 (2)	2 (4)
Number of solution (k=4)	4	4	2 (3)	2 (5)
Number of solution (k=5)	>5	>5	>5	4
Robust clusters k	2	2	2	4

(1) Number of patients switched from high cluster to low cluster is 46.
(2) Number of patients switched from high cluster to middle/low clusters is 101.
(3) Number of patients switched from highest cluster to middle clusters is 45. The lowest cluster is stable.
(4) Number of patients switched from high and middle clusters to low cluster is 31.
(5) Number of patients switched between the two high clusters (from 3 to 4) is 14. The two low clusters (1 and 2) are stable.

Table 4. Calinski–Harabasz pseudo-F statistics from EVS-based EQ-5D-5L indices

	All Patients	MSK	Specialist Nursing	Community Rehabilitation
Stopping rule clusters k*	6-16	6-13	7-18	8-16
Number of solution (k=2)	2 (1)	1	1	1
Number of solution (k=3)	2 (2)	3	4	5
Number of solution (k=4)	2 (3)	3	1	5
Number of solution (k=5)	5	>5	4	4
Robust clusters k	3	2	4	2

(1) Number of patients switched from high cluster to low cluster is 37.
(2) Number of patients switched from middle cluster to low cluster is 13. The high cluster is stable.
(3) Number of patients switched from the lowest cluster to the other three clusters is 346.

The WSS and pseudo-F index statistics are never robust to different initial values for the whole sample when $k > 5$. For clusters $k \leq 5$, the larger the clusters, the more distinct they are. The last row of Tables 3 and 4 show the robust value when $k \leq 5$, based on the pseudo-F index.

Table 3 suggests that the solution for the MVS index for all patients is unique, and therefore robust, for each starting point when $k = 2$, but for $k = 3$, 4 or 5, there are more than two unique solutions. For MSK, there are two solutions when $k = 2$; 46 patients were in different clusters in the two solutions, comprising 0.23% of the MSK sample. For specialist nursing, there is a unique solution for two clusters. For three and four clusters, each has two solutions, and the proportion of patients who switched between different clusters is 3.46% for three clusters and 1.54% for four clusters. For community rehabilitation, three and four clusters are both robust. They both have two solutions, with few patients switching between clusters when comparing the two solutions: 0.46% for three clusters and 0.21% for four clusters. Four clusters is therefore slightly more robust.

Table 4 suggests that for the EVS index for all patients, the most robust number of clusters is three. Two, three and four clusters each had two solutions, but three clusters reported the smallest number of switchers (0.13% for two clusters; 0.04% for three; and 1.18% for four). For MSK, two clusters was robust as this gave a unique solution for all 51 starting values. There were more than two solutions for greater numbers of clusters. For specialist nursing, four clusters was robust. Two and four clusters both generated unique solutions, but four clusters had more distinct partitions according to the pseudo-F statistic. For community rehabilitation, two clusters generated a unique solution, but greater numbers of clusters did not generate any robust solutions.

Using k-medians instead of k-means gave similar results, reported in Appendices 1 and 2, and did not demonstrate greater robustness.

The summary statistics for each cluster of the MVS- and EVS-based indices are reported in Table 5 and Table 6. The statistics of clusters for all patients are only directly comparable between the MVS and the EVS for the MSK patients, as the others have different numbers of clusters. An explanation for these different numbers is that the EVS has a narrower range and is more uniformly distributed than the MVS. With large numbers of observations, some might be grouped into a neighbouring cluster and are less easily identified. For the MSK patients, the thresholds of indices between clusters are higher for the EVS [0.569, 0.571] than the MVS [0.461, 0.463]. There are also slightly more patients in the lowest cluster for the MVS, and fewer in the highest cluster. The difference in the means of the clusters is very similar, though of course the EVS has a smaller range than the MVS.

Table 5. Summary statistics for clusters from the MVS-based indices

	All Patients			MSK Patients			Specialist Nursing Patients			Community Rehabilitation Patients				
	K=2			K=2			K=2			K=4				
Cluster	1	2	Total	1	2	Total	1	2	Total	1	2	3	4	Total
N	9048	20266	29314	5394	14208	19602	791	2131	2922	922	1966	2216	1686	6790
min	-0.594	0.445	-0.594	-0.594	0.463	-0.594	-0.594	0.439	-0.594	-0.594	0.114	0.419	0.660	-0.594
max	0.444	0.906	0.906	0.461	0.906	0.906	0.434	0.906	0.906	0.112	0.417	0.659	0.906	0.906
mean	0.209	0.681	0.535	0.236	0.688	0.564	0.171	0.698	0.556	-0.047	0.275	0.562	0.758	0.445
median	0.249	0.691	0.617	0.290	0.703	0.642	0.206	0.708	0.635	-0.021	0.277	0.568	0.740	0.511
SD	0.183	0.105	0.256	0.180	0.098	0.238	0.189	0.121	0.274	0.128	0.081	0.066	0.065	0.277
skewness	-1.008	-0.210	-1.036	-1.119	-0.325	-1.227	-0.768	-0.194	-1.041	-1.194	-0.049	-0.457	0.545	-0.593
kurtosis	3.755	2.306	3.548	4.070	2.430	4.163	3.206	2.019	3.452	4.704	1.969	2.205	2.310	2.703
range	1.038	0.461	1.500	1.055	0.443	1.500	1.028	0.467	1.500	0.706	0.303	0.240	0.246	1.500

Table 6. Summary statistics for clusters from the EVS-based indices

	All Patients				MSK Patients			Specialist Nursing Patients					Community Rehabilitation Patients		
	K=3				K=2			K=4					K=2		
Cluster	1	2	3	Total	1	2	Total	1	2	3	4	Total	1	2	Total
N	4256	8510	16548	29314	5108	14494	19602	253	584	899	1186	2922	2075	4715	6790
min	-0.281	0.375	0.692	-0.281	-0.281	0.571	-0.281	-0.281	0.256	0.548	0.782	-0.281	-0.281	0.502	-0.281
max	0.374	0.69	0.951	0.951	0.569	0.951	0.951	0.252	0.547	0.78	0.951	0.951	0.5	0.951	0.951
mean	0.189	0.56	0.824	0.655	0.345	0.796	0.678	0.098	0.409	0.687	0.876	0.657	0.273	0.727	0.589
median	0.212	0.566	0.823	0.728	0.387	0.813	0.752	0.122	0.404	0.696	0.874	0.731	0.309	0.728	0.652
SD	0.146	0.091	0.071	0.242	0.183	0.095	0.234	0.127	0.087	0.062	0.05	0.253	0.17	0.117	0.249
skewness	-0.954	-0.32	0.038	-1.146	-0.947	-0.353	-1.301	-0.902	-0.084	-0.444	-0.05	-1.034	-0.859	-0.004	-0.867
kurtosis	3.269	1.882	2.004	3.809	3.299	2.282	4.28	2.876	1.712	2.223	1.686	3.417	3.172	2.173	3.209
range	0.655	0.315	0.258	1.232	0.85	0.379	1.232	0.533	0.291	0.232	0.168	1.232	0.781	0.449	1.232

5. Discussion

The patients whose data were analysed in this paper differ considerably from those analysed by Parkin *et al* (2016). Our comparison of the three- and five-level versions of the EQ-5D therefore assumes that each dataset demonstrates general characteristics of the EQ-5D indexes, rather than of particular patient groups. There was no obvious evidence that EQ-5D-5L profiles themselves resulted in clustering, unlike the EQ-5D-3L profiles. Both studies found non-normal distributions of values, and as with the three-level data, our study found two distinct clusters, but only for the MVS.

Although clustering was found for both the MVS and EVS, they generated different clusters. The results were much more clear for the MVS in the sense of resulting in two optimal clusters, and as suggested were similar to the clustering of the EQ-5D-3L. A possible explanation is that the MVS inherits several important characteristics of the EQ-5D-3L value set, in particular the range of possible index values i.e. between -0.594 and 1.

Cluster analysis proved to be a useful exploratory tool, but a limitation of its use in our study is that arbitrary judgements are involved in deciding the number of robust clusters. For the EVS index for all patients, there was no $k \leq 5$ that gave a unique solution. There were two solutions for two, three or four clusters. We regarded three clusters as a robust solution because it reported the smallest number who switched between different clusters (0.04%).

This study demonstrates again the importance of undertaking careful exploratory analysis of EQ-5D data before its use in different applications, such as health technology assessment and health care management processes involving patient reported outcome measurement. Any statistical techniques used should take account of features of the distribution of the data such as clustering, to make sure that inferences drawn are valid and reliable.

Heterogeneity of patient response is observed in many situations (Willke *et al*, 2012; Greenfield *et al*, 2007; Kravitz *et al*, 2004). Individual patients might self-select into specific treatments based on observed and unobserved characteristics that cause patients to respond to the same treatment differently (Basu *et al*, 2007). A further potential use of cluster analysis of EQ-5D data, as used in this study, is to provide a means of identifying distinct groups of patient pre-treatment and post-treatment, and to use that information to predict which patients might benefit the most from treatment and to investigate if there are groups of patients where it appears treatment is less successful.

References

Appleby, J., Devlin, N. and Parkin, D., 2016. *Using Patient Reported Outcomes to Improve Health Care*. Wiley. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1118948602.html.

Basu, A., Heckman, J.J., Navarro-Lozano, S. and Urzua, S., 2007. Use of instrument variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health Economics*, 16(11), pp. 1133-1157.

Calinski, T. and J. Harabasz., 1974. A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), pp.1–27.

Craig, B.M., Pickard, A.S. and Lubetkin, E.I., 2014. Health problems are more common, but less severe when measured using newer EQ-5D versions. *Journal of Clinical Epidemiology.* 67(1), pp.93–99.

Devlin, N. and Brooks, R., 2016. EQ-5D past, present and future. *Journal of Applied Health Economics and Health Policy*. (forthcoming).

Devlin, N., Shah, K., Feng, Y., Mulhern, B. and van Hout, B., 2016. Valuing health related quality of life: an EQ-5D-5L value set for England. OHE Research Paper. London: Office of Health Economics.

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

Dilts, D., Khamalah, J. and Plotkin, A. 1995. Using cluster analysis for medical resource decision making. *Medical Decision Making,* 15(4), pp. 333–47.

Duda, R.O., Hart, P.E. and Stork, D.G., 2001. Pattern Classification. 2nd ed. New York: Wiley.

Feng, Y., Devlin, N. and Herdman, M., 2015. Assessing the health of the general population in England: how do the three- and five-level versions of EQ-5D compare? *Health and Quality of Life Outcomes*, 13(171).

Greenfield, S., Kravitz, R., Duan, N. and Kaplan, S., 2007. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. *The American Journal of Medicine*, 120(4A), pp. S3–S9.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M.F., Kind, P., Parkin, D., Bonsel, G. and Badia, X., 2011. Development and preliminary testing of the new five-level version of EQ- 5D (EQ-5D-5L). *Quality of Life Research*, 20(10), pp.1727-1736.

Kravitz, R.L., Duan, N. and Breslow, J., 2004. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Quarterly*, 82(4), pp. 661–687.

Makles, A., 2012. Stata tip 110: How to get the optimal k-means cluster solution. *The Stata Journal,* 12(2), pp. 347–351.

McLachlan, G.J., 1992. Cluster analysis and related techniques in medical research. *Statistical Methods Medical Research*, 1(1), pp. 27–48.

Milligan, G.W. and Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a dataset. *Psychometrika*, 50(2), pp.159–179.

Parkin, D., Devlin, N. and Feng, Y., 2016. What determines the shape of an EQ-Index distribution? *Medical Decision Making*. [Epub ahead of print]

Ramos-Goñi, J.M., Pinto-Prades, J.L., Cabasés, J.M. and Rivero-Arias, O., 2015. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Medical Care*. [Epub ahead of print]

Szende, A., Oppe, M. and Devlin, N., 2007. EQ-5D value sets: inventory, comparative review and user guide. Dordrecht: Springer.

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 in Health*, 15, pp.708-715.

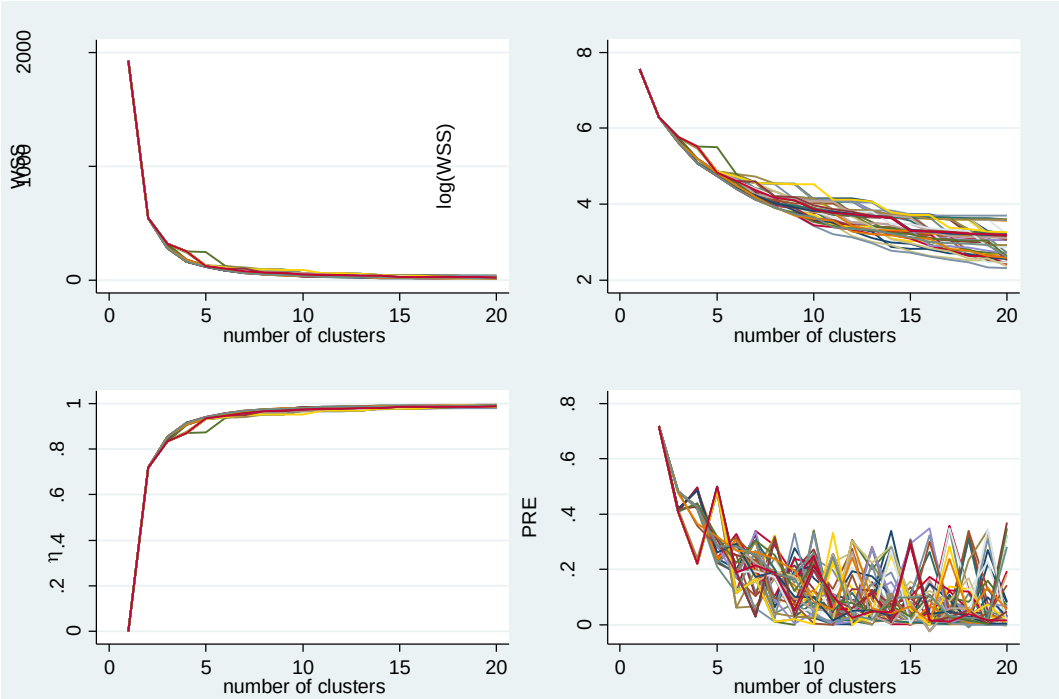
Versteegh, M.M., Vermeulen, K.M., Evers, S.M.A.A., de Wit, G.A., Prenger, R., Stolk, E.A., 2016. Dutch Tariff for the Five-Level Version of EQ-5D. *Value in Health,* 19(4) pp.343-352.

Willke, R.J., Zheng, Z., Subedi, P., Althin, R. and Mullins, C.D., 2012. From concepts, theory, and evidence of heterogeneity of treatment effects to methodological approaches: a primer. *BMC Medical Research Methodology*, 12(185).

Xie, F., Pullenayegum, E., Gaebel,K., Bansback, N., Bryan, S., Ohinmaa, A., Poissant, L., and Johnson, J.A., 2016. Canadian EQ-5D-5L Valuation Study Group. A Time Trade-off derived Value Set of the EQ-5D-5L for Canada. *Medical Care*, 54(1), pp.98-105.

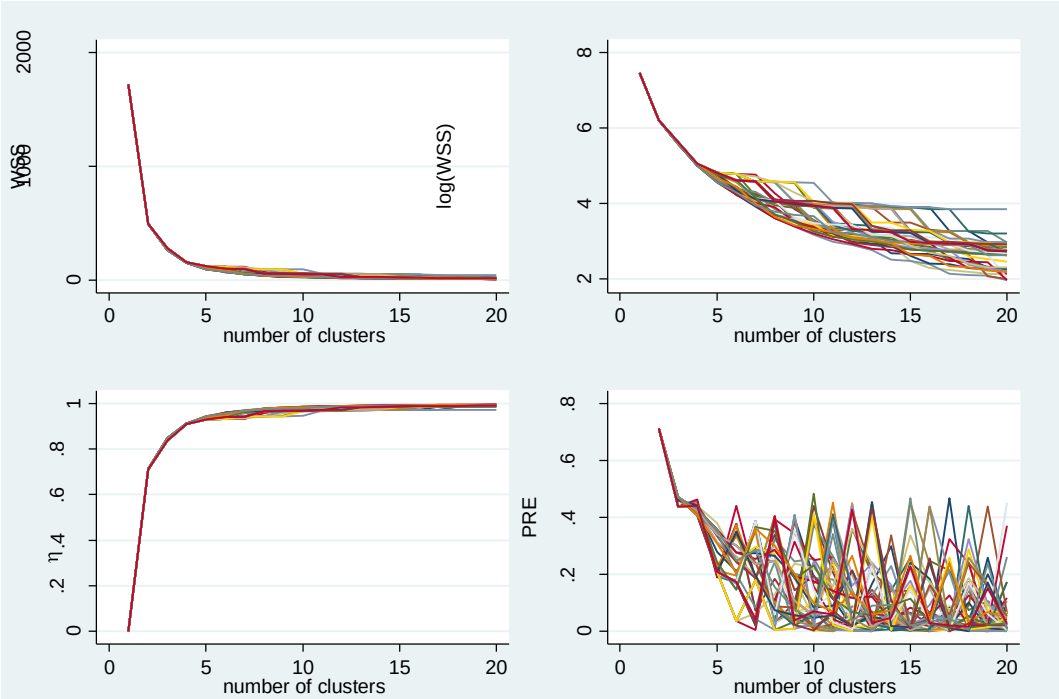
Appendix

Appendix 1. K-medians cluster analyses statistics for MVS-based EQ-5D-5L values for all patients



Kmedian algorithm. 50 random draws of initial k values from sample, and 1 initial values from equal partition

Appendix 2. K-medians cluster analyses statistics for EVS-based EQ-5D-5L values for all patients



Kmedian algorithm. 50 random draws of initial k values from sample, and 1 initial values from equal partition