Chapter 2
Analysis of EQ-5D Profiles

The aims of this chapter are

- to demonstrate a variety of analyses that can be performed on the profile data generated from the EQ-5D instruments: EQ-5D-3L, EQ-5D-5L and EQ-5D-Y;
- to explain methods that can be used to describe EQ-5D profile data in cross-sectional (collected at a single point in time) and longitudinal (describing changes over time) designs; and
- to consider the advantages and limitations of each method, and outline in which decision contexts insights from them might be useful.

Profile data form the cornerstone of analyses of EQ-5D data and, in many cases, are likely to be the primary focus of interest. In this chapter, we provide an overview of methods that can be used to describe the profile data from respondents at a given point in time, and to describe the changes in profiles between different points in time.

Even when the ultimate goal of analysis is to generate EQ-5D values and to estimate quality-adjusted life years (QALYs), analysis of profile data provides important insights and should always be the starting point for analysts. For example, summarising EQ-5D patient data simply as values obscures the underlying information about which aspects of their health have been most affected by their condition, or improved by treatment. To know about that, you need to look at the data that respondents have given you: the boxes they ticked on an EQ-5D questionnaire.

The methods presented here need not be treated as alternatives, but rather as complementary. Although they are illustrated using EQ-5D data, and in some cases developed specifically for the analysis of EQ-5D profile data, these same methods could just as readily be applied to other generic or condition specific health status or patient reported outcome (PRO) measures.

It should be noted that we do not cover inferential statistics, either hypothesis testing or estimation, as this book is not intended as a statistical primer and we assume that readers will be able to apply appropriate inference procedures where required. For example, we describe contingency tables, to which measures of association such as a $\chi 2$ test could be applied.

2.1   Cross-Sectional Analysis: Describing Health at a Point in Time by Dimension and Level

Exploratory data analysis (EDA) of EQ-5D data, including the use of simple descriptive statistics, is undervalued, and often underreported in papers that contain more complex econometric and psychometric analyses. This is bad practice and wasteful of information, because EDA not only generates information that helps in interpreting more complex analyses, but also generates information about health within populations and about the properties of the EQ-5D which is valuable in itself.

Describing health at the most detailed level possible for the EQ-5D can be done very simply, by reporting the number and percentage of patients reporting each level of problem on each dimension of the EQ-5D. An example of this is shown in Table 2.1, which shows EQ-5D-3L data provided by patients before and after hip surgery, using data from a pilot study for the Patient Reported Outcome Measures (PROMs) programme in the English National Health Service (NHS) (Devlin et al. 2010).

This very simple table provides some important information. For example, before hip surgery, 420 of these patients (95.7% of the sample) reported a level 2 problem on mobility, but none reported a level 3 problem. The reason is that Level 3 on the EQ-5D-3L mobility dimension is 'confined to bed'—and even patients with very poor mobility because of hip problems aren't confined to bed. That is a problem with the EQ-5D-3L—as has been pointed out previously (Oppe et al. 2011). This issue has been corrected in the EQ-5D-5L (Herdman et al. 2011), where the most severe problem with mobility is 'unable to walk about', and is an important advantage of the 5L over the 3L (Janssen et al. 2018).

The information on the types of problems experienced by a sample of patients at any given point in time can be simplified still further by collapsing levels together, to create just two categories: the number and percentage of patients reporting no problems (level 1), and the number reporting any level of problems (levels 2 and 3 for the 3L, and levels 2, 3, 4 and 5 for the 5L). This can also be seen in Table 2.1. For example, before surgery, mobility problems are common in these patients, as might be expected: only 4.3% of these patients had no problems with mobility. Of the 95.7% of patients who reported having at least some problem on mobility before surgery, all reported a level 2, as noted above. However, problems on other dimensions are just as prevalent: 99.8% of patients reported at least some problem with pain and discomfort, and 96.6% at least some problem with usual activities. Over 40% of these patients also reported problems with anxiety and depression—something that might be missed by condition specific instruments focused on mobility and function-related issues specific to hips, such as the Oxford Hip Score.

Examining the profile data by each dimension and level in this manner is a good starting point to understanding the nature of the health problems reported in the data you have collected. However, there are limitations to this way of reporting the data. Because the focus is on the frequency of observations in each level *within* each dimension, it doesn't tell us how these problems combine in the people reporting

Table 2.1 Frequency of levels by dimension of 'some problems', before and after elective hip surgery in the English NHS

Level	Mobility		Self-care		Usual activities		Pain and discomfort		Anxiety and depression	
	Pre-op	Post-op	Pre-op	Post-op	Pre-op	Post-op	Pre-op	Post-op	Pre-op	Post-op
1	19 (4.3%)	239 (54.4%)	168 (38.6%)	319 (73.3%)	15 (3.4%)	199 (45.6%)	1 (0.2%)	219 (50.5%)	240 (55.5%)	349 (80.8%)
2	420 (95.7%)	200 (45.6%)	264 (60.7%)	115 (26.4%)	347 (79.6%)	221 (50.7%)	240 (55.3%)	200 (46.1%)	183 (42.4%)	74 (17.1%)
3	0 (0%)	0 (0%)	3 (0.7%)	1 (0.2%)	74 (17.0%)	16 (3.7%)	193 (44.5%)	15 (3.4%)	9 (2.1%)	9 (2.1%)
Total[a]	439 (100%)	439 (100%)	435 (100%)	435 (100%)	436 (100%)	436 (100%)	434 (100%)	434 (100%)	432 (100%)	432 (100%)
Number reporting some problems[b]	420 (95.7%)	200 (45.6%)	267 (61.4%)	116 (26.6%)	421 (96.6%)	237 (54.4%)	433 (99.8%)	215 (49.5%)	192 (44.5%)	83 (19.2%)
Change in numbers reporting problems	−220		−151		−184		−218		−109	
% change in numbers reporting problems	−52%		−57%		−44%		−53%		−57%	
Rank of dimensions in terms of % changes	3		1		4		2		1	

[a]Results are for those who responded to both the pre- and the post-operative questionnaire. 84% of respondents to the pre-operative EQ-5D also responded to the post-operative EQ-5D

[b]'Some problems' = levels 2 + 3

Source Devlin et al. (2010)

them. For example, are the people who report a level 3 on Anxiety and Depression also the same people who report a level 3 on Usual Activities? For this reason, it is also important to examine the way that observed levels of problems on each dimension combine into EQ-5D profiles, which is covered in Sect. 2.3.

2.2 Longitudinal Analysis: Describing Changes in Health Between Two Time Points by Dimension and Level

In addition to describing health states at any one point in time, if you have collected EQ-5D profile data at more than one time point, you are likely also to be interested in describing the changes between them—for example, before and after surgery, or between various time points in a clinical trial, compared to baseline. This too can be done at the level of the EQ-5D dimensions, as is also shown in Table 2.1.

'Eyeballing' the differences in numbers and percentages of patients in each of the levels tells us about the nature of the changes in health that resulted from surgery. For example, the results in Table 2.1 show there were quite striking improvements in patients' Anxiety and Depression, Self-care and Pain and Discomfort—not just Mobility. And because of the issue with level 3 Mobility noted above, whereby the worst level of problem these patients were likely to report on mobility was level 2, the only improvements to mobility that were possible as a result of hip replacement surgery were from 'some' to 'no' problems. This issue with the use of the EQ-5D-3L to measure health outcomes from hip surgery would not have been apparent if these patients' data had been analysed just in terms of EQ-5D values.

It can however be difficult readily to get an overall picture of improvements, even for these relatively simple EQ-5D-3L data. As with the analysis of cross-sectional data, this does not summarise the extent of improvement across dimensions. As noted in Sect. 2.1, one way of handling this is to collapse the levels into just two categories: no problems and some problems. The shift between these two categories provides a simpler way of capturing change. The change in health between time points, reported in this manner, provides a way of summarising the overall extent to which patients go from any level of problem to no problem within each dimension. This may be useful in some contexts, but it has some limitations as an indicator of improvement because of the loss of information caused by aggregation of levels. It doesn't capture improvements other than shifts to no problem, so other improvements that may be of value to patients, for example from extreme to moderate problems, are not captured. That means, that if applied to the EQ-5D-5L, the advantages of its more refined descriptive system will be lost.

2.3   Cross-Sectional Analysis: Describing Health at a Point in Time Using Profiles

While describing the number and percentage of observed levels within each dimension (as in Table 2.1) gives very useful information dimension-by-dimension, it does not tell you anything about the way these problems are *combined* in the health states reported by patients.

One of the most simple and instructive things you can do with an EQ-5D profile data set is to report the cumulative frequency of these profiles. This will reveal the extent to which your observations are evenly distributed over many profiles, or instead concentrated on a relatively small number of health profiles.

The results can sometimes be quite surprising. For example, in Table 2.2 we show the cumulative frequency of self-reported EQ-5D-3L profiles reported by 7294 respondents in the 2012 Health Survey for England. In this example, the great majority of respondents self-reported their health using only a small number of profiles. The top three most frequently reported profiles represented almost three quarters of the respondents.

In contrast, Table 2.3 shows the cumulative frequency of profiles reported by 996 respondents from the general public in the EQ-5D-5L value set study for England for their self-reported health on the EQ-5D-5L. This shows, in comparison to Table 2.2, a larger number of unique health states observed in this data set, and the observations are less concentrated on a small number of states. A large proportion of observations are accounted for by profile 11111 (no problems on any dimension) in both data sets, which is not surprising given that both samples comprise members of the general public, many of whom would not regard themselves as ill. But in general, this 'ceiling effect' is somewhat less in the EQ-5D-5L data (Devlin et al. 2018). Obviously, the

Table 2.2 Prevalence of the 10 most frequently observed self-reported health states and frequency of reporting of the worst possible health state in EQ-5D-3L

Health states	Frequency (%)	Cumulative frequency (%)
11111	4096 (56.2)	56.2
11121	855 (11.7)	67.9
11112	496 (6.8)	74.7
11122	241 (3.3)	78.0
21221	224 (3.1)	81.0
21121	222 (3.0)	84.1
21222	138 (1.9)	86.0
11221	103 (1.4)	87.4
11222	67 (0.9)	88.3
22221	64 (0.9)	89.2
…		
33333	4 (0.1)	100

Source Feng et al. (2015)

Table 2.3 Prevalence of the 10 most frequently observed self-reported health states and frequency of reporting of the worst possible health states in EQ-5D-5L

Health States	Frequency (%)	Cumulative Frequency in (%)
EQ-5D-5L		
11111	474 (47.6)	47.6
11121	93 (9.3)	56.9
11112	46 (4.6)	61.6
11131	22 (2.2)	63.8
21121	21 (2.1)	65.9
11122	21 (2.1)	68.0
21221	19 (1.9)	69.9
11123	13 (1.3)	71.2
21111	11 (1.1)	72.3
11221	11 (1.1)	73.4
…		
55555	0 (0.0)	100

Source Feng et al. (2015)

states observed and their cumulative frequency will differ from data set to data set, but in general the EQ-5D-5L yields less concentrated data, reflecting the advantages of the larger number of response options.

Understanding these patterns of observations in your data is important for three reasons:

(i) The way self-reported health problems are combined may be useful, as a complement to clinical information, for understanding and planning for patients' treatment needs.

(ii) The combination of problems into health profiles determines the distribution of EQ-5D values data. For example, Parkin et al. (2016) show that the clustering of observations on particular EQ-5D-3L profiles contributes to the unusual 'two group' distribution that is often seen in EQ-5D-3L values data.

(iii) The characteristics of the distribution of problems at baseline may have important implications for the potential for and nature of health improvements that can be observed at later time points.

Looking at the cumulative frequency is a simple and effective way of getting an insight into the distribution of health profiles in a data set. However, a limitation is that it does not provide a summary statistic that allows us readily to (a) describe how good or bad the health states are, or (b) the extent to which the observations cluster on just a few health states, or are evenly spread out over the available heath states described by the descriptive system. Having a summary statistic to characterise the degree to which there is clustering or dispersion of observed health states is useful, especially if one wanted to compare this characteristic, for example to find out whether there are changes in the distribution of profile data from a group of patients observed at

different time points, or between EQ-5D profile data from patients with different conditions.

2.4 Longitudinal Analysis: Describing Changes in Health Between Two Time Points Using Profiles

Descriptive analyses of profile data such as Table 2.1 can be very useful, but they contain a lot of information and sometimes an overall summary is required. One way of summarising profile data is to generate a single number for each profile using weights, for example using value sets. However, as noted in Chap. 1, this introduces possible problems of information loss and bias. The good news is that there are ways of summarising changes in EQ-5D health status without using value sets, just using the data that respondents have given you.

2.4.1 The Paretian Classification of Health Change (PCHC)

Devlin et al. (2010) introduced a way of summarising changes in profile data called the Paretian Classification of Health Change (PCHC). The approach is based on the principles of a Pareto improvement in Welfare Economics, drawing an analogy with the challenge of summing up changes in utility of different individuals, where utility can be measured only in ordinal terms. The idea is simple: an EQ-5D health state is deemed to be 'better' than another if it is better on at least one dimension and is no worse on any other dimension. And an EQ-5D health state is deemed to be 'worse' than another if it is worse in at least one dimension and is no better in any other dimension. Using that principle to compare a person's EQ-5D health states between any two time-points, there are only four possibilities:

(i) Their health state is better
(ii) Their heath state is worse
(iii) Their health state is the same
(iv) The changes in health are 'mixed': better in at least one dimension, but worse in at least one other.

Applying this to the English NHS PROMs pilot hip replacement data, we found that under 5% had no change, 82% had improved health, under 5% had worse health, and under 10% had a 'mixed' change (Devlin et al. 2010). In other words, this simple analysis provides a very clear summary of what is happening to patients' health because of hip surgery—without relying on value sets. It also highlighted important differences in the benefits from hip surgery, compared with the other types of elective surgery analysed in the English NHS PROMs pilot, shown in Tables 2.4 and 2.5. Looking at Table 2.4, hip replacement operations were by far the best in terms of success in reducing the number of patients who had problems, with knee

Table 2.4 Changes in health for five surgical procedures according to the PCHC

	Hip	Knee	Hernia	Veins	Cataract
No change	21 (4.7%)	45 (10.0%)	127 (29.5%)	72 (27.1%)	335 (47.1%)
Improve	356 (82.0%)	329 (73.3%)	203 (47.2%)	148 (55.6%)	149 (21.0%)
Worsen	18 (4.2%)	34 (7.6%)	71 (16.5%)	34 (12.8%)	188 (26.4%)
Mixed change	39 (9.0%)	41 (9.1%)	29 (6.7%)	12 (4.5%)	39 (5.5%)
Total	434	449	430	266	711

Source Devlin et al. (2010)

Table 2.5 Changes in health state for three conditions according to the PCHC, taking account of those with no problems

	Hernia	Veins	Cataract
Number with problems (% of those with problems)			
No change	53 (14.9%)	29 (13.0%)	99 (20.8%)
Improve	203 (57.0%)	148 (66.4%)	149 (31.4%)
Worsen	71 (19.9%)	34 (15.2%)	188 (39.6%)
Mixed change	29 (8.2%)	12 (5.4%)	39 (8.2%)
Total with problems	356 (82.8%)	223 (83.8%)	475 (66.8%)
No problems	74 (17.2%)	43 (16.2%)	236 (33.2%)

Source Devlin et al. (2010)

replacement operations a close second. Hernia and varicose vein repairs were much less successful, and cataract removals had a very low success rate, with more patients getting worse than improving—although the last of these should be interpreted carefully because the EQ-5D may not be capturing the kind of benefits that cataract operations provide. The numbers of patients who worsened or had no change show the same pattern.

One problem with this analysis is that 'No change' is confounded when patients record no problems according to any of the dimensions before treatment, because they are, according to the EQ-5D, healthy patients whose only alternative would be for their condition to worsen as a result of treatment. Recording no problems at all is rare for patients who have conditions serious enough to require a joint replacement but may occur for conditions whose need for treatment may not be fully captured by their EQ-5D profile. Table 2.5 shows for the three conditions to which this applies the PCHC taking into account those with no problems before surgery. In each case, this shows a slightly better performance than suggested by Table 2.4.

The advantage of the PCHC is that it provides a high-level summary of the nature of changes in health reported by patients, without the need to introduce any external scoring system or preference weighting.

The limitations of the PCHC are:

(i)   It focuses on whether there is improvement or worsening in self-reported health, and does not account for the magnitude of those changes. It does not differentiate between small improvements and big improvements (e.g., both a shift from level 5 to level 4, and a shift from level 5 to level 1, are counted as improvements).

(ii) It takes no account of whether the changes occur in dimensions that matter a lot to people or in dimensions that may be considered less important.

(iii) The PCHC will not be informative in cases where mixed changes dominate the changes in health self-reported by patients.

The PCHC can be extended to give information about the composition of differences between profiles according to how dimensions and levels differ. These are illustrated using newer data on hip replacement patients in the English NHS PROMs programme that was instituted following the pilot study referred to earlier, using simple graphs. They also show how data can be compared at different time periods. This could be adapted to compare, for example, patients in different populations.

First, Fig. 2.1 shows the PCHC for three years in graphical form.

Figure 2.2 shows which dimensions were improved for those patients whose PCHC category was 'Improved'

This shows that improvements were spread over all dimensions, but were most frequently found in Pain and Discomfort, followed by Usual Activities and Mobility,
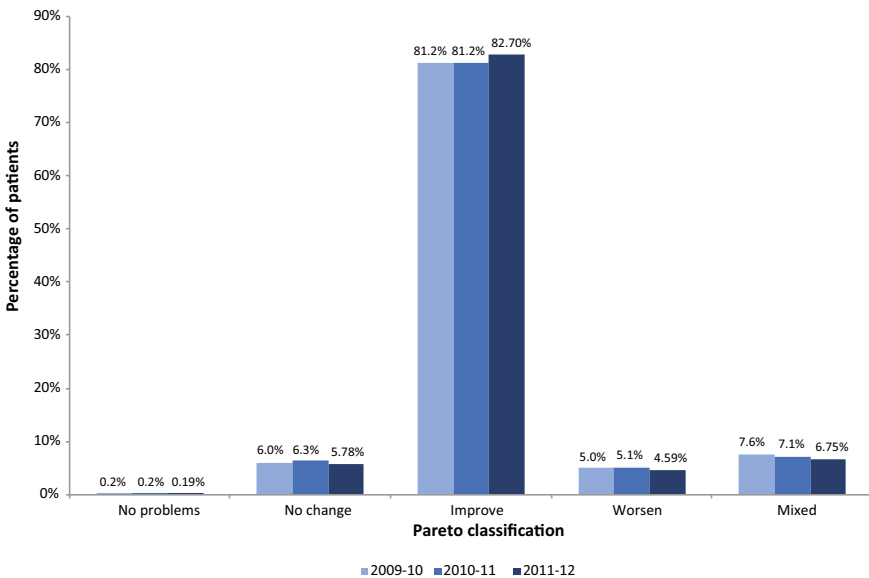

Fig. 2.1 The PCHC for hip replacement patients in the English NHS, 2009–12
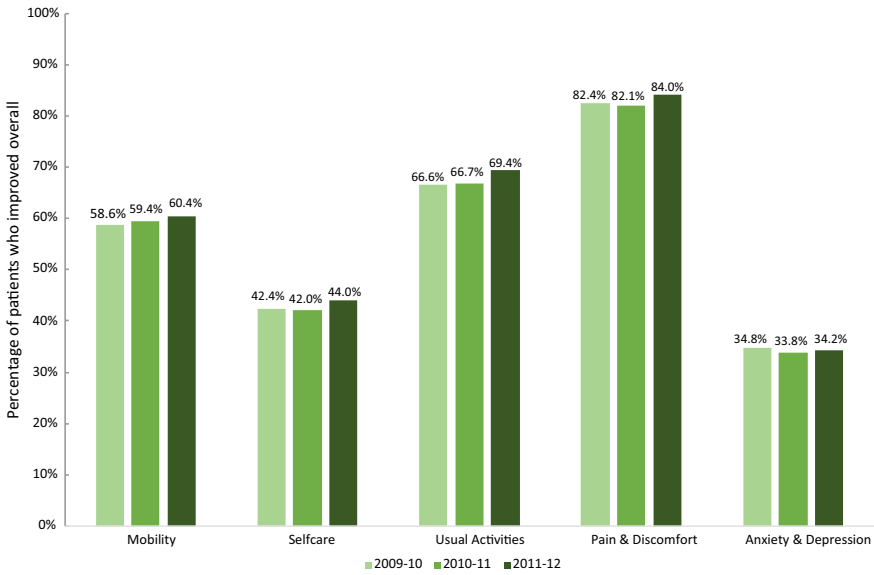
Fig. 2.2 Percentage of hip replacement patients who improved overall, by the dimensions in which they improved, English NHS 2009–2012

with Self-care and Anxiety and Depression improving for less than 50% of those who improved overall.

Figure 2.3 shows which dimensions were worsened for those patients whose PCHC category was 'Worsened'.

This shows the opposite pattern to improvements. Worsening health was spread over all dimensions, but was most frequently found in Anxiety and Depression and Self-Care followed by Usual Activities, with Pain and Discomfort and Mobility getting worse for less than 20% of those whose health was worse overall.

Figure 2.4 shows a comparison of PCHC 'Mixed' patients, which is more complicated because it involves both worsening and improving in different dimensions.

For the EQ-5D-3L, it is possible to show every possible change in every dimension. Each dimension can change in one of three ways—no change, improvement or worsening—each of which has three possible specific level changes, resulting in 9 categories for each dimension. Table 2.6 shows how these can be displayed for the hip replacement data.

This shows that the dominant change for Mobility, Usual Activities and Pain and Discomfort is an improvement from level 2 to level 1, but for Self-care and Anxiety and Depression it is no change from 'no problems.' Within change categories, it is notable that in each dimension improvements are dominated by a change from level 2 to level 1; that improvements from level 3 to level 1 and worsening from 1 to 3 are rare, reflecting the rarity of level 3 observations in the data set; and worsening from 2 to 3 is the most common amongst those who worsened overall in Usual Activities and Pain and Discomfort.
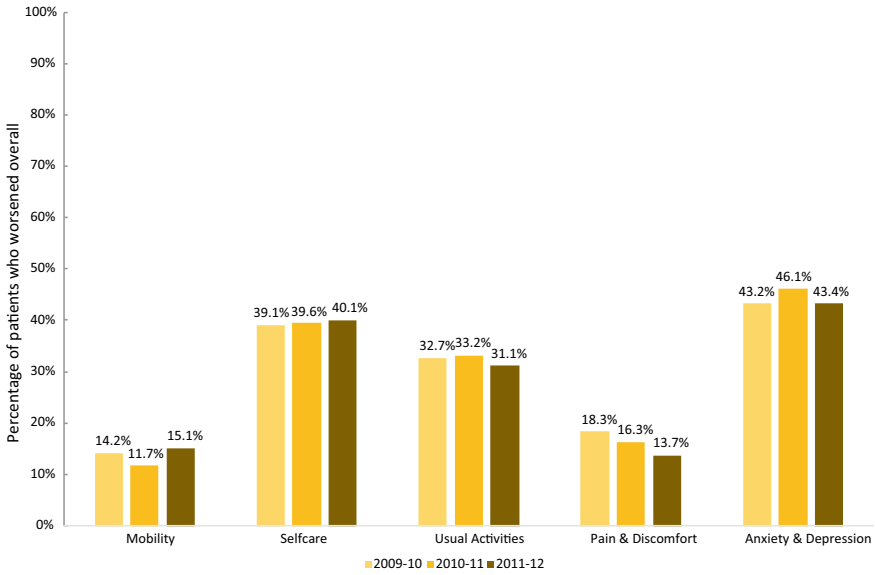
Fig. 2.3 Percentage of hip replacement patients whose health worsened overall, by the dimensions in which they worsened, English NHS 2009–12
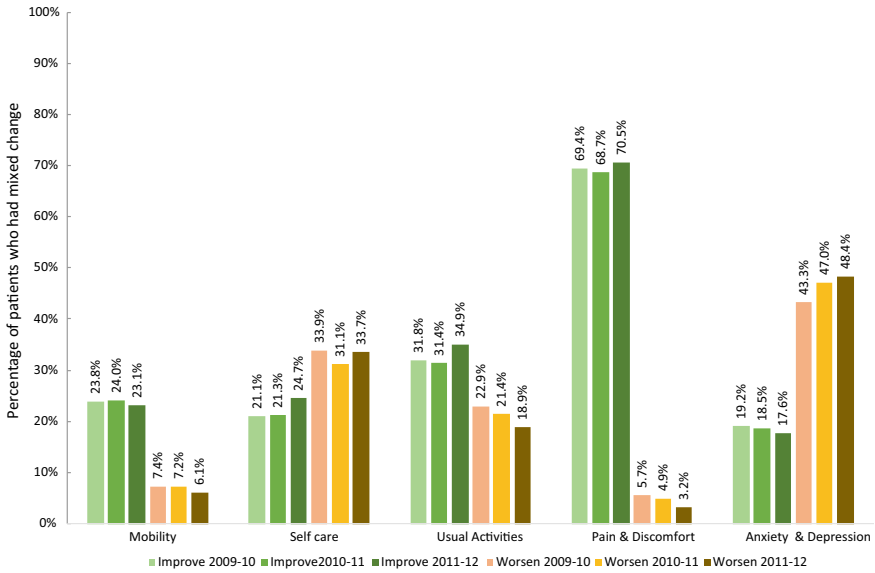

Fig. 2.4 Percentage of hip replacement patients who had a mixed change overall, by the dimensions in which they improved and worsened, English NHA 2009–12

Table 2.6 Changes in levels in each dimension for hip patients, NHS PROMs, 2009–10, percentages of total and of type of change

Change type	Mobility		Self-care		Usual activities		Pain and discomfort		Anxiety and depression	
	% total	% type	% total	% type	% total	% type	% total	% type	% total	% type
No change										
1–1	5.34	10.8	**40.2**	**67.6**	4.73	11.8	0.72	2.71	**52.8**	**81.5**
2–2	43.9	**89.1**	19.0	32.0	32.9	**82.0**	22.9	**86.5**	11.3	17.4
3–3	0.02	0.04	0.24	0.40	2.50	6.24	2.86	10.8	0.73	1.13
Better										
2–1	**49.0**	**99.1**	35.0	**97.2**	**39.4**	**69.6**	**33.6**	**46.5**	25.5	**85.6**
3–2	0.35	0.71	0.67	1.87	11.0	19.4	21.1	29.2	2.06	6.92
3–1	0.09	0.18	0.35	0.97	6.21	10.9	17.5	24.3	2.21	7.43
Worse										
1–2	1.21	**95.5**	4.11	**91.1**	1.29	38.4	0.22	16.5	4.26	**78.4**
2–3	0.06	4.52	0.33	7.25	1.97	**58.6**	1.11	**82.7**	0.93	17.1
1–3	0.00	0.00	0.08	1.68	0.10	3.06	0.01	0.85	0.25	4.55

% total = % of all in the relevant dimension; largest category highlighted in bold
% type = % of all in the change type in the relevant dimension; largest category highlighted in bold

Unfortunately, it is much more difficult to display the same analysis for the 5L version, as there are 25 possible categories for each dimension.

2.4.2   The Probability of Superiority

Buchholz et al. (2015) introduced a nonparametric effect size measure, the probability of superiority (PS), to analyse paired samples of EQ-5D profile data in the context of assessing changes in health in terms of improvement or deterioration. This measure was initially recommended by Grissom and Kim (2012). For each dimension, the number of patients with positive changes is divided by the total number of matched pairs (i.e. the number of respondents scoring EQ-5D at both time-points). To account for patients with no changes, that is 'ties', half the number of ties is added to the numerator. PS is therefore the probability that within a randomly sampled pair of dependent scores, the score obtained at follow-up will be smaller than the score obtained at baseline. It ranges from 0 to 1 and is

- <0.5 if more patients deteriorate than improve,
- = 0.5 if the same number of patients improve and deteriorate or do not change and
- >0.5 if more patients improve than deteriorate.

This is a further, useful way of examining the nature of change in EQ-5D data. A limitation is that it focuses on changes at the dimension level, rather than on how this combines at the patient level.

2.4.3   Health Profile Grid (HPG)

A further way of summarising changes in health in an EQ-5D data set is the Health Profile Grid (HPG), also introduced by Devlin et al. (2010). The HPG relies on profiles being ordered from best to worst. This can be done using a value set, a scoring system based on equally weighted dimensions and levels, or a scoring system based on the EQ VAS predicted from the profile (see Chap. 4).

The HPG plots the profiles between any two points in time. The example shown in Fig. 2.5, again taken from the English NHS PROMs pilot, shows profiles before and six months after hip replacement surgery. The rank ordering is determined by the EQ-5D-3L values according to the value set for the United Kingdom (Dolan et al. 1997). The PCHC category for each profile change is also shown.

The location of each point shows improvement and worsening according to the profiles' rank order. The 45° line represents 'no change'; the further above the line, the greater the improvement in health; below the line means health has worsened. The pattern of observations in the HPG in Fig. 2.5 suggests that most patients experience benefit from hip replacement surgery, as the observations lie predominantly
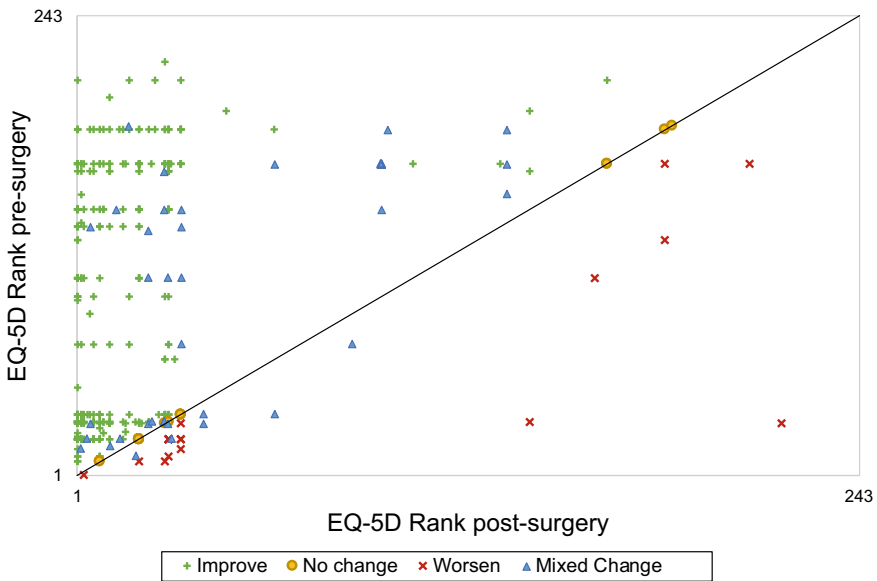

Fig. 2.5 Health profile grid for hip operations, English NHS

above the 45° line. There is a spread of health profiles from less to more severe before surgery, but a much narrower distribution after surgery, concentrated in the least severe profiles, with some outliers. The PCHC category adds to this by identifying cases where overall improvement and worsening of the patients' 'before' and 'after' profiles according to their rank are 'Mixed Change', that is they include both improvements in at least one dimension and worsening in at least one other. In these data, every mixed change case included only one dimension which changed in the opposite direction to the overall change according to the profiles' rank.

By contrast, the HPG shown in Fig. 2.6, for the English NHS PROMs pilot cataract surgery data, shows a much more mixed picture of improvements and worsening. The immediately obvious observation is that similar numbers improved and worsened. However, another feature is that most of those with the worst health profiles before surgery improved and most of those with the worst profiles after surgery had amongst the least severe health profiles before surgery. Unlike the clear-cut conclusions that may be drawn from the hip HPG, such a pattern suggests further investigation is required into the impact of cataract operations on patients' health-related quality of life (HRQoL).

Presenting the profiles in this manner can suggest clusters of patients, characterised by the nature of their profiles at time point 1, and the direction and magnitude of the change between the time-points. However, it is important not to rely on visual inspection alone to identify clusters, because some of the gaps that are apparent simply identify EQ-5D health profiles that are very infrequently observed, for example states having no problems in four dimensions and the worst state in the
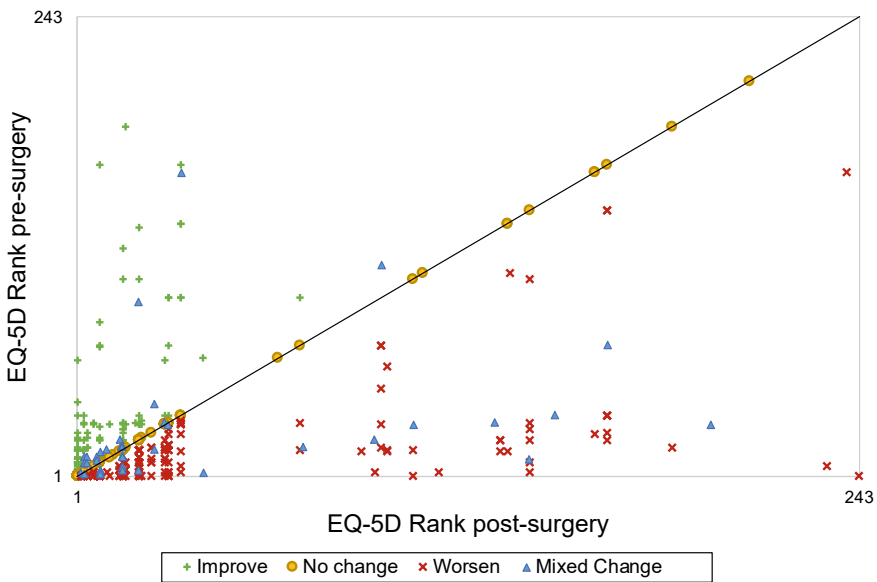

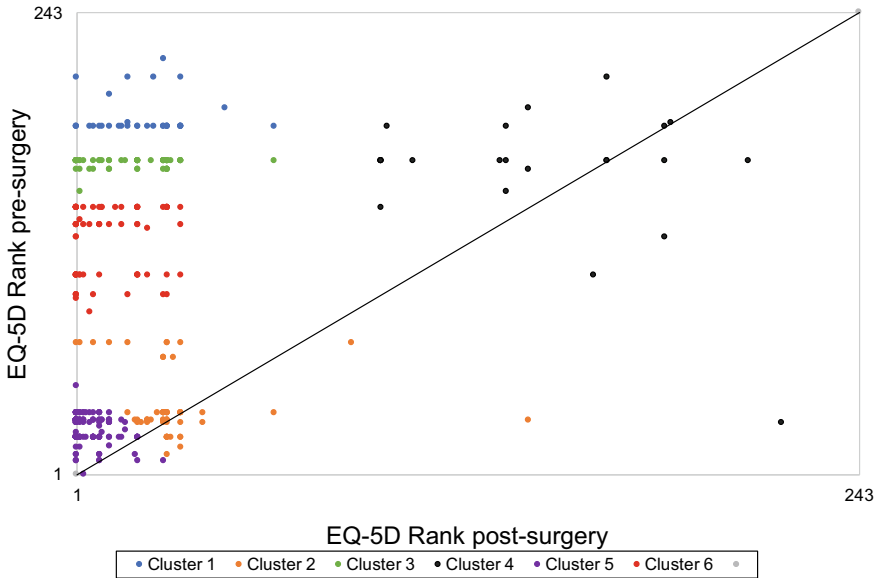**Fig. 2.6** Health profile grid for cataract operations, English NHS

Fig. 2.7 Health profile grid showing clusters of changes in health for NHS hip replacement patients, using the k-means procedure

other. It is essential to test for these formally using statistical cluster analysis techniques. An example, with clusters identified using a k-means procedure, is shown in Fig. 2.7.

The numbers represent the 6 different clusters of patients identified. Most of the clusters seem to be identified as similar because of the patients similar pre-surgery profiles. Cluster 4 is of more interest, identifiable as the patients with worst health profiles after surgery. Also of interest is the comparison of clusters 2 and 5, with similar, relatively less severe profiles before surgery but with cluster 2 having more severe profiles after surgery. These observations could form the basis of further investigation into whether or not these are real clusters of clinical importance.

It is to possible to improve the appearance of the HPG and reduce the problem of artefactual gaps by including only those health states found within the data. It is also possible to take this further by including only the most frequently found profiles. In many data sets, only a few very common profiles are found, along with many rarer cases, so restricting the analysis to profiles covering, for example, 90% of all observations would be informative.

The advantage of the HPG is that it provides a ready means of displaying and examining the changes in health within a sample of patients. A limitation of the HPG is that it relies upon having a valid and appropriate means of ranking the EQ-5D profiles. The method used to rank the profiles may affect the HPG and the statistical identification of clusters.

2.5 Summarising the Severity of EQ-5D Profiles

It is sometimes useful to summarise the overall 'severity' of EQ-5D health states, by means other than generating weighted scores such as values. Because these involve information loss and hidden assumptions about the aggregation of dimensions and levels, they should be used with care.

2.5.1 The Level Sum Score (LSS)

It is possible to summarise a profile by calculating a Level Sum Score (LSS), sometimes misleadingly referred to as the 'misery score'. This simply adds up the levels on each dimension, treating each level's conventional label (1, 2 or 3) as if it were a number rather than simply a categorical description.

The best EQ-5D health state involves having no problems on any dimension and is conventionally represented by the label 11111. Treating the level labels as numbers, the best possible score is $(1 + 1 + 1 + 1 + 1) = 5$. Similarly, the most severe problem on any dimension has the label 3 for the EQ-5D-3L, so the LSS for the worst health state is $(3 + 3 + 3 + 3 + 3) = 15$. Every other health state on the EQ-5D-3L will have a level sum score between 5 (the best) and 15 (the worst 15), and as these are integer there are 11 possible scores; the larger the score, the worse the health state. For the EQ-5D-5L, the range is between 5 and 25 and there are 21 possible scores.

The LSS has been used as a crude measure of severity to gauge the validity of values obtained in valuation for studies for different health states. Figure 2.8 shows the relationship between the English value set for the EQ-5D-5L and the LSS (Devlin et al. 2018). This shows that, as the LSS increases (states get worse), the values decline.

However, the LSS has some important limitations as a means of summarising health states across dimensions and levels:

 (i) It's a very crude summary score—for example, the very different EQ-5D-3L profiles 22222, 33211 and 11233 all have the same level sum score (LSS = 10). The Dutch values for these profiles are 0.569, 0.350 and 0.009 respectively (Lamers et al. 2006).
(ii) Within the LSS scores, the weighted index values derived from profiles have very wide and overlapping ranges.
(iii) Each score contains a very different number of potential profiles: for example, in the EQ-5D-3L, LSS = 5 and LSS = 15 have just one profile each, but LSS = 10 contains 51 profiles. For the 5L, there are 381 profiles with LSS = 15, but just 5 profiles with a LSS = 6.
(iv) Giving equal weight to the dimensions and the difference between levels means the LSS is not free from value judgements—it makes a specific assumption about their relative importance (Parkin et al. 2010).
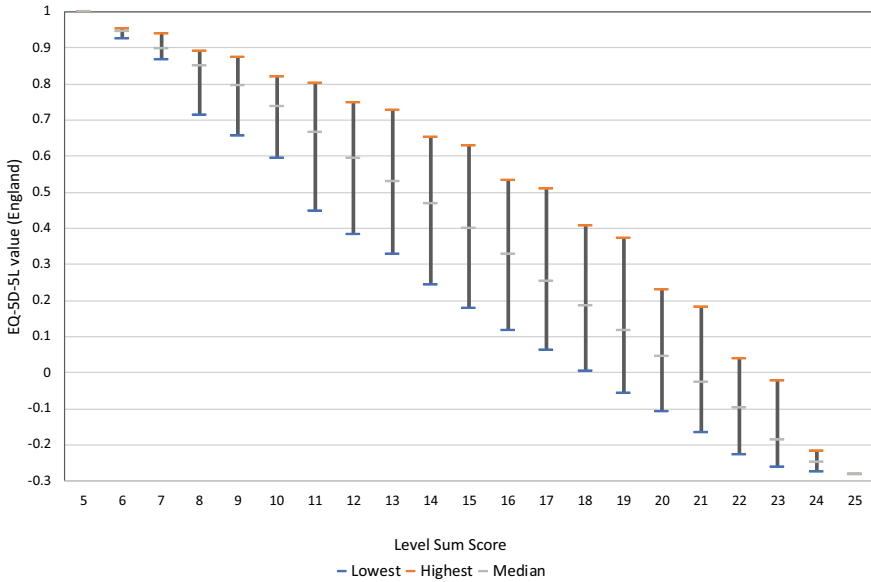
Fig. 2.8 EQ-5D-5L values (English value set) plotted against the LSS

These issues can be seen below, with respect to the EQ-5D-5L. Table 2.7 shows all possible LSSs for the EQ-5D-5L. It also shows descriptive statistics for the English value set for the EQ-5D-5L for all the different LSSs for the EQ-5D-5L. Although the mean and median values relate reasonably well to the order of the LSS, it does show big differences in the standard deviation. Importantly, it shows the overlap between the range of values for the different level summary scores. For example, the range for LSS = 15 includes the mean values of LSS = 12 and LSS = 18 and the lower or upper range respectively of LSS = 10 and LSS = 21. This issue can also be seen in Fig. 2.8. For these reasons, it is wrong to treat the LSS as ordinal.

2.5.2 The Level Frequency Score (LFS)

An alternative, although rarely used, means of summarising profile data is the level frequency score (LFS). The measure was proposed by Oppe and de Charro (2001) and used there to demonstrate the distribution of the EQ-5D-3L profiles in their data on the effects on HRQoL of a helicopter trauma team. The method characterises each health state by the frequency of levels at 1, 2 or 3 (for the EQ-5D-3L) or the frequency of levels at 1, 2, 3, 4 and 5 on the EQ-5D-5L. For example, in the EQ-5D-5L, the full health profile 11111 has 5, 1 s, no level 2, 3, 4 and 5 s, so the LFS is 50000; the worst health profile is 00005; profiles such as 31524 and 53412 would be 11111; 20 profiles such as 13211 have a LFS of 31100.

Table 2.7 Summary statistics for the EQ-5D-5L values (English value set) by all the different LSSs

Sum score	Number	Mean	Standard deviation	Median	Minimum	Maximum	Range
5	1	1.000	–	1.000	1.000	1.000	0.000
6	5	0.942	0.011	0.945	0.924	0.951	0.027
7	15	0.898	0.024	0.896	0.866	0.939	0.074
8	35	0.844	0.039	0.850	0.714	0.890	0.176
9	70	0.783	0.056	0.795	0.656	0.874	0.219
10	121	0.722	0.063	0.737	0.594	0.819	0.225
11	185	0.660	0.068	0.667	0.447	0.802	0.355
12	255	0.595	0.074	0.593	0.384	0.747	0.363
13	320	0.530	0.079	0.530	0.329	0.728	0.400
14	365	0.463	0.081	0.467	0.241	0.652	0.410
15	381	0.396	0.083	0.399	0.179	0.628	0.449
16	365	0.327	0.085	0.329	0.118	0.533	0.415
17	320	0.258	0.085	0.254	0.062	0.509	0.446
18	255	0.189	0.083	0.186	0.003	0.407	0.403
19	185	0.118	0.083	0.118	−0.057	0.372	0.430
20	121	0.045	0.079	0.045	−0.107	0.228	0.335
21	70	−0.025	0.073	−0.026	−0.165	0.181	0.346
22	35	−0.098	0.068	−0.099	−0.226	0.037	0.263
23	15	−0.173	0.066	−0.185	−0.261	−0.024	0.237
24	5	−0.245	0.026	−0.246	−0.276	−0.218	0.058
25	1	−0.281	–	−0.281	−0.281	−0.281	0.000

Oppe and de Charro used the LFS to show the way in which the EQ-5D-3L values data observed in their data (using the UK EQ-5D-3L value set) were distributed over the various EQ-5D-3L profiles (see Table 2.8).

The distribution of EQ-5D-5L profiles by LFS is provided in an Appendix to this chapter.

2.6   Analysing the Informativity of EQ-5D Profile Data

2.6.1   Shannon Indices

Shannon's indices, originally developed to analyse the information content of strings of text, are widely used in the ecology literature to measure how many species are observed and how evenly animals, or plants are spread over the various categories. It has also been applied widely in assessing distributional characteristics of the EQ-5D

Table 2.8   Number of observations in the LFS according to the UK EQ-5D-3L values

Value	014	023	041	050	122	131	140	212	221	230	311	320	410	500	Total
−0.484	1														1
−0.166		1													1
−0.016			1												1
−0.003				3											3
0.030				2											2
0.055						1									1
0.082			2												2
0.088						3									3
0.101								2							2
0.150					1										1
0.189			1												1
0.255									1						1
0.291											1				1
0.293						2									2
0.329									1						1
0.516				6											6
0.585							2								2
0.587							9								9
0.620							15								16
0.656										2					2
0.689										13					13
0.691										25					25
0.710										1					1
0.725												5			5
0.727												15			15
0.743										2					2
0.744												1			1
0.760												9			9
0.796													13		13
0.812												5			5
0.814												3			3
0.848													3		3
0.850													2		2
0.883													4		4
1.000														33	33
Total	1	1	4	6	6	6	27	2	2	43	1	38	22	33	192

Source Taken from a EuroQol scientific plenary paper which preceded the subsequent journal articles

(Buchholz et al. 2018), where the categories of interest are EQ-5D profiles and we are interested in a summary measure of how evenly respondents to EQ-5D questionnaires are spread over the profiles defined by the descriptive system. The main application of the Shannon indices has been to compare informational richness and evenness of dimensions, either comparing the EQ-5D-3L with the EQ-5D-5L or to compare similar dimensions between different generic health status instruments (Janssen et al. 2007). It is also possible to apply the Shannon indices to distributions of health profiles.

The Shannon index is defined as:

$$H' = -\sum_{i=1}^{C} p_i \log_2 p_i$$

where H' represents the absolute amount of informativity captured, C is the total number of possible categories (levels or profiles), and $p_i = n_i/N$, the proportion of observations in the ith category ($i = 1,…, C$), where n_i is the observed number of scores (responses) in category i and N is the total sample size. The higher the index H' is, the more information is captured by the dimension or instrument. In the case of a uniform (rectangular) distribution (i.e., $p_i = p*$ for all i), the optimal amount of information is captured and H' has reached its maximum (H'max) which equals $\log_2$ C. If the number of categories (C) is increased, H'max increases accordingly, but H' will only increase if the newly added categories are actually used. The Shannon Evenness index (J') exclusively reflects the evenness (rectangularity) of a distribution, regardless of the number of categories, and is defined as: $J' = H'/H'$max. Variance of the Shannon index can be calculated as described by Janssen et al. (2007) and accordingly standard errors and 95% confidence intervals can be calculated.

The Shannon indices are purely descriptive measures of the informational richness and evenness of a classification system and have no relation to the content, meaning, or clinical relevance of what the instrument aims to measure. Both the Shannon index and the Shannon Evenness index are needed to make a useful interpretation of the measurement scale.

2.6.2 Health State Density Curve (HSDC)

Zamora et al. (2018) introduced a graphical means of depicting the nature of the distribution of EQ-5D profiles, the health state density curve (HSDC). This draws on an analogy with the Lorenz curve in describing an income distribution. The cumulative frequency of health states is compared against the cumulative frequency of the sample or population. A 45° line means that the observed health states are completely evenly spread across the sample: 10% of the sample accounts for 10% of the health states; 50% of the sample accounts for 50% of the health states, and so on.
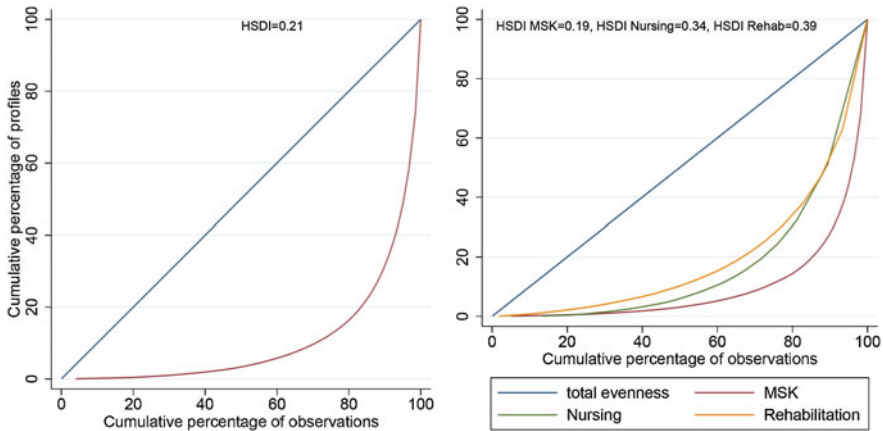
Fig. 2.9 HSDC for EQ-5D-5L profiles from Cambridgeshire NHS patients

A concentrated distribution—that is, where relatively few profiles are reported and are common to a large proportion of the sample—will be show as a curve which lies below the 45° line. The more unevenly distributed the profile data, the further below the diagonal line the HSDC will be. In the extreme, where just one profile is reported by all members of the sample, the HSDC will take a right-angled shape.

Figure 2.9 shows the HSDC for patients from three groups of patients, and overall, from Cambridgeshire NHS in the UK. This shows that for all patients, observed profiles are not evenly distributed, that is a small number of profiles accounts for a relatively large share of the observations. The musculoskeletal patients had the most concentrated data.

The HSDC provides a simple means of illustrating this property of a profile data set, in a manner that facilitates comparisons between data sets. It has limitations. As with Lorenz curves, where two curves cross (as is the case with rehabilitation and nursing data shown in Fig. 2.9), there is no unequivocal way of declaring one data set to be more concentrated than another. It also does not tell us which profiles are the most commonly self-reported. Therefore, the HSDC is best seen as a complement to the information from the cumulative frequency of profiles.

2.6.3 Health State Density Index (HSDI) and Other Related Indices

In the analysis of income distribution, the Lorenz curve is often accompanied by the Gini coefficient, which describes the extent of inequality which is apparent as the area between the diagonal line and the curve, divided by the entire area underneath the diagonal. In a similar way, an index can be calculated to summarise the inequality of observed health state profiles. Zamora et al. (2018) introduce a broadly similar

summary measure, the Health State Density Index (HSDI). HSDI has a value of 1
where there is total equality, that is where there are the same number of patients in
each profile, and HSDI $= 0$ for total inequality, that is where one profile accounts
for all the observations.

The HSDI allows the degree of concentration in self-reported health to be
compared both between different sets of patients and between different instruments,
for example the 3 and 5 level versions of the EQ-5D. Zamora et al. (2018) use
the HSDC to compare the EQ-5D-3L and EQ-5D-5L, their respective HSDIs indi-
cating the advantages of the 5L in differentiating between patients and yielding less
concentrated data.

The specific properties of the HSDI may be compared with the Shannon' indices.
Each performs somewhat differently as a measure in capturing specific aspects of
the distribution of patients' data, such as the concentration over the most common
states, and the influence of 'rare' states. For example, the Shannon index (absolute
and relative) is not sensitive to random variations but decreases slowly with "rare
health states". The HSDI decreases slowly with random variations and is strongly
affected by infrequently observed health states with large decreases towards zero
(total inequality). For more detail see Zamora et al. (2018).

Appendix: Analysis of the LFS for the EQ-5D-5L

For the EQ-5D-5L, the LFS has a total of 102 possible scores, from 00005 (for the
worst profile 55555) through to 50000 (no problem on any dimension, state 11111).
Like the LSS, a problem with the LFS is that LFSs contain an uneven number of
profiles. For example, LFS $= 50000$ and LFS $= 00005$ each contain 1 profile, whereas
LFS $= 11111$ (meaning: any health profile containing one level 1, one level 2, one
level 3, one level 4 and one level 5) represents 120 different EQ-5D-5L profiles.
Table 2.9 is a full list of the possible values for the LFS.

Table 2.9 Distribution of the EQ-5D-5L profiles by LFS

LFS	Freq.	%	Cum. (%)	LFS	Freq.	%	Cum. (%)	LFS	Freq.	%	Cum (%)
11111	120	3.84	3.84	20120	30	0.96	63.36	00230	10	0.32	91.52
01112	60	1.92	5.76	20201	30	0.96	64.32	00302	10	0.32	91.84
01121	60	1.92	7.68	20210	30	0.96	65.28	00320	10	0.32	92.16
01211	60	1.92	9.6	21002	30	0.96	66.24	02003	10	0.32	92.48
02111	60	1.92	11.52	21020	30	0.96	67.2	02030	10	0.32	92.8
10112	60	1.92	13.44	21200	30	0.96	68.16	02300	10	0.32	93.12
10121	60	1.92	15.36	22001	30	0.96	69.12	03002	10	0.32	93.44
10211	60	1.92	17.28	22010	30	0.96	70.08	03020	10	0.32	93.76
11012	60	1.92	19.2	22100	30	0.96	71.04	03200	10	0.32	94.08
11021	60	1.92	21.12	00113	20	0.64	71.68	20003	10	0.32	94.4
11102	60	1.92	23.04	00131	20	0.64	72.32	20030	10	0.32	94.72
11120	60	1.92	24.96	00311	20	0.64	72.96	20300	10	0.32	95.04
11201	60	1.92	26.88	01013	20	0.64	73.6	23000	10	0.32	95.36
11210	60	1.92	28.8	01031	20	0.64	74.24	30002	10	0.32	95.68
12011	60	1.92	30.72	01103	20	0.64	74.88	30020	10	0.32	96
12101	60	1.92	32.64	01130	20	0.64	75.52	30200	10	0.32	96.32
12110	60	1.92	34.56	01301	20	0.64	76.16	32000	10	0.32	96.64
20111	60	1.92	36.48	01310	20	0.64	76.8	00014	5	0.16	96.8
21011	60	1.92	38.4	03011	20	0.64	77.44	00041	5	0.16	96.96
21101	60	1.92	40.32	03101	20	0.64	78.08	00104	5	0.16	97.12
21110	60	1.92	42.24	03110	20	0.64	78.72	00140	5	0.16	97.28
00122	30	0.96	43.2	10013	20	0.64	79.36	00401	5	0.16	97.44
00212	30	0.96	44.16	10031	20	0.64	80	00410	5	0.16	97.6
00221	30	0.96	45.12	10103	20	0.64	80.64	01004	5	0.16	97.76
01022	30	0.96	46.08	10130	20	0.64	81.28	01040	5	0.16	97.92
01202	30	0.96	47.04	10301	20	0.64	81.92	01400	5	0.16	98.08
01220	30	0.96	48	10310	20	0.64	82.56	04001	5	0.16	98.24
02012	30	0.96	48.96	11003	20	0.64	83.2	04010	5	0.16	98.4
02021	30	0.96	49.92	11030	20	0.64	83.84	04100	5	0.16	98.56
02102	30	0.96	50.88	11300	20	0.64	84.48	10004	5	0.16	98.72
02120	30	0.96	51.84	13001	20	0.64	85.12	10040	5	0.16	98.88
02201	30	0.96	52.8	13010	20	0.64	85.76	10400	5	0.16	99.04

(continued)

Table 2.9 (continued)

LFS	Freq.	%	Cum. (%)	LFS	Freq.	%	Cum. (%)	LFS	Freq.	%	Cum (%)
02210	30	0.96	53.76	13100	20	0.64	86.4	14000	5	0.16	99.2
10022	30	0.96	54.72	30011	20	0.64	87.04	40001	5	0.16	99.36
10202	30	0.96	55.68	30101	20	0.64	87.68	40010	5	0.16	99.52
10220	30	0.96	56.64	30110	20	0.64	88.32	40100	5	0.16	99.68
12002	30	0.96	57.6	31001	20	0.64	88.96	41000	5	0.16	99.84
12020	30	0.96	58.56	31010	20	0.64	89.6	00005	1	0.03	99.87
12200	30	0.96	59.52	31100	20	0.64	90.24	00050	1	0.03	99.9
20012	30	0.96	60.48	00023	10	0.32	90.56	00500	1	0.03	99.94
20021	30	0.96	61.44	00032	10	0.32	90.88	05000	1	0.03	99.97
20102	30	0.96	62.4	00203	10	0.32	91.2	50000	1	0.03	100

Table 2.10 shows how the LFS could be used to analyse the characteristics of an EQ-5D-5L value set, using data from the English value set (Devlin et al. 2018). It shows the mean and median values for each LFS.

Table 2.10 Summary statistics of EQ-5D-5L values by LFS

LFS	Mean	Median	LFS	Mean	Median	LFS	Mean	Median
00005	−0.285	−0.285	02120	0.362	0.359	12101	0.543	0.535
00014	−0.247	−0.246	02201	0.461	0.452	12110	0.581	0.588
00023	−0.209	−0.209	02210	0.499	0.508	12200	0.718	0.717
00032	−0.171	−0.170	02300	0.637	0.637	13001	0.564	0.551
00041	−0.133	−0.134	03002	0.307	0.314	13010	0.602	0.609
00050	−0.095	−0.095	03011	0.345	0.348	13100	0.740	0.737
00104	−0.109	−0.099	03020	0.383	0.378	14000	0.762	0.760
00113	−0.071	−0.071	03101	0.483	0.473	20003	0.229	0.221
00122	−0.033	−0.031	03110	0.521	0.529	20012	0.267	0.262
00131	0.005	0.008	03200	0.659	0.659	20021	0.305	0.300
00140	0.043	0.037	04001	0.505	0.490	20030	0.343	0.349
00203	0.067	0.062	04010	0.543	0.553	20102	0.405	0.413
00212	0.105	0.093	04100	0.680	0.680	20111	0.443	0.448
00221	0.143	0.140	05000	0.702	0.702	20120	0.481	0.476
00230	0.181	0.184	10004	−0.028	−0.011	20201	0.581	0.570
00302	0.243	0.247	10013	0.010	0.019	20210	0.619	0.628

(continued)

Table 2.10 (continued)

LFS	Mean	Median	LFS	Mean	Median	LFS	Mean	Median
00311	0.281	0.285	10022	0.048	0.051	20300	0.756	0.756
00320	0.319	0.316	10031	0.086	0.082	21002	0.426	0.439
00401	0.418	0.408	10040	0.124	0.113	21011	0.464	0.466
00410	0.456	0.462	10103	0.148	0.138	21020	0.503	0.498
00500	0.594	0.594	10112	0.186	0.181	21101	0.602	0.591
01004	−0.087	−0.073	10121	0.224	0.222	21110	0.640	0.651
01013	−0.049	−0.049	10130	0.262	0.264	21200	0.778	0.777
01022	−0.011	−0.007	10202	0.324	0.331	22001	0.624	0.612
01031	0.027	0.024	10211	0.362	0.366	22010	0.662	0.673
01040	0.065	0.054	10220	0.400	0.393	22100	0.800	0.797
01103	0.088	0.083	10301	0.499	0.490	23000	0.821	0.819
01112	0.126	0.116	10310	0.537	0.544	30002	0.486	0.495
01121	0.164	0.157	10400	0.675	0.674	30011	0.524	0.526
01130	0.202	0.206	11003	0.170	0.164	30020	0.562	0.556
01202	0.264	0.269	11012	0.208	0.203	30101	0.662	0.647
01211	0.302	0.307	11021	0.246	0.243	30110	0.700	0.711
01220	0.340	0.335	11030	0.284	0.289	30200	0.838	0.838
01301	0.440	0.428	11102	0.345	0.357	31001	0.683	0.670
01310	0.478	0.486	11111	0.383	0.388	31010	0.721	0.737
01400	0.616	0.616	11120	0.421	0.415	31100	0.859	0.861
02003	0.110	0.103	11201	0.521	0.510	32000	0.881	0.883
02012	0.148	0.144	11210	0.559	0.566	40001	0.743	0.726
02021	0.186	0.177	11300	0.697	0.696	40010	0.781	0.793
02030	0.224	0.229	12002	0.367	0.378	40100	0.919	0.920
02102	0.286	0.293	12011	0.405	0.407	41000	0.940	0.942
02111	0.324	0.328	12020	0.443	0.436	50000	1.000	1.000

The following chart (Fig. 2.10) shows how the EVS and LFS are related. As with the LSS, this gives an indication of the general validity of a value set, in that there are no patterns that indicate that a value set takes perverse or other undesirable characteristics.
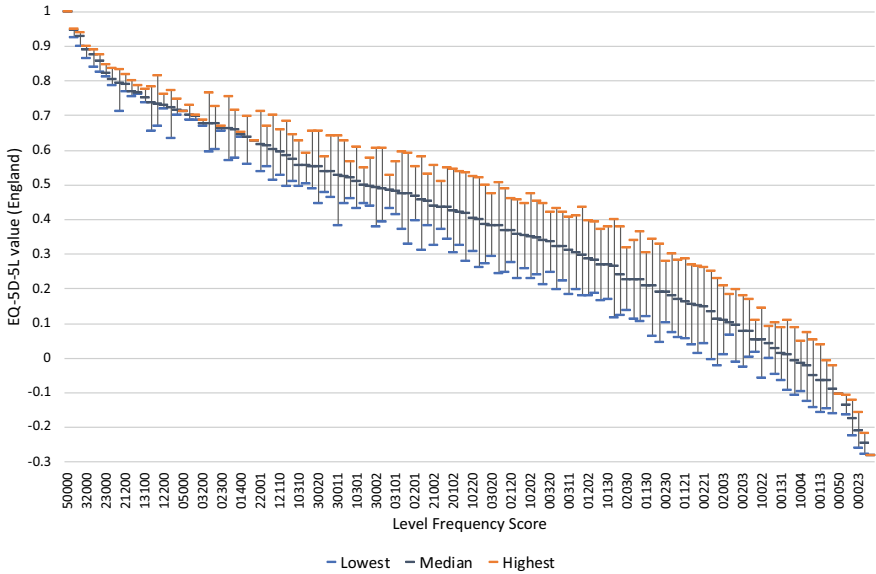
Fig. 2.10 EQ-5D-5L values (English value set) plotted against the LFS

References

Buchholz I, Janssen MF, Kohlmann T, Feng YS (2018) A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. Pharmacoecon 36(6):645–661

Buchholz I, Thielker K, Feng YS, Kupatz P, Kohlmann T (2015) Measuring changes in health over time using the EQ-5D 3L and 5L: a head-to-head comparison of measurement properties and sensitivity to change in a German inpatient rehabilitation sample. Qual Life Res 24(4):829–835

Devlin N, Parkin D, Browne J (2010) Using the EQ-5D as a performance measurement tool in the NHS. Health Econ 19(8):886–905

Devlin N, Shah K, Feng Y, Mulhern B, van Hout B (2018) Valuing health-related quality of life: an EQ-5D-5L value set for England. Health Econ 27(1):7–22

Dolan P (1997) Modeling valuations for EuroQol health states. Med Care 35(11):1095–1108

Feng Y, Devlin N, 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 Qual Life Outcomes 13:171

Grissom RJ, Kim JJ (2012) Effect sizes for research: univariate and multivariate applications, 2nd edn. Taylor & Francis, New York

Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel GJ, Badia X (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 20(10):1727–1736

Janssen MF, Birnie E, Bonsel GJ (2007) Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a US general population survey using Shannon's indices. Qual Life Res 16(5):895–904

Janssen MF, Bonsel GJ, Luo N (2018) Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. Pharmacoecon 36(6):675–697

Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJV (2006) The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 15(10):1121–1132

Oppe M, Devlin N, Black N (2011) Comparison of the underlying constructs of the EQ-5D and Oxford Hip Score: implications for mapping. Value Health 14(6):884–891

Oppe S, De Charro FT (2001) The effect of medical care by a helicopter trauma team on the probability of survival and the quality of life of hospitalised victims. Accid Anal Prev 33(1):129–138

Parkin D, Devlin N, Feng Y (2016) What determines the shape of an EQ-5D index distribution? Med Decis Making 36(8):941–951

Parkin D, Rice N, Devlin N (2010) Statistical analysis of EQ-5D profiles: does the use of value sets bias inference? Med Decis Making 30(5):556–565

Zamora B, Parkin D, Feng Y, Bateman A, Herdman M, Devlin N (2018) New methods for analysing the distribution of EQ-5D observations. OHE Research Paper. www.ohe.org/publications/new-methods-analysing-distribution-eq-5d-observations