What determines the shape of an EQ-5D index distribution?

David Parkin (NHS South of England and City University), Nancy Devlin and Yan Feng (Office of Health Economics)

1. Introduction

The EQ-5D is internationally one of the most widely-used health-related quality of life instruments (Brooks, 1996). It is therefore very important that we obtain a good understanding of the characteristics of EQ-5D data. Distributions of EQ-5D index values in patient and general populations typically have a non-normal distribution divided into two distinct groups. (This is often described as 'bi-modal', but we will argue that this is a misleading label.) In examining the distribution of health states within a particular population, it is useful to know to what extent its shape is determined by the way that the EQ-5D index is constructed rather than by the true distribution of ill-health. In addition, the EQ-5D index is constructed from two separate elements – a classification system used to create health state profiles and a set of weights applied to profiles – and it is useful to know the relative importance of these factors in shaping health state distributions.

This paper examines the determinants of the shape of EQ-5D distributions, and in particular the origins of the 'two groups' distribution. We analyse data from the English NHS PROMs programme (hip and knee replacements and varicose vein and hernia repairs) and from a study of two chronic conditions (asthma and angina). The distributions of EQ-5D index values are compared with distributions from which weights have been stripped, and the profile data are decomposed into their constituent dimensions and levels to see how they influence each of those distributions. They are also compared with the distributions of condition-specific indexes assessed for the same patients and of the EQ-VAS. The conclusions suggest practical ways in which researchers should analyse EQ-5D index data to obtain richer results than are conventionally reported.

2. Analysing EQ-5D data

The EQ-5D is used widely in economic analyses, population health surveys and, more

recently, for routine assessment of patients' health, for example the NHS Patient Reported Outcome Measures (PROMs) programme. The EQ-5D instrument comprises two self-report elements: the EQ-5D self-classifier, where respondents tick boxes to indicate which of three levels of problems (no, some, extreme) they have on each of five dimensions (mobility, self care, usual activities, pain & discomfort, and anxiety & depression) to create a health 'profile'; and the EQ-VAS, where respondents rate their overall health on a visual analogue scale from 0 (worst health imaginable) to 100 (best health imaginable).

Both the profile data and the EQ-VAS provide valuable data about the patient's view of their own health, and can themselves be the focus of analysis. For example, Devlin, Parkin and Browne (2010) demonstrate a number of ways in which profile data can be analysed, including a Health Profile Grid and a Paretian Classification of Health Change. However, by far the most common way of analysing data from the EQ-5D is to use index values to summarise the profile data. These index values provide, for each of the 243 states described by the EQ-5D, a value on a scale anchored at 1 (full health) and 0 (a state as bad as being dead), with values < 0 indicating states worse than being dead. Typically, these values are obtained for a sub-set of the 243 states from surveys of the general public, using stated preference methods to find out their views about how good or bad the states are in their opinion. Those preference data are then used to model the values for all 243 states. The resulting 'value sets', often called 'tariffs', result from and are influenced by choices about whose values are relevant, for example the general public or patient populations, what methods are used to elicit preferences, for example Time Trade Off, Visual Analogue Scale or Standard Gamble, and how the data are modelled.

There are clear normative grounds for using index values to summarise EQ-5D profile data where the purpose is to estimate QALYs in economic evaluation. For example, it is often argued that it is the views of the general public that are relevant in this context, as taxpayers and potential users of the NHS. However, index values are also widely used in other sorts of applications, where that rationale may not be relevant, for example, the Health Survey for England reports all EQ-5D data in terms of index values. This is probably because converting profile data into a single index number is convenient: single numbers are easier to analyse than profiles comprising multiple dimensions. It is also likely that this is a result of misunderstandings about what the index values represent; for example, that they are like the 'scoring systems' used in condition specific instruments, which are developed as an integral

part of the instrument.

However, there are some important concerns about the use of index values. Parkin, Devlin and Rice (2010) show that the use of value sets to summarise EQ-5D profiles introduces an exogenous source of variance, which can bias statistical inference. For example, conclusions about whether there are statistically significant differences in the health of two regions, or over time or between two arms of a clinical trial will be influenced by which value set is used, and what its particular properties are. More generally, there is no such thing as a 'neutral' value set: any way of weighting EQ-5D profile data will exert an influence on results[1].

Given that EQ-5D is such a widely used instrument and that its data are so often summarised using the index, obtaining a good understanding of the characteristics of index-weighted EQ-5D data is very important. One particular issue, which has been widely identified as problematic for the use of the EQ-5D, is that the distribution of values is non-normal and has what is often described as a 'bi-modal' shape. The main concern has been estimation rather than hypothesis testing, with the suggestion that the unusual shape might result in a non-normal distribution of residuals when the EQ-5D is the dependent variable in regression analysis.

A further problem is that a bi-modal shape might suggest that there are in fact two separable patient populations that should be analysed separately. It might be that many patient populations comprise distinct groups, and the EQ-5D is capable of picking out two from many of them. Or it might be that the EQ-5D profile tends to divide patients into groups, even though they are from a common distribution. Or that the weighting of EQ-5D profiles tends to divide patients into groups, even though they are from a common distribution. Knowing which of these possibilities is the source of the observed distributions in patients' data is crucial to their proper analysis and interpretation.

3. The distribution of EQ-5D index values

Studies that have identified this issue – which we review below - often suggest that the distribution of EQ-5D index values is bi-modal or tri-modal, but those labels are misleading.

[1] This same point applies equally to the scoring systems of other health measures, both generic and condition specific and includes measures that simply sum responses with equal weight.

A better description is that EQ-5D data appear to fall into two groups with an identifiable gap in index values between them. We will refer to these as the 'high cluster' and 'low cluster'. The 'tri-modal' shape arises if there are many observations of people who have no problems according to the EQ-5D classification. These form a third group, again with an identifiable gap in values from the high cluster. Although the size of the gap in value terms between no problems (11111 = 1) and the next best health state (11211 = 0.883) is important, the reasons for the existence of a single value 'group' at 1 is obvious, and its mode is trivially determined. The interesting questions concern the other two groups and the gap between them, which has been identified as being around 0.5 (Versteegh *et al*, 2010) or 0.45 (Brazier *et al*, 2004; Hernández Alava, Wailoo and Ara, 2012). The reason why 'bi-modal' is a misleading description for this phenomenon is that the modes of the two groups are not their most interesting feature; the groups do not always have a single local mode; and in practice these modes are never actually identified, reported or analysed.

This feature of the distribution of EQ-5D values has been reported in studies of a diverse range of conditions. There have been many in arthritis (Fransen and Edmonds, 1999, Conner-Spady and Suarez-Almazor, 2003; Russell *et al*, 2003; Marra *et al*, 2004; Scott *et al*, 2007; Harrison *et al*, 2008; Harrison *et al*, 2009; Lillegraven, Kristiansen and Kvien, 2010; Versteegh *et al*, 2010 Gaujoux-Viala *et al*, 2011; Gaujoux-Viala *et al*, 2012) and in orthopaedic conditions and treatments including herniated lumbar disc (Jansson *et al*, 2005), common spinal conditions for which surgery is indicated (McDonough and Grove, 2005), total knee replacement (Xie *et al*, 2007), lumbar spinal stenosis (Jansson *et al*, 2009), all elective orthopaedic operations (Jansson and Granath, 2010) and hip arthroplasty (Paulsen *et al*, 2012). Studies of other conditions include breast cancer (Conner-Spady *et al*, 2001), chronically ill patients undergoing haemodialysis (Gerard *et al*, 2004), lower back pain, chronic obstructive pulmonary disease, irritable bowel syndrome, leg ulcer and osteoporosis (Brazier *et al*, 2004; Hernández Alava, Wailoo and Ara, 2012), menopausal women and healthy older women (Brazier *et al*, 2004), pregnant women with chronic energy deficiency (Shaheena and Lindholm, 2006), HIV (Huang *et al*, 2008), postmenopausal women (Langdahl *et al*, 2009) multiple myeloma and non-Hodgkin's lymphoma (Versteegh *et al*, 2010), asthma, chest pain, Clodronate, hormone replacement therapy, leg reconstruction and varicose veins (Hernández Alava, Wailoo and Ara, 2012) and multiple sclerosis (Hawton *et al*, 2012).

The reason for the gap has often been alleged to be that the decrement in the EQ-5D index between levels 2 and 3 is relatively large compared with that between levels 1 and 2 (Fransen and Edmonds, 1999). More specifically, the so-called 'N3' term (Dolan, 1997) used to calculate the EQ-5D index in the UK and elsewhere has been implicated (Conner-Spady *et al*, 2001; Brazier *et al*, 2004; Hernández Alava, Wailoo and Ara, 2012). This gives a large decrement to the index score if there is a level 3 state in any dimension.

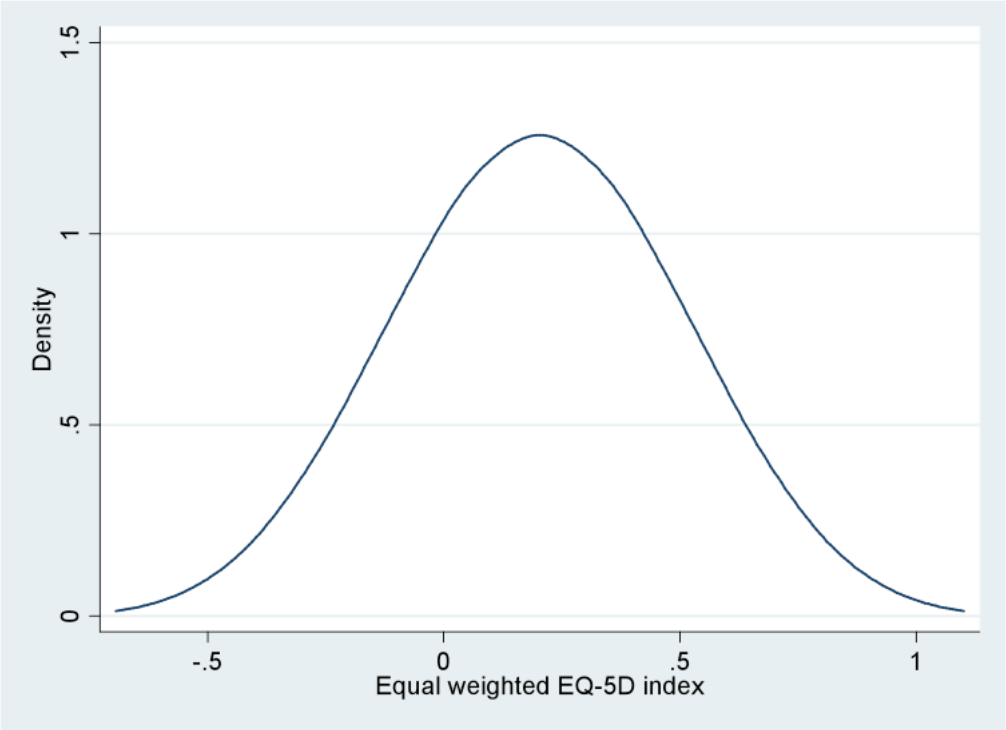
Related to this is evidence that in mapping between the EQ-5D and other health indexes, the EQ-5D may 'overestimate' the scores for more severe EQ-5D health states. Rowen, Brazier and Roberts (2009) hypothesized that predictions are poor for more severe states, defined as EQ-5D index <0.5, because they all have at least one dimension at the most severe level and the EQ-5D model uses an N3 term. They tested the importance of the N3 term by re-estimating the EQ-5D model without it using the same data and methods as the Dolan (1997) original. Although the predictions for more severe health states were better, they still appeared to be overestimated.

Versteegh *et al* (2010) tested the hypothesis that N3 in itself does not generate a 'bi-modal' distribution, by generating a random set of EQ-5D cases with an equal distribution of answers across the dimensions. They claimed that the resulting index scores were normally distributed (though it would be more accurate simply to say that the distribution did not have the two-groups-and-gap shape) suggesting that N3 is not the sole cause of that shape. They concluded that although N3 is a factor in the 'bi-modal' distribution and 'overestimation' in mapping states whose values are <0.5, these are also due to the fact that there are fewer observed responses at level 3 than at level 1 or 2 and only a few states are observed.

One of the points that this work clarifies is the issue of whether the existence of two groups is entirely an artefact, or in some cases or in some way does identify different patient groups. Jansson and Granath (2010) repeat an earlier assertion in Jansson *et al* (2009) that "We strongly believe that it is the structure of the instrument that causes this phenomenon rather than the fact that it appears to highlight 2 sub-groups of patients." One way to examine this is to seek external validation of the groups identified by the EQ-5D distribution. Hawton *et al* (2012) mapped a condition specific measure, the MSWS-12, with the EQ-5D. To test the specifications of their mapping models they calculated median MSWS-12 scores for two groups defined by a cut-point, which was the EQ-5D score closest to 0.5.

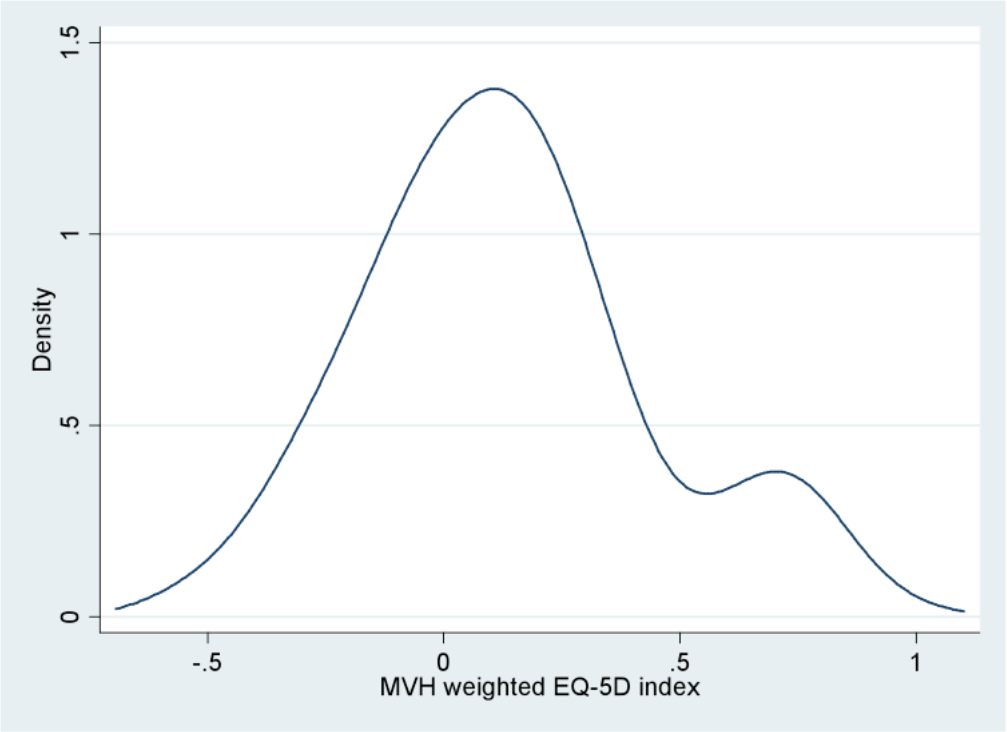
The effect of index weighting can readily be shown by contrasting the shape of distributions of health states with and without index weighting. As an illustrative exercise, an artificial data set was created, comprising one observation of each of the 243 health states described by the EQ-5D, from 11111 (no problems in any dimension) to 33333 (severe problems on all dimensions). Each health state was assigned index values using the UK MVH value set derived from the Dolan (1997) model and an 'equally weighted' scoring system, calculated by summing the level numbers (1, 2 or 3) over all dimensions, producing a number from 5 to 15. The equal weighting index was then converted so that it has the same range (-0.594 to 1) as the MVH value set, using a simple linear transformation. The two resulting distributions were smoothed using identical kernel density estimation functions. (This produces out-of-range predictions, but that is not important for our purely illustrative purposes.) Figures 1 and 2 show the smoothed frequency distributions of the resulting index values.

Figure 1. The distribution of equally and MVH weighted scores for all 243 EQ-5D profiles in the EQ-5D descriptive system



Mean	0.203	Median	0.203	Mode	0.203
Std. Dev.	0.292	Skewness	0.000	Kurtosis	2.700

Normality confirmed by skewness/kurtosis tests



Mean	0.137	Median	0.109	Mode	0.107
Std. Dev.	0.311	Skewness	0.437	Kurtosis	2.95
Local minimum	0.559		Local mode	0.700	

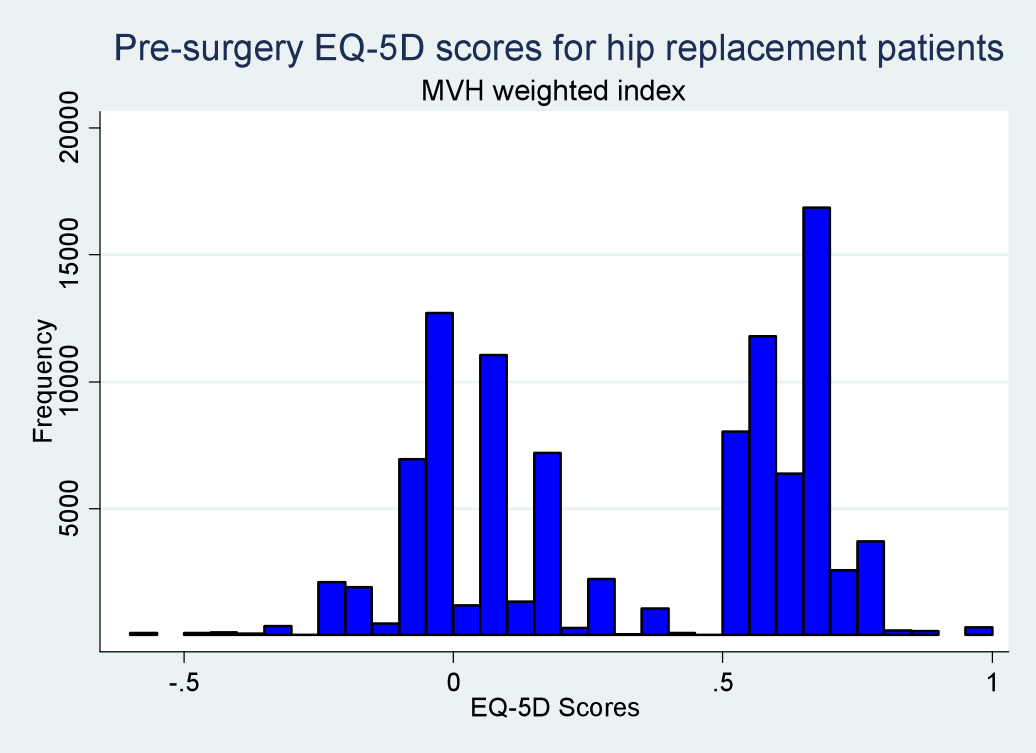
Non-normality confirmed by skewness/kurtosis tests.

On the face of it, this might suggest that equally weighted data are 'naturally' normally distributed, whereas MVH weighted data are 'naturally' skewed and bi-modal. But is this *purely* a product of index weighting or are there other factors involved?

4. Exploring EQ-5D distributions using real data

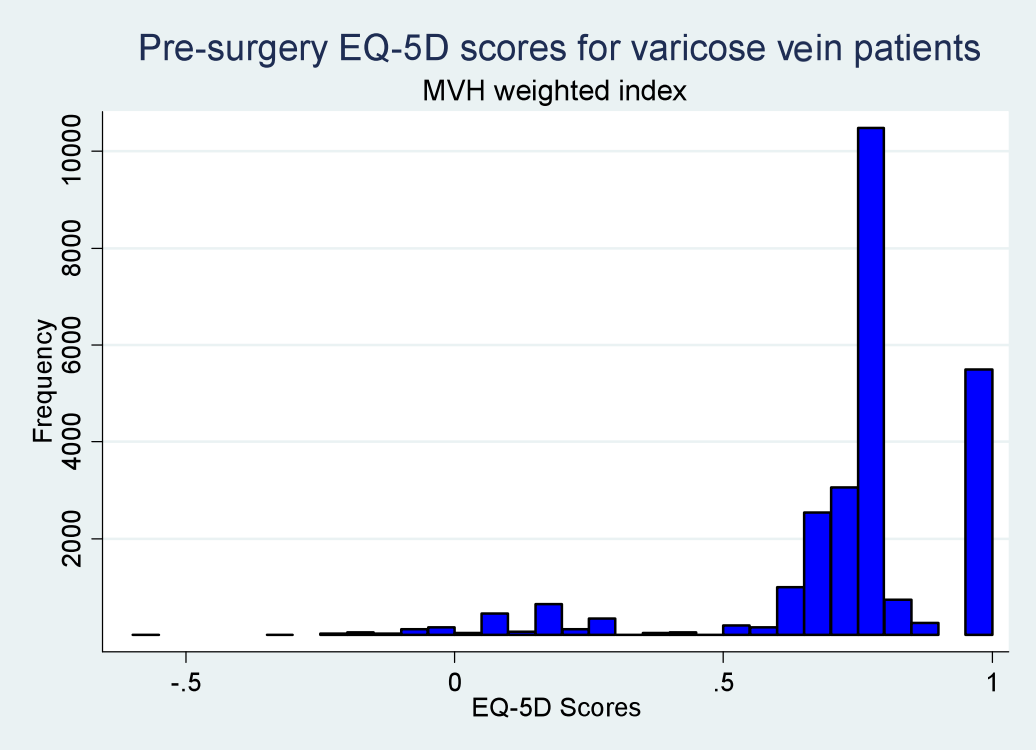
We explored this issue using data from the NHS PROMs programme on four elective surgical procedures (hip and knee replacements and varicose vein and groin hernia repairs) and two old primary care data sets on angina and asthma patients (Eccles *et al*, 2002). Our main focus in this paper is the hip replacement patient data. Figure 2 shows the distribution of MVH-weighted EQ-5D index scores for patients prior to hip replacement surgery, which clearly shows a two-groups-with-gaps distribution. As suggested, the upper 'group', consisting of all patients who reported no problems on any EQ-5D dimension, is of less importance for our purposes.

Figure 2: Distribution of pre-surgery EQ-5D index scores for hip replacement patients



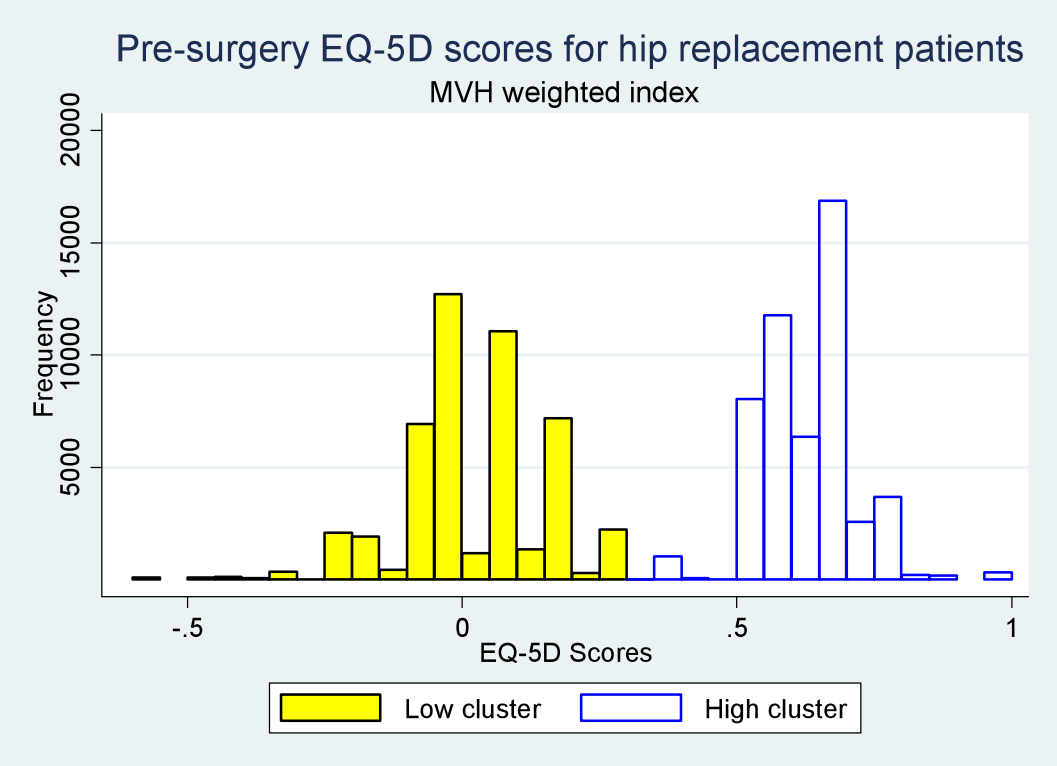
Those undergoing varicose veins surgery have very different underlying health characteristics to those of hip replacement patients. However, as Figure 3 demonstrates, a three-group-and gap distribution is apparent for varicose vein patients.

Figure 3: Distribution of pre-surgery EQ-5D index scores for varicose vein patients



Our first task was to identify more rigorously the two groups suggested by the histograms. Simple inspection of the hip replacement histogram suggests that the two could be defined as being above or below the EQ-5D index value 0.5. An alternative is to use a clustering technique. Using a simple kmeans clustering procedure with two groups identifies a different dividing line, lying between the values 0.313 and 0.329. However, only 1 180 out of 99 447 observations are affected by this, as may be seen from Figure 4, which divides the observations according to the kmeans-derived clusters.

Figure 4. High and low clusters of EQ-5D scores for hip replacement patients



The index scores are derived from profiles, so is there any clustering in the profiles? First, we can examine the different dimensions separately to understand the characteristics of this patient population. Table 1 shows the percentage of responses in each level of each of the dimensions.

Table 1: Percentage of responses in different dimensions and levels of the EQ-5D for pre-surgery hip replacement patients

Level	Mobility	Self-care	Usual activities	Pain & discomfort	Anxiety & depression
1	6.24%	66.17%	8.69%	1.08%	60.15%
2	93.34%	32.84%	76.12%	58.32%	35.23%
3	0.42%	0.98%	15.19%	40.61%	4.62%

These data suggest that in none of the dimensions is there a distribution across all three levels. There is very little difference between patients with respect to mobility (MO); they almost always record level 2 and very rarely record level 3. In each of the other dimensions, there are two levels that dominate. For self care (SC) and anxiety & depression (AD) these are levels 1 and 2, and for usual activities (UA) and pain & discomfort (PD) they are levels 2 and 3. In each case, the less severe of the two levels has the largest numbers.

We can therefore rule out differences in mobility as a cause of the two groups. To see whether the other dimensions are individually or in combination the cause, we can examine the distribution of profiles, as follows.

Table 2. Distributions of the EQ5D profile between high cluster and low cluster

Profile	Index	Number	Within cluster		Overall	
			%	Cumulative %	%	Cumulative %
High cluster						
21221	0.691	27412	43.14	43.14	23.95%	23.95%
21222	0.62	10040	15.80	58.94	8.77%	42.22%
22221	0.587	6613	10.41	69.35	5.78%	54.55%
22222	0.516	5525	8.70	78.05	4.83%	64.98%
Low cluster						
21231	0.159	10867	21.34	21.34	9.49%	33.44%
22232	-0.016	7502	14.73	36.08	6.55%	48.77%
21232	0.088	6413	12.59	59.51	5.60%	60.15%
22231	0.055	5518	10.84	59.51	4.82%	69.80%
22332	-0.074	4177	8.20	67.71	3.65%	77.23%
22331	-0.003	2135	4.19	71.90	1.87%	83.27%
21331	0.101	1921	3.77	75.68	1.68%	84.95%
21332	0.03	1551	3.05	78.72	1.36%	86.31%

This table shows only the most frequently observed profiles. Between them, these 12 profiles account for 86% of all profiles. The four within the high cluster account for 78% of profiles in that cluster, and the eight within the low cluster account for 79% of profiles in that cluster.

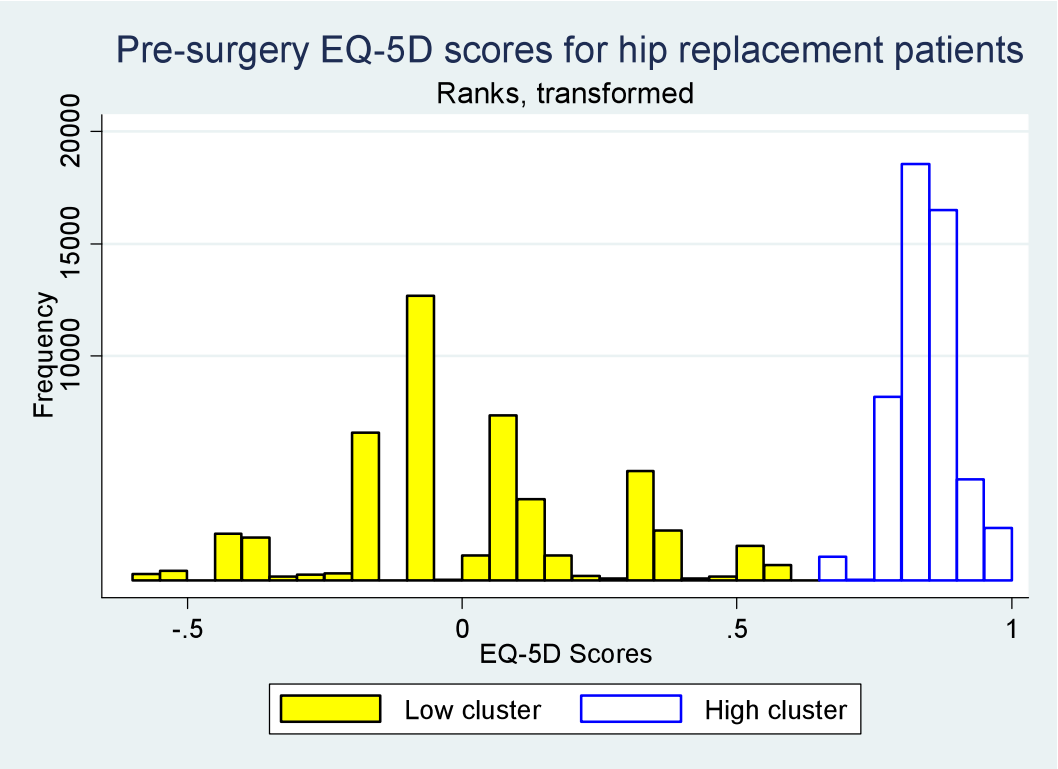
All twelve profiles have, as suggested by the earlier figures, MO = 2. The four main profiles in the High cluster all have UA = 2 and PD = 2. They are only distinguished by whether they have SC = 1 or 2 and AD = 1 or 2. The eight profiles in the low cluster are all one of the four high cluster profiles, but with PD = 3 and UA=3.

The implication is that the difference between the two groups is simply in the dimensions of PD and UA – the low score cluster has people who experience more pain and discomfort and have more restrictions on their usual activities than those in the high score cluster. However, there is a complicating factor, because the difference within these dimensions is between levels 2 and 3 rather than 1 and 2, and the presence of one or more level 3 gives additional decrements in scores within the MVH value set. This is because the differences between levels 2 and 3 in each dimension are greater than those between levels 1 and 2 and also because of the N3 term. The question remains whether it is the difference between the dimensions alone that generates the clusters or the fact that the low score cluster has more level 3 observations.

For varicose veins patients, the difference between the two groups is as clear. Almost all in both groups report SC=1. Almost all patients in the high cluster report PD=2 and do not have a level 3 in any dimension. Almost all of those in the low cluster report PD=3 and a few report a level 3 in dimensions other than SC. Again, the fact that the difference is in level 3 is a complication.

One way to examine this is to change the weights used in calculating the EQ-5D scores from a profile and see if the grouping remains. There is of course no such thing as a truly ‘unweighted’ score and, as noted earlier, there is no ‘neutral’ set of weights that can be used for this purpose. As with the early analysis, it is possible to give equal weighting to levels and dimensions, but this over-smoothes the data into 15 categories, giving a very weak test of the effect of more specific weights. A better alternative is to convert the set of weights into ranks. This retains the level of detail, in that every profile has an individual score, but removes the impact of size differences in the relative weighting of levels and dimensions, including the level 3 factor. The result of this is shown in Figure 5 for hip replacement patients.

Figure 5. Distribution of pre-surgery EQ-5D rank scores for hip replacement patients by cluster



In this figure, the ranks have been transformed into a variable with the same scale as the MVH EQ-5D index – this is simply to make direct visual comparisons easier and has no impact on the results. Although the division into two groups is less obvious, because of the wider spread of the data in the low cluster, it is nevertheless there.

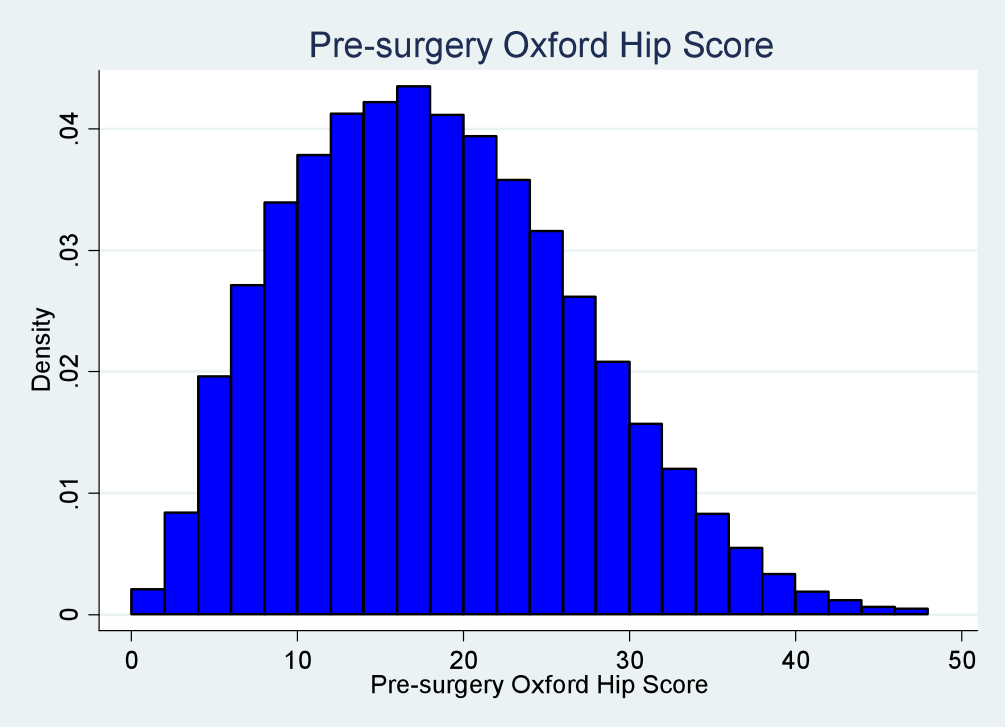
A possible conclusion is therefore:

- The division into two groups is a result of differences between groups of patients that are identified by the EQ-5D classification system in key dimensions of health.
- This distribution is reinforced by the weighting system, which generates the large gap between the two groups in index values.

A final question is whether the two-group distribution is reflecting true differences in the underlying health of patient populations or is an artefact of the EQ-5D classification system. It is not possible to answer that question directly, but it is possible in this case to explore it using additional data. The NHS PROMs programme data also include condition specific health state instruments, in this case the Oxford Hip Score (OHS). Is the two-group

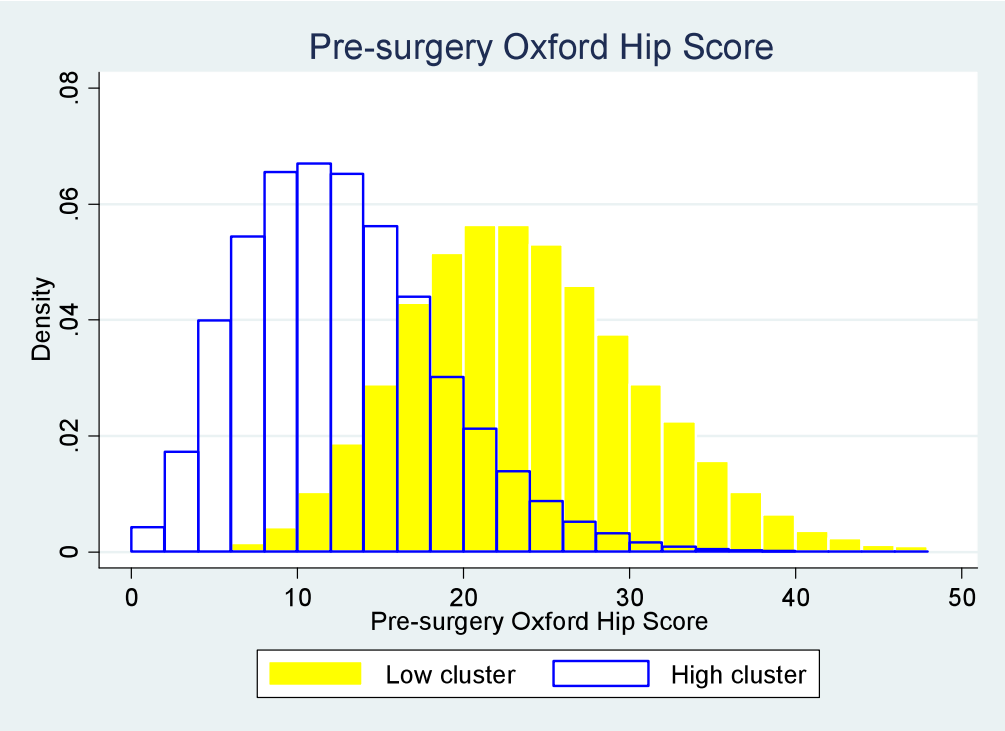
distribution apparent in those data? Figure 6 suggests not.

Figure 6. Pre-surgery Oxford Hip Score distribution for hip replacement patients



This might suggest that the EQ-5D clusters are indeed an artefact. However, if we apply the EQ-5D clusters to the OHS data, in the same way as Hawton *et al* (2012), a different picture emerges, as shown in Figure 7.

Figure 7. Pre-surgery Oxford Hip Score distribution for hip replacement patients, divided by EQ-5D clusters



This is consistent with an overall OHS distribution that is actually made up of the overlapping distributions of two groups. If so, it suggests that the OHS is not sensitive enough to discriminate between these different groups of patients. It is possible that a proper weighting system, rather than simply adding up ranks within each dimension, might enable the OHS to have more discriminatory power. But it also suggests that the EQ-5D currently has more discriminatory power.

A conclusion from these analyses is that better descriptions of EQ-5D index distributions are required, which take account of what appears to be a natural tendency for EQ-5D index data to generate a two-groups-and-gap distribution. Table 3 shows descriptive statistics for all four of the PROMs procedures and the two chronic conditions that we analysed.

Table 3: Distributions for six conditions according to high and low clusters

	Low cluster					High cluster				
	Mean	Median	Mode (Profile)	Std. Dev.	Range	Mean	Median	Mode (Profile)	Std. Dev.	Range
Hip	.019	-.003	.159 (21231)	.128	-.594 to .313	.631	.62	.691 (21221)	.087	.329 to .883
Knee	.052	.055	.159 (21231)	.122	-.594 to .345	.655	.691	.691 (21221)	.084	.362 to .883
Varicose veins	.130	.159	.088 (21232)	.140	-.594 to .436	.755	.796	.796 (11121)	.064	.487 to .883
Hernia	.183	.189	.159 (21231)	.147	-.594 to .452	.755	.796	.796 (11121)	.065	.485 to .883
Angina	.088	.088	-.016 (22232)	.144	-.594 to .383	.692	.691	0.620 (21222)	.095	.414 to .883
Asthma	.086	.088	.088 (21232)	.159	-.484 to .383	.722	.725	.796 (11121)	.093	.414 to .883

These kind of simple descriptive statistics give a far clearer picture of the distribution of EQ-5D index scores than are usually reported. This table also serves further to emphasise that the mode is not the key feature of the groups or the best descriptor of their distributions, suggesting that the label 'bi-modal' should not be used to describe the typical shape of an EQ-5D index distribution.

5. Conclusions

Non-normal distributions of index weighted EQ-5D data featuring two distinct groups of patients are commonly observed in patient populations. This has implications for statistical analysis and modelling of those data. The reasons for this shape are that the EQ-5D classification system picks out differences between patients with the same condition in respect of dimensions that are mainly observed at level 2 or 3. The weights commonly used to calculate the index exacerbate this grouping by placing a larger weight on level 3 observations, and generates a noticeable gap in index values between the groups.

A further factor involved in this, which is not analysed in this paper, is that in general only a few of the 243 potential EQ-5D states are observed with any great frequency. Devlin, Parkin

and Browne (2010) reported that in a large and diverse data set just 22 of the 243 EQ-5D profiles covered 90% of all health states observed and 161 profiles were not found at all. One reason for this is that profiles that contain very great differences in levels between dimensions are rarely observed, for example profiles having four level 1 dimensions and one level 3. It is therefore not unexpected that patients' health states form groupings for particular conditions, since extreme variations from a condition-specific typical EQ-5D profile will not be observed.

The analysis that we have carried out is on the original 3 level version of the EQ-5D. It will be interesting to see whether similar issues apply to the five-level EQ-5D-5L. We suggest that it will be important to examine EQ-5D-5L data using methods similar to those used here.

One recommendation from this analysis is that it is very important and informative to undertake exploratory data analysis on EQ-5D data, and that the analytical methods used for this may be simple. As we have argued elsewhere (Devlin, Parkin and Browne, 2010; Parkin, Devlin and Rice, 2010), concentrating on the EQ-5D index in effect obscures useful information about health states and may even produce misleading information. We suggest that this exploratory approach will enable us better to analyse EQ-5D data for comparison and inference purposes, and help in developing more accurate mapping between different health measures.

Although we have concentrated on the EQ-5D, our analytical approach also applies to any health status index that uses the weighted profile approach. This includes both generic and condition-specific measures, and also indexes that are calculated without explicit weights, such as the Oxford Hip Score. Indeed, it is arguable that a measure such as the Oxford Hip Score is far more wasteful of useful information than the EQ-5D index and that it positively obscures important differences between patients. There are potentially 244 150 625 (5^{12}) different Oxford Hip Score profiles. Its simple scoring system reduces this to 49 categories, involving a huge loss of information. Of course, the vast majority of those profiles would never be observed, but it is likely that far more would be observed than 49 and that the differences between those profiles will be of interest.

Acknowledgements

The NHS Information Centre provided the PROMs data. We would like to thank participants in seminars at the Universities of Exeter and Manchester for their helpful comments and for providing useful references. All opinions expressed in this paper are the responsibility of the authors alone and not of the NHS South of England or the OHE.

References

Brazier, J., Roberts, J., Tsuchiya, A. and van Busschbach, J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics*. 2004;13: 873–884

Brooks R (1996). EuroQol: the current state of play. *Health Policy* 37(1):53-72

Conner-Spady, B., Cumming, C., Nabholtz, J-M., Jacobs, P. and Stewart, D. Responsiveness of the EuroQol in breast cancer patients undergoing high dose chemotherapy. *Quality of Life Research* 2001; 10: 479-486

Conner-Spady, B. and Suarez-Almazor, M.E. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Medical Care* 2003; 41(7): 791–801.

Devlin, N, Parkin D, Browne J. (2010) Patient Reported Outcomes in the NHS: new methods for analysing and reporting EQ-5D data. *Health Economics* 19(8):886-905.

Dolan, P. Modeling Valuations for EuroQol Health States. *Medical Care*, 1997; 35(11): 1095-1108

Eccles, M, McColl, E, Steen, N, Rousseau, N, Grimshaw, J, Parkin, D and Purves, I. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. *British Medical Journal*, 2002, 325: 941-947

Fransen, M. and Edmonds, J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. *Rheumatology*, 1999; 38(9): 807-813.

Gaujoux-Viala, C., Rat, A.C., Guillemin, F., Flipo, R.M., Fardellone, P., Bourgeois, P. and Fautrel, B. Comparison of the EQ-5D and the SF-6D utility measures in 813 patients with early arthritis: results from the ESPOIR cohort. *Journal of Rheumatology*. 2011; 38(8): 1576-84.

Gaujoux-Viala, C., Rat, A.C., Guillemin, F., Flipo, R.M., Fardellone, P., Bourgeois, P. and Fautrel, B. Responsiveness of EQ-5D and SF-6D in patients with early arthritis: results from the ESPOIR cohort *Annals of the Rheumatic Diseases* 2012; 71: 1478–1483.

Gerard, K., Nicholson, T., Mullee, M., Mehta, R. and Roderick, P. EQ-5D versus SF-6D in an

Older, Chronically Ill Patient Group. *Applied Health Economics and Health Policy*, 2004; 3(2): 91-102

Harrison, M.J., Davies, L.M. Bansback, N.J., et al. The validity and responsiveness of generic utility measures in rheumatoid arthritis: a review. *Journal of Rheumatology* 2008;35:592–602

Harrison, M.J., Davies, L.M. Bansback, N.J., McCoy M.J., Verstappen, S.M.M., Watson, K. and Symmons, D.P.M. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. *Quality of Life Research* 2009; 18: 1195–1205

Hawton, A., Green, C., Telford, C,J., Wright, D.E. and Zajicek, J.P. The use of multiple sclerosis condition-specific measures to inform health policy decision-making: mapping from the MSWS-12 to the EQ-5D. *Multiple Sclerosis* 2012; 18(6): 853-861

Hernández Alava, M., Wailoo, A. J. and Ara, R. Tails from the Peak District: Adjusted Limited Dependent Variable Mixture Models of EQ-5D Questionnaire Health State Utility Values. *Value In Health* 2012; 15 550–561

Huang, I-C., Frangakis, C., Atkinson, M.J., Willke, R.J., Leite, W.L., Vogel, W.B. and Wu, A.W. Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. *Health Services Research* 2008; 43(1): 327–39.

Jansson, K-A. and Granath, F. Health-related quality of life (EQ-5D) before and after orthopedic surgery. *Acta Orthopaedica* 2010; 82(1): 82-89.

Jansson KÅ, Németh G, Granath F, Blomqvist P. Health-related quality of life in patients before and after surgery for a herniated lumbar disc. *Journal of Bone Joint Surgery (Br)* 2005;87-B:959-64.

Jansson, K-A., Németh, G., Granath, F., Jönsson, B. and Blomqvist, P. Health-related quality of life (EQ-5D) before and one year after surgery for lumbar spinal stenosis. *Journal of Bone Joint Surgery (Br)* 2009; 91(2): 210-6.

Langdahl, B.L., Rajzbaum, G., Jakob, F., Karras, D., Ljunggren, O., Lems, W.F., Fahrleitner-Pammer, A., Walsh, J.B., Barker, C., Kutahov, A. and Marin, F. Reduction in Fracture Rate and Back Pain and Increased Quality of Life in Postmenopausal Women Treated with Teriparatide: 18-Month Data from the European Forsteo Observational Study (EFOS). *Calcified Tissue International* 2009; 85: 484–493

Lillegraven, S., Kristiansen, I. S. and Kvien, T. K. Comparison of utility measures and their relationship with other health status measures in 1041 patients with rheumatoid arthritis. *Annals of the Rheumatic Diseases* 2010; 69(10):1762-7

McDonough, C. M. and Grove, M. R. Comparison of EQ-5D, HUI, and SF-36-Derived Societal Health State Values among Spine Patient Outcomes Research Trial (SPORT) Participants *Quality of Life Research*. 2005; 14(5): 1321–1332.

Marra, C. A., Esdaile, J. M., Guh, D., Kopec, J. A., Brazier, J. E., Koehler, B. E. et al.). A

comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Medical Care* 2004; 42(11): 1125–1131.

Parkin D., Devlin N., Rice N. Statistical analysis of EQ-5D profiles: does the use of value sets bias inference? *Medical Decision Making* 2010;30:556-565.

Paulsen, A., Pedersen, A. B., Overgaard, S. and Roos, E. M. Feasibility of 4 patient-reported outcome measures in a registry setting. A cross-sectional study of 6,000 patients from the Danish Hip Arthroplasty Registry. *Acta Orthopaedica* 2012; 83 (4): 321–327

Rowen, D., Brazier, J. and Roberts, J. Mapping SF-36 onto the EQ-5 D index: How reliable is the relationship? *Health and Quality of Life Outcomes* 2009; 7(27).

Russell, A.S., Conner-Spady, B., Mintz, A., Mallon, C. and Maksymowych, W. P. The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *The Journal of Rheumatology* 2003; 30(5): 941–947.

Scott, D.L., Khoshaba, B., Choy, E.H. and Kingsley, G.H. Limited correlation between the Health Assessment Questionnaire (HAQ) and EuroQol in rheumatoid arthritis: questionable validity of deriving quality adjusted life years from HAQ. *Annals of the Rheumatic Diseases* 2007; 66(11): 1534-7.

Shaheena, R. and Lindholm, L. Quality of life among pregnant women with chronic energy deficiency in rural Bangladesh *Health Policy* 2006; 78(2–3): 128–134

Versteegh, M. M., Rowen, D., Brazier, J. E. and Stolk, E. A Mapping onto EQ-5D for patients in poor health. *Health and Quality of Life Outcomes* 2010; 8:141

Xie F, Li S C, Luo N, Lo N N,Yeo S J,Yang KY, Fong KY, Thumboo J. Comparison of the EuroQol and short form 6D in Singapore multiethnic Asian knee osteoarthritis patients scheduled for total knee replacement. *Arthritis and Rheumatism* 2007; 15: 57 (6): 1043-9.