Section 1: Introduction

IN cost-utility analysis, a preference based measure of health is required to conduct economic appraisals of competing health care interventions from a society-wide perspective. However, there are several instruments available to researchers working in this area. Each has strong theoretical foundations, and differ in their health classification systems, method of eliciting preferences and scoring functions. Studies that have compared their properties have found that the measures are all valid and viable in their own way (Macran et al 2003) but utilities for similar health states are can vary dramatically (Kopec & Willison 2003).

Two commonly used measures of quality of life are the Euroqol EQ5D and the SF36. The Euroqol is simple to use and can detect large differences in health (Dolan et al 1995, Brazier et al 1999). However, it has been criticized for having poor sensitivity to improvements in conditions associated with low morbidity (Brazier et al 2002) and for being unable to detect small changes in health (Brazier et al 1999). The SF36 is a very widely used generic measure of HRQoL, but is not designed for economic evaluation since it does not explicitly incorporate preferences into its scoring algorithm. A preference-based measure of health (SF6D) has recently been derived from the SF36. It has a large descriptive system and it is claimed to have a greater degree of sensitivity and to be less skewed than the EQ5D. However, early studies have identified a number of concerns (Longworth 2003, O'Brien 2003). The model is known to over-predict the value of the poorest health states (Brazier 2002) and may not be sensitive to changes in conditions of high morbidity (Longworth). There is also concern over whether the SF6D has compromised the descriptive richness and sensitivity of the SF36 (Brazier 2002) and whether the assumptions of multi-attribute utility theory, on which preference-based measures of health are founded, can be applied to this measure (O'Brien 2003).

This paper investigates whether , in the context of a community-based population of patients receiving physiotherapy interventions for neck pain, the choice of instrument may change the results or conclusions of an economic evaluation. The dataset in the SPRINTER trial included EQ-5D and SF36 as outcome measures. The aims of this

paper are to show how the estimates of QALYs and cost-utility can change depending on the choice of outcome measure. The paper then looks whether such differences can be ascribed to the ability of the health state descriptive systems to detect meaningful differences in health or from the values estimated for similar health states being different.

Section 2: Properties of Euroqol EQ5D and Short-Form 6D health state classification systems and index measures

The SF6D preference based measure of health is derived from the SF36 by a two-stage process.

First, eleven questions are taken from the SF36 to define the six dimensions of the SF6D (Physical functioning, role limitation, social functioning, pain , mental health and vitality). Each dimension has between four and six levels allowing 18000 (6x4x5x6x5x5) health states to be used. The procedure to map the SF36 to the six dimensions is quite complex and is described in Appendix 1. The estimation of the $SF6D_{index}$ used a linear additive utility model. This imposes the assumption that there is no interaction for preferences among the attributes at all. For example, a change from level 1 to level 2 on the pain dimension would reduce overall utility by 0.052, regardless of the levels on the other dimensions. This in turn requires that the questions used in the survey are phrased in such a way as to avoid interactions between dimensions. For example, questions should not ask about how pain has affected physical functioning. A closely related assumption is structural independence, that is, whether it is conceivable for an individual to function at any level on that attribute, regardless of the levels on the other attributes, so that all combinations of the levels in the system are possible.

The second stage is to apply the preference tariff. The tariff was estimated using the standard gamble method to obtain utility values from 836 respondents from the UK general public. 249 out of the possible 18000 health states were chosen to be valued. A total of six states per subject were valued relative to an upper anchor of no dysfunction (level 1 on all dimensions) and a lower anchor of the lowest levels for all

dimensions. The lower anchor was then valued relative to 'full health' and death. The second lottery values were used to normalise the first lottery values to the conventional (0,1) scale of death to full health. A linear additive utility model was fit by ordinary least squares using the mean valuation of the health states as the dependent variable and with 25 SF6D item-levels and a dummy variable indicating whether one or more items was 'severe' as covariates. The most severe SF6D state is valued at 0.291. Brazier et al (2002) report that negative observations, which represent states valued at worse than death, did occur in 7% of observations in the valuation survey but the model over-predicted the value of the poorest health states. The algorithm is available from John Brazier.

The EQ5D is a much simpler health state classification than the SF6D because each of the five dimensions (mobility, self care, usual activity, pain and depression) is assessed by a unique question, with one of three possible responses. The instrument and its tariff is shown in Appendix 2. The tariff was estimated using the time trade-off method from a sample of 3395 respondents of the UK general population. A linear additive utility model was estimated at the individual level using generalised least squares, using the item levels and a term which identified whether any dimension was at its most severe level as covariates. The most severe EQ5D state is valued at –0.594.

The SPRINTER trial dataset

The SPRINTER trial was designed to compare the cost-effectiveness of a brief intervention (BI) for physiotherapy with usual physiotherapy (UC) for patients in a community setting with neck pain. 139 participants were randomised to BI and 129 to UC. Patients were asked to complete the EQ5D, SF36 and use of healthcare resource questionnaires at baseline, three months and twelve months post-randomisation.

Section 3: Methods

QALYs were constructed from utility scores by the 'area under the curve' method (Matthews et al 1990). Incremental mean costs, QALYs and the incremental cost-effectiveness ratio were calculated to investigate whether the choice of instrument affected the estimate of cost-effectiveness between the two interventions (based on expected values).

The EQ5D and SF6D instruments are designed to measure HRQoL on the same scale, with zero representing death and one representing full health. Negative values indicate states valued by the general population as worse than death. It is important for the purposes of economic evaluation that if there are two possible instruments for assessing health-related utility, then a particular value generated by one instrument should have an identical interpretation on the other instrument. That is, the index values generated by the two instruments should both be well correlated and in agreement. To see why they must agree, consider a hypothetical situation in which there are two treatments A and B that can be given to a severely ill patient, and a choice of two instruments (EQ5D and SF6D) with a high degree of correlation, where one index can be mapped to the other by some monotonic transformation. If a patient is given A, their condition does not change over the time horizon of the analysis. If they are given B, they die immediately for some reason. If utility is measured using the SF6D instrument, the patients utility is positive and the patient receives positive QALYs in treatment A and zero in B. If utility is measured using the EQ5D, utility is negative and so the patient receives negative QALYs in treatment A and zero in B. Therefore if costs are the same in A and B, using SF6D will favour A and using EQ5D will favour B. Therefore the choice of treatment will be different depending only on the measure of utility chosen. This perverse result will always be possible if there is a choice of utility measure where values for one can be mapped to the other under a non-identical but positive linear transformation. Therefore if the two measures generate systematically different utility values for the same patient at the same point in time, then the instruments cannot be measuring the same thing. The level of agreement and correlation was visually assessed using a scatterplot (Longworth 2003, O'Brien 2003).

The sensitivity of the instrument to measure changes in the health of the patient over time has two separate aspects. Firstly, the instrument must accurately record the health state of each individual, and allow the recorded health state to change in response to the individuals actual health. Secondly, the tariff must accurately estimate societies preferences with regard to that health state.

With respect to the ability of the instrument to accurately record the individuals health state, Longworth (2003) and Brazier (1998) consider floor and ceiling effects. A ceiling effect occurs when a population gives a high proportion of scores for some dimension at the least severe level, and if present will make the instrument insensitive to further improvements in health for the population. A floor effect occurs when the population under consideration gives a high proportion of their responses at the lowest level for one or more dimensions of an instrument. If this is the case, the instrument will be insensitive to further decrements in the health of this population.

Floor and ceiling effects are estimated for the SPRINTER population by calculating the proportion of responses at the worst and best possible level for each dimension of each instrument. Other authors seem to have declared a ceiling or floor effect when more than 80% of responses are at the highest or lowest level.

The construct validity of the preference-based index can be assessed by investigating whether it correlates with a known or hypothesised measure of health. Brazier et al (1998) compared the expected distribution of total scores with a diagnosis of chronic physical problems, but this is not available in the SPRINTER dataset. The SF36 contains an question on self assessed health, which has five possible responses ranging from 'excellent' to 'poor'. A crude assessment of construct validity can be made by comparing self-assessed health with the $SF6D_{index}$, though recognising that self assessed health need not necessarily correlate with preferences for health from a societal perspective.

The association between self assessed health and the preference value generated by each instrument is assessed using 'box plots'. These graphs are able to concisely summarise and compare the distributions.

A further assessment of sensitivity can be made by examining the frequency with which individuals change their reported EQ5D or SF6D state over time, given that they report an improvement in self assessed health over the same period. A high proportion of patients reporting an unchanged EQ5D health state but improved self-assessed health between these periods would indicate a lack of sensitivity in the instrument.

To test this aspect of sensitivity of the EQ5D, the most frequently used EQ5D health states reported at twelve months post-randomisation were compared with the state reported at baseline for the EQ5D instrument by each individual. There are too many states reported using the SF6D to make this last test practical for the SF6Dinstrument.

The design of both instruments assumes a linear additive model, and therefore independence between utility attributes. To determine whether a linear model is appropriate, we compute a correlation matrix between the questions used in the instrument for both EQ5D and SF6D (O'Brien 2003).

Section 4: Results

Table 1 shows summary statistics for the two measures at baseline and month 12. The mean value of the $SF6D_{index}$ is greater than the $EQ5D_{index}$ at both follow-ups, with a lower variance and a smaller range. 40% of the sample reported a general improvement in self-assessed health over 12 months, 13 % reported a deterioration and 47% reported no change. Similar trends were seen when analysed by treatment group. The mean change in utility over 12 months measured using the $EQ5D_{index}$ is an increase of 0.004, and an increase of 0.029 using the $SF6D_{index}$. Table 2 shows the cost-effectiveness of the interventions based on an analysis of expected values. Usual care is more expensive than BI. Using the $EQ5D_{index}$, UC is slightly more effective, and has an ICER of £18709. Using the $SF6D_{index}$, BI is slightly more effective and dominates based on expected values.

Figure 1 shows the distribution of each measure at baseline. Neither measure appears well-correlated with self-assessed health, although the relatively small number of

scores for the $EQ5D_{index}$ below 0.5 appear to be associated with lower levels of self-assessed health.

Figure 2 shows a scatterplot of the $EQ5D_{index}$ and $SF6D_{index}$ for each patient at baseline and 12 months post-randomisation. The $EQ5D_{index}$ and $SF6D_{index}$ are well-correlated with each other, but they cannot be said to agree (the fitted values do not coincide with the 45° line). Less than 2% of the utility scores estimated using the $EQ5D_{index}$ are negative. None of the $SF6D_{index}$ are negative since the value of the worst possible state using the tariff is 0.291. There are clear gaps in the number line where no score is possible using the $EQ5D_{index}$.

Table 3 shows that the EQ5D may have a ceiling effect for the mobility and self care dimensions with this population, and no floor effect. There is a ceiling effect for the 'Role limitation' dimension of the SF6D. . There is no evidence of a floor effect for any of the SF6D dimensions. Examination of the responses for individual elements of the SF36 shows a high proportion of participants reported either 'no problems in ability to bathe and dress' (Question SF3j), or 'limited a lot in vigorous activity' (Question SF3a) or both but this does not translate into either a floor or ceiling effect for the physical functioning dimension as a whole.

Table 4 shows changes in health state measured with the EQ5D instrument between baseline and month 12. Eight health states were used by 82% of respondents at month 12. Relatively few (15%) participants reported the same health state in both follow-up surveys.

Correlations between the answers to the SF6D survey form are significant at the 1% level for all elements except 'ability to bathe and dress oneself' vs 'severity of depression', and 'ability to bathe and dress oneself' vs 'vitality'. Correlation coefficients between the dimensions of the EQ5D survey were significant for all except the 'pain' dimension.

Section 5: Discussion

The $SF6D_{index}$ and $EQ5D_{index}$ both aim to measure preference values for health on a scale where zero represents death and one represents full health. Therefore they

should be measuring the same thing, and the choice of instrument should not influence the conclusions of an economic evaluation. This paper has shown that, when applied to the SPRINTER trial population, then usual care is cost-effective if the maximum willingness to pay is greater than £19000 when using the EQ5D instrument but brief intervention dominates when using the SF6D instrument.

One reason explored in the literature for the lack of agreement between the two measures is that the EQ5D is not sensitive to changes at the upper end of the distribution of health and the SF6D is not sensitive to changes at the lower end. Longworth et al, looking at patients receiving liver transplants, have noted a floor effect for the role limitation dimension of SF6D with 42% of patients reporting the most severe level. They found an absence of floor effect for any dimension of the EQ5D. Brazier et al (1998), looking at a population representatively drawn from GP practice lists, found a ceiling effect for the older version of the EQ5D across all dimensions. The findings from this study of patients receiving physiotherapy for neck pain do not indicate any evidence of floor effects for either measure in this population, but do indicate a ceiling effect for the EQ5D mobility and self care, and for the role limitation dimension of the SF6D.

This study adds to evidence that the role limitation dimension of the SF6D is susceptible to a floor effect when the population of patients is fairly ill or a ceiling effect when the population has low morbidity. Our study, however, shows that the ceiling effect is not because the questions which make up this dimension are inappropriate. Taken on their own, neither question exhibits a ceiling effect. The questions are 'how much of the time were you limited in the kind of work or other activity due to physical health' (SF4c) and 'how much of the time did you accomplish less than you would like due to emotional problems' (SF5b). Answers can range from 1 (all of the time) to 5 (none of the time). If patients do not answer 'all of the time' to both questions, they are allocated to the least severe level of the role limitation dimension. This is highly likely in a community sample even though relatively few patients answered 'none of the time' to either of these questions. A similar argument may explain why Longworth et al (2003) found a floor effect for this dimension in their relatively ill population. The two questions are highly correlated and therefore patients who are limited in their role all of the time in physical health

may also feel limited because of emotional problems. These patients will be allocated to the most severe level of the dimension. It may be that an alternative method to construct the role limitation dimension can be devised, that would not produce such severe floor or ceiling effects.

However, the absence of a floor effect is not necessarily a strength of the EQ5D instrument. The phrasing of the mobility and self care questions suggests that the use of the most severe state will only be appropriate for a very small minority of severely morbid patients. This has led to a high proportion of the EQ5D states being valued at worse than death by the general population, but they are unlikely to be used in practice. Since there are relatively few possible health states in the EQ5D compared with other instruments such as SF6D and HUI3, in principle this may further undermine the ability of the EQ5D to be sensitive to changes in HRQoL. Since only level 1 and level 2 are available in practice for mobility and self care, this may explain the presence of such a pronounced ceiling effect.

Brazier et al (1993) found that just 10 health states accounted for 95% of observations in their community-based population, using an older version of the EQ5D. In this population, eight EQ5D health states account for 82% of those used over all follow-up periods. However, the frequency with which participants register the same health state in subsequent follow-ups is very low. This indicates that patients are able to register changes in their health state using the EQ5D.

The spacing of the tariff values of the $EQ5D_{index}$ along the number line is uneven. It is not possible to register an index value between 1.000 and 0.883, nor between 0.487 and 0.516. This accounts for the tri-modal, non-symmetrical distribution of the $EQ5D_{index}$ seen in Figure 2. The SF6D covers the number line much more evenly over its more limited range. There is no reason why the indices should cover the number line evenly, only that the index should accurately reflect society's preferences. However, it is likely that the gaps in the number line seen in the EQ5D indicate a lack of health states for the use of patients with better health.

This study does not present unambiguous evidence either for or against the assertion that the EQ5D classification system is insensitive to changes in health status in a

community population. On the one hand, patients do not tend to record the same health state at different follow ups, which indicates the instrument is allowing patients to record changes in health status. On the other hand there are relatively few health states available for patients with better health, particularly in the self care and mobility dimensions. The SF6D has more health states. It may be possible to improve the ceiling effect of the role limitation dimension of the SF6D by changing the way the levels of this dimension are constructed from the SF36 questions. The EQ5D might be improved by making the lowest levels of the mobility and self care dimensions less severe.

The index measures of both instruments tend to correlate with self-assessed health though it is not possible to say whether one performs better than the other. The SF6D is known to over-value poor health states. It is clear from the analysis of correlations that there are significant interaction effects between the questions used to construct the dimensions of the SF6D, and that the linear additive model is not valid. The questions of the SF36 appear designed to encourage interactions (for example, question 8 asks 'How much did pain interfere with your normal work?' encompassing pain and physical functioning). The estimation of interaction effects would require a much larger sample size than that taken to estimate the version used here. The lack of orthogonality may be a reason why the model shows low predictive ability.

The two instruments aim to measure the same thing, that is, a preference-based measure of health. They are clearly not generating equivalent values for patients at the same point in time. This is not unexpected given the differences in both the classification system and the estimation of the tariff. The implication is that the choice of instrument is important and can change the decision about which health technology to adopt. In order that this choice is not arbitrary, the analysis plan for any study should specify which in advance which instrument is to be used to measure health outcome, and guidance from NICE and other decision makers should specify which instrument is preferred.

From the evidence presented here, both systems could be improved by changes in their construction. This would necessitate re-estimation of the preference weights. It is also unlikely that the linear additive model is suitable especially for the SF6D

system, although the inclusion of ad-hoc 'fixes' such as the additional decrement for one or more severe states appears to improve model predictive behaviour.

References

Brazier J, Usherwood T et al. Deriving a preference-based single index from the UK SF-36 Health Survey. Journal of Clinical Epidemiology 51(11) (1998) 1115-1128

Brazier J, Deverill M, Green C et al. A review of the use of health status measures in economic evaluation. Health Technol Assess 1999;3(9)

Brazier J, Roberts J, Deverill M. The estimation of a preference based measure of health from the SF36. Journal of Health Economics 21 (2002) 271-292

Dolan P, Gudex C et al. A social tariff for Euroqol: results from a UK general population survey. Centre for Health Economics Discussion Paper 138 (1995), University of York

Kopec J, Willison K. A comparative review of four preference-weighted measures of health related quality of life. Journal of Clinical Epidemiology 56 (2003) 317-325

Longworth L, Bryan S. An empirical comparison of EQ5D and SF6D in liver transplant patients. Health Economics 2003 (in press)

Macran S, Weatherly H, and Kind P, "Measuring Population Health: A Comparison of Three Generic Health Status Measures." Medical Care, Vol. 41, No. 2, February, 2003, pp 218-231

Matthews, JNS, Altman, D, and Campbell, MJ. Analysis of serial measurements in medical research. *British Medical Journal* 1990;300:230-235

O'Brien B, Spath M et al. A view from the bridge: agreement between SF-6D utility algorithm and Health Utilities Index. Health Economics 2003 (in press)

Figures and Tables

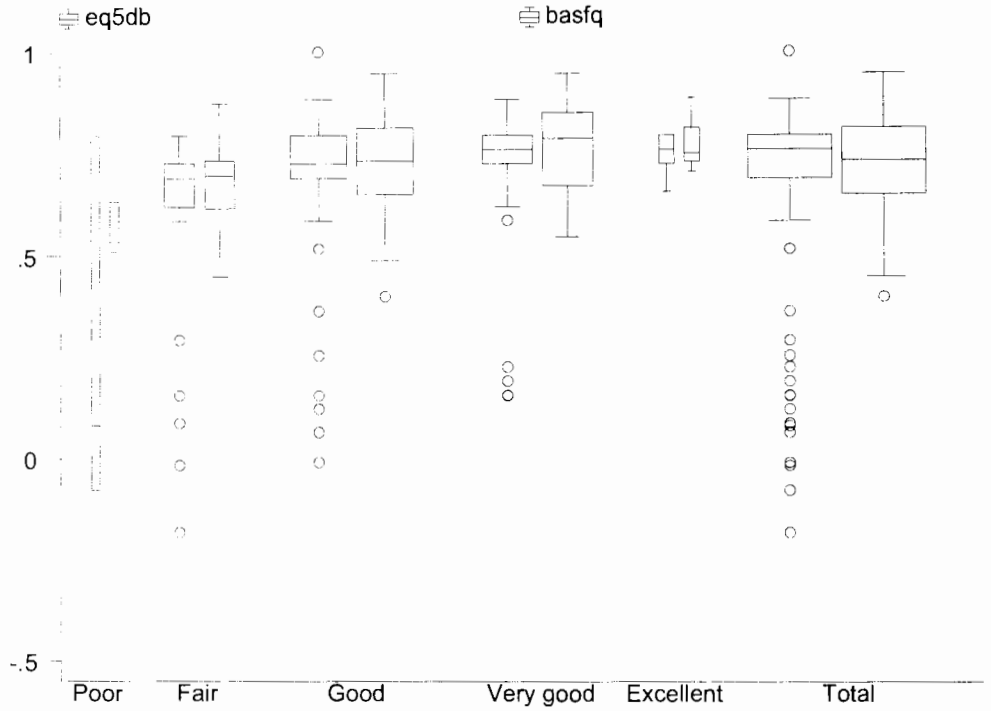
Table 1: Summary statistics for $EQ5D_{index}$ and $SF6D_{index}$ at baseline and follow up

	$EQ5D_{index}$		$SF6D_{index}$	
	Mean	(SD)	Mean	(SD)
Baseline	0.696	(0.170)	0.735	(0.103)
Month 3	0.712	(0.191)	0.761	(0.115)
Month 12	0.700	(0.221)	0.760	(0.133)

Table 2: Cost-effectiveness based on expected values using $EQ5D_{index}$ and $SF6D_{index}$ as outcome measures

	Incremental costs	Incremental QALYs ($EQ5D_{index}$)	ICER (EQ)	Incremental QALYs ($SF6D_{index}$)	ICER (SF)
Mean	58	0.0031	18709	-0.014	Dominated
95% CI	(-2,120)	(-0.03,0.04)		(-0.04,0.09)	

Figure 1: Distributions of $EQ5D_{index}$ and $SF6D_{index}$ at baseline by self-assessed health status



The line in the middle of the box is the median. The box extends from the 25th to the 75th percentile. The lines emerging from the box (whiskers) extend to the lower and upper adjacent values (25th percentile - 1.5 x IQR to 75th percentile + 1.5 x IQR). Observed points more extreme than the adjacent values (outside values) are individually plotted. The width of the box is proportional to the square root of the number of observations.

Figure 2: Scatter plot and predicted regression of paired $EQ5D_{index}$ and $SF6D_{index}$ utility scores at baseline and month 12 post-randomisation

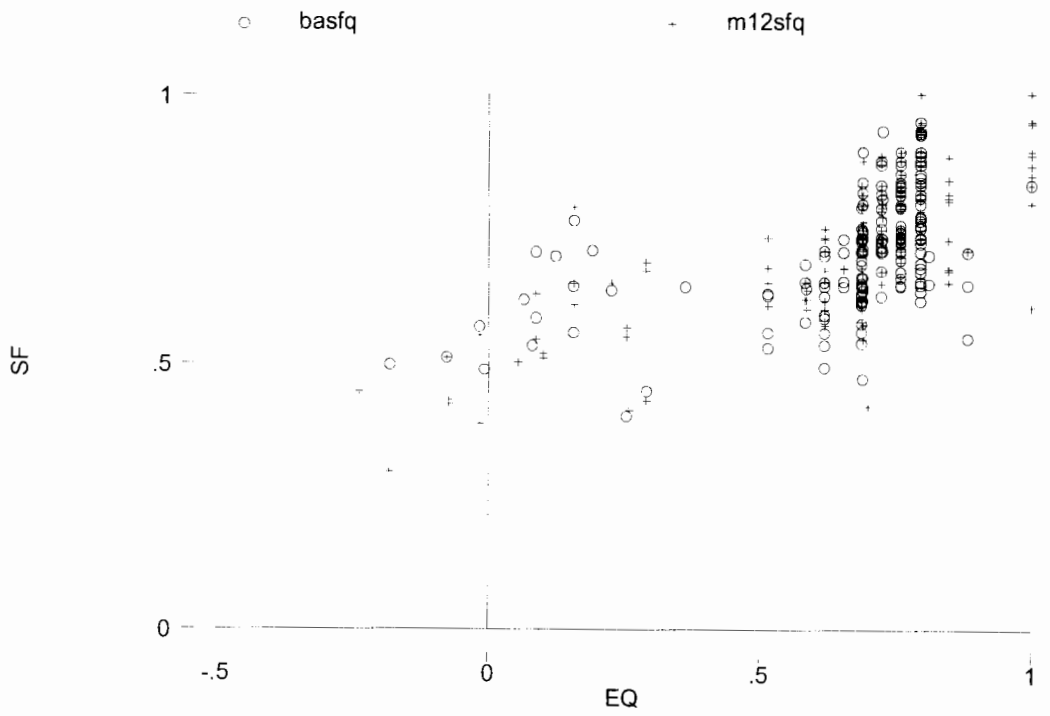


Table 3: Number (%) of respondents at baseline reporting best and worst levels of each dimension for SF6D and EQ5D (n=268)

	EQ5D				SF6D	
Dimension	Best	Worst	Dimension	Survey	Best	Worst
Mobility	**220 (82)**	**-**	Physical	*3a*	*24 (9)*	*104 (39)*
Self care	**253 (94)**	**-**		*3b*	*123 (46)*	*21 (8)*
Usual	**103 (38)**	**3 (1)**		*3j*	*228 (85)*	*5 (1)*
Pain	**4 (1)**	**14 (5)**	**Physical**	**Dimension**	**21 (8)**	**5 (1)**
Depression	**166 (62)**	**5(2)**	Role	*4c*	*67 (25)*	*2 (0)*
				5b	*126 (47)*	*0 (0)*
			Role	**Dimension**	**260 (97)**	**6 (2)**
			Social	**10**	**100 (37)**	**2 (1)**
			Pain	*7*	*4 (1)*	*3 (1)*
				8	*24 (9)*	*6 (2)*
			Pain	**Dimension**	**1 (0)**	**6 (2)**
			Mental	*9b*	*130 (49)*	*2 (1)*
				9f	*97 (36)*	*1 (0)*
			Mental	**Dimension**	**70 (26)**	**3 (1)**
			Vitality	**9e**	**9 (3)**	**32 (12)**

Table 4: Changes in health state measured with EQ between baseline and month 12 (n=268)

Most frequently reported states at month 12									
State	11111	11112	11121	11122	11221	11222	21221	21222	Total
Tariff	1.000	0.848	0.796	0.725	0.760	0.689	0.691	0.620	
11111	*								0
11121		*							0
11121	9	3	16*	4	7	3	1	6	49
11122	1	1	4	1*	4	1		1	13
11212			1		1				2
11221	7	2	8	6	6*	8	5	3	45
11222	4	1	8	2	3	2*	6	2	28
11223			1						1
11231			1						1
11232								1	1
11233				1					1
11291		2							2
12221							1		1
12331							1		1
21121			1					1	2
21122		1	1						2
21221	2		4		4		2*	1	13
21222	1	1	4	1	1			1*	9
21231				1					1
21232					1				1
21233			1						1
21322				1					1
22221			1						1
22222	1		2			1			4
22233					1				1
Total	25	11	53	17	28	15	16	16	181

*participants who reported no change in EQ state over 12 months

Appendix 1 : Procedure to derive SF6D levels from SF36 answers for each SF6D dimension

The SF36 questions used to derive the SF6D

			Levels
Phys	*SF3a*	*Limited in vigorous activity*	*3*
	SF3b	*Limited in moderate activity*	*3*
	SF3j	*Bathing and dressing yourself*	*3*
Role	*SF4c*	*Limited in work due to physical health*	*5*
	SF5b	*Limited in work due to emotional problems*	*5*
Social	*SF10*	*Limited in social activity*	*5*
Pain	*SF7*	*Bodily pain*	*6*
	SF8	*Limited in work due to pain*	*5*
Mental	*SF9b*	*Nervous*	*5*
	SF9f	*Depressed*	*5*
Vitality	*SF9e*	*A lot of energy*	*5*

Note that a higher SF36 score indicates a better health outcome, while a higher SF6D level indicates a worse health outcome. The SF36 question response is in *italics*, the corresponding SF6D dimension level is in **bold.**

1. Physical functioning dimension

		SF3b		
	Level	*3*	*2*	*1*
SF3j	*3*	**1 or 2 (*)**	**3**	**4**
	2	**5**	**5**	**4**
	1	**6**	**6**	**6**

* *if SF3a = 1 then* **Role =1**, *or* **Role=2** *otherwise*

2. Role limitation dimension

		SF4c	
SF5b	*Level*	*>=2*	*1*
	>=2	**1**	**2**
	1	**3**	**4**

3. Social dimension

	SF10				
Level	*5*	*4*	*3*	*2*	*1*
	1	**2**	**3**	**4**	**5**

4. Pain dimension

	SF7					
	Level	*5*	*4*	*3*	*2*	*1*
SF8	*5*	**1**	**2**	**2**	**2**	**2**
	4	**3**	**3**	**3**	**3**	**3**
	3	**4**	**4**	**4**	**4**	**4**
	2	**5**	**5**	**5**	**5**	**5**
	1	**6**	**6**	**6**	**6**	**6**

5. Mental health dimension

	SF9b					
	Level	*5*	*4*	*3*	*2*	*1*
SF9f	*5*	**1**	**2**	**3**	**4**	**5**
	4	**2**	**2**	**3**	**4**	**5**
	3	**3**	**3**	**3**	**4**	**5**
	2	**4**	**4**	**4**	**4**	**5**
	1	**5**	**5**	**5**	**5**	**5**

6. Vitality dimension

	SF9b				
Level	*5*	*4*	*3*	*2*	*1*
	1	**2**	**3**	**4**	**5**

Appendix 2 : EQ5D questionnaire

Level	1	2	3
Mobility	*No problem*	*Some problem*	*Confined to bed*
Self care	*No problem*	*Some problem*	*Unable to wash or dress myself*
Usual activity	*No problem*	*Some problem*	*Unable to perform usual activity*
Pain / discomfort	*No pain*	*Moderate*	*Extreme*
Anxiety/depression	*No problem*	*Moderate*	*Extreme*