

# Assessing the performance of the EQ-VAS in the NHS PROMs programme

Yan Feng · David Parkin · Nancy J. Devlin

Accepted: 17 September 2013 / Published online: 1 October 2013  
© Springer Science+Business Media Dordrecht 2013

## Abstract

**Objectives** The study aims to increase knowledge about the performance of the EuroQol-visual analogue scales (EQ-VAS) in the UK NHS patient-reported outcome measures (PROMs) programme, which covers groin hernia, hip and knee replacement and varicose vein surgery, and make suggestions for improved collection, coding and analysis of data.

**Methods** Four hundred scanned images of matched before-and-after EQ-VAS PROMs responses were selected at random. These were classified according to the different ways in which they were completed. Patient-level PROMs programme data linked to Hospital Episode Statistics for all patients from April 2009 to February 2011 were used to analyse the relationship between the EQ-VAS and the EQ-5D profile, index-weighted profile and condition-specific instruments. The linked PROMs and HES data comprise 331,951 anonymised patient records.

**Results** A large majority (95 %) of EQ-VAS responses were completed in an unambiguous way, but only a minority (45 %) conformed strictly to the instructions given, posing challenges for data coding. The EQ-VAS data have a predictable and consistent relationship with the EQ-5D profile, although the correlations between the EQ-VAS and other measures of patient-reported health, both

before and after surgery and in the change between them, are weak.

**Conclusions** EQ-VAS data might be improved by providing better guidance on collection and coding. It is argued that the observed differences in results from EQ-VAS and other measures of health reflect the fact that it measures a broader underlying construct of health, arguably providing a means of summarising overall health that is closer to the patient's perspective.

**Keywords** Outcomes research · EQ-5D · Visual analogue scales · Patient-reported outcomes

## Introduction

The NHS patient-reported outcome measures (PROMs) programme introduced in April 2009 is a significant development in the routine collection and use of patient-reported outcome information. Data, including the EQ-5D and condition-specific measures, are collected from all National Health Service (NHS) patients in England undergoing four elective surgical procedures, both before and after surgery. The range of conditions for which PROMs data are collected will be extended gradually, including long-term conditions and also incorporated into a new GP Patient Survey. Longer-term PROMs collection will be rolled out across all NHS services wherever practicable [1]. This would mean several million patients will complete the EQ-5D in England each year.

Results from the PROMS initiative are reported on a Department of Health website, actively disseminated to NHS organisations and used in a wide number of decision-making contexts. For example, comparisons of changes in patient health from before to after surgery are used as one

---

Y. Feng · N. J. Devlin (✉)  
Office of Health Economics, 7th Floor, Southside,  
105 Victoria Street, London SW1E 6QT, UK  
e-mail: ndevlin@ohe.org

Y. Feng  
e-mail: yfeng@ohe.org

D. Parkin  
Department of Primary Care and Public Health Sciences, King's  
College London, London, UK

indicator of hospitals' performance [2]. Those performance indicators are also available to patients to help them to choose the hospital where they will have their operation. NHS commissioners also use the data in evaluating effectiveness and cost-effectiveness of services [3].

The EQ-5D has two parts. The EQ-5D self-classifier asks patients to describe their health in terms of the level of problems (“no”, “some” or “extreme”) on each of five dimensions, giving a health “profile”. The EQ-VAS is a vertical visual analogue scale that takes values between 100 (best imaginable health) and 0 (worst imaginable health), on which patients provide a global assessment of their health. The EQ-VAS is reproduced in Appendix 1.

The EuroQol Group, which developed and owns the copyright on the EQ-5D, recommends that both of these parts be used [4]. The data can be analysed and reported in terms of the profile itself, an index number derived from the profile using a standard set of weights, or the EQ-VAS. These can be reported as levels before and after surgery and the difference between the two [5].

Because the PROMs programme will increasingly drive important decisions in the NHS, the data are coming under close scrutiny. Two particular concerns have been expressed about the nature of the EQ-VAS data that have been collected. First, the contractor to the Department of Health that collects these data, Quality Health, alleges that patient questionnaires present various “irregularities” or “difficulties” in accurately coding the EQ-VAS (personal communication). Secondly, the EQ-VAS data appear to yield quite different results with respect to the effect of surgery on patient health, compared with both the EQ-5D index and condition-specific scores. For example, smaller proportions of patients are observed to have an improvement in their health and substantially higher proportions apparently worsen [2].

The NHS Information Centre's explanation for these differences is that the EQ-VAS captures aspects of quality of life that are not related to the patient's condition or the outcomes of surgery:

The variation in improvement seen for each of these scoring mechanisms may be partly due to their nature. The EQ-VAS score asks patients to score their health on the day that they complete the questionnaire and therefore provides an indication of the patient's health that may not necessarily be associated with the condition for which they underwent surgery and may be affected by factors other than healthcare... The EQ-5D Index score reflects general health status, capturing condition specific issues in a broad way, but is more disaggregated than the EQ-VAS [2].

However, the suggestion that patients think specifically about their surgery-related health problems while

completing the EQ-5D profile and the condition-specific questions, but not when completing the EQ-VAS, is not supported by any evidence. On the contrary, this seems unlikely because all of the instruments are administered at the same time, within the same questionnaire and within the same context, namely the specific healthcare intervention that they will receive or have received.

In view of the widespread use of the EQ-VAS in clinical and epidemiological studies, it is perhaps surprising that there are few publications that report on and discuss these issues. It seems unlikely that they have not been encountered, so they may simply be the kind of methodological issues that are often left out of published papers. It may be because the resource implications of dealing with them are less visible in a typical trial or cohort study than in a large-scale routine data collection exercise such as PROMs, so the issues have not been publicly raised. However, a study of the use of PROMs in the Danish Hip Arthroplasty Registry [6] reported that its automated processing of electronically scanned patient questionnaires failed for the EQ-VAS about 3 times as often as other questionnaires.

These issues raise fundamental questions about the role and use of EQ-VAS. This paper aims to improve our understanding of EQ-VAS data, leading to better ways of collecting, coding and analysing them. It investigates the causes of the alleged problems with EQ-VAS data in the PROMs programme and the extent to which these account for the observed differences between EQ-VAS, profile and index-weighted profile data. In particular, the paper analyses

1. The different ways in which patients complete the EQ-VAS and how this is affected by their characteristics;
2. How the different ways of completing the EQ-VAS are currently handled in coding the data in the PROMs programme and other applications, and how these deal with variations from the way intended by the questionnaire instructions; and
3. The relationship between the EQ-VAS, the EQ-5D profile and other summary score data in the NHS PROMs programme, as a way of examining how the differences between these as measures of patient-reported health arise.

We begin by reviewing the theoretical literature about visual analogue scales, what they measure, and how they are used in the measurement of health status. Following descriptions of data, methods and results, we consider the implications of our findings for the use and analysis of EQ-VAS data, the interpretation of results from the PROMs programme and potential implications for the current design of EQ-VAS.

## What does the EQ-VAS measure?

Visual analogue scales (VAS) have been used in psychological research for nearly a century, dating from early experimentation with use of a “graphic rating scale” [7, 8]. They ultimately derive from psychophysics, notably Fechner in 1860 [9]. This is concerned with “the way in which people perceive and make judgements about physical phenomena, such as the length of a line, the loudness of a sound or the intensity of a pain: psychophysics investigates the characteristics of the human being as a measurement instrument” [10]. It is concerned with the subjective judgment of phenomena that can be measured objectively. An extension of this is psychometrics, the application of psychophysical methods to measuring qualities for which there is no physical scale, which is the basis for measuring subjective assessments in health and social sciences.

VAS became widely used in the 1960s, following the work of Aitken [11] and others, who used them as a single-item approach to the measurement of mood. He argued that “words may fail to describe the exactness of the subjective experience” and advocated use of VAS to measure feelings. Subsequently, VAS were developed for a wide range of research and clinical applications, including mood, suicidal intent, depression, anxiety, dyspnoea, craving for cigarettes, quality of sleep, functional abilities, acute pain, chronic pain, nausea, grip, disability and vigour [12, 13]. The VAS became used as a measure of health-related quality of life from the 1970s, following Priestman and Baum’s [14] study of cancer patients.

Advantages noted for VAS include simplicity, ease of administration and scoring, and suitability for frequent and repeated use. Studies generally report high levels of validity and reliability [12], including when used to measure quality of life [15]. However, Streiner and Norman [16] noted that some studies suggest that the VAS is not always considered simpler than alternatives, such as “adjectival rating scales” that use verbal descriptors along a continuum instead of simply labelling the end points and that illiterate and older people can experience difficulties in completing a VAS.

The EuroQol Group’s use of the EQ-VAS to seek an overall measure of health status might be seen as part of this wider tradition of VAS measurement. However, the specific form, wording and presentation of the EQ-VAS to measure self-reported health came about indirectly. Its primary function was not to assess health status for its own sake, but to act as a warm-up task for the valuation of EQ-5D health profiles using a VAS. Early research on valuing EQ-5D profiles mainly used paper questionnaires in which people were asked to value several profiles using a VAS. The VAS was presented as a vertical line, marked from 100

(best imaginable health state) to 0 (worst imaginable health state) in the centre of each page, with 4 profiles presented in boxes to the left and right of it. Respondents were asked to draw a line from each box to the VAS to indicate how good or bad each is in their opinion.

This provenance of the EQ-VAS is reflected in crucial aspects of its design. For example, people were asked to draw a line to the VAS from the box marked “Your own health state today” to prepare them for the subsequent valuation task, which used the same procedure. The valuation task used that device, instead of the more usual marking a point on the VAS, to ensure that the values for several different profiles could be recorded on the same VAS without ambiguity. The EQ-VAS, which requires a single line to be drawn, was a relatively simple way to get people used to the idea. Similarly, the vertical orientation, scale demarcation, numbering and end points were all determined by the requirements of the VAS valuation task.

Other special characteristics of the EQ-VAS are important. A VAS is often simply a straight line of specified length with verbal descriptors at each end stating the meaning attached to the end points. The EQ-VAS has such descriptors, but also demarcates the line in units of ones and tens, and places number labels on the multiples of tens. Formally, this is a “numerical rating scale”, though such scales often do not have end point descriptors.

The labels used to describe the end points of a VAS are especially important [16]. The EQ-VAS labels may mean different things to different people completing it, which may “attenuate the comparison of scores” [17]. Early studies conducted by the EuroQol Group using convenience samples identified various issues, such as occasional misinterpretation of “best imaginable” to mean how easily the state could be imagined [18]. However, to our knowledge, there never has been any investigation into the way the EQ-VAS end points are defined by respondents in a non-experimental context and how this affects the way in which they respond to it.

The end points of a VAS measuring the intensity of a single phenomenon, such as pain, may run from zero intensity, such as “no pain”, to an upper intensity limit, such as “as painful as can be”. However, the end points for the EQ-VAS are “worst imaginable health state” and “best imaginable health state”. It is possible to argue that these are two distinct concepts, in which case the EQ-VAS might be described as a bi-polar scale. Such scales are more difficult for subjects to understand than unipolar scales [15, 19].

Both the underlying purpose and stimuli provided by the EQ-VAS differ in important ways from the EQ-5D profile. The EQ-5D profile was developed to produce a short, easily self-completed measure of a common core of dimensions of health-related quality of life, capable of yielding a single index value for any health state defined by

it [20–22]. The EQ-VAS seeks a respondent's overall rating of their health. Any aspects of health-related quality of life that matter to respondents, not just those contained in the five EQ-5D dimensions, will influence the way that overall health is described on the EQ-VAS. For example, it is commonly observed that some respondents who describe themselves as having no problems in any of the dimensions of the EQ-5D provide an EQ-VAS rating of their health that is less than 100 [23].

A comparison of respondents' self-reported EQ-VAS with their weighted EQ-5D profile index may suggest different results simply because of these extra-dimensional considerations. This also may happen because the index weightings reflect stated preferences elicited from members of the general public asked to imagine those health states, rather than the views and values of people who are experiencing them. It may be, for example, that valuations by those currently experiencing health states take into account any adaptation that they have made to mitigate their underlying health state, for example pain relief or mobility aids, or other ways they have found to cope with it. As alternative means of providing a single summary score of patient health, there are therefore key differences in *what* is being valued in each case, as well as:

1. The methods by which they are obtained. For example, the EQ-VAS has a lower limit score of 0, whereas index weightings are obtained by a variety of methods, known to produce different results, which involve anchoring at dead = 0 and allow weights <0, reflecting states worse than dead.
2. Whose views are represented? For example, it is known that there are differences between patients' experienced utility and the general public's affective forecasts of utility in health states they have not experienced [24, 25]. Furthermore, conclusions about similarities or statistically significant differences between patients' EQ-VAS and their index-weighted EQ-5D profiles will depend on *which* set of weights are applied to those profiles, as each set of weights has its own properties [26].

These considerations affect empirical comparisons between EQ-VAS data and EQ-5D profiles and indexes. In making such comparisons, it is also assumed that the numerical values given to the EQ-VAS and EQ-5D index behave as if they have a cardinal scale and are interpersonally comparable, such that it is meaningful to calculate descriptive statistics for the data, such as means, and to apply statistical procedures, such as correlation and regression analysis. Whether or not this assumption is justified is beyond the scope of this paper to consider. However, the PROMs programme implicitly makes this assumption. To ensure that our analysis is consistent with

and relevant to this context, we make the same assumption explicitly.

## Data and methods

### Analysis of response types

The alleged problems with coding EQ-VAS data in the PROMs programme were investigated using a sample of completed EQ-VAS forms. The aim was to establish the frequency of responses that did not follow the instructions given on the form; to analyse similarities and differences in the way that respondents who did not follow the instructions completed them; to understand how such responses are currently coded; and to draw out potential implications for the design, analysis and interpretation of EQ-VAS.

Quality Health, a contractor to the English NHS PROMs programme, provided us with scanned images of matched before-and-after EQ-VAS responses of a randomly selected sample of 200 patients across all four elective surgical procedures, giving a total of 400 images. These data were anonymous and contained no means of linking them to other data sets. The data included some background characteristics, namely age, sex and operation type.

We constructed a classification of different ways in which respondents completed the EQ-VAS. Two of the authors (Feng and Devlin) independently examined the images and proposed a list of these. Both used a "constant comparisons" approach, examining responses sequentially until no new completion types emerged. The third author (Parkin) examined the lists, and led a process that agreed a final classification by consensus. This was used to categorise all responses across the entire sample.

The way that each of our identified response types are currently handled in practice was examined by consulting the coding manuals for EQ-VAS data used by the NHS PROMs programme and, for comparison, guidance on EQ-VAS data provided by the EuroQol Group.

### Analysis of EQ-VAS in the PROMs dataset

We also analysed patient-level NHS PROMs programme data linked to Hospital Episode Statistics (HES) data, provided by the NHS Information Centre. This covered all cases for the four elective procedures covered by PROMs from 1 April 2009 to 28 February 2011, comprising 331,951 anonymised patient records.

The variables used in the analysis include the type of surgery performed and all of the PROMs data both before (Q1) and after (Q2) treatment. The PROMs data were the index-weighted EQ-5D scores, EQ-5D profile, EQ-VAS and scores for the condition-specific instruments, the

Oxford Hip Score (OHS), Oxford Knee Score (OKS) and Aberdeen Varicose Vein Score (AVVS). The OHS and OKS range from 0 (worst) to 48 (best). The AVVS ranges from 100 (worst) to 0 (best).

Regression analysis, using ordinary least squares, explored the relationship between the EQ-VAS and the EQ-5D profile. The independent variables represent the five dimensions of the EQ-5D profile. Dummy variables were used to represent levels 2 and 3 within each dimension, with level 1 as the comparison baseline. We also tested for differences between the level 2 and level 3 coefficients.

## Results

### Analysis of response types

The initial classification identified 15 different ways in which respondents completed the EQ-VAS. However, the differences between some of these types of response were too small to warrant distinguishing from each other. A reduced classification had six key EQ-VAS completion types, described in Table 1 along with the frequency with which they were observed in the Q1 and Q2 responses. Type I is completion in the intended way; examples of types II–V are provided in Appendix 2. In our “Discussion” section, we speculate about the reasons why these different types of response may have arisen.

Only type I responses strictly follow the instructions on the EQ-VAS and can be labelled as correct. However, in addition to these, response types II and III also unambiguously identify a single number. The reason why type III is unambiguous is that, as discussed below, the respondents were clearly attempting to indicate a number by taking more literally the idea, expressed as an analogy in the completion instructions, that the VAS is a thermometer. They were therefore drawing an analogy of how a line of mercury, or other liquid, would look for a particular temperature.

There were no significant differences in the proportions of correct and unambiguous responses according to age and sex. The same was true for condition type, except that those with varicose veins were slightly less likely to complete the first questionnaire correctly compared with other conditions. However, they were as likely to complete it unambiguously and to complete the second questionnaire both correctly and unambiguously.

There was a difference between the ways in which correct and unambiguous completion proportions changed from Q1 to Q2. A significantly greater proportion completed Q2 correctly than completed Q1 (McNemars’ test,  $p = 0.0169$ ). However, there was no significant change in the proportion completing them unambiguously ( $p = 0.089$ ).

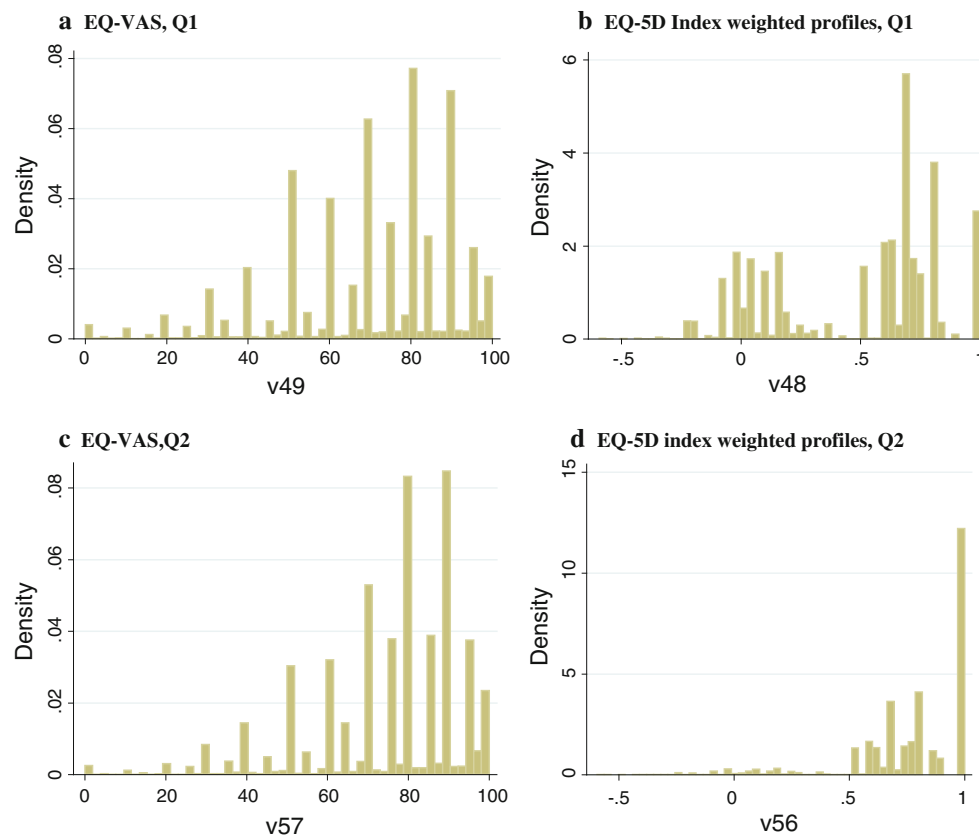
**Table 1** EQ-VAS response types and the frequency with which these are observed, in a randomly selected sample of 400 matched before-and-after responses in the NHS PROMs programme

EQ-VAS response type	Number of responses to Q1 ( $n = 199^*$ )	Number of responses to Q2 ( $n = 200$ )
I. Drew a line from the box towards the EQ-VAS, sometimes touching or crossing it. This is the way that the EuroQol Group intends the EQ-VAS to be completed	79 (39.7 %)	100 (50.0 %)
II. Indicated precisely a horizontal level on the VAS, but did not draw a line to it. For example, ticks, crosses, lines, arrows, asterisks on or beside the VAS, or a tightly drawn circle around a specific number or tick mark	72 (36.2 %)	54 (27.0 %)
III. Drew a vertical line extending from 0 up to a point parallel with a point on the VAS	17 (8.5 %)	26 (13.0 %)
IV. Drew a vertical line parallel to the VAS, but not extending from 0, or circled an area of the VAS. This indicated a range rather than a single point	10 (5.0 %)	7 (3.5 %)
V. Gave an unclear response. For example, multiple markings on the VAS or vertical lines drawn from 100 downwards	1 (0.5 %)	2 (1.0 %)
VI. Left the form blank	20 (10.1 %)	11 (5.5 %)

\* There was one missing Q1 response because of an image copying error

### Current approaches to coding EQ-VAS data

The coding procedures for EQ-VAS used in the NHS PROMs programme are provided in Appendix 3 with, for comparison, the current coding procedures noted by the EuroQol Group (Boxes 1 and 2, respectively). Both coding procedure guides cover instances where the line from the box goes towards the EQ-VAS, but does not touch or cross it. The PROMs guide has a procedure for type III responses, but these are not mentioned in the EuroQol Group guide. Similarly, the PROMs guide has procedures for handling ranges, but the EuroQol Group guide does not. However, the PROMs procedures for a range, what we have called a type IV response, may not be entirely consistent. The procedure for a range indicated by a vertical line is to record the lowest value. However, where there is a mark on the scale that implicitly describes a range, for example, a circle, the mid-point is recorded.



**Fig. 1** The distribution of EQ-VAS and index-weighted profiles in Q1 and Q2, all procedures combined **a** EQ-VAS, Q1, **b** EQ-5D index-weighted profiles, Q1, **c** EQ-VAS, Q2, **d** EQ-5D index-weighted profiles, Q2

### Analysis of the HES data

The data contained both Q1 and Q2 questionnaires, but not everyone completed the EQ-VAS on both occasions. There were 331,951 respondents to Q1, 294,249 of whom completed the EQ-VAS. Of these, 159,697 also completed it in Q2. A total of 17,862 patients did not complete the EQ-VAS in Q1, but did so in Q2. Therefore, 294,249 EQ-VAS responses are available for analysis from Q1, 177,559 for Q2 and 159,697 for both, which enables us to analyse changes from Q1 to Q2.

One of the key comparisons that we will use is between the EQ-VAS and the weighted profile score. Figure 1a–d shows their distributions. A feature of EQ-VAS data is digit preference, whereby responses cluster around tens and to a lesser extent fives. The EQ-VAS distribution is unimodal. However, a feature of many EQ-5D-weighted profile data distributions is that they fall into two separable groups [27]. Taking account of this difference is important when comparing the two scores.

Of equal importance for our comparisons is the EQ-5D profile itself. Overall, comparing patients' EQ-5D profiles between Q1 and Q2 demonstrates the overall positive effects of surgery. As might be expected, there is a

reduction in the proportion of patients reporting a level three (“extreme problems”) and a level two (“some problems”), and an increase in the proportion reporting no problems, on every dimension following surgery. Devlin, Parkin and Browne [5] discussed the difficulty in summarising overall changes in EQ-5D profiles and proposed for this the Paretian Classification of Health Change (PCHC), which classifies patients as either having no EQ-5D problem (11111) both before and after surgery; the same (imperfect) health at both points in time; or improved, worse or mixed changes in health. We have used the PCHC to compare the performance of the different index numbers according to the patterns demonstrated by the profiles themselves. Table 2 reports, for each PCHC category, the average change in EQ-VAS, the mean change in the EQ-5D-index-weighted profiles and the mean changes in the condition-specific scores.

For those reporting an EQ-5D profile of 11111 both before and after surgery, the small negative mean change in the EQ-VAS contrasts with the 0 change (by definition) in the EQ-5D index and improvements in the hip, knee and varicose veins condition-specific scores. Similarly, those with identical EQ-5D profiles before and after surgery, but worse than 11111, show a small, negative mean change in

**Table 2** The relationship between the Paretian Classification of Health Change (PCHC) and mean changes in EQ-VAS scores, EQ-5D index, OHS, OKS and VV score

PCHC category	Mean change in				
	EQ-VAS	EQ-5D Index	OHS	OKS	VV
No EQ-5D problems	−0.499 (−0.713, 0.286)	0	7.11 (5.63, 8.60)	8.28 (6.80, 9.76)	−6.58 (−6.86, −6.31)
No change	−2.17 (−2.45, −1.89)	0	10.5 (10.2, 10.8)	8.16 (7.97, 8.36)	−5.23 (−5.63, −4.84)
Improved	7.377 (7.26, 7.50)	0.410 (0.409, 0.412)	21.9 (21.8, 22.1)	17.8 (17.7, 17.9)	−10.2 (−10.0, −10.5)
Worsen	−9.24 (−9.55, −8.94)	−0.212 (−0.216, −0.210)	4.36 (3.99, 4.749)	2.54 (2.31, 2.78)	−2.75 (−3.20, −2.31)
Mixed change	−1.48 (−1.83, −1.13)	0.168 (0.164, 0.1734)	15.6 (15.3, 15.9)	11.7 (11.5, 12.0)	−7.04 (−7.74, −6.35)

“No change” excludes those with no EQ-5D problems both before and after surgery

95 % confidence interval in parentheses

**Table 3** Correlations between EQ-VAS score and EQ-5D index, OHS score, OKS score and VV score

	EQ-5D index	OHS	OKS	Aberdeen VV score
EQ-VAS (Q1)	0.453 (0.450, 0.456)	0.379 (0.374, 0.385)	0.383 (0.378, 0.388)	−0.271 (−0.282, −0.259)
EQ-VAS (Q2)	0.613 (0.610, 0.616)	0.585 (0.580, 0.591)	0.559 (0.554, 0.564)	−0.317 (−0.331, −0.302)
Change in EQ-VAS (Q2 minus Q1)	0.328 (0.323, 0.332)	0.337 (0.329, 0.344)	0.298 (0.290, 0.305)	−0.130 (−0.147, −0.113)

Correlations and numbers in each of the columns relate to Q1, Q2 or Q2 minus Q1, as relevant to each row

95 % confidence interval in parentheses

EQ-VAS, compared to (again, by definition) no change on the EQ-5D index, and improvements on the condition-specific instruments. This observation also applies to mixed changes in health status.

In contrast, for EQ-5D profiles that either improved or worsened, the mean changes in the EQ-VAS work in the same (and expected) direction as the changes in the EQ-5D index. For patients whose health is unequivocally worse using the PCHC, each of the condition-specific instruments contradicts that by reporting a small improvement in mean health scores.

Tables 3 and 4 further explore the relationship between the EQ-VAS, EQ-5D index and condition-specific instruments. Table 3 shows the correlations between these for Q1, Q2 and the change between Q1 and Q2. The correlations between the EQ-VAS and the EQ-5D index and each of the condition-specific summary scores are stronger after surgery than before, but greater than the corresponding correlation between the change in the EQ-VAS and the change in the other summary scores. Table 4 shows the correlations between the EQ-5D index and condition-specific scores, which are considerably stronger than for the EQ-VAS and all statistically significant.

One of the allegations made about EQ-VAS data is that patients' responses are so influenced by personal and non-health-related contextual factors that they have no consistency and no relation to underlying health status. We explored this by examining the extent to which respondents' EQ-VAS scores could be predicted from their EQ-

5D profile. We estimated a simple regression model in which the VAS score was the dependent variable and the levels in each dimension of the EQ-5D were binary independent variables. We applied this model to Q1 data, Q2 data and pooled Q1 and Q2 data. For consistency, we used only data from respondents who completed both Q1 and Q2. Table 5 presents these results for Q1 data. Very similar results were obtained for Q2 and pooled Q1 and Q2 data.

The adjusted  $R^2$  suggests that respondents' EQ-5D profiles only partially explain their EQ-VAS scores. However, the binary variable coefficients are all in the expected direction and are highly statistically significant. Moreover, they are consistent in each dimension, so that the coefficients on level 3 are all higher than the coefficients of level 2. The differences between the level 2 and level 3 scores are all significant ( $p < 0.05$ ).

We repeated this procedure using the two Oxford hip and knee score instruments, and again found that the items within them produced a reasonable and always consistent model, although not as good as the EQ-5D profile model.

## Discussion

The concerns raised about the EQ-VAS in the NHS PROMs programme have not been widely expressed elsewhere. Nevertheless, the concerns need to be investigated, and our findings suggest ways in which EQ-VAS data

**Table 4** Correlation coefficients showing the relationship between EQ-5D index, HR score, KR score and VV score

	Oxford_HR	Oxford_KR	Aberdeen_VV
EQ5D_Index before	0.742 (0.739, 0.744)	0.709 (0.706, 0.711)	−0.415 (−0.425, −0.405)
EQ5D_Index after	0.766 (0.763, 0.769)	0.773 (0.769, 0.776)	−0.477 (−0.490, −0.465)
EQ5D_Index changes	0.630 (0.625, 0.635)	0.590 (0.585, 0.595)	−0.313 (−0.327, −0.297)

Correlations and numbers in each of the columns relate to Q1, Q2 or Q2 minus Q1, as relevant to each row

95 % confidence interval in parentheses

**Table 5** The relationship between the EQ-VAS and the dimensions/levels of the EQ-5D

	Coefficient	Standard Error
Mobility level 2	−5.1151	0.1412
Mobility level 3	−10.6205	0.8807
Self-care level 2	−6.5424	0.1098
Self-care level 3	−10.5095	0.5634
Usual activities level 2	−3.4104	0.1439
Usual activities level 3	−7.6327	0.2009
Pain and discomfort level 2	−2.3947	0.1608
Pain and discomfort level 3	−6.6688	0.1929
Anxiety and depression level 2	−7.8700	0.1001
Anxiety and depression level 3	−15.2284	0.2524
Constant	86.3203	0.1298

Number of observations = 154,890.  $R^2 = 0.2672$ , adjusted  $R^2 = 0.2672$ ,  $F = 5,647.66$ ,  $p = 0.00005$

All coefficients significantly different from 0 at the 0.0005 level

might be improved by better collection and coding procedures.

The EQ-VAS has an advantage over the EQ-5D and condition-specific profile data because those consist of multiple data items and are more prone to being unusable because of missing items. However, the EQ-VAS has the disadvantage of being more challenging to complete, compared to the “tick box” responses required of the profiles, potentially leading to more unusable responses. In our sample, a large majority (95 %) of those completing the EQ-VAS did so in an unambiguous way, suggesting that most respondents understood that they were expected to indicate a single number. However, many did not understand exactly *how* they were supposed to indicate it. Better instructions and appropriate coding rules therefore could substantially increase the number of usable responses.

We do not know why respondents who responded unambiguously did not follow the instructions fully, but we can speculate. It is possible that they understood what was required of them and attempted to provide it, but used only parts of the instructions. The instructions are detailed and written in a style that many people would find difficult. They work as a linear narrative, but people may not treat

them that way, may assign more importance to some parts than others, find some parts easier to understand than others, find some parts more memorable than others and even find individual parts inconsistent with each other. Those with type II responses obeyed the first sentence of the second paragraph of instructions, which asks them to indicate their health on the scale. They simply ignored the more specific instruction in the second sentence to draw a line from the box. Similarly, type III responses are, as suggested, essentially a drawing of how a specific temperature would appear on a thermometer. Respondents clearly took the message from the first sentence of the first paragraph that the scale was analogous to a thermometer and gave their response in that way, ignoring the more detailed instructions below that.

If this argument is correct, then despite the variance in the way that respondents have completed the EQ-VAS the unambiguous data are not only usable, but also consistent with each other.

Nevertheless, fewer than half of respondents complete the EQ-VAS in the way that the instructions are designed to produce. Further, the current guidance provided by the EuroQol Group on the coding of these data does not address many of the common forms of completion adopted by respondents. This has the potential to result either in unnecessary data wastage or to different users adopting different practices for interpreting and coding these data. It is unclear to what extent these data coding issues apparent with the EQ-VAS in the PROMs programme also are evident in other applications and previous studies. We were unable to find any examples of these issues being documented or reported in published papers. However, it seems unlikely that these problems are new or restricted to the PROMs programme.

A particular difficulty is when respondents indicate a range, which was the case for around 4 % of those in our data set. If this arises because the respondents do not understand the instructions, it may be that the incidence of this could be reduced by instructions that specifically ask respondents to indicate one number only, or not to provide a range. However, it may be more likely that respondents are unwilling to provide a single number, reflecting their uncertainty about what it should be. In this case, it is most



important to have coding rules that reflect what in principle should be recorded as the single value, rather than a simple pragmatic rule. These should certainly be consistent between different ways in which ranges are recorded.

All of these issues could be addressed by providing improved guidance on coding EQ-VAS data or revisiting the instructions regarding the EQ-VAS. Arguments for considering a change to the EQ-VAS task include

1. The current instructions are a historical by-product of the initial role of the EQ-VAS as a warm up to subsequent VAS valuation tasks, whereas the EQ-VAS is now a fundamental element of the EQ-5D. It is not now necessary to use the box-and-line device, for example.
2. Only a minority of respondents complete the task using all of the instructions that they are given
3. The description of the EQ-VAS as being “a bit like a thermometer” may become less and less relevant as traditional thermometers are replaced by digital displays of temperature readings
4. The increasing use of digital and web-based versions of the EQ-5D has already led to a substantial shift away from the current instructions and format, and
5. An alternative format already exists, in the EQ-VAS used in the new five-level version of the instrument, the EQ-5D-5L [28]. In that instrument, the respondent is asked to mark the EQ-VAS with a cross, and to note the corresponding number in a box. As an increasing number of translations of that version of the instrument become available, this raises the possibility of adopting the same approach for the three-level version of the EQ-5D.

Concerns had emerged from the NHS PROMs programme that the EQ-VAS was not adequately reflecting the health gain for patients resulting from surgery and was therefore a less useful and appropriate measure of health change than the EQ-5D profile or condition-specific instruments. However, our analysis of the EQ-VAS data from the PROMs programme suggests the following.

First, the EQ-VAS has a predictable relationship with the EQ-5D profile. The models estimated from the EQ-5D profile data have a well behaved and consistent ordering of coefficients on the levels of each dimension. Indeed, the models estimated from the NHS PROMs data produce a more consistent relationship between the profile and the EQ-VAS than previously reported [23, 29], with similar explanatory power.

Secondly, some of the difference between the NHS PROMs results reported in terms of the index-weighted profile and the results in terms of the EQ-VAS are attributable to the characteristics of the particular weightings within the EQ-5D index. For example, our model of the EQ-VAS shows that the highest coefficient is for extreme anxiety/depression. The same finding was reported by

Hardman et al. [29]. This contrasts with the weights of the UK EQ-5D index, derived from the general public rather than patients, where the decrements in the index are largest for extreme pain and discomfort [30]. These differences between the views of patients and the general public about what aspects of health impact the most on health-related quality of life provide at least part of the explanation for differences in PROMs results suggested by EQ-VAS and index-weighted EQ-5D profiles.

Nevertheless, our results confirm the observation in PROMs reports that there are clear differences between the EQ-VAS and the index-weighted EQ-5D and condition-specific profiles. There is a moderate correlation between the EQ-VAS and other measures of patient-reported health both before and after surgery, with a slightly weaker correlation between the *change* in the EQ-VAS and the *change* in these other PROMs instruments. In essence, the EQ-VAS is measuring a broader underlying construct than the EQ-5D profile or the condition-specific instruments. This does not mean that the data it produces are less meaningful or useful. Indeed, in applications where the *patients'* view of their overall health is the measurement goal, the EQ-VAS is *prima facie* more appropriate than the use of EQ-5D profile data weighted by general public preferences. Moreover, compared to EQ-VAS scores, condition-specific instruments not only provide a very partial account of overall health, but also have weights, either explicit or implicit, that reflect neither patient nor public preferences, but solely the judgements of a small number of surgeons.

As noted earlier, we have found no papers that investigate how patients or members of the general public interpret the end points of the EQ-VAS and how this may affect the manner in which they self-report their health on it. A report on a survey of EuroQol Group members' understanding of the “intended meaning” of the EQ-5D items revealed somewhat different ways of thinking about the meaning of best and worst imaginable health state [31]. It is unknown whether wider and more disparate interpretations of these concepts are evident across population or clinical subgroups. This is a surprising gap in the knowledge base of the EQ-5D. Given the role of the EQ-VAS in the EQ-5D instrument, and the important policy decisions these data may inform, a better understanding of this—including how the conceptualisation of those end points might shift due to changes in expectations, health or social circumstances—is desirable.

**Acknowledgments** The authors are grateful to Tony Culyer, Reg Race and David Nuttall for helpful comments on an earlier draft, to Tim Benson of Abies Ltd for pointing out an important error in one of the tables in an earlier version and to Quality Health for providing access to patients' completed EQ-VAS pages and information on their coding procedures. Access to that information was granted under the terms of the Department of Health's governance arrangements for the

NHS PROMs programme. The patient-level PROMs data and linked Hospital Episodes Statistics data were provided by the NHS Information Centre. Funding for this study was provided by the EuroQol Foundation.

**Conflict of interest** David Parkin and Nancy Devlin are members of the Department of Health's stakeholder reference group for the

NHS PROMs programme. Both are also members of the EuroQol Group.

## Appendix 1: The EQ-VAS

To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is today.

Your own  
health state  
today

Best  
imaginable  
health state

100

90

80

70

60

50

40

30

20

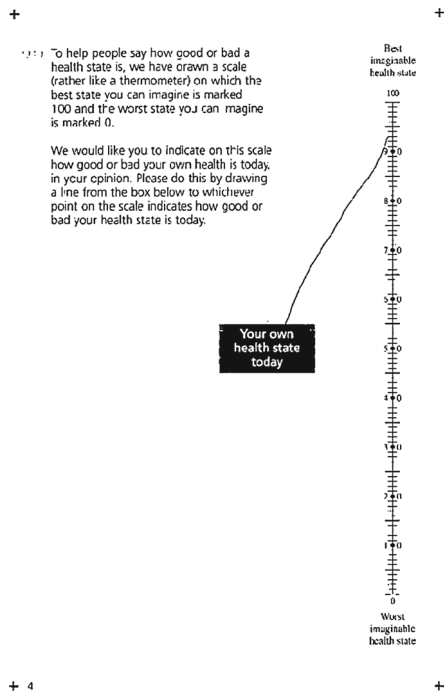
10

0

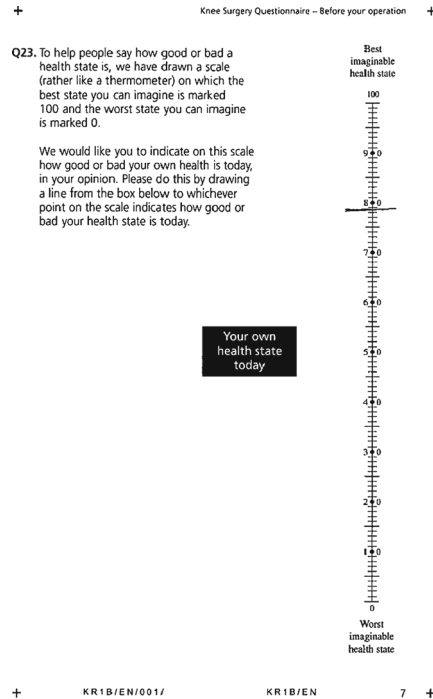
Worst  
imaginable  
health state

Appendix 2: Examples of each completion type

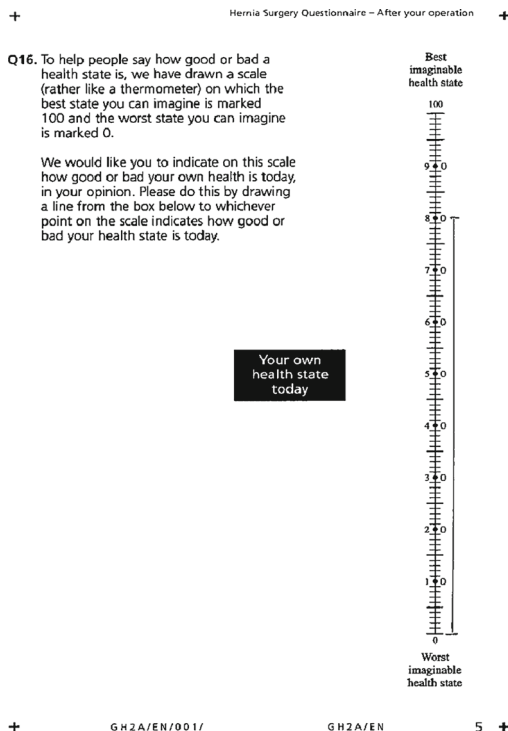
I. Follows EuroQol Group instructions



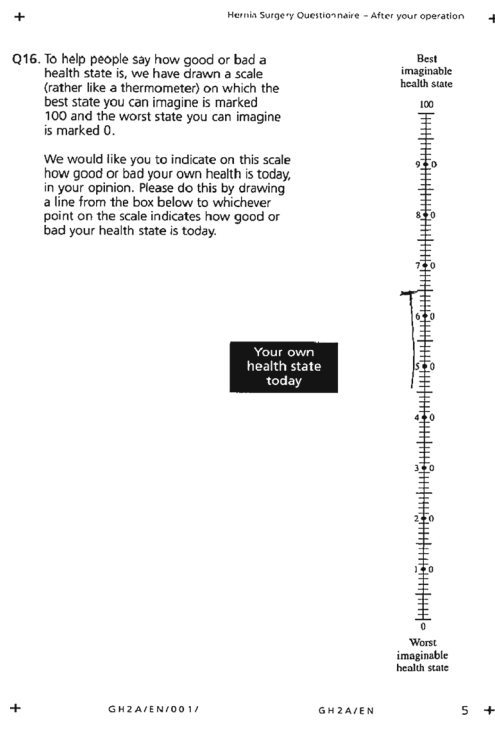
II. Other means of indicating a precise position on the VAS



III. Thermometer completion



IV. Response indicate sarange



### Appendix 3: Coding guides and procedures

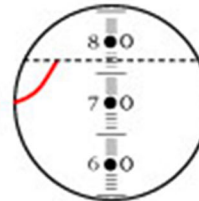
See Boxes 1 and 2.

## Scoring the EQ VAS

The EQ VAS should be scored as follows:



For example this response should be coded as 77



Even though the line does not cross the VAS this response can still be scored by drawing a horizontal line from the end point of the response to the VAS. In this example the response should be coded as 77

Missing values should be coded as '999'.

Ambiguous values (e.g. the line crosses the VAS twice) should be treated as missing values.

#### Box 1 EQ-VAS coding procedures used in the NHS PROMs programme

1. When completed correctly, the question will be coded as the value where the line crosses the VAS
2. If a line has been drawn from the box but does not actually meet the VAS, then the verifier will code by scoring the end of the line in relation to the scale
3. Where a patient has circled/drawn a mark on the scale itself, the responses will be coded at the value of the central point/mark
4. If a patient has drawn a line from the bottom of the scale to a point further up the page, the question will be coded as the highest point relative to the VAS
5. Where a patient has drawn a line indicating a range of health status, e.g., from 20 to 55, our verifiers will code as the lowest point
6. Where there is doubt due to multiple lines or marks, the question will be left blank (Coded as 999). Source: [32]

#### Box 2 Guidance on EQ-VAS data provided by the EuroQol Group

The respondent rates his/her health state by drawing a line from the box marked “Your health state today” to the appropriate point on the EQ VAS

Sometimes, respondents tend to rate their health state by placing a mark on the scale instead of drawing a line. There is no reason why this could not be interpreted as a valid response

If the line does not cross the scale, the value horizontally opposite where the line stops should be taken and not where it would be if hypothetically extended. It is important to ensure that the respondent is not prompted in any way by the administrator and that it is the respondent’s own rating of health-related quality of life that is being recorded

In order to achieve comparable results, it is necessary to adhere to the standard text and instructions and layout of EQ-5D. This is especially relevant for EQ VAS as this is a graphical representation of the value of health (it is important for example that the scale should be a standard 20 cms)

A three-digit number between 000 and 100 is read off the scale, from the exact point where the line crosses the scale, for example, 046 or 069. For comparative purposes, we recommend that: missing response is coded as “999”; ambiguous response is coded as “888” Source: [4]

## References

1. DH (Department of Health, UK). (2010). *Equity and excellence: Liberating the NHS*. London: Stationery Office.
2. IC (NHS Information Centre). (2011). Finalised Patient Reported Outcome Measures (PROMs) in England. Pre- and post-operative data. Experimental statistics. London: Health and Social Care Information Centre. Available at <http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=1583>. Apr 2009–Mar 2010 (Accessed 13 April 2012).
3. Devlin, N., & Appleby, J. (2010). *Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision making*. London: King's Fund and Office of Health Economics.
4. EuroQol Group. (2011). EQ-5D-5L User Guide. Rotterdam: EuroQol Group. Available at <http://www.euroqol.org/eq-5d/publications/user-guide.html>. (Accessed 12 Apr 2012).
5. 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.
6. Paulsen, A., Pedersen, A. B., Overgaard, S., & Roos, E. M. (2012). 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*, 83(4), 321–327.
7. Hayes, M. H. S., & Patterson, D. G. (1921). Experimental development of the graphic rating method. *Psychology Bulletin*, 18, 98–99.
8. Freyd, M. (1923). The graphic rating scale. *Journal of Educational Psychology*, 14(2), 83–102.
9. Herrnstein, R. J., & Boring, E. G. (Eds.). (1965). *A source book in the history of psychology*. Cambridge: Harvard University Press.
10. McDowell, I. (2006). *Measuring health: A guide to rating scales and questionnaires*. New York: Oxford University Press.
11. Aitken, R. C. B. (1969). Measurement of feelings using visual analogue scales. *Proceedings of the Royal Society Medicine*, 62(10), 989–993.
12. McCormack, H., Horne, D., & Sheather, S. (1988). Clinical applications of visual analogue scales: A critical review. *Psychological Medicine*, 18, 1007–1019.
13. Wewers, M., & Lowe, N. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, 13(4), 227–236.
14. Priestman, T. J., & Baum, M. (1976). Evaluation of quality of life in patients receiving treatment for advanced breast cancer. *The Lancet*, 1(7965), 899–900.
15. Boer, A. G., van Lanschot, J. J., Stalmeier, P. F., et al. (2004). Is a single item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Quality of Life Research*, 13(2), 311–320.
16. Streiner, D., & Norman, G. (2008). *Health measurement scales: A practical guide to their development and use*. New York: Oxford University Press.
17. Torrance, G., Feeny, D., & Furlong, W. (2001). Visual analog scales: Do they have a role in the measurement of preferences for health states? *Medical Decision Making*, 21(4), 329–334.
18. Busschbach, J., Hensing, D., & de Charro, F. (2005). Chapter 7: Observations on one hundred students filling in the EuroQol questionnaire. In Kind, P., Brooks, R., & Rabin, R. (eds). *EQ-5D concepts and methods: A developmental history*. Dordrecht: Springer.
19. Nyren, O. (1988). Visual analogue scale. In M. Hersen & A. Bellack (Eds.), *Dictionary of behavioural assessment techniques*. New York: Pergamon Press.
20. EuroQol Group. (1990). EuroQol- a new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199–203.
21. Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, 37(1), 53–72.
22. Kind, P., Brooks, R., & Rabin, R. (Eds.). (2005). *EQ-5D concepts and methods: A developmental history*. Dordrecht: Springer.
23. Whyne, D., & The TOMBOLA Group. (2008). Correspondence between EQ-5D health state classifications and EQ VAS scores. *Health and Quality of Life Outcomes*, 6, 94–103.
24. Brazier, J., Akehurst, R., Brennan, A., et al. (2005). Should patients have a greater role in valuating health states? *Applied Health Economics and Health Policy*, 4(4), 201–208.
25. Mann, R., Brazier, J., & Tsuchiya, A. (2009). A comparison of patient and general population weightings of EQ-5D dimensions. *Health Economics*, 18(3), 363–372.
26. Parkin, D., Rice, N., & Devlin, N. (2010). Statistical analysis of EQ-5D profiles: Does the use of value set bias inference? *Medical Decision Making*, 30(5), 556–565.
27. Parkin, D., Feng, Y., & Devlin, N. (2013). What determines the shape of an EQ-5D index distribution? Paper presented at the Health Economics Study Group meeting, Exeter University. Jan 2013.
28. Herdman, M., Gudex, C., Lloyd, A., et al. (2011). Development and preliminary testing of a new five level version of the EQ-5D: The EQ-5D-5L. *Quality of Life Research*, 20(10), 1727–1736.
29. Hardman, G., Kind, P., & Macran, S. (2002). Living with a VASectomy: Exploring the relationship between EQ-5D responses and the EQ-5D Vas? Paper presented at the 19th Plenary Meeting of the EuroQol Group, York. Available at <http://www.euroqol.org/uploads/media/Proc02York24Hardman.pdf>. (Accessed 6 Mar 2012).
30. Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care*, 35(11), 1095–1108.
31. Fox-Rushby, J., & Selai, C. (2003). What concepts does the EQ-5D measure? Intentions and interpretations. In R. Brooks, R. Rabin, & F. de Charro (Eds.), *The measurement and valuation of health status using EQ-5D: A European perspective*. Dordrecht: Kluwer Academic Publications.
32. Health, Quality. (2010). *Internal document*. Chesterfield: Quality Health.