

Research article

Open Access

What is the value of social values? The uselessness of assessing health-related quality of life through preference measures

Luis Prieto* and José A Sacristán

Address: Health Outcomes Research Unit. Eli Lilly & Co. Spain. Av. de la Industria, 30. 28108 Alcobendas SPAIN

Email: Luis Prieto* - prieto_luis@lilly.com; José A Sacristán - sacristan_jose@lilly.com

* Corresponding author

Published: 29 April 2004

Received: 18 December 2003

BMC Medical Research Methodology 2004, 4:10

Accepted: 29 April 2004

This article is available from: <http://www.biomedcentral.com/1471-2288/4/10>

© 2004 Prieto and Sacristán; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The use of preference-based measures in the evaluation of health outcomes has extended considerably over the last decade. Their alleged advantage over other types of general instruments in the evaluation of health related quality of life (HRQOL), supposedly lies in the fact that preference measures incorporate values or utilities that reflects the value of social preferences through health states. The objective of this study was to determine whether the use of social preference weights or utilities makes any real difference when calculating scores for the Euroqol (EQ5-D) questionnaire, a HRQOL preference-based measure.

Methods: Responses to the EQ5-D of a sample of 10,972 patients from 10 countries enrolled in an observational study of the treatment of schizophrenia in Europe were used for this purpose. Two different methods of scoring the EQ-5D were compared: 'weighting the items' of the questionnaire through the UK official weight coefficients, and 'non-weighting the items'. Pearson's, Spearman's, and two-way mixed parametric intraclass correlation coefficients were used to estimate the association of the scores obtained in both ways.

Results: The association between weighted and unweighted Euroqol scores was extremely high (Pearson's $r = 0.91$), as was the association between their ranks (Spearman's $\rho = 0.93$). The intraclass correlation coefficient obtained (0.89) also suggested that the concordance between the score distributions was prominent.

Conclusions: A non-weighted approach to score the EQ5-D is enough to explain a high proportion of variance in scores obtained through the use of utilities. The differential contribution of weights based on population preference values is therefore minimal and, in our opinion, negligible.

Background

The use of *preference-based measures* in the evaluation of health outcomes has extended considerably over the last decade [1-4]. Their advantage over other types of general instruments in the evaluation of health related quality of life (HRQOL), such as the SF-36 [5], the Sickness Impact Profile [6], and the Nottingham Health Profile [7], sup-

posedly lies in the fact that preference measures result in a single numerical index that reflects the value of social preferences through health states [8]. These numerical indices are finally used to calculate the Quality Adjusted Life Years (QALYs) [9] required to effect cost-utility studies [10] (cost-utility analysis is a form of economic evaluation that focuses particular attention on the quality of the

health outcomes produced by health programmes or treatments) [11].

Broadly speaking, the preference-based approach assumes that the social value or 'utility' of a health state is the same as the value of the quality of life of those individuals who are in it [12]. The value or utility of a health state is expressed on a scale from 0 to 1, where 0 is the utility of the state 'dead' and 1 is the utility of the state 'perfect health'. The lower the quality of life associated with a health state, the lower is its utility score on this scale [13].

Preference-based measures (or 'utility measures', as they are normally called) [2] may take the form of multi-attribute questionnaires [13]. Currently there are three main multi-attribute questionnaires available: the Quality of Well-Being (QWB) [14], the Health Utilities Index (HUI) [15], and the EuroQol (EQ-5D) [16]. These questionnaires can be considered as simple classification systems based on the degree of limitation that individuals indicate for different health dimensions, such as mobility, pain, emotional aspects and social functioning. The answers provided by the patients to each of these dimensions allows the analyst to transform the scores – often referred to as "health profiles" – into a single utility number [13]. The transformation algorithms are based on theoretical fundamentals [17] and on previous research in which one or more valuation techniques (i.e. Standard Gamble (SG), Time Trade-Off (TTO), and Rating Scale (RS)) [13] were used to measure directly the preferences of individuals (usually from community surveys) for different "health profiles".

Transformation algorithms basically consist of weighting each of the answers provided to the items or classification system domains, by means of a coefficient determined by the social preferences obtained empirically by means of a population sample. The final utility score is obtained from a more or less complex combination of the resulting weighted values. The description of the health state of a subject may be summarised, in theory at least, as a health index that reflects the social preferences of the population.

The alleged characteristics of the utility approach make it quite attractive as a measure of HRQOL independent of its usefulness in economic evaluation studies [8]. Utility measures also provide a mechanism for making broad comparisons across an array of clinical settings as well as in the context of assessing population health quality [8]. Many studies have already used utility measures to quantify the impact of interventions on HRQOL or to characterise the severity profile of any health problem [1]. Even the SF-36, which may be the paradigm of measures developed under a non-preference based approach, has been tested for use as a utility measure [18].

Despite the relative merits of the theoretical arguments made by proponents of the utility approach, in our opinion the empirical issues concerning the assessment of HRQOL through preference based measures remains unresolved. The assumed advantage of weighting each item individually is that valuations for all possible combinations of single health descriptors, and thus all possible health states, are elicited. Nevertheless, the crucial question in seeking social preference weights for items is how much difference it makes to use these differential weights to calculate the composite utility score for a given health state. As some authors have theorised, [19] it would make a difference if the weighted and unweighted scores on the multi-attribute questionnaire did not correlate highly. On the evidence so far, the use of differential weights seldom makes an important difference [20,21].

The aim of this study is to determine whether the use of social preference weights or utilities makes any real difference when calculating scores in the *preference-based measures* in the evaluation of HRQOL.

Methods

In order to carry out this study, we chose one of the more popular preference-weighted health state classification questionnaires in Europe, the Euroqol [16]. The Euroqol group has devoted considerable attention and effort to the weighting of its items and has investigated a broad range of modelling approaches for this purpose [22,23].

In order to test the usefulness of utilities in scoring the EuroQol Descriptive system (EQ-5D), we developed a parallel unweighted scoring rule solely based on patients' answers to EQ-5D; in this way it is possible to determine the value of any health state defined by the questionnaire without having to use any type of social preference for the items. We then compared the EQ-5D scores obtained using the unweighted scoring rule with 'official' weighted scores. These weights were determined on a random sample of the non-institutionalised adult population of the United Kingdom [22].

The EQ-5D descriptive system

The instrument (see Table 1) contains a description of the health state in 5 dimensions or items: Mobility, Self-care, Usual activities, Pain/discomfort, and Anxiety/Depression. The items are always presented in the same order and there are three levels of severity for each item: 1 (No problems), 2 (Some problems) and 3 (Unable to do/Extreme problems). For each item, the respondent must indicate the level of severity that best describes his/her personal health state at the time of giving the answers. The subject's global health state is finally defined as the combination of the level of problems described for each of the five dimensions contained in the EQ-5D. The assemblage

Table 1: EuroQol Descriptive system

Mobility
 1. No problems walking
 2. Some problems walking about
 3. Confined to bed

Self-care
 1. No problems with self-care
 2. Some problems washing or dressing self
 3. Unable to wash or dress self

Usual activities
 1. No problems with performing usual activities (e.g. work, study, housework, family or leisure activities)
 2. Some problems with performing usual activities
 3. Unable to perform usual activities

Pain/discomfort
 1. No pain or discomfort
 2. Moderate pain or discomfort
 3. Extreme pain or discomfort

Anxiety/depression
 1. Not anxious or depressed
 2. Moderately anxious or depressed
 3. Extremely anxious or depressed

Each composite health state has a five digit code number relating to the response provided to each dimension. The composite state 22322, for example, corresponds to a subject who has *some problem walking about (2)*, *some problems washing or dressing self (2)*, *is unable to perform usual activities (3)*, *has moderate pain or discomfort (2)*, and *is moderately anxious or depressed (2)*.

Table 2: EuroQol Scoring Formula based on UK Coefficients (Weights)

DIMENSION	COEFFICIENT (Weight)
Constant	0.081
<i>Mobility</i>	
Level 1	0
Level 2	0.069
Level 3	0.314
<i>Self-care</i>	
Level 1	0
Level 2	0.104
Level 3	0.214
<i>Usual activities</i>	
Level 1	0
Level 2	0.036
Level 3	0.094
<i>Pain/discomfort</i>	
Level 1	0
Level 2	0.123
Level 3	0.386
<i>Anxiety/depression</i>	
Level 1	0
Level 2	0.071
Level 3	0.236
N3	0.269

Scoring Formula EuroQol weighted scores are calculated by subtracting the relevant weight coefficients from 1 (Perfect health). The constant term is used if there is any item with a response greater than level 1. The N3 term is used if any item is at level 3. For example, the algorithm for computing the score for the health state **21223** is: $1 - (0.081 + 0.069 + 0 + 0.036 + 0.123 + 0.236 + 0.269) = 0.186$

Table 3: EuroQol Scoring Formula based on Unweighted Coefficients

DIMENSION	COEFFICIENT (Unweighted)
<i>Mobility</i>	
Level 1	0
Level 2	1
Level 3	2
<i>Self-care</i>	
Level 1	0
Level 2	1
Level 3	2
<i>Usual activities</i>	
Level 1	0
Level 2	1
Level 3	2
<i>Pain/discomfort</i>	
Level 1	0
Level 2	1
Level 3	2
<i>Anxiety/depression</i>	
Level 1	0
Level 2	1
Level 3	2

Scoring Formula The proposed unweighted scores are calculated by subtracting the sum of relevant unweighted coefficients, divided by 10, from 1 (Perfect health). The resulting values are linearly transformed to the same scale (min = -0.59, max = 1) of the weighted scores by multiplying them for 1.59 (the range) and adding -0.59 (the lowest possible value). For example, the algorithm for computing the score for the health state **21223** is: $[(1 - ((1 + 0 + 1 + 1 + 2)/10)) \cdot 1.59 - 0.59 = 0.205$

of the 3 values for the 5 dimensions results in a five digit number that classifies the subject in one of the 243 possible combinations ($3^5 = 243$). For example, health state 21223 corresponds to an individual who has some problem walking about, no problems with self-care, some problems with performing usual activities, has moderate pain or discomfort, and is extremely anxious or depressed. It should be noted that the numerals have no arithmetic properties and should not be used as a cardinal score.

Scoring weights for the EQ-5D descriptive system

Health states defined by the EQ-5D may be eventually converted to a single summary or composite index by applying scores from a standard set of values (or preferences) derived from general population samples [22]. Over the past few years, the EuroQol Group has been engaged in several research projects exploring this issue. Values have been elicited for different subsets of EQ-5D health states from respondents in Canada, Denmark, Finland, Germany, Japan, Netherlands, New Zealand, Slovenia, Spain, Sweden, UK, US and Zimbabwe.

The UK weights and scoring function for the EQ-5D are depicted in detail in Table 2. Scores are calculated by subtracting the relevant weight coefficients from 1 (Perfect health). The constant term is used if there is any dysfunction at all (any item with a response greater than 1). The

N3 term is used if any item is at level 3. The weight for each item is selected based on the level of response provided by the individual. The algorithm for computing the score (or 'tariff', as referred to by some) is straightforward. Table 2 shows the step by step procedure to obtain the EQ5-D index score for the example described above, where the respondent had the health state 21223. As it can be seen, the resulting score for this particular health state is: 0.186.

Unweighted scoring rule

In order to eliminate the possible influence that social preferences may have on the EQ-5D index score, we developed an unweighted scoring rule based solely on answers provided by subjects to the Descriptive system. Our proposal consisted of assigning values 0, 1 and 2 to answer options 1, 2 and 3, respectively (Table 3). Values 0, 1 and 2 represent the simplest option for scoring the EQ-5D as if it were a non-preference based measure. The constant and N3 terms were not considered for the calculation. As in the case of the weighted scores, scores were calculated by subtracting the sum of relevant unweighted coefficients from 1. In this case, the sum of the coefficients must first be divided by 10, in order to be able to express results on a scale in which the value of 1 represents the level of perfect health. The resulting values are linearly transformed to the same scale (min = -0.59, max = 1) as the weighted

Table 4: Descriptives and Measures of Association of Weighted and Unweighted EuroQol Scores.

	Weighted Scores		Unweighted Scores
Valid cases		9,991	
Mean (95% CI)	0.56 (0.56–0.57)		0.50 (0.50–0.51)
SD	0.32		0.31
Minimum	-0.59		-0.59
Maximum	1		1
Spearman's ρ		0.91*	
Pearson's r		0.93*	
ICC (95% CI)		0.89* (0.85–0.92)	

P < 0.0001

scores by multiplying them by 1.59 (the range) and adding -0.59 (the lowest possible value). In short, in order to obtain the final index this unweighted scoring strategy simply combines the answers provided by the subjects to each of item of the questionnaire. Table 3 offers an example in calculating the unweighted score for a health state equivalent to 21223 described above. The resulting value (0.205) shows a difference of 0.019 units in comparison with the value obtained by means of weighted scoring of items (0.186, Table 2). This difference accounts for just over 1% of the whole possible range for the final score (1.59).

The study sample

Considering that the objective of the study was to compare the weighted and unweighted scores of the EQ-5D, the analysis described below could be performed on any sample of subjects answering the EQ-5D (results would be invariant). In our case, and fundamentally for availability reasons, we used answers to the EQ-5D of patients included in an ongoing 3-year, prospective, observational study of the treatment of schizophrenia in Europe. The primary objective of the study is to assess the costs and outcomes of antipsychotic treatment of schizophrenia using antipsychotics. This study is being conducted in 10 European countries (Denmark, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain and the UK) A total of 10972 patients were enrolled. Baseline data collection was conducted via a core data collection form that included an self-administered version of the EQ-5D Descriptive system. Details of the design of the study have been presented elsewhere [24].

Comparison of preference-weighted and unweighted EQ-5D scores

To assess the relevance of the social preference weights (utilities) when analysing EQ-5D scores, the two different methods of scoring the questionnaire, weighted and unweighted, were compared.

Basic descriptive statistics of both score distributions were provided. The comparison of the two scoring alternatives was also performed through a paired design involving weighted and unweighted scores: Spearman's (ρ), Pearson's (r) and two-way mixed parametric intraclass (ICC) correlation coefficients were used to estimate the association of the scores obtained in both ways. A graphical comparison approach (scatterplot) was additionally used to illustrate the degree of association between the scores obtained by the two methods. Analyses were done with SPSS® for Windows®, v. 10.1.3.

Results

In the study, valid answers were obtained for all the items in the questionnaire in a total of 9,991 patients.

Table 4 details the descriptive statistics associated with the weighted and unweighted scores in the EQ-5D, and the result of association tests between the two. The great similarity in average scores obtained in both methods is of particular note (0.50 and 0.56).

The correlation coefficient estimates were excellent (Table 4). The association between weighted and unweighted EuroQol scores was extremely high (Pearson's $r = 0.91$), as was the association between their ranks (Spearman's $\rho = 0.93$). The intraclass correlation coefficient obtained (0.89) also suggests that, apart from a high association, the concordance between the weighted and unweighted score distributions was prominent.

The scatterplot in Figure 1 compares graphically the two scoring strategies. The extent to which the points of the scatterplot fall along a line reveals the degree of association of the scoring options. Scatterplots are often a good way of displaying data. Often, however, two or more observations will have the same values on the variables being graphed. When this happens, the points are graphed on top of each other, and it cannot be told from the scatterplot how many data points each symbol on the

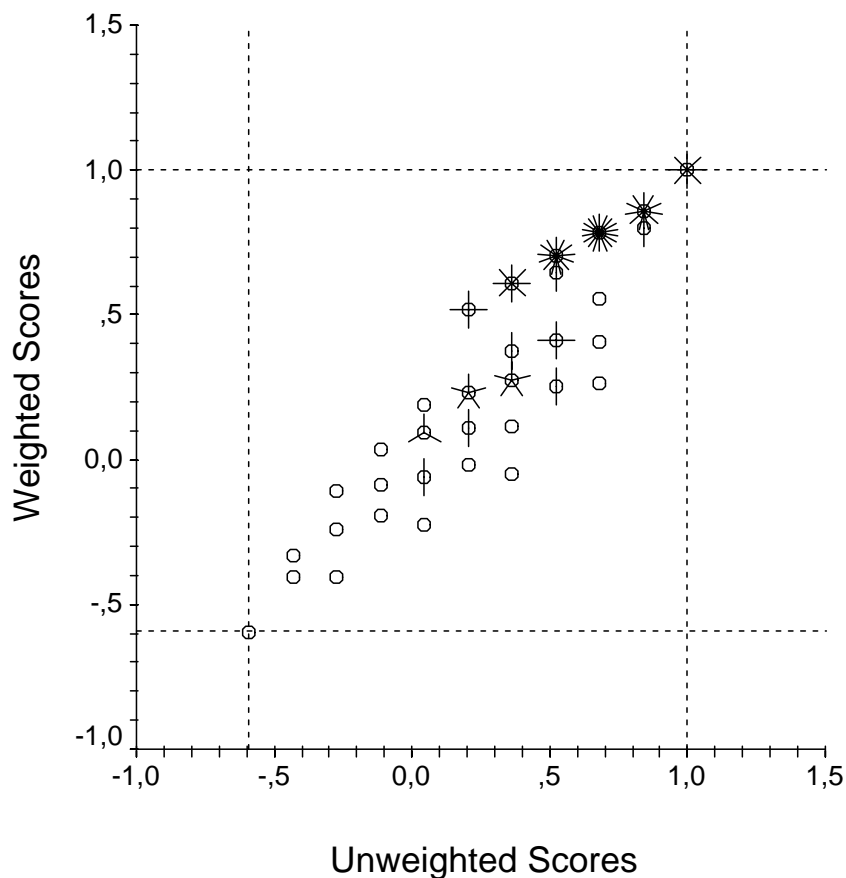


Figure 1

Scatterplot illustrating the degree of association between the weighted and the unweighted EuroQol scores. A small line, called a petal, is added to each point on the scatterplot to indicate how many observations each point represents; each petal symbolizes 100 cases. Dotted lines mark the maximum and minimum limits of the scales, that is to say, the range of possible scores.

graph represents. In order to solve this problem, a small line, called a petal, is added to each point on the scatterplot to indicate how many observations each point represents. In Figure 1 each petal symbolises 100 cases. As the figure shows, all the points lie quite well along a line ($R^2 = 0.83$), placing the weighted and unweighted scores close enough to conclude that they are more than comparable.

Discussion

The results of this study reveal that a simple combination of arbitrary values assigned to the items of the EQ5-D Descriptive system is enough to explain a high proportion of variance in scores obtained through the use of utilities. The differential contribution of weights based on population preference values is therefore minimal, and in our opinion, negligible.

The supposed advantages obtained from the use of the utility approach to measure HRQOL no longer stand if it is possible to generalise these results. The EQ-5D, and therefore, all preference based multi-attribute questionnaires supported by analogous scoring rules, would provide information that is conceptually comparable with information from any non-preference based HRQOL measure (such as the SF-36 or the Nottingham Health Profile (NHP)). As in the case of the non-preference based measures, scores obtained from evaluation of HRQOL through preference-based measures are fundamentally a direct reflection of the answers provided by the individuals to the items in the questionnaire. The results presented here demonstrate, yet again, that weighting answers to the items in the instruments does not imply a significant difference in the final score.

Although this fact seriously questions the conceptual fundament of evaluating HRQOL through preference measures, it does not necessarily jeopardise the validity or reliability of results obtained through the use of such instruments. In any event, it is still necessary to subject results to the scrutiny of their basic psychometric properties through standard methods [19]

To the contrary of the argument put forward by Brazier and Deverill in 1999 [25], our view is that the psychometric (read 'non-preference based') and economic approaches (read 'preference based') are not different in relation to conventional measurement criteria because they seek to measure the same concepts. Both in preference based measures and in non-preference based measures alike, primary interest lies in locating the responding individuals at different points on a theoretical linear continuum representing possible levels of HRQOL. For this purpose, total scores are computed by assuming that point values assigned to each possible response to the items form a numerical scale with the properties of order and equal units. Item scoring weights might be assigned by an arbitrary decision of the scale developer, but, as we have already mentioned, this action seldom makes an important difference. Thus, the sole purpose of preference and non-preference based measures is to "scale" the subjects based on their responses (weighted or not) to the items. Torgerson called this scaling approach "Subject-centred approach", where the systematic variation in the reactions of the subjects to the items is attributed to individual differences in the subjects [26]. The items, also called "stimuli" in psychometric jargon, are considered as replications: adding or deleting stimuli from the same stimulus-population at random would have no effect on procedure or results other than those due to the usual sampling fluctuations [26].

What is indeed true is that the weights used in preference based measures are obtained through a different scaling approach. Torgerson called this "Stimulus-centred" or "Judgement approach" [26]. In the "Stimulus-centred" approach the immediate purpose of the assessment is to scale the stimuli, which alone are assigned scale values. Valuation techniques like SG, TTO and RS form part of this *modus operandi*. The systematic variation in the reactions of the subjects to the items or stimuli is attributed to differences in the stimuli with respect to a designated attribute. Adding subjects chosen at random from the same population, or deleting subjects at random, would have no effect on either the procedure or the results other than the usual sampling fluctuations [26].

Although it is obvious that the weights used in preference measures are taken from a Stimulus-centred scaling

approach, their action mechanism and the results they provide are clearly defined as "Subject-centred".

In the light of the results presented in this study, we believe that it is time to review the conceptual fundamentals related to the evaluation of HRQOL through preference measures. Do these measures really differ from traditional non-preference based measures? Does it make any sense to go on using them? As Feeny argued in another context [7], one reason for going on using preference based measures is that these instruments provide a single summary score of outcomes that facilitate their interpretation and integration of the same in formulae to calculate the cost-effectiveness ratio in economic evaluations of health interventions. Although this argument may be attractive at first sight, it should not be accepted without further thought. The single summary score produced by preference based measures is the result of a simple combination of the different dimensions that are contained in the instrument. In the case of the EQ-5D, this implies the integration of dimensions such as anxiety/depression and physical mobility, which may not initially appear to be closely related, but may be combined for the purpose of a common, second degree dimensionality that could be called 'General Health'. But, if we permit the integration of disparate dimensions in the EQ-5D, for example, then why should we not permit it in other non-preference based profile measures such as the SF-36? In fact, the authors of the SF-36 have already empirically explored this possibility [27].

Another reason for going on using preference based measures, also put forward by Feeny [7], is the integration of the concept of mortality and morbidity in the scores for these scales (in conventional utility scales the state of being dead is assigned a score of 0 and perfect health is assigned a score of 1). On this matter our view is more radical: the scale produced by HRQOL preference based measures is an interval scale, not a ratio scale; this means that the numerical values assigned by the scale are totally arbitrary, and 0 does not imply an 'absolute lack' of HRQOL. The problem with the argument that HRQOL does have a natural zero as death, is that there can be states worse than death [28], and these states require a score as well. In fact, to respond to this need, the score algorithm of the EQ-5D assigns the value -0.59 to the worst state of health possible when it uses the UK weights. The integration of the concept of mortality is therefore a fallacy, which is even more untenable considering the marginal contribution of social preferences on the scores of preference based measures like the EQ-5D.

Regardless of which theoretical arguments or personal preferences are used for a given type of measure (preference based versus non-preference based), the final out-

come regarding the utility of preference based measures must be a result of the evaluation of the quality of information provided by the scores. The determination of the reliability and validity of scores from this type of scales using standardised methods is therefore absolutely essential.

The results of this study also questions the multiple efforts dedicated to obtaining specific national weighting for instruments such as EQ-5D. Given the uselessness of utilities in scoring the EQ-5D, the only point in effecting this type of activity is to obtain rankings or 'league tables' that permit trans-cultural comparison of different health states. In any event, we doubt that the high cost of such effort is really worth it.

It is unquestionable that the concept of utility applied to HRQOL evaluation has played a crucial role in the development of a new discipline linked to the standardised evaluation of the impact of health on the subjective perception of individuals. The evaluation of HRQOL through preference measures has permitted the concept of Quality-Adjusted Life Year (QALY) to be extended as a measure of the value of health outcomes. QALY, in turn, has allowed the numerical representation of the value of health through a single index combining individuals' quantity and quality of life. This hallmark has been outstanding in the definition of certain type of economic analyses (i.e. Cost-Utility).

Conclusions

However, the findings in this study imply a new starting point. Facts show that social preferences do not substantially modify scores on scales that are simply calculated from the combination of the answers provided to their items. The supposed advantages of the preference based measures in comparison with other less sophisticated measures in health states are not so, and it is yet be determined what their differential use is, and whether it really exists.

The debate on the convenience or otherwise of using social preferences in the evaluation of health states is far from being solved. In theory, in government-financed health systems, social decisions are responsible for allocation of resources. However, the supposed objectivity of social preference measures should not neglect the fact that many conceptual, ethical and methodological problems have yet to be solved, and the majority of instruments used have not been designed for planning or allocation of resources.

The patient is becoming the core of the health system. In medicine there is now a concern for the measurement of variables that interest the patient, and it is increasingly

important to have a good knowledge of the characteristics of the same, in order to be able to individualise interventions. Probably, optimised allocation of resources should include the identification of all the patient's peculiarities, including his/her own perception of health. In this context, and with the limitations described earlier, we should be asking ourselves what is the true value of society deciding on the health states of individuals (this is not based on conventional clinical variables, for example). Furthermore, if, as this study shows, social preferences do not make any real difference to the scores provided by the individuals themselves, then maybe this is the right moment to leave the "stimulus-centred approach" to one side, and to focus on the "subject-centred approach".

Competing interests

Luis Prieto and José A. Sacristan are Eli Lilly & Co. employees. Nevertheless, This paper reflects the point of view of the authors and do not necessarily reflects the opinion of Eli Lilly & Co.

Authors' contributions

LP and JAS actively participated in the discussions that lead to the present manuscript. LP run the statistical analysis. Both authors participated in writing the manuscript and read and approved the final one.

References

1. Neumann PJ, Goldie SJ, Weinstein MC: **Preference-based measures in economic evaluation in health care.** *Annu Rev Public Health* 2000, **21**:587-611.
2. Garrat A, Schmidt L, Mackintosh A, Fitzpatrick R: **Quality of life measurement: bibliographic study of patient assessed health outcomes.** *BMJ* 2002, **324**:1417.
3. Sanders C, Egger M, Donovan J, Tallon D, Frankel S: **Reporting on quality of life in randomized controlled trials: bibliographic study.** *BMJ* 1998, **317**:1191-1194.
4. Kind P, Dolan P, Gudex C, Williams A: **Variations in population health status: results from a United Kingdom national questionnaire survey.** *BMJ* 1998, **316**:736-41.
5. Ware JE Jr, Gandek B: **Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project.** *J Clin Epidemiol* 1998, **51**:903-12.
6. Bergner M, Bobbitt RA, Carter WB, Gilson BS: **The Sickness Impact Profile: development and final revision of a health status measure.** *Med Care* 1981, **19**:787-805.
7. Hunt SM, McKenna SP, McEwen J, Williams J, Papp E: **The Nottingham Health Profile: subjective health status and medical consultations.** *Soc Sci Med* 1981, **15**:221-229.
8. Feeny D: **A utility approach to the assessment of health-related quality of life.** *Med Care* 2000, **38**(9 Suppl):II151-II154.
9. Rosser R: **From health indicators to quality adjusted life years: technical and ethical issues.** In *Measuring the outcomes of medical care* Edited by: Hopkins A, Costain D. London: Royal College of Physicians of London; 1990:1-17.
10. Greenberg D, Pliskin JS: **Preference-based outcome measures in economic evaluations. A 20-year overview.** *Int J Technol Assess Health Care* 2002, **18**:461-6.
11. Torgerson D, Raftery J: **Economics notes: measuring outcomes in economic evaluations.** *BMJ* 1999, **318**:1413.
12. Nord E: **Cost-value analysis in health care: making sense out of QALYs.** Cambridge: Cambridge University Press; 1999.
13. Drummond ME, O'Brien B, Stoddart GL, Torrance GW: **Methods for the economic evaluation of health care programs.** 2nd edition. Oxford: Oxford University Press; 1997.

14. Kaplan RM, Ganiats TG, Sieber WJ, Anderson JP: **The Quality of Well-Being Scale: critical similarities and differences with SF-36.** *Int J Qual Health Care* 1998, **10**:509-20.
15. Torrance GW, Furlong W, Feeny D, Boyle M: **Multi-attribute preference functions. Health Utilities Index.** *Pharmacoeconomics* 1995, **7**:503-20.
16. Rabin R, de Charro F: **EQ-5D: a measure of health status from the EuroQol Group.** *Ann Med* 2001, **33**:337-43.
17. Keeny R, Raiffa H: **Decisions with multiple objectives: preferences and value tradeoffs.** 2nd edition. Cambridge: Cambridge University Press; 1993.
18. Brazier J, Roberts J, Deverill M: **The estimation of a preference-based measure of health from the SF-36.** *J Health Econ* 2002, **21**:271-92.
19. Nunnally JC, Bernstein IH: **Psychometric Theory.** New York: McGraw-Hill; 1994.
20. Jenkinson C: **Why are we weighting? A critical examination of the use of item weights in a health status measure.** *Soc Sci Med* 1991, **32**:1413-6.
21. Prieto L, Alonso J, Viladrich MC, Anto JM: **Scaling the Spanish version of the Nottingham Health Profile: evidence of limited value of item weights.** *J Clin Epidemiol* 1996, **49**:31-8.
22. Dolan P: **Modeling valuations for EuroQol health states.** *Med Care* 1997, **35**:1095-108.
23. Dolan P, Roberts J: **Modelling valuations for Eq-5d health states: an alternative model using differences in valuations.** *Med Care* 2002, **40**:442-6.
24. Prieto L, Novick D, Sacristan JA, Edgell ET, Alonso J: **A Rasch model analysis to test the cross-cultural validity of the EuroQoL-5D in the Schizophrenia Outpatient Health Outcomes Study.** *Acta Psychiatr Scand Suppl* 2003, **416**:24-9.
25. Brazier J, Deverill M: **A checklist for judging preference-based measures of health related quality of life: learning from psychometrics.** *Health Econ* 1999, **8**:41-51.
26. Torgerson WS: **Theory and Methods of Scaling.** New York: John Wiley and Sons; 1958.
27. Keller SD, Ware JE Jr, Bentler PM, Aaronson NK, Alonso J, Apolone G, Bjorner JB, Brazier J, Bullinger M, Kaasa S, Lepelge A, Sullivan M, Gandek B: **Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project. International Quality of Life Assessment.** *J Clin Epidemiol* 1998, **51**:179-88.
28. Macran S, Kind P: **"Death" and the valuation of health-related quality of life.** *Med Care* 2001, **39**:217-27.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/4/10/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

