Exploring health preferences in sociodemographic and health related groups through the paired comparison of the items of the Nottingham Health Profile

Luis Prieto, Jordi Alonso

Abstract

Background—Preference weighted measures of health related quality of life are necessary for cost effectiveness calculations involving quality of life adjustment. There are conflicting data about the influence of factors such as sociodemographic and health related variables on health preferences.

Study objective—The relative values attached to the items of the Spanish version of the Nottingham Health Profile (NHP) were assessed to make comparisons across social and health subgroups.

Design and participants—Preference values were obtained in sets of 250 to 253 persons (total n=1258) using the method of paired comparisons after all possible pairs of NHP items had been presented to respondents for judgement of severity. χ^2 Tests and Spearman's correlations among item ranks were calculated.

Main results—Findings show that preferences elicited with the method of paired comparisons are consistent and independent of the sample from which they are obtained (mean correlation coefficients across subgroups range from 0.87 to 0.96). **Conclusion**—The evaluation of health did not seem to be related to sociodemographic variables (gender, age, social class) or to the health status of the respondents, suggesting that health preferences are stable across different populations.

(J Epidemiol Community Health 2000;54:537-543)

Health Services
Research Unit, Institut
Municipal
d'Investigació Mèdica
(IMIM), Barcelona,
Spain
L Prieto
J Alonso

Facultat de Psicologia i
Ciències de l'Educació
Blanquerna,
Universitat Ramon
Llull, Barcelona, Spain
L Prieto

Correspondence to:
Dr Prieto, Institut Municipal
d'Investigació Mèdica
(IMIM), C/ Dr Aiguader, 80,
E-08003, Barcelona, Spain
(lprieto@imim.es)

Accepted for publication
19 December 1999

The relative values that people attach to different states of health and illness are important for both resource allocation and clinical decision making.¹⁻⁴ Preference information about health is usually obtained by asking respondents to assign values or ratings to imagined specific health states. Several standard scaling methodologies for determining these preferences have been proposed (for example, standard gamble, time trade off, rating scale, magnitude estimation),⁵⁻⁸ but agreement has not been reached regarding their validity and reliability.^{2 9-12} Given the different nature of the scaling methods (for example, the manner in which the judgement stimulus is presented and the mode of measuring the raters' response) the agreement of results is not always expected.¹³⁻¹⁵

There is controversy about whether, in addition to the scaling method, there are other aspects that may affect preferences.¹⁶⁻²⁰ Some authors have suggested that sociodemographic characteristics of people as well as the experience with illness may influence the judgement task.²¹ Early work by Sackett and Torrance²² showed that age and health status of the respondent were associated with the utilities attached to health scenarios. Similarly, Worsley²³ reported that gender, social class and education influence lay valuations of a variety of aspects of health. Bucquet *et al*²⁴ also found differences in preference judgements according to sociodemographic groups (for example, gender and whether "Patient" or "Non-patient") when comparing the item weights of a generic measure of subjective health status, the Nottingham Health Profile.²⁵⁻²⁷ Dolan showed that current health status has an effect on the valuations attached to different health states, with those in poorer health generally giving higher valuations.²⁸ On the other hand, a study by Llewellyn-Thomas *et al*²⁹ found no differences in the valuation of a health status before and after the patients entered that particular status. Other studies report no influence of sociodemographic or health status variables on health state valuations.³⁰⁻³²

As has been pointed out,²⁸ many existing studies might be inconsistent because their results are largely based on small numbers of observations and on non-valid and/or reliable measures. It is also possible that discordant results are related to the nature of the responses required by the majority of scaling methods³³ used in these studies.

To assess whether the sociodemographic characteristics or current health status of people have any influence on the way health preferences are expressed, this study compared the relative values attached to each of the items of the Spanish version of the Nottingham Health Profile (NHP),^{34 35} with a fairly large number of observations, and considering a straightforward scaling method based on comparative judgements: the method of paired comparisons.^{8 33 36} The paired comparison method is one of the simplest methods for preference assessment and has been used in a variety of applications.^{33 37} Preferences were compared between age, gender, social class and different health status groups. Analyses were performed without an a priori hypothesis about the degree of relation between preferences across these groups.

Methods

THE NHP

The NHP was originally developed in the United Kingdom and is now widely used in Europe.^{24–27 38 39} It contains 38 items of Yes/No response describing several health problems on six different dimensions (Energy, Pain, Emotional Reactions, Sleep, Social Isolation and Physical Mobility). Weights for the original NHP English version were obtained using the method of paired comparisons proposed by Thurstone.^{33 36} The Spanish weights obtained with the same procedure were similar to those of the original English version. It was concluded that similarity of item weights with the original English version supported the cross cultural validity of the adapted questionnaire.⁴⁰ Similar results were obtained when Swedish and French adaptations of the questionnaire were contrasted with the source version.^{24 38}

SUBJECTS

A convenience stratified quota sample of 1258 subjects was recruited from several settings in the city of Barcelona, Spain.⁴⁰ The quota method was used to ensure a sufficient representation of people within each category of some sociodemographic variables that have been shown to be relevant in recognising the variations in perception of health: gender, age, social class and health status. The study designed to obtain the French NHP weights also took advantage of this sampling technique in a comparable way.²⁴ The sample was sought in similar recruitment centres to those used in the production of the French, the British, and the Swedish NHP weights.^{25 36 38} Specifically, subjects were obtained from these places: three public elderly nursing homes (Gràcia, Sarrià, Eixample), a rehabilitation centre (Peracamps), and several inpatients and outpatients clinics from three hospitals (Mar, Valle Hebron, Barcelona) and two primary health care centres (Sant Andreu, Drassanes). Inpatients and outpatients showed acute and chronic conditions, minor and severe injuries or disabilities. At the outpatient clinics, interviews were conducted with both patients and non-patients who were accompanying friends, relations, etc. People were recruited according to the following strata (quota): gender (50% male), age (33% 20–39 years, 33% 40–64 years and 33% 65+ years), socioeconomic class (50% class I–III—professional, intermediate or skilled occupations—, and 50% class IV–V—partly skilled or unskilled occupations) and health status (“Patient”: 50%, or “Non-patient”: 50%). Social class was assessed according to a categorisation based on the Spanish National Classification of Occupations that follows the British Registrar General Classification.^{41 42} This classification has been recommended by the Spanish Society of Epidemiology.⁴³ The level of health of the participants was assessed by considering their NHP responses. For circumstances related with the design of a previous study,⁴⁰ NHP data were only available for the first 643 respondents (51% of the total). To assure comparability with previous studies^{24 38} participants were

classified as “Patients” or “Non-Patients”. A person was considered to be a “patient” if the subject was going to be visited by a physician because of sickness the same day or if the subject had been visited during the week before the interview (general check ups or visits seeking prescriptions were not considered as medical visits). People were also classified as “Patients” if they reported a medical visit in the past three months attributable to any of nine chronic conditions (arthrosis or rheumatism, bronchitis or asthma, diabetes, migraine, paralysis, chronic back pain, blood circulation problems, problems with nerves or depression) listed in a screening questionnaire. Within each of the six NHP dimensions, people were classified as “Scorers in the dimension X” if they had responded “Yes” at least to one item of such dimension when responding to the questionnaire, or as “Non-scorers in the dimension X” if they had responded “No” to all the items of this given dimension. People were also classified in two groups considering whether they suffered or not any of the nine chronic conditions described before.

Three interviewers were involved in the enrolment of subjects. Each interviewer received training designed to standardise the interview. The interviewer introduced herself and explained that research was being conducted into the problems people have in their daily lives when they are unwell. Interviewees were invited to participate and no attempt to pressure them was made. Refusals were replaced by new consecutive subjects.

JUDGEMENT TASK

According to the paired comparisons procedure, all possible combinations of two different items in a given dimension of the Spanish version of the NHP were presented for judgement of preference.³³ To decrease respondent burden, every person was asked to judge the items of one dimension only, except for the Energy and Sleep dimensions (containing three and five items, respectively), which were judged by the same people. For each dimension, pairs of items were shown in random order and reversed for 50% of each sub-sample. The position of each item on the sheet presented (top or bottom) was also randomised. For each pair of items subjects were asked the following question “Which of these items do you consider to be more severe or more difficult to live with?”. On average, 252 people judged the items of each dimension (range 250–253).

Frequencies of participants indicating that each item was “more severe” than the others were calculated and rank orderings of severity for each item within its dimension were finally obtained. Although Thurstone’s procedure allows to convert these frequencies, expressed in an ordinal scale, into a scale with higher measurement properties (interval),³³ it was not done as previous research brought into question these results when applied to the NHP data.^{40 44 45} Although the method of paired comparisons has great practical simplicity it does not always produce a set of appropriate data. One problem occurs when an observer or

some observers are particularly bad judges, or are poorly motivated to take the care required to produce consistent comparisons. A second problem occurs if the experimenter asks too much of the observers (the items might be so close together that distinguishing between them became purely a guessing game) or if the quality under examination is not unidimensional. When any one of these difficulties is present, the reported judgements may contain intransitivities or inconsistencies called circular triads. The assessment of the internal consistency of the task of paired comparisons was carried out by counting the total number of inconsistencies each judge produced in the course of making choices among all possible pairs of items.⁴⁰ An inconsistency or circular triad occurred whenever intransitive pairwise choices were formed. Given all the possible pair combinations of three items, A, B and C, a circular triad was counted when a subject judged that item A was more severe than B, B more severe than C, but C more severe than A. The coefficient of consistency (ζ) proposed by Kendall and Smith⁴⁰ was calculated to indicate the amount of intransitivity in the comparison judgements made by each person. The coefficient of consistency ζ can range between 0, indicating the maximum number of circular triads, and 1 indicating the absence of any circular triads. To discover the probability of obtaining a given value of ζ under the hypothesis that the subject's judgements were made at random, a significance χ^2 test for the coefficient of consistency ζ was also calculated for each person.

Variations in the number of inconsistencies committed by each judge were also explored by sociodemographics (gender, social class, age group and patients status).

ANALYSIS

Results on each paired comparison of items were contrasted by sociodemographic and health status groups through χ^2 tests. The number of "discordances" produced across comparisons was calculated next. A "discordance" was counted whenever a paired comparison showed a significantly different proportion of item preference.²⁴ Discordances were considered statistically significant at the 1% level ($p < 0.01$). At this level of significance, data were powerful enough to detect minimal changes: to detect a medium effect size of 0.30 at the $\alpha = 0.01$ significance level by a χ^2 contingency test with $df = 1$, it was found that for $n = 198$ the power ($1 - \beta$) was 0.95 and for $n = 267$ the power was 0.99. Threshold levels of significance were additionally adjusted for multiple comparisons in each set of k χ^2 values ($k = 3, 10, 28$ and 36) by Bonferroni correction. This measure considerably reduces the possibility of obtaining spurious differences in any analysis involving repeated testing of data.

The rank orderings of the NHP items across sociodemographic and health status groups were compared by means of Spearman's correlation coefficients.

Results

Table 1 shows sociodemographic characteristics of the 1258 study participants. Subjects appear homogeneously distributed along the strata defined by the quota sampling criteria. The exposed stratified recruitment of judges across sociodemographic variables was preserved within each of the NHP dimension groups. Thus, the sociodemographic composition of each NHP dimension group maintained the same proportion of participants for each variable category as shown in table 1. The p values (χ^2 tests) calculated for each cross tabulation between the "NHP dimension being judged" variable with each of the other sociodemographic variables involved in table 1 (gender, social class, age group and patients status) ranged between 0.976 and 0.999.

An example of how rank orderings were obtained is presented in table 2. Using the subgroup of people who judged the Sleep dimension ($n = 252$), a matrix was formed including the participants (observed frequencies) in the sample indicating that each item was judged "more severe" than the others. Item SL1 "I take tablets to help me to sleep" was considered more severe than item SL2 "I'm waking in the early hours of the morning" (SL1 > SL2) by 211 subjects out of 252. Conversely, item SL2 was considered more severe than item SL1 (SL2 > SL1) by 41 subjects. The sum of each column of frequencies was finally obtained and the corresponding rank orderings calculated. The rank order of each column provides a measure of severity for each item compared with all other items within each dimension.

Results showed that all the subjects in the sample did not make a significantly ($p > 0.05$) large number of circular triads, that is all of them showed a certain degree of consistency in

Table 1 Distribution of sociodemographic variables

	Non-patient		Patient		Total
	Class I-III	Class IV-V	Class I-III	Class IV-V	
Male					
20-39	53	50	51	55	209
40-64	51	52	50	50	203
≥65	53	49	49	51	202
Female					
20-39	54	56	53	55	218
40-64	52	49	57	60	218
≥65	49	50	54	55	208
Total	312	306	314	326	1258

Table 2 Rank orderings of the items of the Sleep dimension of the Spanish NHP

	Item judged as more severe				
	SL1	SL2	SL3	SL4	SL5
SL1	—	41	111	57	84
SL2	211	—	157	199	215
SL3	141	95	—	68	114
SL4	195	53	184	—	190
SL5	168	37	138	62	—
Sum	715	226	590	386	603
Rank order	1	5	3	4	2

SL1 = "I take tablets to help me to sleep"; SL2 = "I'm waking in the early hours of the morning"; SL3 = "I lie awake for most of the night"; SL4 = "It takes me a long time to get to sleep"; SL5 = "I sleep badly at night".

Table 3 Severity rank orderings of the items of the Spanish NHP

NHP items and dimensions	Rank ordering
<i>Energy</i>	
I'm tired all the time	1
I soon run out of energy	2
Everything is an effort	3
<i>Pain</i>	
I have unbearable pain	1
I'm in constant pain	2
I have pain at night	3
I'm in pain when I walk	4
I'm in pain when I'm standing	5
I'm in pain when I'm sitting	6
I'm in pain when going up or down stairs	7
I find it painful to change position	8
<i>Emotional reactions</i>	
I feel that life is not worth living	1
I feel as if I'm losing control	2
Things are getting me down	3
I'm feeling on edge	4
Worry is keeping me awake at night	5
I wake up feeling depressed	6
These days seem to drag	7
I lose my temper easily these days	8
I've forgotten what it's like to enjoy myself	9
<i>Sleep</i>	
I take tablets to help me sleep	1
I sleep badly at night	2
I lie awake for most of the night	3
It takes me a long time to get to sleep	4
I'm waking in the early hours of the morning	5
<i>Social isolation</i>	
I feel I am a burden to people	1
I feel lonely	2
I feel there is nobody I am close to	3
I'm finding it hard to make contact with people	4
I'm finding it hard to get on with people	5
<i>Physical mobility</i>	
I'm unable to walk at all	1
I need help to walk about outside	2
I can only walk about indoors	3
I find it hard to reach for things	4
I have trouble getting up and down stairs	5
I find it hard to dress myself	6
I find it hard to stand for long	7
I find it hard to bend	8

their judgements despite a lack of perfection (the coefficient of consistency ζ ranged from 0.15 to 1). No statistical significant differences in the number of inconsistencies committed by judges were found between the groups defined by the sociodemographic variables. The distribution of number of triads was also standardised and subjects with values greater than $z=1.64$ ($p<0.05$) were identified. The exclusion of the analysis of those subjects with significantly higher values of inconsistency (11%) did not show any difference with the results presented here.

Table 4 Proportion (%) of the item preference in the Sleep dimension of the Spanish NHP by sociodemographic and health status groups

Paired comparisons	Sociodemographic characteristics								Health status							
	Gender		Age				Social class		Patient		Sleep dimension		Chronic conditions			
	Male (n=125)	Female (n=127) p	20-39 (n=85)	40-64 (n=84)	65+ (n=83)	p	IV-V (n=127)	I-III (n=125) p	Yes (n=122)	No (n=130) p	Non-scoring (n=64)	Scoring (n=62) p	No (n=160)	Yes (n=62) p		
SL1 > SL2	83	84 *	88	87	76	*	80	87 *	84	84 *	77	89 *	84	84 *		
SL1 > SL3	55	57 *	56	60	52	*	56	56 *	53	59 *	53	53 *	56	55 *		
SL1 > SL4	76	79 *	85	77	70	*	74	81 *	74	81 *	66	89 †	76	79 *		
SL1 > SL5	64	69 *	74	73	53	*	65	68 †	60	73 *	58	69 *	67	66 *		
SL2 > SL3	41	35 *	39	43	31	*	44	31 *	38	38 *	39	39 *	39	36 *		
SL2 > SL4	23	19 *	22	13	28	*	23	19 *	22	20 *	23	19 *	22	20 *		
SL2 > SL5	17	13 *	11	13	20	*	17	13 *	18	11 *	20	11 *	16	12 *		
SL3 > SL4	73	73 *	87	69	63	*	69	77 †	72	74 *	64	89 †	74	72 *		
SL3 > SL5	60	50 *	64	50	51	*	52	58 *	52	57 *	53	58 *	56	53 *		
SL4 > SL5	29	20 *	20	25	29	*	28	21 *	20	30 *	22	24 *	21	30 *		

SL1 = "I take tablets to help me to sleep"; SL2 = "I'm waking in the early hours of the morning"; SL3 = "I lie awake for most of the night"; SL4 = "It takes me a long time to get to sleep"; SL5 = "I sleep badly at night". * $p>0.01$, † $p<0.01$, ‡ $p<0.001$.

KEY POINTS

- The relative values that people attach to different states of health and illness are important for both resource allocation and clinical decision making.
- The method of paired comparison allows an appropriate assessment of the differences of preferences across groups of people.
- The paired choices strategy was applied to the items of the six dimensions of the Nottingham Health Profile.
- Health preferences elicited with the method of paired comparisons did not seem to be related to sociodemographic variables or to the health status of the respondents.

Table 3 shows all the NHP items and their corresponding rank ordering per dimension for the Spanish sample. These ranks conform exactly to those obtained according to the Spanish NHP weights.⁴⁰

The percentage of item preferences in the Sleep dimension of the Spanish NHP by sociodemographic and health status groups are presented in table 4. Few significant differences ($p < 0.01$) were found by age, gender or social class level, but none lower than the 0.1% level of significance. No differences were found among "Patients" and "Non-patients" and Chronic conditions status, only two discordances were found when comparing items among "Non scorers" and "Scorers" in the Sleep dimension by health status. The frequency of item SL1 "I take tablets to help me to sleep" being judged more severe than item SL4 "It takes me a long time to get to sleep" was statistically lower ($p<0.01$) among "Non NHP scorers" (66%) than among "NHP Scorers" (89%). Similar results were found when comparing SL3 and SL4. By age groups, significant differences were found for pairs SL1-SL5 and SL3-SL4.

Analogous results were obtained with the rest of the NHP dimensions at the 1% level. Few (1 to 3) discordances were found by sociodemographic characteristics and health status (table 5). Emotional Reactions presented the highest number of discordances (3 of 36 possible comparisons). In contrast, of the

Table 5 Number of discordances* in the paired comparisons of the Spanish NHP items by sociodemographic and health status groups

NHP dimension	Judges	Paired comparisons	Number of discordances					
			Sociodemographic characteristics			Health status		
			Gender	Age	Social class	Patient status	NHP scoring	Chronic conditions
Energy	252	3	0	1	0	0	0	0
Pain	253	28	1	1	0	0	2	0
Emotional reactions	253	36	3	1	3	0	1	1
Sleep	252	10	0	2	0	0	2	0
Social isolation	250	10	0	0	0	0	0	0
Physical mobility	250	28	0	3	1	0	1	1

*A discordance was counted whenever a paired comparison showed a significantly different proportion of item preference ($p < 0.01$).

Table 6 Spearman's correlation coefficients between the rank orderings of the Spanish NHP items by sociodemographic and health status groups

	NHP dimensions							Row mean
	Energy	Pain	Emotional reactions	Sleep	Social isolation	Physical mobility		
Men/women	1	1	0.85	0.90	0.97	0.88		0.93
Age 20–39/40–64	0.50	0.98	0.94	0.90	0.97	0.92		0.87
40–64/65+	0.50	1	0.87	1	0.97	0.91		0.88
20–39/65+	1	0.98	0.87	0.90	0.90	0.92		0.93
Class IV–V/Class I–III	1	1	0.90	0.90	0.90	0.98		0.95
Patients/non-patients	1	1	0.98	0.90	0.90	1		0.96
Non-scoring/scoring in dimension	1	0.98	0.70	0.90	0.90	0.96		0.91
With/without chronic conditions	1	1	0.93	1	0.90	0.95		0.96
Column mean	0.88	0.99	0.88	0.93	0.93	0.94		0.92

10 comparisons among Social Isolation items, none showed discordances. Considering the sociodemographic variables and the health status of participants, age was associated with the most important discordance. Using the Bonferroni correction as needed, no discordances were identified in the paired comparisons.

Spearman's correlation coefficients between the rank orderings of the Spanish NHP items by sociodemographic and health status groups are shown in table 6. Mean correlation coefficients for the whole sample ranged from 0.88 for Energy to 0.99 for Pain. When considering sociodemographic and health status variables, mean correlation coefficients ranged from 0.87 for age groups 20–39/40–64 to 0.96 for "Patients" and "Chronic Conditions". The average of the mean correlation coefficients in table 6 was 0.92. All the reported correlation coefficients were statistically significant ($p < 0.05$).

Discussion

By means of a paired comparison task, the study evaluated the degree of variation of health preferences (described by the items of the Spanish version of the NHP) across different sociodemographic and health status groups of subjects. Results show that preferences for health did not appear to be related to sociodemographic variables such as gender, socioeconomic class, or age group. They were not associated either with the health status of the respondents.

Although some of the paired comparisons resulted in a significantly different proportion of item preference among socioeconomic and health status categories, the relative amount of these discordances over the total number of comparisons led us to conclude that preferences for health did not seem to be related to

sociodemographic or health status variables. Although age groups showed the "highest" amount of dissension, this disagreement only supposed eight significant discordances over 115 different paired comparisons (less than 7%). Bonferroni corrected probability values indicated that this level of discordance is insufficient to conclude that demographic characteristics are relevant.

The assignment of values (preferences) to different states of health and illness must be viewed as a classic problem of measurement involving the construction of a scale with a continuous unit of measurement, that is to say, a scale with interval or ratio properties.⁴⁶ A considerable number of procedures have been devised to determine interval scale values of a series of health states (for example, standard gamble, time trade off, rating scale, magnitude estimation).^{6 47–49} Given that these proposed techniques comprise quite different cognitive tasks, and given that they change along many dimensions (for example, level of abstraction, use of numbers versus qualitative judgements, time horizons, anchor points) it should not be surprising that they produce different results.^{21 47 50 51} As Tversky and Kahneman showed,¹⁶ changes of perspective often reverse the relative apparent size of objects and the relative desirability of options. Differences in valuations attributable to the personal characteristics of respondents might be insignificant when compared with the differences that might arise from the heterogeneity of the proposed scaling methods. The effects of demographic variables on health perceptions observed in other studies might be attributable to measurement conditions (for example, the way health states were described, scores generated, and surveys administrated).

The simple ordering of a health state series in an ordinal scale, although it reduces the amount of information compared with that provided by continuous scales, is much more straightforward than an interval scale. We do believe that the method of paired comparison can be viewed as a valid simplification of the judgement process of people, allowing an appropriate assessment of the differences of preferences across groups of people. The criterion for order we used is based on the proportion of times any health state is designated as possessing more of the attribute than any other health state, and there is considerable evidence that the rank order obtained is substantially invariant with respect to the different experimental methods that might be used (paired comparisons, ranking, single-stimulus rating, or sorting into successive intervals to obtain the data) for the same purpose.³³

Our study suggests that preferences, elicited with the paired comparisons method, are consistent and independent of the sample they are obtained from. The generalisation of these findings must be nevertheless undertaken with caution because of a number of limitations in the study design.

The paired choices strategy was applied only to the items within a dimension. The response task is cognitively simple and natural but the burden because of the need of comparing all possible pairs might be considerable. This is a limitation of the paired comparison procedure. This handicap can be partially overcome using paired comparison procedures designed to manage incomplete matrices of preferences.⁴⁶

Another limitation of the paired comparison procedure could be attributable to the particular instructions given to the participants in the judgement task: subjects were asked to indicate which of the items in a pair was "more severe or more difficult to live with". In a sense, the question is double barrelled (there are two questions being asked at the same time) and, strictly speaking, there is no guarantee either that all the individuals used the same part of the question nor that each part have been answered similarly by the respondents. In such a case, the items involved in the judgement task might have been ordered in a different way given the part of the question considered. Nevertheless, we are convinced that the two possible different meanings involved in the question ("more severe" or "more difficult to live with") are not sufficiently different to define a distinct attribute of measure when used independently.

It could be also argued that single items from the NHP do not describe health states. As test items become the operational definition of the construct being measured, and the intention of the authors of the original NHP was "to capture and record accurately some aspects of the feelings and perceptions of lay people with respect to their health status,"⁵² we do believe that, although the items may not be considered a full scenario of a health state, they should legitimately be considered as specific health states.

Given the characteristics of the recruiting methods, participants for this study may have resulted in groups with more similar preferences that might be true of the population as a whole. Results of this study should be replicated in more representative samples and designs.

More evidence concerning the validity of the method of paired comparisons to value health would be also necessary to confirm the invariant results observed; otherwise, the stability of the observations could be considered just to be a consequence of a design bias introduced by a scaling procedure insensitive to the real variation of these preferences. Unfortunately, there is no independent objective way (a gold standard) to establish what such preferences are supposed to be.²⁸ However, the convergent validity of the method could be assessed by examining the extent to which other methods produce the same results with the same individuals making judgements about the same health states.

The authors are grateful to Mr Dave McFarlane for his editorial assistance.

Funding: this research was supported by the Fondo de Investigación Sanitaria (FIS) (Expdte 96/0776). Additional support was received by the Generalitat de Catalunya (CIRIT/1995 SGR 00434) and the Facultat de Psicologia i Ciències de l'Educació Blanquerna, Universitat Ramon Llull.

Conflicts of interest: none.

- 1 Froberg DG, Kane RL. Methodology for measuring health-state preferences-IV. Progress and a research agenda. *J Clin Epidemiol* 1989;42:675-85.
- 2 Stiggelbout AM, Kiebert GM, Kievit J, et al. The "utility" of the Time Trade-Off method in cancer patients: feasibility and proportional trade-off. *J Clin Epidemiol* 1995;48:1207-14.
- 3 Kapp MB. Enforcing patient preferences. Linking payment for medical care to informed consent. *JAMA* 1989;261:1935-49.
- 4 Sutherland HJ. Assessing patients' preferences. *Med Decis Making* 1995;15:286-7.
- 5 Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econom* 1986;5:1-30.
- 6 Froberg DG, Kane RL. Methodology for measuring health-state preferences-I. Measurement strategies. *J Clin Epidemiol* 1989;42:345-54.
- 7 Froberg DG, Kane RL. Methodology for measuring health-state preferences-II. Scaling methods. *J Clin Epidemiol* 1989;42:459-71.
- 8 Patrick DL, Erickson P. *Health status and health policy*. New York, NY: Oxford University Press, 1993.
- 9 Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, et al. The measurement of patients' values in medicine. *Med Decis Making* 1982;2:449-62.
- 10 Kievert GM, Stiggelbout AM, Leer JH, et al. Test-retest reliabilities of two treatment-preference instruments in measuring utilities. *Med Decis Making* 1993;13:133-40.
- 11 Katz JN, Phillips CB, Fossel AH, et al. Stability and responsiveness of utility measures. *Med Care* 1994;32:183-8.
- 12 Danis M, Garrett J, Harris R, et al. Stability of choices about life-sustaining treatments. *Ann Intern Med* 1994;120:567-73.
- 13 Kaplan RM, Bush JW, Berry CC. Health status index: category rating versus magnitude estimation for measuring levels of well-being. *Med Care* 1979;17:501-25.
- 14 Leighton Read J, Quinn RJ, Berwick DM, et al. Preferences for health outcomes. Comparison of assessment methods. *Med Decis Making* 1984;4:315-29.
- 15 Revicki DA. Relationship between health utility and psychometric health status measures. *Med Care* 1992;30:MS274-82.
- 16 Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science* 1981;211:453-8.
- 17 Kahneman D, Tversky A. The psychology of preferences. *Sci Am* 1982;246:160-73.
- 18 Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, et al. Describing health states. Methodologic issues in obtaining values for health states. *Med Care* 1984;22:543-52.
- 19 Guyatt GH, Cook DJ. Health status, quality of life, and the individual. *JAMA* 1994;272:630-1.
- 20 Veit CT, Rose BJ, Ware JE. Effects of physical and mental health on health-state preferences. *Med Care* 1982;20:386-401.
- 21 Froberg DG, Kane RL. Methodology for measuring health-state preferences-III. Population and context effects. *J Clin Epidemiol* 1989;42:585-92.

- 22 Sackett DL, Torrance GW. The utility of different health states as perceived by general public. *J Chron Dis* 1978;31:697-704.
- 23 Worsley A. Lay person's evaluation of health: an exploratory study of an Australian population. *J Epidemiol Community Health* 1990;44:7-11.
- 24 Bucquet D, Condon S, Ritchie K. The French version of the Nottingham Health Profile. A comparison of items weights with those of the source version. *Soc Sci Med* 1990;30:829-35.
- 25 Hunt SM, McKenna SP, McEwen J, *et al.* A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health* 1980;34:281-6.
- 26 Hunt SM, McKenna SP, McEwen J, *et al.* The Nottingham Health Profile: subjective health status and medical consultations. *Soc Sci Med* 1981;15A:221-9.
- 27 Hunt SM, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. *J R Coll Gen Pract* 1985;35:185-8.
- 28 Dolan P. The effect of experience of illness on health state valuations. *J Clin Epidemiol* 1996;49:551-64.
- 29 Llewellyn-Thomas H, Sutherland HJ, Thiel EC. Do patients' evaluations of a future health state change when they actually enter that state? *Med Care* 1993;31:1002-12.
- 30 McQuellon RP, Muss HB, Hoffman SL, *et al.* Patient preferences for treatment of metastatic breast cancer: a study of women with early-stage breast cancer. *J Clin Oncol* 1995;13:858-68.
- 31 Dalen H, Williams A, Gudex C. Lay people's evaluations of health: are there variations between different subgroups?. *J Epidemiol Community Health* 1994;48:248-53.
- 32 Badia X, Fernández E, Segura A. Influence of socio-demographic and health status variables on evaluation of health states in a Spanish population. *Eur J Public Health* 1995;5:87-93.
- 33 Nunnally JC. *Psychometric theory*. New York, NY: McGraw-Hill, 1978.
- 34 Alonso J, Antó JM, Moreno C. Spanish version of the Nottingham Health Profile: translation and preliminary validity. *Am J Public Health* 1990;80:704-8.
- 35 Alonso J, Prieto L, Antó JM. The Spanish version of the Nottingham Health Profile: a review of adaptation and instrument characteristics. *Qual Life Res* 1994;3:385-93.
- 36 McKenna SP, Hunt SM, McEwen J. Weighting the seriousness of perceived health using Thurstone's method of paired comparisons. *Int J Epidemiol* 1981;10:93-7.
- 37 Hadorn DC, Hays RD, Uebbersax J, *et al.* Improving task comprehension in the measurement of health state preferences. A trial of informational cartoon figures and a paired comparison task. *J Clin Epidemiol* 1992;45:233-43.
- 38 Hunt SM, Wicklund I. Crosscultural variation in the weighting of health statements: a comparison of English and Swedish valuations. *Health Policy* 1987;8:227-35.
- 39 Hunt SM, Alonso J, Bucquet D, *et al.* Cross-cultural adaptation of health measures. *Health Policy* 1991;19:33-4.
- 40 Prieto L, Alonso J, Viladrich MC, *et al.* Scaling the Spanish version of the Nottingham Health Profile: evidence of limited value of item weights. *J Clin Epidemiol* 1996;49:31-8.
- 41 Domingo A, Marcos J. Propuesta de un indicador de clase social basado en la ocupación (Proposal of a social class indicator based on the occupation). *Gac Sanit* 1989;3:320-6.
- 42 Alonso J, Pérez P, Sáez M, *et al.* Validez de la ocupación como indicador de la clase social según la clasificación del British Registrar General (Validity of occupation as an indicator of social class, according to the British Registrar's General classification). *Gac Sanit* 1997;11:205-13.
- 43 Grupo de Trabajo de la Sociedad Española de Epidemiología. *La medición de la clase social en ciencias de la salud*. Barcelona: SG Editores, 1995.
- 44 Kind P. A comparison of two models for scaling health indicators. *Int J Epidemiol* 1982;11:271-5.
- 45 Jenkinson C. Why are we weighting? A critical examination of the use of item weights in a health status measure. *Soc Sci Med* 1991;32:1413-16.
- 46 Torgerson WS. *Theory and methods of scaling*. New York, NY: John Wiley, 1958.
- 47 Schwartz S, Richardson J, Glasziou PP. Quality-adjusted life years: origins, measurements, applications, objections. *Aust J Public Health* 1993;17:272-8.
- 48 Boyle MH, Furlong W, Feeny D, *et al.* Reliability of the Health Utility Index-Mark III used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. *Qual Life Res* 1995;4:249-57.
- 49 Torrance GW, Feeny DH, Furlong WJ, *et al.* Multiattribute Utility function for a comprehensive health status classification system. Health Utility Index Mark 2. *Med Care* 1996;34:702-22.
- 50 Burrows C, Brown K. QALYs for resource allocation: probably not and certainly not now. *Aust J Public Health* 1993;17:278-86.
- 51 Smith R, Dobson M. Measuring utility values for QALYs: two methodological issues. *Health Econ* 1993;2:349-55.
- 52 European Group for Quality of Life and Health Measurement. *European guide to the Nottingham Health Profile*. Richmond: Brookwood Medical Publications, 1993.