

Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey

Oriol Cunillera · Ricard Tresserras · Luis Rajmil · Gemma Vilagut · Pilar Brugulat · Mike Herdman · Anna Mompарт · Antonia Medina · Yolanda Pardo · Jordi Alonso · John Brazier · Montse Ferrer

Accepted: 8 March 2010 / Published online: 31 March 2010
© Springer Science+Business Media B.V. 2010

Abstract

Purpose To compare the EQ-5D, SF-6D, and SF-12 in terms of their capacity to discriminate between groups defined by relevant socio-demographic and health characteristics in a general population survey.

Methods Data were obtained from the 2006 Catalan Health Interview Survey, a representative sample ($n = 4,319$) of the general population of Catalonia (Spain). Effect sizes (ES) and Receiver Operating Characteristic (ROC) curves were calculated to evaluate the instruments' capacity to distinguish between groups based on socio-demographic variables, recent health problems, perceived health, psychological distress, and selected chronic conditions.

Results All instruments showed a similar discriminative capacity between groups based on socio-demographic variables, recent medical visit (ES = 0.47–0.55), activity limitations (ES = 0.92–0.98), perceived health (ES = 0.97–

1.33), and psychological well-being (ES = 1.17–1.57). Effect sizes between respondents with and without any of fourteen selected chronic conditions were large (0.76–1.04) for 4, moderate (0.55–0.74) for 8, and small (0.17–0.39) for two on the EQ-5D index. A similar pattern was observed for the SF-12 but ES were predominantly moderate (7 conditions) or small (6 conditions) on the SF-6D.

Conclusions The EQ-5D and SF-12 were largely comparable in estimating the health burden of chronic conditions, recent health problems, and social inequalities. The SF-6D was less sensitive than the EQ-5D index and SF-12, particularly for physical chronic conditions.

Introduction

Evaluating the health of the general population is one of the specific applications proposed for multi-attribute

O. Cunillera · L. Rajmil · G. Vilagut · M. Herdman · J. Alonso
CIBER en Epidemiología y Salud Pública (CIBERESP),
Barcelona, Spain

O. Cunillera · L. Rajmil · G. Vilagut · M. Herdman ·
Y. Pardo · J. Alonso
Institut Municipal d'Investigació Mèdica (IMIM-Hospital del
Mar), Health Services Research Unit, Barcelona, Spain

R. Tresserras · P. Brugulat · A. Mompарт · A. Medina
Direcció General de Planificació i Avaluació del Departament de
Salut de la Generalitat de Catalunya, Barcelona, Spain

L. Rajmil
Catalan Agency for Health Technology Assessment and
Research, Barcelona, Spain

J. Alonso
Universitat Pompeu Fabra, Barcelona, Spain

J. Brazier
Sheffield Health Economics Group, University of Sheffield,
Sheffield, UK

M. Ferrer (✉)
Institut Municipal d'Investigació Mèdica (IMIM-Hospital del
Mar), Research Unit (IMIM-Hospital del Mar), Barcelona, Spain
e-mail: mferrer@imim.es

M. Ferrer
CIBER en Epidemiología y Salud Pública (CIBERESP),
Health Services, Barcelona, Spain

M. Ferrer
Universitat Autònoma de Barcelona, Aiguader 88, 08003
Barcelona, Spain

measures of health status and well-being [1]. Such instruments are increasingly used in national health surveys alongside the well-established single question of perceived general health [2]. Generic multi-attribute instruments are suitable for this type of application as they can be used in healthy individuals as well as those with different conditions. These generic instruments can in turn be classified as psychometric profile measures or econometric index measures [3]. Psychometric measures generate scores on different health dimensions (profile). Econometric measures provide a single global score (index) which incorporates societal preferences for health states (utilities) that can be used to calculate quality-adjusted life years for use in economic evaluations.

Given its low respondent burden, the EQ-5D has been widely used in population surveys [4–10]. Its descriptive system, composed of only five items with 3 response options for each, is used to generate a preference-based index [11]. On the other hand, the length of the most commonly used profile measure, the Short Form-36 Health Survey (SF-36) [12–15], may have limited its use in population health surveys. However, the development of the SF-12 [16], an abbreviated version with only 12 items, considerably reduced respondent burden making it more suitable for use in large health surveys while still providing the two component summary scores (Physical and Mental) of the original SF-36 [16].

The Short Form-6 Dimension (SF-6D) is an econometric index with an empirically derived system of health state preference values developed first from the SF-36 [17] and then from the SF-12 [18]. Given that the SF-12 questionnaire can generate both an index (the SF-6D) and profile (Physical and Mental Component Summaries), it may provide a useful alternative to the EQ-5D for use in health surveys. However, the advantages of each measure in this context remain unclear.

Several studies have compared the EQ-5D with the SF-6D index and/or SF-12 summaries in different patient groups [19–29], though evidence on the comparative performance of this type of questionnaire requires cumulative results from different settings and types of study. Nevertheless, there have been much fewer head to head comparisons in health surveys [2, 6, 12, 30–32]. Five studies in the general population compared EQ-5D and the SF-12 [2, 6, 30–32] but only one also considered the SF-6D, though the comparison was restricted to the group of individuals in good health [31]. All of these studies found a high proportion of respondents with the best possible scores (ceiling effect) on EQ-5D, suggesting limited discriminative capacity at mild levels of morbidity, while the SF-6D has been shown to suffer from floor effects in patients with severe conditions [29]. To the best of our knowledge, no studies are available which compare the

EQ-5D with the SF-6D index and the SF-12 profile in terms of their power to differentiate between groups defined by their socio-economic characteristics or illness profile in the general population as a whole.

The aim of this study was to evaluate the adequacy of the EQ-5D, SF-6D, and SF-12 version 2 for use in population health surveys by comparing their ceiling effects and capacity to discriminate between groups defined by socio-demographic and health characteristics in a general population sample.

Methods

Study population

Data used in this study came from the 2006 Catalan Health Interview Survey (CHIS), a cross-sectional health survey carried out in Catalonia [33, 34], an Autonomous Community in the north-east of Spain with about 7.2 million inhabitants. A representative sample of individuals from the non-institutionalized general population was surveyed [35]. Participants were selected using a multi-stage, random sampling strategy. In the initial sampling stage, the number of interviews required in each of Catalonia's 37 health care governing districts (HGDs) was established based on population size. The number of interviews estimated per HGD ranged from 400 to 650 for a sampling error at this level of $\pm 5\%$ ($p = q = 0.5$). In the case of Barcelona, 305 interviews were planned for each of the city's 10 municipal areas. In a second stage, representative municipalities were chosen at random for each HGD after stratifying by number of inhabitants. In a third stage, participants from each municipality were selected by simple random sampling from the Catalan census register [36]. The study was approved by the Consultants' Committee of Confidential Information Management (CATIC) from the Catalan Health Department, according to the 2000 revision of the Helsinki Declaration.

To minimize non-response, when selected sampling units could not be located or declined to participate, they were replaced by units with the same characteristics in terms of age group, sex, and neighbourhood. Respondents were replaced as a result of change of address (15.1%), erroneous data (11.2%), refusal to participate (10.8%), absence (5.1%), or death (1.3%). Computer-assisted personal interviews were administered by trained interviewers in the respondents' homes, from December 2005 to July 2006 ($n = 18,126$).

The EQ-5D has been included in all editions of the CHIS performed to date. The inclusion of the SF-12 in a random sub-sample of the 2006 CHIS consisting of 4,319

participants aged 15 or over provided the necessary data for the present head to head comparison.

Both questionnaires were interviewer administered. The EQ-5D was administered towards the beginning of the survey and the SF-12 at the end. In the sub-sample used for the present study, the same sampling methods were applied as in the overall sample to ensure representativeness of the Catalan general population. Within this sub-sample, statistical power to detect small differences (effect size = 0.3) between groups with and without selected chronic conditions was calculated for the least prevalent condition (stroke). Given a standard deviation of 0.27 for the EQ-5D Index and 0.17 for SF-6D in the overall sample, and a significance level of 0.05, statistical power was calculated to be 0.80 for both measures.

Measurement instruments

The EQ-5D is a brief, multi-attribute, generic, preference-based health status measure [11, 37]. Its descriptive system covers five dimensions of health (mobility, self-care, usual activities, pain or discomfort, and anxiety or depression) with three levels of severity in each dimension (no problems, some problems, extreme problems). The instrument therefore defines 243 distinct health states from all the possible combinations of dimensions and levels of severity (i.e. 3^5). In this study, the Spanish version of the EQ-5D and Time Trade Off preference values from the Catalan general population were used [38], thereby producing a single preference-based index ranging from 1 (best health status) to negative values (health states valued as worse than death, being the minimum -0.6533), where 0 is equal to death.

The SF-12 [39] is composed of 12 questions covering eight dimensions of health: Physical Functioning, Role Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role Emotional and Mental Health. In this study, version 2 of the SF-12, which has previously been adapted into Spanish [40], was used. Version 2 of the SF-12 has greater comparability among linguistic adaptations and uses a five-level response scale for all items, except those in the Physical Functioning dimension which retained the original 3-point Likert scale. Scores for the two component summaries (physical and mental component summaries, PCS-12 and MCS-12) were calculated using a standardized procedure to obtain a mean of 50 and a standard deviation of 10 based on USA general population data [39].

The SF-6D [17] is a utility index based on a descriptive system composed of items from six of the eight SF-36 dimensions (role physical and role emotional were combined and general health was omitted). It can be derived from the SF-36 or SF-12 though the number of items

included differs (11 and 7 items, respectively). The SF-6D derived from the SF-12 [18] was used in this study. It classifies individuals into 7,500 health states, of which 241 identified by a fractional factorial design were valued by a representative sample of 611 members of the UK general population. The utility algorithm and the standard gamble preference utilities applied in this study were developed by Brazier et al. [18]. Health state values on the SF-6D are anchored to 1 (perfect health) and 0 (death), and theoretical values for health states generated by the instrument range from 1 to 0.345.

Gender, age, income, and social class were chosen to test the instruments' capacity to discriminate by socio-demographic groups, as it has frequently been shown that there are differences in health on these variables. Women have consistently been shown to have worse health status than men in many different populations [41–43] despite, paradoxically, having a higher life expectancy. A deterioration in health status with age has also been shown on the physical domains of health, though older adults often report better psychological health than younger adults [44]. Those with low income or in less-privileged social classes also tend to report poorer health status [45].

Social class was assigned according to the respondent's most recent occupation (or the occupation of the head of the household in the case of those who were looking after the home) using an adapted version of the British Registrar General's Social Classes [46]: classes I and II (managerial and free-lance professionals), class III (skilled non-manual occupations), class IV (skilled manual workers), and class V (non-skilled manual workers).

Recent health problems were assessed by asking respondents about visits to a health care professional and limitation of activities during the previous two weeks. Limitation of activities was computed as a dichotomous variable which assumed limitation in individuals responding affirmatively to any of three questions: having to spend at least half a day in bed; being unable to perform household tasks or go to work or school; or having to reduce daily activities such as walking, sports, play, and/or shopping as a result of health problems.

Perceived health was assessed by the SF-12 question "In general, how would you rate your health: Excellent, Very good, Good, Fair, or Poor?" Psychological distress was used as the indicator of mental health. It was measured using the 12-item version of the General Health Questionnaire-GHQ [47], a screening instrument designed to detect diagnosable psychiatric disorders, which has shown adequate psychometric properties [48]. The global GHQ-12 score was computed following the method proposed by Goldberg with all responses on the 4 point Likert scale being dichotomized to (0-0-1-1). A total GHQ score ≤ 2 was considered to indicate absence of psychological

distress and ≥ 3 points to indicate risk of psychological distress [47, 48].

The CHIS also included a checklist of 27 common chronic conditions. Respondents were asked ‘Do you suffer from or have you suffered from any of the following chronic conditions?’ and had to answer Yes or No for each condition. A summary indicator of physical health was derived from the checklist based on the number of reported chronic physical conditions and after excluding mental disorders. Respondents were classified into one of four categories by quartiles (none, 1 or 2, 3 or 4, and 5 or more physical conditions), as well as by presence or absence of fourteen specific chronic conditions considered the most clinically relevant in each medical specialty.

Data analysis

To obtain corrected standard errors from the complex sample survey, analyses were performed using the SU-DAAN version 8.0 statistical package [49]. Standard errors (SE) and test significance were estimated using the Taylor series method and Without Replacement design, correcting by finite units in each stage. A weighting factor which took into account age, gender, and municipality was applied to restore the representativeness of the population in Catalonia.

Sample characteristics were described by calculating the raw number of individuals in each group and the weighted percentage. Ceiling and floor effects (proportion of respondents with the best and worst possible theoretical scores, respectively) were calculated for the EQ-5D, SF-6D, and SF-12 component summaries.

To evaluate the capacity to discriminate between known groups, group mean scores were compared using one-way analysis of variance. Effect sizes were computed as the difference between the mean of the groups divided by the pooled standard deviation [50]. Due to the complex sampling, the pooled standard deviation was estimated from the corrected standard errors and the weighted number of individuals in the groups. The extreme-split formula for the variance [50] was used to calculate 95% Confidence Intervals (95% CI) for the effect sizes. For variables with more than two categories (i. e. age and social class), the effect size between extreme groups was calculated. General guidelines define an effect size of 0.2 as small, 0.5 as moderate, and 0.8 as large [51, 52]. This classification was used to interpret differences in the discriminative capacity of the instruments studied. A relevant difference was defined as one in which different instruments showed different categories of effect size for the same between-group comparison, as this would imply disagreement on the amount of health burden. For example, we would consider

a relevant difference to exist if the SF-6D showed a large effect size and the EQ-5D a moderate effect size for the same between-group comparison.

The discriminative properties of the econometric indexes were also compared using receiver operating characteristic (ROC) curves [12, 19, 53]. In this analysis, the performance of the EQ-5D index and the SF-6D index was evaluated against two external indicators of health status: ‘‘Perceived Health’’ and ‘‘Reported chronic physical conditions’’. These two indicators were dichotomized using all possible cut-off points. The utility measure generating the largest area under the ROC curve (AUC) was regarded as the most sensitive at detecting differences in the external indicator. F-ratios of the significance test for the AUC were referenced to 1.0 for the SF-6D Index. A value over 1.0 would therefore indicate that the EQ-5D Index was more efficient than the SF-6D at detecting differences between groups.

Results

Table 1 shows the sample characteristics for the 4,319 participants in the CHIS SF-12 subsample. Half of the respondents were women, almost 40% were in social class IV, and 41.4% reported having 3 or more chronic physical conditions. However, only 13.9% of the sample reported limitation of activities in the previous two weeks and 88.2% presented no risk for psychological distress according to their GHQ scores. The most frequent chronic condition was low back pain and stroke had the lowest prevalence with weighted percentages of 28.2 and 1.3%, respectively.

The mean scores of the EQ-5D and SF-6D indexes (Table 2) were 0.88 and 0.82, respectively. Median scores were higher than means, indicating a skewed distribution for both indexes which was particularly noticeable on the EQ-5D. Floor effects were negligible on all of the measures. The EQ-5D Index showed a considerable ceiling effect, with 59.7% of individuals having the highest possible score. This compares to a ceiling effect of 18.3% and 0% on the SF-6D and SF-12 summaries, respectively. The ceiling effect of the EQ-5D and SF-6D indices varied substantially across different groups (Table 3). For example, ceiling effects on the EQ-5D Index ranged from 89% in individuals who reported no chronic conditions to 20% in those with ≥ 5 chronic physical conditions. The corresponding figures for the SF-6D were 27% and 7%, respectively.

Table 4 shows mean scores and effect sizes for groups based on socio-demographic variables and health indicators. Statistically significant differences ($P < 0.01$) by age and social class were found for all between-group

Table 1 Sample characteristics of the Catalan Health Interview Survey (CHIS) SF-12 subsample (2006)

	<i>n</i>	%	SE
Gender			
Female	2166	50.2	0.9
Male	2153	49.8	0.9
Age			
15–44 years	2152	53.5	0.9
45–64 years	1268	28.9	0.8
65–74 years	504	10.0	0.5
75 years and over	395	7.7	0.5
Social class			
I–II (managerial and free-lance professionals)	849	20.7	0.7
III (skilled non-manual occupations)	1143	28.3	0.8
IV (skilled manual workers)	1835	39.5	0.8
V (non-skilled manual workers)	492	11.5	0.6
Income			
<600 €/month	1044	29.2	0.9
601–1,200 €/month	1330	39.1	1.0
1,201–1,800 €/month	692	21.6	0.8
>1,800 €/month	309	10.1	0.6
Medical visit last 2 weeks			
Yes	927	23.1	0.8
No	3392	76.9	0.8
Restriction last 2 weeks			
Yes	585	13.9	0.6
No	3734	86.1	0.6
Perceived health			
Excellent	337	8.0	0.5
Very good	1012	24.1	0.8
Good	1918	44.9	0.9
Fair	841	18.4	0.7
Poor	211	4.7	0.4
Reported chronic physical conditions			
None	1013	25.5	0.8
1 or 2 conditions	1402	33.2	0.9
3 or 4 conditions	864	18.8	0.7
5 or more conditions	1040	22.6	0.7
Psychological distress			
Yes	491	11.8	0.6
No	3828	88.2	0.6

n are unweighted, percentages are weighted, and standard errors (SE) were estimated using the Taylor series method and with replacement design. There were no missing data on any variables except income, where data were missing for 944 individuals (18.2%); no imputation was used and percentages for this variable are calculated for the total of valid responses

comparisons on both indexes and the SF-12 Physical Component Summary. Small differences were found by gender, with effect sizes (ES) ranging from 0.23 to 0.35. Differences between the upper and lower age groups were

large for both indexes and the SF-12 Physical component summary (ES = 0.75–1.60). Small to moderate differences were observed on the three instruments between groups defined by social class (ES = 0.18–0.54), income (ES = 0.27–0.53), and medical visits (ES = 0.31–0.55). Effect sizes between groups differing on activity limitations, perceived health, number of chronic physical conditions, and psychological distress were large on all three instruments.

Mean scores and effect sizes for groups with and without specific chronic conditions are shown in Table 5. The highest effect sizes were found for arthritis (1.09 on SF-12 PCS and 1.05 on EQ-5D), while chronic allergies presented the lowest effect sizes. Effect sizes on the EQ-5D index were large for 4 conditions, moderate for 8 conditions, and small for asthma and chronic allergies. Most effect sizes on the SF-6D were moderate or small (7 and 6 conditions, respectively); a large effect size was only observed when comparing groups based on presence of mental disorders. Differences between SF-6D and EQ-5D effect sizes were statistically significant in five of the six conditions for which the SF-6D was less sensitive than the EQ-5D. This is because the 95% CI of the difference did not include the zero for hypertension (0.06–0.27), myocardial infarction (0.07–0.65), arthritis (0.27–0.48), low back pain (0.17–0.36), and skin problems (0.01–0.33).

The pattern of effect sizes across physical chronic conditions shown by the SF-12 Physical component summary was similar to that of the EQ-5D, with differences being observed on only four conditions; two in favour of the SF-12 (diabetes ES = 0.78 vs. 0.72 and cancer ES = 0.84 vs. 0.74, large vs. moderate), and two in favour of the EQ-5D (skin problems ES = 0.55 vs. 0.39 and migraine ES = 0.57 vs. 0.43, moderate vs. small). However, the differences were not statistically significant. Finally, effect sizes for the SF-12 Mental component summary were small for all conditions except mental disorders, which showed large effect sizes.

Table 6 shows the areas under the receiver operating characteristic curves (AUC) for the EQ-5D and SF-6D indexes when detecting differences in Perceived Health and Reported chronic physical conditions. F-ratios ranging from 1.03 to 3.57 indicated that EQ-5D was more efficient in detecting differences between groups in almost all cases.

Discussion

This comparative study of the EQ-5D, SF-6D, and SF-12, all short instruments which are likely to be increasingly used as indicators of population health status, confirmed the considerable ceiling effect of the EQ-5D index observed previously in general population health surveys

Table 2 Distribution of EQ-5D and SF-6D indexes and SF-12 summary scores on the 2006 CHIS SF-12 subsample

	No. of items	Theoretical range	Observed range	Floor effect (%)	Ceiling effect (%)	Mean (SD)	Median [IQR]
EQ-5D	5	−0.65 to 1.00	−0.65 to 1.00	0.12	59.69	0.88 (0.27)	1.00 [0.87, 1.00]
SF-6D	6	0.35 to 1.00	0.35 to 1.00	0.18	18.26	0.82 (0.17)	0.86 [0.71, 0.93]
SF-12 PCS	12	0.65 to 80.73	7.83 to 68.48	0	0	48.88 (11.34)	52.17 [44.49,55.13]
SF-12 MCS	12	−8.81 to 81.65	4.61 to 74.77	0	0	53.52 (11.58)	55.78 [49.09,60.77]

SD standard deviation, *IQR* interquartile range

Table 3 Ceiling effect for the EQ-5D and SF-6D by socio-demographic, physical and mental indicators, in the 2006 CHIS SF-12 subsample

	EQ-5D		SF-6D	
	<i>n</i>	%	<i>n</i>	%
Age				
15–44	1587	73.03	468	20.66
45–64	667	53.27	227	17.48
65–74	182	34.55	67	15.26
75 and over	101	23.51	27	8.35
Number of chronic physical conditions				
None	909	89.47	288	26.68
1 condition	614	77.60	169	20.67
2 conditions	389	62.63	124	19.43
3 or 4 conditions	414	47.35	138	16.96
5 or more conditions	211	19.63	70	7.08
Psychological distress				
Yes	104	21.46	4	0.92
No	2433	64.78	785	20.57
Chronic conditions				
Hypertension (<i>n</i> = 917)	362	37.99	118	12.86
Myocardial infarction (<i>n</i> = 105)	32	29.26	13	12.34
Arthritis (<i>n</i> = 993)	185	18.67	74	7.33
Low back pain (<i>n</i> = 1277)	406	31.40	140	12.21
Asthma (<i>n</i> = 259)	114	43.51	28	12.25
Chronic bronchitis (<i>n</i> = 284)	87	28.39	28	10.29
Chronic allergies (<i>n</i> = 697)	380	54.40	109	16.40
Skin problems (<i>n</i> = 327)	125	35.74	40	12.23
Diabetes (<i>n</i> = 285)	83	32.00	27	12.06
Thyroid disease (<i>n</i> = 186)	62	32.64	12	4.28
Stroke (<i>n</i> = 65)	22	36.19	3	7.17
Migraine (<i>n</i> = 845)	297	35.71	92	11.02
Mental disorders (<i>n</i> = 773)	154	19.66	43	6.00
Cancer (<i>n</i> = 139)	39	26.63	15	10.55

Number and percentage of respondents reporting highest theoretical score (1 = best health status). *n* are unweighted and percentages weighted

[2, 6, 30, 31]. However, both the EQ-5D and the SF-12 showed a good capacity to distinguish between groups defined by socio-demographic and health variables, and

neither was found to be clearly superior, with the results depending on which groups were compared. The SF-6D did, however, generally show poorer discriminative capacity for chronic physical conditions.

The EQ-5D index presented an extremely large ceiling effect compared with the SF-6D index (59.7% and 18.3%, respectively, in this sample). Previous studies in general population samples have also reported ceiling effects of approximately 50% for the EQ-5D index [2, 6, 12, 30, 31]. Furthermore, the ceiling effect observed among individuals reporting specific chronic conditions was also substantial (over 20% in the majority of cases). In contrast, the ceiling effect on the SF-6D was lower than the frequently quoted threshold of 15% [54] in all conditions except chronic allergies (16%). The difference in the ceiling effects may be due, at least in part, to the different recall periods of the EQ-5D ('today') and the SF-6D ('last 4 weeks'), as a longer recall period may give more scope for a respondent to include small transient effects on health which might not be picked up by the EQ-5D [55]. The SF-12 component summaries showed no ceiling or floor effects because they were obtained by summing scores of health dimensions after applying positive and negative weights to each dimension [39].

Only two recent studies have directly compared the EQ-5D and the SF-6D indexes in the general population [12, 31]. Most previous head to head studies only compared the EQ-5D index with the SF-12 profile [2, 6, 30]. In our study, the SF-6D index showed poorer discriminative capacity than the EQ-5D and the SF-12. These findings are consistent with those reported by Bharmal and Thomas [31], who compared the discriminative capacity of the SF-6D with the SF-12 in individuals reporting full health on the EQ-5D. However, findings from the Health Survey for England [12] showed that the SF-6D derived from the SF-36 was better than EQ-5D at discriminating between groups. The SF-6D had greater areas under the curve (AUC) than EQ-5D, and it was 30.9%–100.4% and 10.4%–45.6% more efficient than the EQ-5D at detecting differences in self-reported health status and illness, disability or infirmity, respectively. In contrast, in our study, the EQ-5D presented a significantly higher AUC than the SF-6D. Although differences in effect sizes between the SF-6D and

Table 4 Discriminative capacity for the EQ-5D and SF-6D indexes and the SF-12 version 2 summary scores in the 2006 CHIS SF-12 subsample

	EQ-5D		SF-6D		SF-12 PCS		SF-12 MCS	
	Mean (SE)	Effect size (95% CI)	Mean (SE)	Effect size (95% CI)	Mean (SE)	Effect size (95% CI)	Mean (SE)	Effect size (95% CI)
Gender								
Female	0.84 (0.007)	0.301 (0.241, 0.361)	0.79 (0.004)	0.351 (0.291, 0.411)	47.6 (0.27)	0.232 (0.172, 0.291)	51.9 (0.27)	0.289 (0.229, 0.349)
Male	0.92 (0.005)	1.234 (1.114, 1.353)	0.85 (0.003)	0.753 (0.636, 0.870)	50.2 (0.20)	1.604 (1.481, 1.727)	55.2 (0.22)	0.075*(-0.041, 0.190)
Age (years)								
15–44	0.94 (0.004)		0.84 (0.003)		52.1 (0.17)		54.0 (0.23)	
45–64	0.86 (0.008)		0.81 (0.005)		48.1 (0.32)		52.9 (0.35)	
65–74	0.75 (0.017)		0.77 (0.009)		42.2 (0.59)		53.2 (0.62)	
75 and over	0.67 (0.022)		0.72 (0.011)		37.6 (0.78)		53.2 (0.67)	
Social class								
I–II (managerial and free-lance)	0.93 (0.006)	0.488 (0.377, 0.599)	0.84 (0.005)	0.354 (0.244, 0.465)	51.2 (0.32)	0.539 (0.428, 0.651)	54.2 (0.35)	0.176*(0.066, 0.286)
III (skilled non-manual occupations)	0.89 (0.007)		0.82 (0.005)		49.5 (0.32)		53.5 (0.33)	
IV (skilled manual workers)	0.86 (0.007)		0.82 (0.004)		48.2 (0.28)		53.6 (0.29)	
V (non-skilled manual workers)	0.81 (0.016)		0.78 (0.009)		45.6 (0.56)		52.2 (0.60)	
Income								
< 600 € / month	0.78 (0.011)	0.504 (0.383, 0.626)	0.77 (0.006)	0.471 (0.349, 0.592)	45.3 (0.43)	0.526 (0.404, 0.648)	51.2 (0.42)	0.274 (0.153, 0.395)
601–1,200 € / month	0.89 (0.008)		0.82 (0.005)		49.3 (0.30)		54.0 (0.31)	
1,201–1,800 € / month	0.94 (0.005)		0.85 (0.005)		51.6 (0.31)		54.7 (0.37)	
> 1,800 € / month	0.94 (0.009)		0.85 (0.008)		51.9 (0.40)		54.6 (0.52)	
Medical visit last 2 weeks								
Yes	0.77 (0.012)	0.548 (0.476, 0.620)	0.75 (0.007)	0.472 (0.400, 0.543)	44.3 (0.44)	0.551 (0.479, 0.623)	50.8 (0.47)	0.313 (0.242, 0.384)
No	0.91 (0.004)		0.84 (0.003)		50.3 (0.17)		54.3 (0.18)	
Restriction last 2 weeks								
Yes	0.66 (0.018)	0.980 (0.892, 1.069)	0.69 (0.008)	0.921 (0.833, 1.010)	40.0 (0.62)	0.979 (0.891, 1.068)	48.1 (0.64)	0.557 (0.470, 0.644)
No	0.91 (0.003)		0.84 (0.003)		50.3 (0.16)		54.4 (0.17)	
Perceived Health								
Fair or Poor	0.64 (0.013)	1.328 (1.251, 1.404)	0.70 (0.006)	0.971 (0.897, 1.045)	–	–	–	–
Excellent, very good or good	0.95 (0.002)		0.85 (0.002)		–		–	
Reported chronic physical conditions								
None	0.98 (0.002)	1.103 (1.010, 1.195)	0.87 (0.004)	0.895 (0.805, 0.986)	53.4 (0.19)	1.298 (1.203, 1.392)	55.5 (0.30)	0.460 (0.373, 0.547)
1 or 2 disorders	0.94 (0.004)		0.84 (0.004)		51.5 (0.22)		54.5 (0.28)	
3 or 4 disorders	0.88 (0.008)		0.82 (0.006)		49.0 (0.34)		53.5 (0.39)	
5 or more disorders	0.67 (0.013)		0.72 (0.006)		39.8 (0.43)		49.9 (0.46)	
Psychological distress								
Yes	0.62 (0.020)	1.169 (1.073, 1.265)	0.63 (0.008)	1.336 (1.239, 1.433)	43.2 (0.73)	0.585 (0.492, 0.679)	39.8 (0.67)	1.566 (1.468, 1.665)
No	0.91 (0.003)		0.84 (0.002)		49.6 (0.16)		55.4 (0.15)	

One-way analysis of variance to compare groups defined by each variable: all *P*-values were < 0.001, except for those marked with *

Effect sizes were computed as the difference between the means of the extreme groups divided by the pooled standard deviation

SF-12 PCS physical component summary, SF-12 MCS mental component summary, SE standard error, CI confidence interval

Table 5 Discriminative capacity for the EQ-5D and SF-6D Indexes and the SF-12 version 2 summary scores by specific chronic conditions in the 2006 CHIS SF-12 subsample

	EQ-5D			SF-6D			SF-12 PCS			SF-12 MCS		
	Mean (SE)	Effect size (95% CI)		Mean (SE)	Effect size (95% CI)		Mean (SE)	Effect size (95% CI)		Mean (SE)	Effect size (95% CI)	
Hypertension		0.550 (0.474, 0.627)			0.385 (0.309, 0.462)			0.638 (0.561, 0.716)			0.169 (0.093, 0.245)	
Yes	0.76 (0.013)			0.76 (0.007)			43.2 (0.47)			51.9 (0.46)		
No	0.91 (0.004)			0.83 (0.003)			50.2 (0.17)			53.9 (0.19)		
Myocardial infarction		0.762 (0.557, 0.967)			0.400 (0.195, 0.604)			0.979 (0.774, 1.184)			0.017*(-0.187, 0.221)	
Yes	0.68 (0.044)			0.75 (0.020)			38.1 (1.47)			53.3 (1.35)		
No	0.88 (0.004)			0.82 (0.003)			49.1 (0.17)			53.5 (0.18)		
Arthritis		1.053 (0.977, 1.129)			0.679 (0.604, 0.753)			1.094 (1.017, 1.171)			0.295 (0.222, 0.368)	
Yes	0.67 (0.012)			0.73 (0.006)			40.2 (0.44)			50.8 (0.47)		
No	0.93 (0.003)			0.84 (0.003)			51.2 (0.15)			54.2 (0.18)		
Low back pain		0.689 (0.621, 0.757)			0.421 (0.354, 0.488)			0.634 (0.566, 0.701)			0.236 (0.169, 0.302)	
Yes	0.75 (0.011)			0.76 (0.006)			44.0 (0.38)			51.6 (0.39)		
No	0.93 (0.004)			0.84 (0.003)			50.8 (0.17)			54.3 (0.19)		
Asthma		0.390 (0.266, 0.515)			0.312 (0.187, 0.436)			0.345 (0.220, 0.469)			0.250 (0.126, 0.374)	
Yes	0.78 (0.023)			0.77 (0.012)			45.2 (0.84)			50.8 (0.83)		
No	0.88 (0.004)			0.82 (0.003)			49.1 (0.17)			53.7 (0.18)		
Chronic bronchitis		0.555 (0.431, 0.679)			0.487 (0.363, 0.611)			0.621 (0.497, 0.746)			0.298 (0.174, 0.421)	
Yes	0.74 (0.024)			0.74 (0.012)			42.4 (0.78)			50.3 (0.82)		
No	0.89 (0.004)			0.82 (0.003)			49.3 (0.17)			53.7 (0.18)		
Chronic allergies		0.165 (0.084, 0.246)			0.136 (0.055, 0.217)			0.089*(0.008, 0.170)			0.149 (0.068, 0.230)	
Yes	0.84 (0.012)			0.80 (0.007)			48.0 (0.47)			52.1 (0.50)		
No	0.89 (0.004)			0.82 (0.003)			49.0 (0.18)			53.8 (0.19)		
Skin problems		0.547 (0.435, 0.658)			0.379 (0.268, 0.490)			0.393 (0.282, 0.504)			0.257 (0.146, 0.367)	
Yes	0.74 (0.023)			0.76 (0.011)			44.8 (0.76)			50.8 (0.75)		
No	0.89 (0.004)			0.82 (0.003)			49.2 (0.17)			53.8 (0.18)		
Diabetes		0.724 (0.596, 0.852)			0.500 (0.373, 0.627)			0.775 (0.648, 0.903)			0.253 (0.126, 0.380)	
Yes	0.69 (0.025)			0.73 (0.013)			40.7 (0.87)			50.8 (0.86)		
No	0.89 (0.004)			0.82 (0.003)			49.4 (0.17)			53.7 (0.18)		
Thyroid disease		0.676 (0.525, 0.826)			0.489 (0.339, 0.639)			0.535 (0.385, 0.685)			0.413 (0.263, 0.563)	
Yes	0.70 (0.031)			0.74 (0.014)			43.1 (0.99)			49.0 (1.18)		
No	0.89 (0.004)			0.82 (0.003)			49.1 (0.17)			53.7 (0.18)		
Stroke		0.832 (0.567, 1.097)			0.551 (0.287, 0.816)			0.887 (0.622, 1.152)			0.286*(0.021, 0.550)	
Yes	0.66 (0.056)			0.72 (0.026)			39.0 (1.81)			50.2 (2.42)		
No	0.88 (0.004)			0.82 (0.003)			49.0 (0.17)			53.6 (0.18)		

Table 5 continued

	EQ-5D		SF-6D		SF-12 PCS		SF-12 MCS	
	Mean (SE)	Effect size (95% CI)	Mean (SE)	Effect size (95% CI)	Mean (SE)	Effect size (95% CI)	Mean (SE)	Effect size (95% CI)
Migraine		0.569 (0.493, 0.645)		0.473 (0.397, 0.549)		0.434 (0.358, 0.509)		0.426 (0.350, 0.502)
Yes	0.76 (0.013)		0.75 (0.007)		45.0 (0.46)		49.6 (0.50)	
No	0.91 (0.004)		0.83 (0.003)		49.8 (0.18)		54.5 (0.18)	
Mental disorders		1.038 (0.956, 1.120)		0.953 (0.871, 1.034)		0.566 (0.486, 0.646)		1.021 (0.939, 1.103)
Yes	0.66 (0.015)		0.69 (0.007)		43.7 (0.52)		44.6 (0.56)	
No	0.92 (0.003)		0.84 (0.003)		50.0 (0.17)		55.4 (0.15)	
Cancer		0.744 (0.569, 0.919)		0.530 (0.355, 0.704)		0.844 (0.669, 1.019)		0.212*(0.038, 0.387)
Yes	0.68 (0.034)		0.73 (0.017)		39.7 (1.23)		51.1 (1.11)	
No	0.88 (0.004)		0.82 (0.003)		49.2 (0.17)		53.6 (0.18)	

One-way analysis of variance to compare groups with or without each specific condition: all *P*-values were < 0.001, except for those marked with *

Effect sizes were computed as the difference between the means of the groups divided by the pooled standard deviation

SF-12 PCS physical component summary, SF-12 MCS mental component summary, SE standard error, CI confidence interval

EQ-5D for these or similar indicators were not relevant according to our criteria, effect sizes on the EQ-5D were nevertheless higher, supporting the AUC comparison. These contrasting findings may be explained by the fact that the descriptive system of the SF-6D derived from the SF-12 is composed of four fewer items than that derived from the SF-36 (2 physical functioning items, 1 pain item, and 1 mental health item) [17, 18], which could make it less discriminative. Further research is needed to compare these two versions of the SF-6D.

Other differences with previous studies include the fact that the EQ-5D was compared with version 1 of the SF-12 [2, 6, 30, 31]. Although both versions of the SF-12 are similar, version 2 incorporated a five-level response scale instead of the dichotomous response format used for the role dimensions in version 1 [39] and there were minor changes in the mental health and vitality dimensions (6 to 5 options). Our results are therefore not directly comparable with previous studies which used version 1 of the SF-12 [2, 6, 30, 31]. Regarding the discriminative power of the SF-12 version 2, the mental component showed similar discriminative capacity to the EQ-5D on psychological distress and chronic mental conditions, while the physical component showed similar discriminative capacity across socio-demographic categories, and by recent health problems, and number of chronic physical conditions. On the other hand, the SF-12 physical component summary was more sensitive than the SF-6D in several specific conditions (hypertension, myocardial infarction, arthritis, low back pain, diabetes, stroke, and cancer). Furthermore, for six of these seven conditions, differences between SF-6D and SF-12 effect sizes were statistically significant; this can be explained by the fact that the SF-12 component summaries measure more specific aspects of health than the SF-6D, which provides a global health measure, and might not be as sensitive to differences in physical and mental health.

Our study had a number of limitations. The most important was the lack of a gold standard for determining the size of the true difference between the groups studied. However, there was considerable agreement between the EQ-5D and the corresponding SF-12 summary score in terms of the magnitude of between-group effect sizes, whilst the SF-6D generally showed lower effect sizes. This may be evidence for a higher degree of validity for the magnitude estimated by the EQ-5D and the SF-12 than the SF-6D. The estimation of effect size also had some advantages to compare the discriminative capacity of different instruments. Although the effect size, the F-statistic and the area under the ROC curve all provide a means of testing for statistically significant between-group differences in score, only the effect size provides information which facilitates a direct interpretation of the differences. For example, all instruments showed that stroke has an

Table 6 Area under receiver operating characteristic curves (AUC) with 95% confidence intervals (95% CI) and *F*-ratios of EQ-5D and SF-6D indexes to detect differences in perceived health and number of reported chronic physical conditions

	EQ-5D			SF-6D			<i>F</i> -ratio
	AUC	95% CI	<i>F</i> -statistic	AUC	95% CI	<i>F</i> -statistic	
Perceived health							
Excellent versus very good, good, regular, and bad	0.67	0.648, 0.696	158.6	0.63	0.596, 0.656	46.9	3.33
Excellent and very good versus good, regular, and bad	0.70	0.680, 0.711	444.6	0.64	0.621, 0.655	204.6	2.17
Excellent, very good, and good versus regular and bad	0.82	0.805, 0.838	544.2	0.76	0.740, 0.775	525.3	1.03
Excellent, very good, good, and regular versus bad	0.93	0.919, 0.950	446.8	0.88	0.858, 0.906	532.9	0.84
Reported chronic physical conditions							
None versus 1 or more conditions	0.71	0.697, 0.728	602.4	0.64	0.620, 0.657	166.7	3.57
None, 1 or 2 conditions versus 3 or more conditions	0.76	0.745, 0.775	499.6	0.67	0.655, 0.687	285.5	1.75
Less than 5 conditions versus 5 or more conditions	0.81	0.789, 0.822	430.7	0.72	0.701, 0.738	368.0	1.18

F-ratio is referenced to 1.0 for the SF-6D Index. A value higher than 1.0 indicates that the EQ-5D Index is more efficient than the SF-6D in detecting differences between groups

impact on health, but whereas the SF-6D indicated a moderate impact ($ES = 0.55$), the EQ-5D and the SF-12 Physical component summary suggested that the impact was large ($ES = 0.83$ and 0.89 , respectively).

Secondly, preference weights applied to the SF-6D were obtained from a representative sample of the UK general population because no Spanish preferences for this instrument were available. However, only minor differences were found in terms of preferences for EQ-5D health states between Spanish and English respondents [56]. Thirdly, in addition to the preference-based index, the EQ-5D also includes a visual analogue scale on overall health status. This measure has been shown to be more sensitive to between-group differences than the index [30, 31], but it was not included here as our intention was to focus on multi-attribute measures. Fourthly, the perceived health indicator used was an item from the SF-12, and analysis of discriminant capacity based on this variable could not be applied for the SF-12 Component Summaries. However, this analysis was valid for the SF-6D because the item is not used in the calculation of this index. Finally, the sample surveyed is only representative of Catalonia. However, given similarities on national indicators such as life expectancy or Healthy Life Years in the general population between Catalonia, Spain, and other developed countries [57], it is likely that the results found here in terms of relative discriminant capacity will be generalizable to other European countries, USA, or Canada.

One of the original contributions of this study is that, as far as we are aware, this is the first time the discriminative properties of the SF-6D derived from the SF-12 have been compared to the EQ-5D index in a general population health survey. The previous study by Bharmal and Thomas [31] focused on respondents who reported being in full health on these indexes, which might not be the group of

primary interest from a public health perspective. Rather, those with a need for care are of particular importance as regards absolute and comparative access to health care. In conclusion, this study indicates that the EQ-5D is more sensitive to between-group differences than the SF-6D index and may be more appropriate for determining the burden of chronic conditions or for use in economic studies requiring utilities. However, the SF-12 may be a suitable alternative to the EQ-5D in health surveys, because it is generally comparable in estimating burden of disease and provides further complementary information on Physical and Mental components of health.

References

1. Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, *11*(3), 193–205.
2. Macran, S., Weatherly, H., & Kind, P. (2003). Measuring population health: A comparison of three generic health status measures. *Medical Care*, *41*(2), 218–231.
3. Fitzpatrick, R., Fletcher, A., Gore, S., Jones, D., Spiegelhalter, D., & Cox, D. (1992). Quality of life measures in health care. I: Applications and issues in assessment. *British Medical Journal*, *305*(6861), 1074–1077.
4. Luo, N., Johnson, J. A., Shaw, J. W., Feeny, D., & Coons, S. J. (2005). Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. *Medical Care*, *43*(11), 1078–1086.
5. Kind, P., Dolan, P., Gudex, C., & Williams, A. (1998). Variations in population health status: Results from a United Kingdom national questionnaire survey. *British Medical Journal*, *316*(7133), 736–741.
6. Johnson, J. A., & Pickard, A. S. (2000). Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. *Medical Care*, *38*(1), 115–121.

7. Devlin, N., Hansen, P., & Herbison, P. (2000). Variations in self-reported health status: Results from a New Zealand survey. *The New Zealand Medical Journal*, *113*(1123), 517–520.
8. Burstrom, K., Johannesson, M., & Rehnberg, C. (2007). Deteriorating health status in Stockholm 1998–2002: Results from repeated population surveys using the EQ-5D. *Quality of Life Research*, *16*(9), 1547–1553.
9. Saarni, S. I., Harkanen, T., Sintonen, H., Suvisaari, J., Koskinen, S., Aromaa, A., et al. (2006). The impact of 29 chronic conditions on health-related quality of life: A general population survey in Finland using 15D and EQ-5D. *Quality of Life Research*, *15*(8), 1403–1414.
10. Shin, H., & Kim, D. (2008). Health inequality measurement in Korea using EuroQol-5 dimension valuation weights. *Journal of Preventive Medicine and Public Health*, *41*(3), 165–172.
11. Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care*, *35*(11), 1095–1108.
12. Petrou, S., & Hockley, C. (2005). An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Economics*, *14*(11), 1169–1189.
13. Han, T. S., Tjshuis, M. A., Lean, M. E., & Seidell, J. C. (1998). Quality of life in relation to overweight and body fat distribution. *American Journal of Public Health*, *88*(12), 1814–1820.
14. Silvers, K. M., & Scott, K. M. (2002). Fish consumption and self-reported physical and mental health status. *Public Health Nutrition*, *5*(3), 427–431.
15. Schmitz, N., & Kruse, J. (2007). The SF-36 summary scores and their relation to mental disorders: Physical functioning may affect performance of the summary scores. *Journal of Clinical Epidemiology*, *60*(2), 163–170.
16. Ware, J., Jr., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, *34*(3), 220–233.
17. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, *21*(2), 271–292.
18. Brazier, J. E., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care*, *42*(9), 851–859.
19. Stucki, G., Liang, M. H., Fossel, A. H., & Katz, J. N. (1995). Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *Journal of Clinical Epidemiology*, *48*(11), 1369–1378.
20. Walters, S. J., & Brazier, J. E. (2005). Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research*, *14*(6), 1523–1532.
21. Barton, G. R., Bankart, J., Davis, A. C., & Summerfield, Q. A. (2004). Comparing utility scores before and after hearing-aid provision: Results according to the EQ-5D, HUI3 and SF-6D. *Appl Health Econ Health Policy*, *3*(2), 103–105.
22. Fisk, J. D., Brown, M. G., Sketris, I. S., Metz, L. M., Murray, T. J., & Stadnyk, K. J. (2005). A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *Journal of Neurology, Neurosurgery and Psychiatry*, *76*(1), 58–63.
23. Lamers, L. M., Bouwmans, C. A., van Straten, A., Donker, M. C., & Hakkaart, L. (2006). Comparison of EQ-5D and SF-6D utilities in mental health patients. *Health Economics*, *15*(11), 1229–1236.
24. Marra, C. A., Woolcott, J. C., Kopec, J. A., Shojania, K., Offer, R., Brazier, J. E., et al. (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Social Science and Medicine*, *60*(7), 1571–1582.
25. Stavem, K., Froland, S. S., & Hellum, K. B. (2005). Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. *Quality of Life Research*, *14*(4), 971–980.
26. van Stel, H. F., & Buskens, E. (2006). Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease. *Health and Quality of Life Outcomes*, *4*, 20–28.
27. Russell, A. S., Conner-Spady, B., Mintz, A., & Maksymowych, W. P. (2003). The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *The Journal of Rheumatology*, *30*(5), 941–947.
28. Conner-Spady, B., & Suarez-Almazor, M. E. (2003). Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Medical Care*, *41*(7), 791–801.
29. Harrison, M. J., Davies, L. M., Bansback, N. J., McCoy, M. J., Verstappen, S. M., Watson, K., et al. (2009). The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. *Quality of Life Research*, *18*(9), 1195–1205.
30. Johnson, J. A., & Coons, S. J. (1998). Comparison of the EQ-5D and SF-12 in an adult US sample. *Quality of Life Research*, *7*(2), 155–166.
31. Bharmal, M., Thomas, J., & II, I. (2006). Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value in Health*, *9*(4), 262–271.
32. Kontodimopoulos, N., Pappa, E., Niakas, D., & Tountas, Y. (2007). Validity of SF-12 summary scores in a Greek general population. *Health and Quality of Life Outcomes*, *5*, 55–63.
33. Borrell, C., Fernandez, E., Schiaffino, A., Benach, J., Rajmil, L., Villalbi, J. R., et al. (2001). Social class inequalities in the use of and access to health services in Catalonia, Spain: What is the influence of supplemental private health insurance? *International Journal for Quality in Health Care*, *13*(2), 117–125.
34. Artazcoz, L., Benach, J., Borrell, C., & Cortes, I. (2005). Social inequalities in the impact of flexible employment on different domains of psychosocial health. *Journal of Epidemiology and Community Health*, *59*(9), 761–767.
35. Rue, M., Carles, M., Vilapinyo, E., Martinez-Alonso, M., Espinas, J. A., Pla, R., et al. (2008). Dissemination of periodic mammography and patterns of use, by birth cohort, in Catalonia (Spain). *BMC Cancer*, *8*, 336–346.
36. IDESCAT. (2008). Official Catalan Institute of Statistics. Available at: <http://www.idescat.cat>. Accessed October 20, 2008.
37. Brooks, R., & The EuroQol Group. (1996). EuroQol: The current state of play. *Health Policy*, *37*(1), 53–72.
38. Badia, X., Schiaffino, A., Alonso, J., & Herdman, M. (1998). Using the EuroQol 5-D in the Catalan general population: Feasibility and construct validity. *Quality of Life Research*, *7*(4), 311–322.
39. Ware, J. E., Jr., Kosinski, M., Turner-Bowker, D. M., & Gandek, B. (2002). *How to score version 2 of the SF-12 Health Survey (with a supplement documenting version 1)*. Lincoln, RI: QualityMetric, Inc.
40. Vilagut, G., Ferrer, M., Rajmil, L., Rebollo, P., Permanyer-Miralda, G., Quintana, J. M., et al. (2005). El cuestionario de salud SF-36 español: una década de experiencia y nuevos desarrollos. *Gaceta Sanitaria*, *19*(2), 135–150.
41. Alonso, J., Anto, J. M., & Moreno, C. (1990). Spanish version of the Nottingham Health Profile: Translation and preliminary validity. *American Journal of Public Health*, *80*(6), 704–708.
42. Emery, C. F., Frid, D. J., Engebretson, T. O., Alonzo, A. A., Fish, A., Ferketich, A. K., et al. (2004). Gender differences in quality of life among cardiac patients. *Psychosomatic Medicine*, *66*(2), 190–197.
43. Michelson, H., Bolund, C., Nilsson, B., & Brandberg, Y. (2000). Health-related quality of life measured by the EORTC QLQ-C30—reference values from a large sample of Swedish population. *Acta Oncologica*, *39*(4), 477–484.

44. Power, M., Quinn, K., & Schmidt, S. (2005). Development of the WHOQOL-old module. *Quality of Life Research*, *14*(10), 2197–2214.
45. Starfield, B. (2006). State of the art in research on equity in health. *Journal of Health Politics, Policy and Law*, *31*(1), 11–32.
46. Domingo-Salvany, A., Regidor, E., Alonso, J., & Alvarez-Dardet, C. (2000). Una propuesta de medida de la clase social. Grupo de Trabajo de la Sociedad Española de Epidemiología y de la Sociedad Española de Medicina de Familia y Comunitaria. *Atencion Primaria*, *25*(5), 350–363.
47. Goldberg, D. P., & Williams, P. (1988). *A user's guide to the General Health Questionnaire*. London: Institute of Psychiatry.
48. Gispert, R., Rajmil, L., Schiaffino, A., & Herdman, M. (2003). Sociodemographic and health-related correlates of psychiatric distress in a general population. *Social Psychiatry and Psychiatric Epidemiology*, *38*(12), 677–683.
49. Research Triangle Institute. (2004). *SUDAAN language manual, release 9.0*. North Carolina, EEUU: Research Triangle Institute, NC. Available at: <http://www.rti.org/sudaan>.
50. Hedges, L. V., & Olkin, I. (1985). Estimation of a single effect size: Parametric and non parametric methods. In *Statistical methods for meta-analysis* (pp. 75–106). Orlando, FL: Academic Press.
51. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
52. Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. *Medical Care*, *27*(3 suppl), S178–S189.
53. Osborne, R. H., Hawthorne, G., Lew, E. A., & Gray, L. C. (2003). Quality of life assessment in the community-dwelling elderly: Validation of the assessment of quality of life (AQoL) instrument and comparison with the SF-36. *Journal of Clinical Epidemiology*, *56*(2), 138–147.
54. McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research*, *4*(4), 293–307.
55. Bansback, N., Sun, H., Guh, D. P., Li, X., Nosyk, B., Griffin, S., et al. (2008). Impact of the recall period on measuring health utilities for acute events. *Health Economics*, *17*(12), 1413–1419.
56. Badia, X., Roset, M., Herdman, M., & Kind, P. (2001). A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D Health States. *Medical Decision Making*, *21*(1), 7–16.
57. Mathers, C. D., Sadana, R., Salomon, J. A., Murray, C. J., & Lopez, A. D. (2001). Healthy life expectancy in 191 countries, 1999. *The Lancet*, *357*(9269), 1685–1691.