




Validity of the EQ–5D–5L and reference norms for the Spanish population

Gimena Hernandez^{1,2,3} · Olatz Garin^{1,3,4}  · Yolanda Pardo^{1,2,3} · Gemma Vilagut^{1,3} · Àngels Pont^{1,3} · Mónica Suárez⁵ · Montse Neira⁵ · Luís Rajmil¹ · Inigo Gorostiza^{6,7} · Yolanda Ramallo-Fariña^{7,8} · Juan Cabases⁹ · Jordi Alonso^{1,3,4} · Montse Ferrer^{1,2,3}

Accepted: 3 May 2018 / Published online: 16 May 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract

Background and objective The EuroQol 5 dimensions 5 levels (EQ–5D–5L) is the new version of EQ–5D, developed to improve its discriminatory capacity. This study aims to evaluate the construct validity of the Spanish version and provide index and dimension population-based reference norms for the new EQ–5D–5L.

Methods Data were obtained from the 2011/2012 Spanish National Health Survey, with a representative sample ($n = 20,587$) of non-institutionalized Spanish adults (≥ 18 years). The EQ–5D–5L index was calculated by using the Spanish value set. Construct validity was evaluated by comparing known groups with estimators obtained through regression models, adjusted by age and gender. Sampling weights were applied to restore the representativeness of the sample and to calculate the norms stratified by gender and age groups. We calculated the percentages and standard errors of dimensions, and the deciles, percentiles 5 and 95, means, and 95% confidence intervals of the health index.

Results All the hypotheses established a priori for known groups were confirmed ($P < 0.001$). The EQ–5D–5L index indicated worse health in groups with lower education level (from 0.94 to 0.87), higher number of chronic conditions (0.96–0.79), probable psychiatric disorder (0.94 vs 0.80), strong limitations (0.96–0.46), higher number of days of restriction (0.93–0.64) or confinement to bed (0.92–0.49), and hospitalized in the previous 12 months (0.92 vs 0.81).

Conclusions The EQ–5D–5L is a valid instrument to measure perceived health in the Spanish-speaking population. The representative population-based norms provided here will help improve the interpretation of results obtained with the new EQ–5D–5L.

Keywords EuroQol · EQ–5D–5L · Health-related quality of life · Health status · Utilities · Questionnaires · Reference values · Validity

Introduction

Patient-reported outcomes (PROs) have increasingly gained relevance in research, clinical practice, and health planning. Perceived health, health-related quality of life (HRQL), and other PRO constructs provide complementary information to

traditional health indicators based on morbidity and mortality [1, 2]. PROs are essential to describe health in countries after the epidemiological transition, where life expectancy has been steadily increasing and indicators related to mortality may not be sensitive to the expected results of new treatments and public health interventions.

Psychometric HRQL instruments generate scores on several health dimensions (profiles), while econometric instruments generate a single global score or index, which incorporates society's preferences for health states (utilities). This feature makes econometric instruments suitable for cost–utility analysis by calculating quality-adjusted life years (QALYs). The most widely used econometric instrument in the world is the EuroQoL which, since its development in 1991, has been adapted into more than 170 languages and

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11136-018-1877-5>) contains supplementary material, which is available to authorized users.

✉ Olatz Garin
ogarin@imim.es

✉ Montse Ferrer
mferrer@imim.es

Extended author information available on the last page of the article

countries [3]. It is a generic instrument, applicable both in the general population and in patients with different conditions. There is no doubt that its econometric nature, its low administration burden, and its contrasted metric properties are the main reasons for its wide use. However, the high percentage of individuals with the best health state in the EQ-5D has been repeatedly highlighted as a limitation, since this may reduce its capacity to discriminate within good health [4, 5] and its responsiveness in some health areas [6–8].

The EQ-5D-3L is a brief multi-attribute health status measure composed of five questions with Likert response options (descriptive system) and a visual analogue scale (EQ-VAS). The descriptive system covers five dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with three levels of severity in each dimension (no problems, some problems, and extreme problems). To improve the instrument's sensitivity and to reduce ceiling effects, the EuroQol group has developed a new version with more response options, the EuroQol 5 dimensions 5 levels (EQ-5D-5L). This change from 3 to 5 levels [9] implies an increase in the number of possible health states from 243 (3^5) to 3125 (5^5). The studies assessing its metric characteristics suggest that the new EQ-5D-5L version decreases the ceiling effect [10–17], improves its discrimination capacity [9, 11, 14] without reducing its reliability [9, 12, 14, 15], and provides more precise measurement at individual and group levels [18].

The difficulty in interpreting HRQL scores has been identified as one of the main barriers to the widespread use of this type of outcomes [19]. One strategy used to help interpret scores, especially in generic questionnaires, has been providing reference norms based on general population [20–22]. These indicate a standard value that facilitates the interpretation of the questionnaire scores in comparison to what would be expected, according to age and gender [23]. The EuroQol group has published a book with the population reference norms of the EQ-5D original three-level version from 18 countries [24] including the Spanish standards [25]. Reference norms of the five-level version have also been published [26] for Spain. However, as the development of the Spanish value set needed to calculate the EQ-5D-5L index is quite recent, there are neither studies on its validity nor reference norms for our country with this index [27]. The existing publications covering these issues [26, 28] showed results on the five dimensions and the health index constructed with the 3L–5L crosswalk value set.

Moreover, a multistage sampling process, such as that applied in the Spanish National Health Survey (Spanish NHS) [29] with which the EQ-5D-5L Spanish norms were developed, requires the use of weights to obtain representative estimations of the population, and a specific method for complex sample survey designs to estimate associate

errors. Weights assign to each individual their corresponding proportion in the population, to avoid biased estimators. For example, individuals older than 55 years, those retired, and people with a higher number of chronic conditions are over represented when these sample weights are not applied in the 2011–2012 Spanish NHS. Further to weights, as the associated errors for estimators in studies with a multistage sampling process differ from those with a simple random sample, it is necessary to apply a complex sample survey design method to calculate them correctly. However, the already published EQ-5D-5L Spanish norms [26] were calculated without these sample weights from the Spanish NHS nor with any specific method to estimate associated errors for complex sample survey designs. Therefore, these norms may not be representative of the Spanish population.

The main objectives of this study were to evaluate the construct validity and to obtain the reference norms for the dimensions and health index of the EQ-5D-5L in a representative sample of the non-institutionalized Spanish adults.

Methods

Sample selection and design

Data came from the 2011/2012 Spanish NHS. It is a 3-stage sampling, with a first random selection of 2000 census tracts in each autonomous community (stratifying according to the population size of the municipalities), then a random selection of 24,000 households, and a final selection of an individual aged 15 or older and one below this age per home. Information was collected through a computer-assisted personal interview, held in the homes between July 2011 and June 2012. Non-response rate was 33.8% after adding the homes reserved for replacement. Detailed information on the survey and sample construction can be consulted online from the Spanish Ministry of Health, Social Services and Equality [29].

In the 2011/2012 edition, the survey included for the first time the EQ-5D-5L questionnaire, together with the usual battery of questions to evaluate different health-related aspects and socio-demographic characteristics (age, gender, level of studies, marital status, and work situation) among other. For the current study focused on adults, individuals under 18 years of age have been excluded.

The EQ-5D-5L

The EQ-5D-5L's descriptive system is composed of the EuroQol's 5 original dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). The new 5-level Likert-type scales have the following answer options: No problem, slight problems, moderate problems,

severe problems, and extreme problems or unable to perform. The instrument also includes a visual analogue scale (VAS) on general health: “We would like to know how good or bad your health is today.” The descriptive system for the new version was tested in the United Kingdom (UK) and Spain in 2005, as well as the evaluation of the semantic equivalence [30].

Social preference values of the EQ–5D–5L were obtained in parallel in the UK, Canada, Holland, and Spain, using a standardized protocol designed by the EuroQol group [31] to try to reduce heterogeneity to its minimum, since it was widely highlighted in the social preference estimations of the previous 3-level version [32]. The general population studies carried out to elicit the EQ–5D–5L social preferences combine the techniques of time trade-off and discrete choice [31]. Following this protocol, the definitive social preferences were obtained for the Spanish general population [27]. Applying the social preferences to the individual descriptive system answers, a health index was obtained with a range from 1 (perfect health) to negative values (for those health states considered worse than death), 0 being the value assigned to death.

Variables selected to evaluate construct validity

Based on the Spanish NHS content, the following health indicators were selected to evaluate the construct validity through known groups: education, self-reported chronic conditions, mental health measured by the General Health Questionnaire (GHQ), the restrictions of usual activities due to health reasons, and hospitalization in the previous 12 months. Education degree was transformed into a variable with 4 levels according to the number of years of study: less than compulsory secondary education, finished compulsory secondary education, higher secondary or equivalent, and university studies.

In order to have groups with a balanced number of individuals, a 4-category variable was created with the number of self-reported chronic conditions experienced in the last 12 months (out of a list of 30 included in the survey): none, 1 chronic condition, 2–3, and 4 or more. The GHQ mental health questionnaire [33] measures symptoms of anxiety, depression, and/or insomnia [34]. The GHQ short form has 12 items with response options in a 4-point Likert-type scale. To calculate the global score, responses are dichotomized by assigning value 0 to answer options 1–2, and value 1 to options 3–4. A global score ≥ 3 is associated with a high probability of presenting a psychiatric disorder.

The Spanish NHS includes two questions related to activity restriction due to health reasons in the 2 weeks prior to the interview, and one on limitations in the previous 6 months. The first ones gathered information on having had to reduce or limit usual activities during at least half a

day, and having been forced to stay in bed (or in hospital) for more than half a day. Those individuals who answered positively were asked about the total number of days affected, which was dichotomized into 1–7 days and more than 1 week. The self-perceived limitation question, known as the “Global Activity Limitation Indicator” (GALI), was: *For at least the last 6 months, have you been limited because of a health problem in activities people usually do? Strongly limited; limited; not limited* [35]. The NHS also includes a question about the number of days hospitalized in the previous 12 months. This variable has been dichotomized for known groups’ validity evaluation into having been hospitalized or not.

Data analysis

To describe the characteristics of the sample, the crude frequencies and crude and weighted percentages were calculated. The construct validity of the EQ–5D–5L was assessed by comparing known groups, testing the hypotheses established a priori which were derived from the existing literature [17, 20, 21, 36]. Poorer health (reporting problems at dimensions, or lower values in the EQ–5D–5L index and VAS) was expected in those groups with lower education level, as well as a greater number of chronic conditions, a high probability of presenting psychiatric disorder (evaluated with the GHQ), more limitation in daily activities, higher number of days with restriction or in bed, and hospitalization in previous 12 months. It was specifically hypothesized that individuals with a high probability of presenting a psychiatric disorder would more frequently report problems in the anxiety/depression dimension. Strongly limited individuals, people with a higher number of days with restriction or in bed, and those hospitalized in the previous 12 months, were hypothesized to more frequently report problems in mobility, usual activities, and pain dimensions. To test the differences among known groups, general linear models were applied to the health index and VAS score, and a multinomial logit model to the dimensions, in order to adjust by age and gender.

To graphically show the health differences according to age and gender, figures were created with the different results obtained through the EQ–5D–5L for men and women in each age group (in 10-year intervals): percentage of individuals with no problem in each dimension, and boxplots for the health index and the VAS score. Reference norms based on the Spanish population were estimated stratifying by gender and age groups, calculating the percentage and standard error for each level in the dimensions and the deciles, percentiles 5 and 95, mean, standard deviation (SD), and its 95% confidence interval (95% CI) for both the health index and the VAS score.

The analyses were carried out with the statistic package R. In all analyses, sampling weights were applied to guarantee the sample's representativeness. The standard errors were estimated by the Taylor series linearization method for complex sample survey designs.

Results

Table 1 shows the characteristics of the sample of individuals aged 18 years or older from the Spanish NHS ($N=20,587$), both raw and applying sample weights. Half of the sample were women, and the mean age was 48 (SD=18) years. More than half of the respondents were married (57.6%) and 46.0% were working at the time of the survey. Regarding studies, 32% had completed compulsory secondary education, 13% higher secondary education, 15.5% vocational training, and 16% university studies. Approximately one-third of the sample stated they had presented no chronic conditions in the last 12 months, and most of them had not suffered any restriction from usual activity (88.4%) nor been confined to bed (94.4%) in the previous 2 weeks, or had not been hospitalized (91.3%) in the previous 12 months. According to the GHQ, 20.8% of individuals were likely to present a psychiatric disorder.

The most marked differences between the crude and weighted data were observed in the distribution by age groups, for example, 18.2 versus 13.4% in the 25–34 year-old group and 34.6 versus 44.1% among those over 55 years of age. In this sense, the differences in the proportion of retired individuals (20.7 vs 28.6%) and of those with 3 or more chronic conditions (38.9 vs 33.3%) also stand out.

Figure 1 shows the percentage of individuals with no problems in each dimension of the EQ-5D-5L, according to gender and age group. The dimensions of mobility, self-care, and activities presented a similar distribution: $\geq 90\%$ of individuals without problems up to 45–54 years of age, and from then on the percentage diminishes as the group's age increases, especially in women. In the pain dimension, there was a higher percentage of individuals with no problems in the younger groups and the gender differences were larger. The dimension of anxiety/depression, however, was the one that showed the fewest differences among age groups and highest differences according to gender.

Figure 2 shows the boxplots with the medians, percentiles 25 (Pc25) and 75 (Pc75), and extreme values of both the EQ-5D-5L health index and the VAS score. For instance, for the group of women aged 75–84 the index median was 0.82, the Pc25 0.62, Pc75 0.95, and the interquartile range (IQR) 0.33. The outliers are those values from the sample located between $Pc25-3*IQR$ and $Pc25-1.5*IQR$, while extreme outliers are those which differ considerably from the rest of the set (lower than $Pc25-3*IQR$). In this example,

the outliers are those between 0.17 and -0.38 (represented by circles) and the extreme outliers are placed below -0.38 (represented by stars). In both genders, the index clearly showed differences according to age, with values diminishing in the groups with older individuals. When assessing the general health with the VAS, gender differences were less marked and the relationship with age was more linear.

Table 2 shows the percentage of individuals reporting problems by dimension and means of EQ-5D-5L index and VAS, adjusted by age and gender. These construct validity results based on known groups confirmed the a priori hypotheses, with EQ-5D-5L estimators worsening as the variables' categories indicate lower education or more severe health status. All contrasts were statistically significant ($P < 0.001$). The pattern of dimensions presented the expected results, 12.66% of individuals with a high probability of psychiatric disorder reported problems in the anxiety/depression dimension, while only 1.44% reported them in the rest of the sample ($P < 0.001$). Strongly limited individuals, people with a higher number of days with restriction or in bed, reported problems in the mobility dimension (4.20, 1.41, and 2.53%) more frequently than those who were not limited (0.10, 0.11, and 0.13%). This pattern was also observed for usual activities (8.55, 1.75, and 2.63% vs a negligible percentage among non-limited individuals) and pain dimensions (20.38, 11.34, and 13.06 vs $< 2\%$).

The utility index showed the highest differences between extreme groups in the limitation of daily activities in the 6 months prior to the interview (mean: 0.96 vs 0.46) and in the number of days confined to bed (mean: 0.92 vs 0.49). In contrast, the education level showed the smallest differences (mean: 0.94 vs 0.87). The VAS score, similarly to the health index, showed a worse perceived health in the groups with a lower education level (from 79.09 to 71.78), higher number of chronic conditions (84.31–63.16), probable psychiatric disorder (79.09 vs 63.78), strong limitations (79.65–48.66), a higher number of days of restriction (77.46–55.48) or confinement to bed (76.53–47.38), and with hospitalization in the previous 12 months (76.30 vs 67.14).

The EQ-5D-5L reference norms are presented in the annex, including each of the 5 dimensions, the health index, and the VAS score, for the whole sample and separately for women and men, stratified by age groups (18–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75–84, and 85 or older).

Discussion

This is the first study to evaluate the construct validity and to obtain the Spanish reference norms of the EQ-5D-5L health index calculated with the new country-specific value set. Until now, publications assessing construct validity and estimating reference norms of the Spanish version of the

Table 1 Socio-demographic characteristics of adult individuals from the 2011–2012 Spanish National Health Survey (*N* = 20,587)

	<i>n</i> (Raw%)	Wt%	Standard error
Gender			
Men	9412 (45.7%)	48.6	0.412
Women	11,175 (54.3%)	51.4	0.412
Age			
18–24	1236 (6.0%)	8.8	0.283
25–34	2757 (13.4%)	18.2	0.376
35–44	3951 (19.2%)	20.6	0.364
45–54	3574 (17.4%)	17.8	0.337
55–64	3173 (15.4%)	13.7	0.283
65–74	2731 (13.3%)	10.5	0.243
75–84	2350 (11.4%)	7.8	0.210
85 or more	815 (4.0%)	2.6	0.120
Marital status			
Single	5490 (26.7%)	29.9	0.434
Married (or living with a partner)	10,979 (53.4%)	57.6	0.463
Widow/er	2746 (13.4%)	7.6	0.189
Divorced or separated	1351 (6.6%)	4.9	0.169
Work status			
Employed	8736 (42.5%)	46.0	0.474
Unemployed	2623 (12.8%)	14.9	0.357
Retired	5877 (28.6%)	20.7	0.349
Student	753 (3.7%)	5.6	0.231
Disabled	438 (2.1%)	2.1	0.130
Household chores	2083 (10.1%)	10.2	0.272
Other	51 (0.2%)	0.3	0.069
Level of education			
Cannot read nor write	490 (2.4%)	1.9	0.129
Has attended school for at least 5 years	2563 (12.5%)	9.8	0.296
Attended school for 5 or more years without reaching the last course	2596 (12.6%)	10.9	0.309
Compulsory secondary education	6372 (31.0%)	32.2	0.493
Higher secondary education	2403 (11.7%)	13.3	0.333
Intermediate vocational training or equivalent	1753 (8.5%)	9.1	0.262
Higher vocational training or equivalent	1246 (6.1%)	6.4	0.228
University degree/studies	3137 (15.3%)	16.3	0.406
HEALTH STATUS			
Number of chronic conditions, previous 12 months			
None	5530 (26.9%)	31.6	0.492
1 chronic condition	4142 (20.2%)	21.3	0.362
2 chronic conditions	2909 (14.2%)	13.8	0.291
3 chronic conditions	2199 (10.7%)	9.6	0.245
4 chronic conditions	1588 (7.7%)	6.7	0.204
5 or more chronic conditions	4192 (20.4%)	17.1	0.351
Mental health (GHQ-12)			
Not probable	15,779 (77.8%)	79.3	0.408
Probable psychiatric disorder	4499 (22.2%)	20.7	0.408
Limitation in daily activities, previous 6 months			
Strongly limited	887 (4.3%)	3.6	0.154
Limited, but not strongly	3797 (18.4%)	16.5	0.354
Not limited	15,896 (77.2%)	79.9	0.383

Table 1 (continued)

	<i>n</i> (Raw%)	Wt%	Standard error
Number of days with restriction, previous 2 weeks			
0	18,055 (87.9%)	88.4	0.299
1–7	1556 (7.6%)	7.4	0.230
8–14	924 (4.5%)	4.2	0.176
Number of days in bed, previous 2 weeks			
0	19,384 (94.2%)	94.4	0.210
1–7	935 (4.5%)	4.5	0.188
8–14	252 (1.2%)	1.1	0.086
Hospitalization, previous 12 months			
No	18,650 (90.7%)	91.3	0.237
Yes	1910 (9.3%)	8.7	0.237

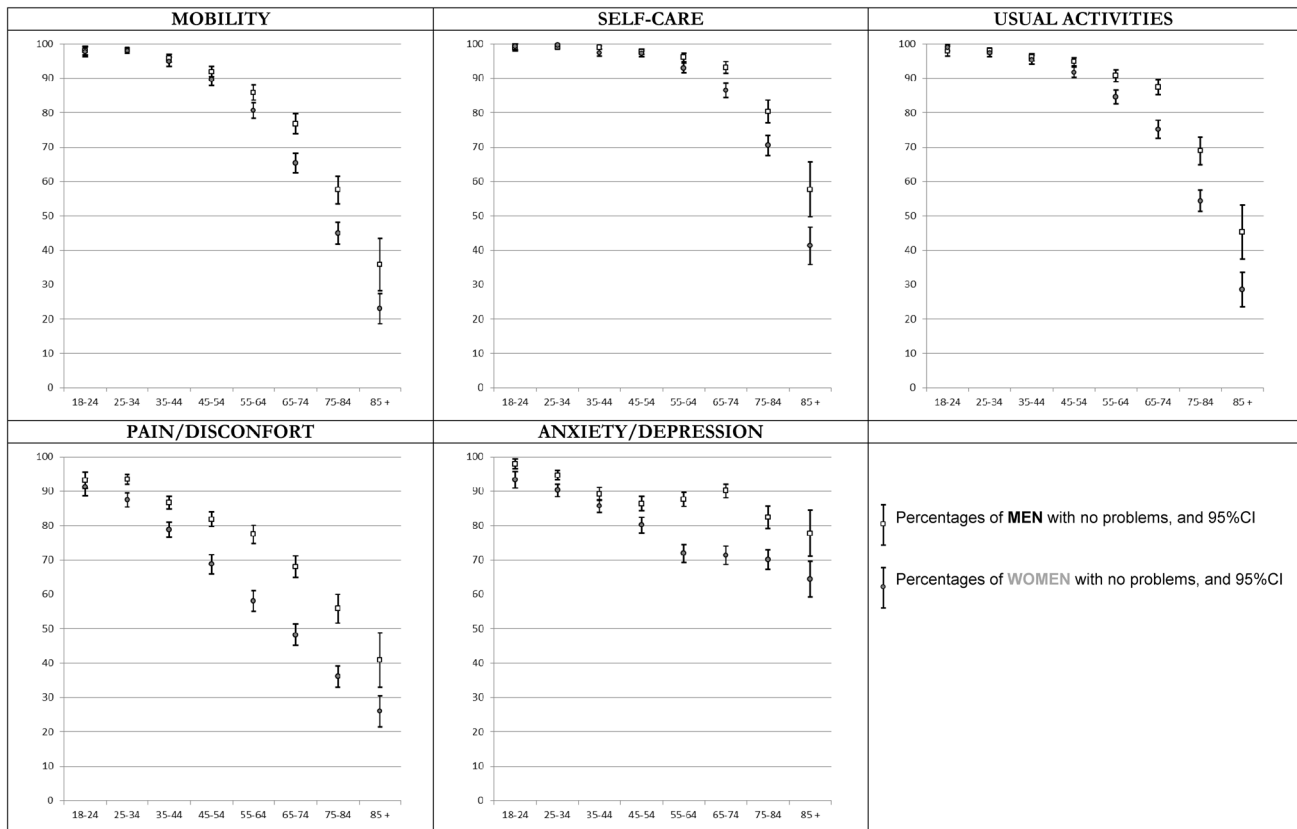


Fig. 1 Percentages of individuals reporting no problems by dimension of EQ-5D-5L index regarding age and gender

EQ-5D-5L showed results regarding the five dimensions and the health index which had been constructed with the 3L-5L crosswalk value set. Moreover, the EQ-5D-5L Spanish population norms are shown for the first time applying sampling weights and methods for complex sample survey designs, and presented with their associated measure of error and percentiles for the EQ-5D-5L index and VAS. As mentioned above, reference norms help to interpret results in instruments that reflect complex and multidimensional

constructs by comparing them to a control group. Having population reference norms in Spain is of fundamental importance, as they will permit estimating the impact of a specific disease, monitoring this impact’s evolution through time, identifying populations that need special attention, and carrying out comparisons among different countries.

The results obtained confirmed the a priori defined hypotheses to evaluate the construct validity of the EQ-5D-5L according to known groups. The magnitudes

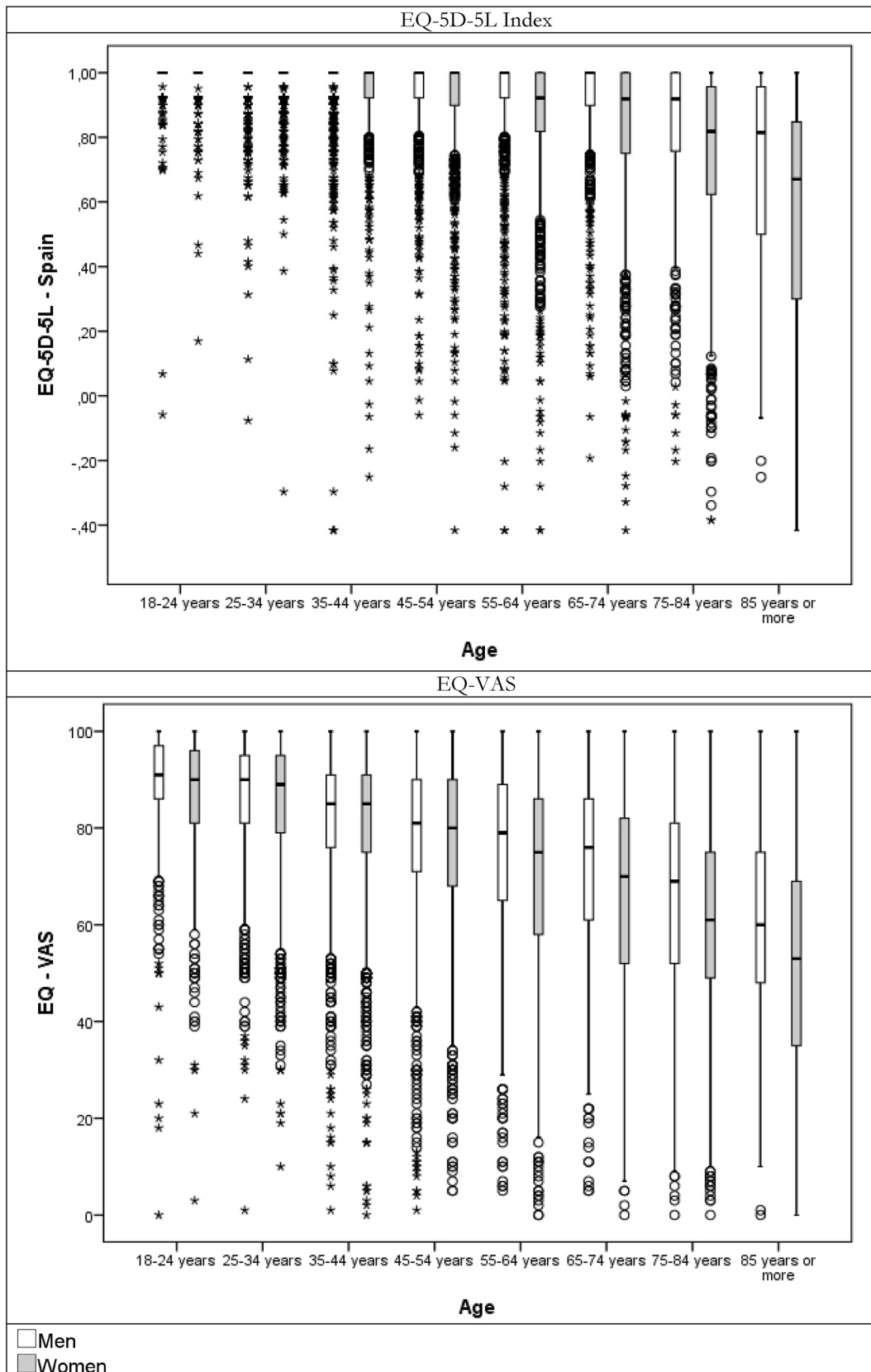


Fig. 2 EQ-5D-5L index and EQ-VAS boxplot, regarding age and gender

Table 2 EQ–5D–5L construct validity based on known groups: percentages of individuals reporting problems by dimension, and means [95% CI] of EQ–5D–5L health index and VAS score

Unweighted <i>n</i> (wt%)	% of individuals reporting problems					Health index	VAS score
	Mobility 3650 (14.2%)	Self-care 1635 (6.1%)	Activity 2839 (11.0%)	Pain 5904 (25.4%)	Anxiety 3407 (14.96%)	Adjusted mean [95% CI]	Adjusted mean [95% CI]
Level of education							
Less than compulsory secondary	0.41%	0.08%	0.43%	3.44%	5.09%	0.87 [0.87–0.88]	71.78 [71.28–72.28]
Compulsory secondary education	0.32%	0.05%	0.29%	2.97%	4.50%	0.91 [0.90–0.91]	75.26 [74.84–75.67]
Higher secondary or equivalent	0.20%	0.04%	0.22%	2.14%	3.30%	0.92 [0.92–0.93]	77.32 [76.85–77.78]
University	0.13%	0.03%	0.13%	1.52%	2.28%	0.94 [0.94–0.95]	79.09 [78.49–79.69]
<i>P</i> value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Number of chronic conditions, previous 12 months							
None	0.11%	0.01%	0.08%	1.26%	2.77%	0.96 [0.96–0.97]	84.31 [83.88–84.75]
1 chronic conditions	0.25%	0.04%	0.26%	4.72%	6.16%	0.96 [0.95–0.96]	80.40 [79.92–80.88]
2–3 chronic conditions	0.42%	0.07%	0.44%	7.87%	11.48%	0.93 [0.93–0.94]	75.55 [75.12–75.98]
4 or more chronic conditions	1.54%	0.24%	1.80%	26.55%	31.47%	0.79 [0.79–0.80]	63.16 [62.72–63.60]
<i>P</i> value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Mental health (GHQ-12)							
Not probable	0.10%	0.02%	0.08%	1.33%	1.44%	0.94 [0.94–0.94]	79.09 [78.85–79.34]
Probable psychiatric disorder	0.43%	0.09%	0.46%	4.95%	12.66%	0.80 [0.80–0.81]	63.78 [63.32–64.24]
<i>P</i> value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Limitation in daily activities, previous 6 months							
Not limited	0.10%	0.02%	0.08%	1.76%	3.62%	0.96 [0.96–0.96]	79.65 [79.41–79.88]
Limited, but not strongly	1.15%	0.27%	1.42%	13.13%	14.21%	0.80 [0.80–0.80]	63.69 [63.20–64.19]
Strongly limited	4.20%	1.88%	8.55%	20.38%	25.15%	0.46 [0.45–0.47]	48.66 [47.62–49.70]
<i>P</i> value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Number of days with restriction, previous 2 weeks							
0 days	0.11%	0.02%	0.09%	1.45%	2.79%	0.93 [0.93–0.93]	77.46 [77.23–77.70]
1–7 days	0.51%	0.10%	0.58%	7.96%	8.64%	0.81 [0.80–0.82]	64.49 [63.68–65.29]
8–14 days	1.41%	0.29%	1.75%	11.34%	10.56%	0.64 [0.63–0.65]	55.48 [54.43–56.54]
<i>P</i> value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Number of days in bed, previous 2 weeks							
0 days	0.13%	0.03%	0.12%	1.62%	2.77%	0.92 [0.92–0.92]	76.53 [76.30–76.77]
1–7 days	0.70%	0.14%	0.92%	9.04%	10.32%	0.76 [0.75–0.77]	60.48 [59.43–61.54]
8–14 days	2.53%	0.42%	2.63%	13.06%	15.79%	0.49 [0.47–0.51]	47.38 [45.28–49.49]
<i>P</i> value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Hospitalization, previous 12 months							
No	0.15%	0.03%	0.14%	1.81%	2.90%	0.92 [0.91–0.92]	76.30 [76.05–76.54]
Yes	0.49%	0.10%	0.56%	4.15%	5.36%	0.81 [0.81–0.82]	67.14 [66.58–67.90]
<i>P</i> value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Percentages and means are adjusted by age and gender

Italics indicates the Confidence Interval

of the differences between extreme groups are quite similar (both in direction and magnitude) to those published for the EQ-5D-3L with data from the 2006 Catalan Health Interview Survey (CHIS) [17]. In our study, the difference between the group with no chronic conditions and the one with 5 or more was 0.17 (0.31 in the CHIS), and the difference between no days and more than 1 week of restriction was 0.30 and 0.43 for bed rest (0.25 in the CHIS for any type of restriction). All these results support the construct validity of the EQ-5D-5L for the Spanish population. Moreover, the majority of the differences identified are greater than 0.061 ± 0.008 , which has been described as the EQ-5D-5L's minimum important difference (MID), indicating their clinical relevance [37].

Reference norms allow the comparison of the observed results in individuals or groups with the general population by identifying in the table of norms the most appropriate gender and age group. The difference between the observed score and the value found in the table provides the deviation from the reference population. For example, the EQ-5D-5L index mean value for women is higher than 0.90 for those under 55 years of age, and below 0.60 for those who are 85 or older. The VAS mean score for men is between 90 and 80 for those under 45 years old, and below 60 for those at the older group (detailed norms are available at the annex).

According to the abundant available evidence on the effect of age and gender on health [38–41], the results of our study (worse HRQL in women and better in younger groups) confirm the need to generate reference norms stratified by these characteristics, such as the ones in this article's annex. In addition, taking into account the relevant precautions due to the cross-sectional nature of the study, the age distribution reflects the worsening of health associated with aging (more pronounced in the dimension of pain/discomfort) except for the mental component, which remains relatively stable throughout life [41].

Our results show lower percentages of individuals with no problems in all dimensions, compared to the Spanish population reference norms of the EQ-5D-3L [24]: 82.5 versus 86.3% in mobility; 92.1 versus 95.9% in self-care; 86.3 versus 88.3% in usual activities; 71.7 versus 77.1% in pain/discomfort; and 83.6 versus 92.2% in anxiety/depression. The EQ-5D-3L norms were obtained from a representative sample of the Spanish population (2001–2003) with a similar sampling approach and characteristics, but with a lower sample size (5473 vs 20,587) and a slightly lower response rate (66.2 vs 78.5%) than the Spanish NHS. Although the lower proportion of individuals without problems is likely due to the expansion from 3 to 5 levels (which allows reporting slight problems), we cannot discard other reasons related to time point or sampling procedures. The distribution in the five EQ-5D-5L dimensions continues to show a marked aggregation of individuals in the best response option (no

problems), but this was expected in the non-institutionalized population.

The smaller ceiling effect obtained with the EQ-5D-5L index has also been described in studies that have compared the two EQ-5D versions in other countries: a 12.5% decrease of ceiling effect in Germany [5, 42], 8.6% in England [10], 5.9% in Italy [16], and 4.5% in South Korea [15]. Focusing on the EQ-5D-5L comparison between countries, our results are closer to the South Korean [15] population than to the German [42], United States [43], or Polish [44] populations. South Korea is the country with the highest prevalence of people answering 'no problems' in mobility (88%), self-care (97%), and usual activities (90%), while Spain is the country showing the highest proportion of individuals without problems in pain (74.6%) and anxiety/depression (85.4).

This is the first time that there are results from the health index of the new EQ-5D-5L in a representative sample of the Spanish population. The means obtained from the EQ-5D-5L health index are very similar to the reference norms of the EQ-5D-3L index [24], and the differences for each age group are very small (below ± 0.02). For instance, the highest difference is observed in the age group of 64–75 years, with means of 0.87 in 5L and 0.89 in 3L reference norms. These differences could be due to the 10-year lapse between both studies [45], the increase in the number of levels in the new version, or the method used to obtain the social preferences: time trade-off [46] in the EQ-5D-3L index and a mixed method in the EQ-5D-5L index [27, 31].

However, the general health VAS results are substantially higher in our study than in the EQ-5D-3L reference norms [24]. The differences are of a larger magnitude in the younger groups (means of 88.2 vs 82.0, in the 18–24-year-old group) and diminish as age increases (means of 69.8 vs 69.0 in the group aged 65–74). As the general health question and the VAS are identical in both versions of the EQ-5D VAS, differences between studies could only be due to the time lapse. In addition, this question on general health is much more global than the EQ-5D descriptive system. Given that it depends on personal values and expectations, social references, and other context factors, it is more probable for the response to vary considerably among generations and groups within the same society, than in the descriptive system with 5 dimensions or the health index.

The previously published reference norms for the EQ-5D-5L [26] offer raw estimators (unweighted) using the 3L-5L crosswalk value set and, even though they are similar, they are not the same as what was obtained using the new EQ-5D-5L Spanish value set, sample weights, and adequate associated errors. For example, the mean of the EQ-5D-5L index in women was 0.855 [26] versus 0.868 (95% CI 0.860–0.876) in the group that was 60–69 years old; 0.780 [26] versus 0.794 (95% CI 0.785–0.803) in the 70–79 years old group; 0.624 [26] versus 0.658 (95% CI

0.647–0.670) in the group aged 80–89; and 0.418 [26] versus 0.523 (95% CI 0.496–0.549) in the group of 90 or more years old.

When interpreting our results, the study's limitations should be considered. The Spanish NHS only includes the non-institutionalized population, thus leaving out of the study those individuals hospitalized, imprisoned, or in senior citizen homes, who have a worse health state and, therefore, resulting in an overestimation of the population's health. This needs to be taken into account when applying the norms in such groups or individuals. Furthermore, the whole Spanish NHS—including the EQ–5D–5L questionnaire—was administered through computer-assisted personal interviews, while self-administration is recommended for the EuroQol. However, a study comparing the interview and the self-completed EQ–5D–3L questionnaire showed little difference between both administration methods [47].

In conclusion, the study has confirmed the construct validity of the new Spanish version of the EQ–5D–5L, and has provided easy-to-use tables with reference norms for all stakeholders (e.g., healthcare planners, researchers, clinicians, patients) in order to interpret results for different purposes, such as establishing optimal goals in clinical management or evaluating changes at the individual and group level. The population norm tables in the annexes reflect granular reporting of descriptive statistics (estimators and their associated error) that facilitate the comparison of EQ–5D–5L results for individuals or specific groups with data for the average person in the general population with a similar age and/or gender.

Acknowledgements We acknowledge Aurea Martin for helping us in the English editing process and supervision of this manuscript.

Author contributions GH analyzed and interpreted the data, drafted, and critically revised the manuscript and did the statistical analysis. OG provided supervision, conceived, and designed the study, and critically revised the manuscript. YP and GV analyzed and interpreted the data, and critically revised the manuscript. AP analyzed and interpreted the data, and did the statistical analysis. MS, MN, LR, IG, YR, and JC interpreted the data and critically revised the manuscript. JA provided supervision, conceived and designed the study, interpreted the data, and critically revised the manuscript. MF obtained funding, provided supervision, conceived and designed the study, interpreted the data, and critically revised the manuscript.

Funding This study has been funded by grants from the Generalitat de Catalunya (2017 SGR 452 and 2014 SGR 748) and the Instituto de Salud Carlos III FEDER, (PI12/00772).

Data availability The Spanish National Health Survey is an official statistic included in the National Statistical Plan. The agency responsible for the survey is the Ministry of Health, Social Services and Equality. The anonymized microdata of the Spanish National Health Survey can be requested for purposes of scientific research.

Compliance with ethical standards

Conflict of interest All authors declare that they have no competing interests.


Ethical approval The Spanish National Health Survey is a statistical operation included in the National Statistical Plan. The agency responsible for the survey is the Ministry of Health, Social Services and Equality and it is performed jointly with the National Statistics Institute according to the 2000 revision of the Helsinki Declaration.

References

- Mishoe, S. C., & Maclean, J. R. (2001). Assessment of health-related quality of life. *Respiratory Care*, *46*, 1236–1257.
- Vodicka, E., Kim, K., Devine, E. B., Gnanasakthy, A., Scoggins, J. F., & Patrick, D. L. (2015). Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007–2013). *Contemporary Clinical Trials*, *43*, 1–9.
- Brooks, R., & The EuroQol Group (1996). EuroQol: The current state of play. *Health Policy*, *37*, 53–72.
- Bharmal, M., & Thomas, J. III (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*, 262–271.
- Hinz, A., Klaiberg, A., Braehler, E., & Konig, H. H. (2006). The Quality of Life Questionnaire EQ-5D: Modelling and norm values for the general population. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *56*, 42–48.
- Xin, Y., & McIntosh, E. (2017). Assessment of the construct validity and responsiveness of preference-based quality of life measures in people with Parkinson's: A systematic review. *Quality of Life Research*, *26*, 1–23.
- Fransen, M., & Edmonds, J. (1999). Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. *Rheumatology*, *38*, 807–813.
- Krahn, M., Bremner, K. E., Tomlinson, G., Ritvo, P., Irvine, J., & Naglie, G. (2007). Responsiveness of disease-specific and generic utility instruments in prostate cancer patients. *Quality of Life Research*, *16*, 509–522.
- Janssen, M. F., Birnie, E., Haagsma, J. A., & Bonsel, G. J. (2008). Comparing the standard EQ-5D three-level system with a five-level version. *Value in Health*, *11*, 275–284.
- Feng, Y., Devlin, N., & Herdman, M. (2015). Assessing the health of the general population in England: How do the three- and five-level versions of EQ-5D compare? *Health and Quality of Life Outcomes*, *13*, 171.
- Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., et al. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: A multi-country study. *Quality of Life Research*, *22*, 1717–1727.
- Kim, S. H., Kim, H. J., Lee, S. I., & Jo, M. W. (2012). Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea. *Quality of Life Research*, *21*, 1065–1073.
- Pickard, A. S., De Leon, M. C., Kohlmann, T., Cella, D., & Rosenbloom, S. (2007). Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. *Medical Care*, *45*, 259–263.
- Jia, Y. X., Cui, F. Q., Li, L., Zhang, D. L., Zhang, G. M., Wang, F. Z., et al. (2014). Comparison between the EQ-5D-5L and the EQ-5D-3L in patients with hepatitis B. *Quality of Life Research*, *23*, 2355–2363.
- Kim, T. H., Jo, M. W., Lee, S. I., Kim, S. H., & Chung, S. M. (2013). Psychometric properties of the EQ-5D-5L in the

- general population of South Korea. *Quality of Life Research*, 22, 2245–2253.
16. Scalone, L., Ciampichini, R., Fagioli, S., Gardini, I., Fusco, F., Gaeta, L., et al. (2013). Comparing the performance of the standard EQ-5D 3L with the new version EQ-5D 5L in patients with chronic hepatic diseases. *Quality of Life Research*, 22, 1707–1716.
 17. Cunillera, O., Tresserras, R., Rajmil, L., Vilagut, G., Brugulat, P., Herdman, M., et al. (2010). Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey. *Quality of Life Research*, 19, 853–864.
 18. Janssen, M. F., Bonsel, G. J., & Luo, N. (2018). Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics*. <https://doi.org/10.1007/s40273-018-0623-8>.
 19. Aaronson, N., Alonso, J., Burnam, A., Lohr, K. N., Patrick, D. L., Perrin, E., et al. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, 11, 193–205.
 20. Alonso, J., Regidor, E., Barrio, G., Prieto, L., Rodriguez, C., & de la Fuente, L. (1998). Population reference values of the Spanish version of the Health Questionnaire SF-36. *Medicina Clinica*, 111, 410–416.
 21. Vilagut, G., Valderas, J. M., Ferrer, M., Garin, O., Lopez-Garcia, E., & Alonso, J. (2008). Interpretation of SF-36 and SF-12 questionnaires in Spain: Physical and mental components. *Medicina Clinica*, 130, 726–735.
 22. Ravens-Sieberer, U., Herdman, M., Devine, J., Otto, C., Bullinger, M., Rose, M., et al. (2014). The European KIDSCREEN approach to measure quality of life and well-being in children: Development, current application, and future advances. *Quality of Life Research*, 23, 791–803.
 23. Schmidt, S., & Pardo, Y. (2014). Normative data. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 4375–4379). Dordrecht: Springer.
 24. Szende, A., Janssen, B., & Cabases, J. (2014). *Self-reported population health: An international perspective based on EQ-5D*. Dordrecht: SpringerOpen.
 25. König, H. H., Bernert, S., Angermeyer, M. C., Matschinger, H., Martínez, M., Vilagut, G., et al. (2009). Comparison of population health status in six European countries: Results of a representative survey using the EQ-5D questionnaire. *Medical Care*, 47, 255–261.
 26. Garcia-Gordillo, M. A., Adsuar, J. C., & Olivares, P. R. (2016). Normative values of EQ-5D-5L: In a Spanish representative population sample from Spanish Health Survey, 2011. *Quality of Life Research*, 25, 1313–1321.
 27. Ramos-Goñi, J. M., Craig, B. M., Oppe, M., Ramallo-Fariña, Y., Pinto-Prades, J. L., Luo, N., et al. (2018). Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value in Health*, 21(5), 596–604.
 28. Martín-Fernández, J., Ariza-Cardiel, G., Polentinos-Castro, E., Sanz-Cuesta, T., Sarria-Santamera, A., & Del Cura-Gonzalez, I. (2017). Explaining differences in perceived health-related quality of life: A study within the Spanish population. *Gaceta Sanitaria*. <https://doi.org/10.1016/j.gaceta.2017.05.016>.
 29. Ministerio de Sanidad, S.S.e.I., & Instituto Nacional de Estadística (2012). *Encuesta Nacional de Salud de España 2011/12*. <http://www.msssi.gob.es/estadEstudios/estadisticas/encuestaNacional/encuesta2011.htm>.
 30. Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, 20, 1727–1736.
 31. Oppe, M., Devlin, N. J., van Hout, B., Krabbe, P. F., & de Charro, F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health*, 17, 445–453.
 32. Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goni, J. M., & Luo, N. (2016). EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics*, 34, 993–1004.
 33. 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, 677–683.
 34. Goldberg, D. P., & Williams, P. (1988). *A user's guide to the General Health Questionnaire*. London: Institute of Psychiatry.
 35. van Oyen, H., Van der Heyden, J., Perenboom, R., & Jagger, C. (2006). Monitoring population disability: Evaluation of a new Global Activity Limitation Indicator (GALI). *Sozial-Und Präventivmedizin*, 51, 153–161.
 36. Schmidt, S., Vilagut, G., Garin, O., Cunillera, O., Tresserras, R., Brugulat, P., et al. (2012). Reference guidelines for the 12-Item Short-Form Health Survey version 2 based on the Catalan general population. *Medicina Clinica*, 139, 613–625.
 37. McClure, N. S., Sayah, F. A., Xie, F., Luo, N., & Johnson, J. A. (2017). Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. *Value in Health*, 20, 644–650.
 38. Hopman, W. M., Towheed, T., Anastassiades, T., Tenenhouse, A., Poliquin, S., Berger, C., et al. (2000). Canadian normative data for the SF-36 health survey. Canadian Multicentre Osteoporosis Study Research Group. *CMAJ*, 163, 265–271.
 39. Burstrom, K., Johannesson, M., & Diderichsen, F. (2001). Swedish population health-related quality of life results using the EQ-5D. *Quality of Life Research*, 10, 621–635.
 40. 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, 477–484.
 41. Power, M., Quinn, K., & Schmidt, S. (2005). Development of the WHOQOL-old module. *Quality of Life Research*, 14, 2197–2214.
 42. Hinz, A., Kohlmann, T., Stobel-Richter, Y., Zenger, M., & Brahler, E. (2014). The quality of life questionnaire EQ-5D-5L: Psychometric properties and normative values for the general German population. *Quality of Life Research*, 23, 443–447.
 43. Craig, B. M., Pickard, A. S., & Lubetkin, E. I. (2014). Health problems are more common, but less severe when measured using newer EQ-5D versions. *Journal of Clinical Epidemiology*, 67, 93–99.
 44. Golicki, D., & Niewada, M. (2017). EQ-5D-5L Polish population norms. *Archives of Medical Science*, 13, 191–200.
 45. 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, 1547–1553.
 46. 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, 311–322.
 47. Puhan, M. A., Ahuja, A., Van Natta, M. L., Ackatz, L. E., & Meiner, C. (2011). Interviewer versus self-administered health-related quality of life questionnaires—Does it matter? *Health and Quality of Life Outcomes*, 9, 30.

Affiliations

Gimena Hernandez^{1,2,3} · **Olatz Garin**^{1,3,4}  · **Yolanda Pardo**^{1,2,3} · **Gemma Vilagut**^{1,3} · **Àngels Pont**^{1,3} · **Mónica Suárez**⁵ · **Montse Neira**⁵ · **Luís Rajmil**¹ · **Inigo Gorostiza**^{6,7} · **Yolanda Ramallo-Fariña**^{7,8} · **Juan Cabases**⁹ · **Jordi Alonso**^{1,3,4} · **Montse Ferrer**^{1,2,3}

Gimena Hernandez
gimenahernandez@gmail.com

Yolanda Pardo
ypardo@imim.es

Gemma Vilagut
gvilagut@imim.es

Àngels Pont
apont@imim.es

Mónica Suárez
msuarez@msssi.es

Montse Neira
mneira@msssi.es

Luís Rajmil
lrajmil@imim.es

Inigo Gorostiza
inigo.gorostizahormaeche@osakidetza.eus

Yolanda Ramallo-Fariña
yolanda.ramallofarina@sescs.es

Juan Cabases
jmcabases@unavarra.es

Jordi Alonso
jalonso@imim.es

¹ Health Services Research Group, IMIM (Hospital del Mar Medical Research Institute), Doctor Aiguader, 88, 08003 Barcelona, Spain

² Universitat Autònoma de Barcelona, Barcelona, Spain

³ Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

⁴ Experimental and Health Sciences Department, Universitat Pompeu Fabra (UPF), Barcelona, Spain

⁵ Subdirección General de Información Sanitaria e Innovación, Spanish Ministry of Health, Social Services and Equality, Madrid, Spain

⁶ Unidad de Investigación, Hospital Universitario Basurto - Osakidetza, Bilbao, Spain

⁷ Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Madrid, Spain

⁸ Fundación Canaria de Investigación Sanitaria (FUNCANIS), Tenerife, Spain

⁹ Departamento de Economía, Universidad Pública de Navarra, Pamplona, Spain