

Quality of Life Research

VALIDITY OF THE EQ-5D-5L AND REFERENCE NORMS FOR THE SPANISH POPULATION --Manuscript Draft--

Manuscript Number:	QURE-D-17-01437R1	
Full Title:	VALIDITY OF THE EQ-5D-5L AND REFERENCE NORMS FOR THE SPANISH POPULATION	
Article Type:	Original Research	
Keywords:	EuroQol; EQ-5D-5L; Health related quality of life; health status; utilities; Questionnaires; reference values; validity.	
Corresponding Author:	Olatz Garin, MPH IMIM-Research Institute Hospital del Mar Barcelona, SPAIN	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	IMIM-Research Institute Hospital del Mar	
Corresponding Author's Secondary Institution:		
First Author:	Gimena Hernandez, M.PH.	
First Author Secondary Information:		
Order of Authors:	Gimena Hernandez, M.PH. Olatz Garin, M.PH., Ph.D Yolanda Pardo, Ph.D Gemma Vilagut, Ph.D Àngels Pont, B.S.C. Mónica Suárez Montse Neira Luís Rajmil, M.PH., Ph.D Iñigo Gorostiza Yolanda Ramallo Juan M. Cabasés Jordi Alonso, Ph.D Montse Ferrer, M.PH., Ph.D	
Order of Authors Secondary Information:		
Funding Information:	Generalitat de Catalunya (2014 SGR 748) Instituto de Salud Carlos III (PI12/00772)	Not applicable Dr. Montse Ferrer
Abstract:	<p>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.</p> <p>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</p>	

	<p>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.</p> <p>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 to 0.79), probable psychiatric disorder (0.94 vs 0.80), strong limitations (0.96 to 0.46), higher number of days of restriction (0.93 to 0.64) or confinement to bed (0.92 to 0.49), and hospitalized in the previous 12 months (0.92 vs 0.81).</p> <p>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.</p>
--	---

Editorial Office

Quality of Life Research

Dear Associate Editor Dr. Nan Luo,

Thank you very much for the very useful suggestions and comments made by the reviewers. We have thoroughly reviewed each comment and have addressed the concerns and suggestions made. Below you will find detailed individual responses to each reviewer's comments.

We are delighted about the possibility of having this manuscript published by Quality of Life Research and welcome any further recommendations you might have.

Please do not hesitate to ask for any additional information you may deem necessary.

Very best regards,

Montse Ferrer & Olatz Garin

On behalf of all authors

Address for correspondence:

Montse Ferrer & Olatz Garin

Health Services Research Group, IMIM (Hospital del Mar Medical Research Institute)

Doctor Aiguader 88 08003 Barcelona, Spain.

Phone: +34 93 3160 740; Fax: +34 93 3160 797;

e-mail addresses: mferrer@imim.es; ogarin@imim.es

POINT BY POINT RESPONSE TO THE REVIEWERS' COMMENTS

Ref. MS: QURE-D-17-01437_R1

“VALIDITY OF THE EQ-5D-5L AND REFERENCE NORMS FOR THE SPANISH POPULATION”

REVIEWER 1

Comment 1: *Introduction, end of para 1:* Especially when we're talking about population health indicators, as in this case, it's probably relevant to talk about public health interventions, not just therapeutic interventions?

Response: Authors totally agree with the reviewer's comment. This suggestion has been included in the new version of the manuscript.

- **Introduction, end of paragraph 1, page 1:** “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.”

Comment 2: *Intro, para 2:* According to the EuroQol Group website, EQ-5D-3L is available in more than 170 languages.

Response: Authors appreciate the reviewer's remark and apologize for the error. We have addressed it suitably.

- **Introduction, paragraph 2, page 1:** “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 countries [3]”

Comment 3: *Intro, para 2:* "...lower discriminatory capacity compared to other econometric indices...". There are a number of comparisons of EQ-5D with other econometric indices which show that it has equally good or better psychometric properties. I would say the picture is not as black and white as you paint it here and that the inclusion of more references and a more nuanced description is probably warranted. Also, note that '...its econometric nature...' can't be a reason for preferring EQ-5D over other econometric measures, as this sentence seems to suggest if you read it carefully from start to finish.

Response: Authors thank the reviewer for this reflection, and have modified the paragraph to address their concerns.

Introduction, paragraph 2, page 1: “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].”

Comment 4: *Intro, para 3: 'discrimination capacity' - be careful with terminology.*

Response: Authors appreciate the reviewer's remark and have changed the term.

- **Introduction, paragraph 3, page 1:** "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)."

Comment 5: *Intro, para 3: before talking about the new version of EQ-5D, you should probably describe the original version?*

Response: Authors thank the reviewer's comment and have described the original EQ-5D version as suggested.

- **Introduction, paragraph 3, page 1:** "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)."

Comment 6: *Intro, para 4, first sentence: do you have a reference to support this claim?*

Response: Authors thank the reviewer for their remark, and have incorporated a supporting reference.

Introduction, paragraph 4, page 2: "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]."

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.

Comment 7: *Intro, para 4, second sentence: I'm not sure we can say that the use of population norms is 'the most common strategy' to help interpret scores. The use of effect sizes or MIDIs is probably equally frequent? It might be better to say 'One strategy used to help interpret scores on generic questionnaires ...'*

Response: Authors appreciate the reviewer's suggestion and have changed the sentence accordingly.

- **Introduction, paragraph 4, page 2:** "One strategy used to help interpret scores, especially in generic questionnaires, has been providing reference norms based on general population [20-22]."

Comment 8: *Intro, para 4, last sentence: Instead of '...the EQ-5D-3L traditional version...', I would say '...the original three-level version of EQ-5D...'*

Response: Authors totally agree with the reviewer and have incorporated the suggested change.

Comment 9: *The last para of the Intro is not very clear in parts and is somewhat clumsily worded. Last sentence, para 3, p6: Not really clear what this means, and the sentence does not really reflect the objectives described in reference 20.*

Response: Based on both reviewers' comments, authors have re-written the mentioned paragraph, and an English native speaker has reviewed it. However, we would like to clarify that reference 20 (Herdman et al. 2011) describes the development of the 5 levels, and the conceptual equivalence between the new Spanish labels and the English ones, but not the country-specific value set needed to calculate the EQ-5D-5L index, which was developed years later (2016).

- **Introduction, last paragraph, page 2:** "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 its 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."

Comment 10: Data analysis: the exploration of validity could have been more ambitious, given the rich source of data available. For instance, what about contrasting scores by level of education or region or use of health services (assuming data was collected for those variables)? Investigating validity using those types of variables would potentially be of particular interest when thinking about health at population level. I think there should also have been more hypotheses at dimension level, e.g. would you have expected to see more problems reported on the anxiety/depression dimension in those with mental health problems, and to what extent (based on previous literature)? Would you expect the usual activities dimension to be the one showing greatest discrimination in respondents reporting restrictions on the GALI?

Response: Following the reviewer's suggestion, authors have clarified the validity hypothesis at dimension level, and added education and hospitalization previous 12 months as variables for contrasting known groups (see new Table 2). Moreover, these results have been adjusted by age and gender as suggested in Comment 12.

- **Data analysis, page 5:** "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 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.”

Results, paragraphs 5 and 6, page 7: “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 less than 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 to 63.16), probable psychiatric disorder (79.09 vs 63.78), strong limitations (79.65 to 48.66), a higher number of days of restriction (77.46 to 55.48) or confinement to bed (76.53 to 47.38), and with hospitalization in the previous 12 months (76.30 vs 67.14).”

Comment 11: *Discussion, first line:* It is not the case that this is the first study to evaluate the construct validity of the Spanish 5L. See Martín-Fernández J et al, 2017. Likewise, it doesn't appear to be the first study to report population norms from this survey (see: Garcia-Gordillo MA, Adsuar JC, Olivares PR. Normative values of EQ-5D-5L: in a Spanish representative population sample from Spanish Health Survey, 2011. Qual Life Res. 2016 May;25(5):1313-21.).

Response: Authors agree with the reviewer that the sentence could be confusing, so it has been modified to be clearer.

- **Discussion, first line, page 8:** “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 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.”

Comment 12: *Discussion, para 2:* I don't think the CHIS results and those in this study are necessarily very similar, e.g. almost double the impact for psychiatric disorders in the CHIS. It's not appropriate to use the MID of 0.07 as a reference here. That MID was derived by averaging across a number of different studies, mostly or all performed in patients, using the 3L and the UK 3L value set. Also, when comparing across groups, e.g. based on number of chronic conditions or possible psychiatric disorders, was any adjustment made to account for possible differences in other characteristics between groups, e.g. by age and/or gender?

Response: Thank you for reviewer's remark regarding differences on the group of probable psychiatric disorder between CHIS and NHS. Following the reviewer's comments comparisons across groups have been adjusted by age and gender (see new Table 2). After looking for MID values of EQ-5D-5L, we selected the 0.061 (+/-0.008) estimated by McClure et al. using algorithms from 6 countries including Spain.

- **Discussion, end of paragraph 2, page 8:** “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].”

Comment 13: *Discussion, para 3:* This paragraph brings little or nothing to the paper. It is very obvious, and hypothetical in the case of the breast cancer patients.

Response: Authors thank the reviewer for this comment; this part has been deleted and the paragraph has been modified following the other reviewer's suggestion.

Comment 14: *Discussion, para 5:* The comparisons with earlier studies in Spain using the 3L are interesting.

- Was the sampling approach and size similar in the earlier studies? More background on that would help interpret the comparisons.
- It would have been interesting to see some comparisons with similar 5L data from other countries as well.
- I don't think lack of sensitivity testing is a limitation here. I'd be more interested to know, for example, whether there were any issues in the generation of Spanish 5L preference weights that the reader should be aware of which might affect the Index values. In other words, are there any factors that might affect the reliability of those values?

Response: Authors thank the reviewer for these suggestions on these different issues.

- The Spanish study with the EQ-5D-3L has been briefly described.
 - Comparison with similar 5L data form other countries has been extended.
 - The limitation about lack of sensitivity testing has been removed, and as there is no reason for suspecting of any factor that might affect the reliability of this study's results, no limitation regarding this issue has been included.
- **Discussion, paragraph 5 and 6, page 9:** "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% vs 86.3% in mobility; 92.1% vs 95.9% in self-care; 86.3% vs 88.3% in usual activities; 71.7% vs 77.1% in pain/discomfort; and 83.6% vs 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 (5,473 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-5Lindex 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%)."

Comment 15: *Conclusion:* 'use of the EQ-5D-5L is recommended'. In what types of studies and settings, based on these results?

Response: Authors appreciate the reviewer's remark, and have modified the conclusion.

- **Discussion, last paragraph, page 11:** "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.

Comment 16: The English is quite poor in places and makes the paper difficult to interpret at times. It should be carefully revised.

Response: Based on the reviewer's comment, an English native speaker has reviewed the last version of the manuscript.

REVIEWER 2

This study involved two major research aims, the first one is to examine the construct validity of EQ-5D-5L, the second one is to provide the population norm for the Spanish population. The major shortcoming of this study is the population norm for the Spanish population using the same data has been published in QoL in 2016. The author argued without adjusting for the demographic weights, the previously published norm could offer biased estimations. Therefore it is important for the author to provide a strong motivation. What kind error could the first population norm cause? Why need a second population norm? Of course you used the newly developed Spanish value set to estimate the index value, but still, you need to provide a strong argument for providing another set of norm values.

Response: Authors agree with the reviewer's observation, and have modified the last paragraph of the Introduction (kindly see also comment 9 of the other reviewer).

- **Introduction, last paragraph, page 2:** "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 its 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."

Comment 1: 'EQ-5D-5L' is a better term for the 'EQ-5D new 5-level version', you can explain what is EQ-5D-5L first, then keep using just EQ-5D-5L. Also for the title, use only EQ-5D-5L.

Response: Following reviewer's suggestion; authors have used just EQ-5D-5L, also for the title.

Comment 2: In the abstract, simply say the EQ-5D-5L index value is computed by using the Spanish value set. You did not explain either time trade off or discrete choice in this manuscript, so drop them.

Response: Authors appreciate the reviewer's comment, and have removed this statement from the abstract.

Comment 3: Redo the citation format in the text. To list the reference in the text, it should be, for example, 'mortality [1,2].' Instead of 'mortality. [1,2]'

Response: Following the reviewer's observation, we have redone the citation format.

Comment 4: Briefly describe what 'levels' are for 3L and 5L.

Response: Authors thank the comments on description of both 3L and 5L made by the reviewers. Following the suggestion of the other reviewer, a short paragraph describing the EQ-5D-3L has been included in the introduction. Therefore, both versions' levels labels are described in the new version of the manuscript.

- **Introduction, paragraph 3, page 1:** "The EQ-5D-3L is a brief [...] with three levels of severity in each dimension (no problems, some problems, and extreme problems)."
- **Methods, paragraph 3, page 3:** "[...]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 [...]"

Comment 5: For the chronic conditions, why 4 categories? The author should explain here the reason for this kind of decision.

Response: Authors thank the reviewer's comment. To be clearer, the original variable has been incorporated in Table 1. The categories were decided in order to have groups with a balanced number of individuals. However, in the first version of the manuscript this decision was made based on the unweighted percentages, and in this reviewed version the categories have been changed to have homogeneous groups based on weighted percentages. Methods, Results and Table 2 have been modified accordingly.

- **Methods, paragraph 6, page 4:** "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."
- **Kindly see also Comment 10 of the other reviewer.**

Comment 6: In table 2, a sample size for each subgroup should be listed.

Response: Following the comment made by the reviewer, the sample size of each subgroup has been incorporated in Table 2.

Comment 7: For the reference norm, you could shortly listed some key values for the readers, for example, the index value for men is XXX..... and the VAS score for women is XX, for detailed norms, check the annex. You don't expect the reader who wants to cite your paper to look for simple values from the annex.

Response: Following both reviewers' suggestions regarding the example given to use population reference norms, the paragraph has been modified.

Discussion, paragraph 3, page 8: "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).”

Comment 8: In the annex, in all titles for all tables, you should state 'by age groups'.

Response: Following this suggestion, the statement “by age groups” has been incorporated at titles for all tables of the annex.

Comment 9: The second paragraph of the discussion included some results, should be moved to the result section.

Response: Authors agree that no results are recommended in the discussion section. However, we used these figures in the discussion just in order to facilitate the comparison between our results and the CHIS survey.

Comment 10: In the discussion, the reduced percentage of health problems between 3L and 5L could also be, e.g. the population health improved over the years. You used two different instruments at two different time points for two different samples; there should be more than one explanation for this reduced percentage

Response: Authors appreciate the reviewer’s comment, and we have modified the paragraph accordingly.

- ***Discussion, paragraph 5, page 9:*** “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% vs 86.3% in mobility; 92.1% vs 95.9% in self-care; 86.3% vs 88.3% in usual activities; 71.7% vs 77.1% in pain/discomfort; and 83.6% vs 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 (5,473 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. “

Comment 11: In the discussion, the author claimed the norm without adjusting the weights could bias the results, but in the end of that paragraph, the author claimed that the results are similar. To declare such difference or no difference, it is better to include statistical test. This should be placed in the introduction of this study, give an example of how unweighted norm could generate biased results.

Response: Authors agree, and to solve this, the last paragraph of the introduction has been changed (kindly see the response to the reviewer’s initial general comment above), and also 95% CI has been incorporated when comparing (in the discussion) the already published norms with the ones calculated using the country-specific value set, sampling weights and methods for complex sample survey designs (to be able to calculate the SE).

- **Discussion, paragraph 9, page 10:** “The previously published reference norms for the EQ-5D-5L dimensions [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] vs 0.868 (95%CI 0.860 – 0.876) in the group that was 60-69 years old; 0.780 [26] vs 0.794 (95%CI 0.785 – 0.803) in the 70-79 years old group; 0.624 [26] vs 0.658 (95%CI 0.647 – 0.670) in the group aged 80-89; and 0.418 [26] vs 0.523 (95%CI 0.496 – 0.549) in the group of 90 or more years old.”

Comment 12: According to the text, there are hardly difference between 3L and 5L. You could move the explanations to earlier part, where there are large differences for percentages of health problems.

Response: Authors thank the reviewer for this advice, and have re-organized the discussion accordingly. In fact though, following the previous suggestion, also paragraph 9 of the discussion has been modified (kindly see the previous response).

VALIDITY OF THE EQ-5D-5L AND REFERENCE NORMS FOR THE SPANISH POPULATION

Authors: Gimena Hernandez, MPH (1,2,3), Olatz Garin, MPH, PhD (1,3,4), Yolanda Pardo, PhD (1,3), Gemma Vilagut, PhD (1,3), Àngels Pont, BSC (1,3), Mónica Suárez (5), Montse Neira (5), Luís Rajmil, PhD (1), Iñigo Gorostiza (6,7), Yolanda Ramallo (7,8), Juan Cabases (9), Jordi Alonso, PhD (1,3,4), Montse Ferrer, MPH, PhD (1,2,3).

Authors' affiliations:

1. Health Services Research Group, IMIM (Hospital del Mar Medical Research Institute). Barcelona, Spain.
2. Departament of Paediatrics, Obstetrics and Gynaecology and Preventive Medicine, Universitat Autònoma de Barcelona. Barcelona, Spain.
3. Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Spain.
4. Pompeu Fabra University UPF. Barcelona, Spain.
5. Subdirección General de Información Sanitaria e Innovación. Spanish Ministry of Health, Social Services and Equality. Madrid, Spain.
6. Hospital Universitario Basurto. 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.

Authors' e-mails:

gimenahernandez@gmail.com; ogarin@imim.es; ypardo@imim.es; gvilagut@imim.es; apont@imim.es; msuarez@msssi.es; mneira@msssi.es; lrajmil@imim.es; inigoramon.gorostizahormaeche@osakidetza.eus; yolanda.ramallofarina@sescs.es; jmcabases@unavarra.es; jalonso@imim.es; mferrer@imim.es

Corresponding authors: Montse Ferrer mferrer@imim.es & Olatz Garin ogarin@imim.es
Health Services Research Group, IMIM (Hospital del Mar Medical Research Institute).
Doctor Aiguader, 88 - 08003 Barcelona, Spain

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

Ethical Statements: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.
Informed consent was obtained from all individual participants included in the study.

Keywords: EuroQol; EQ-5D-5L; Health Related Quality of Life; health status; utilities; questionnaires; reference values; validity.

Running title: EQ-5D-5L Spanish Population Norms

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 to 0.79), probable psychiatric disorder (0.94 vs 0.80), strong limitations (0.96 to 0.46), higher number of days of restriction (0.93 to 0.64) or confinement to bed (0.92 to 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.

[Click here to view linked References](#)

INTRODUCTION

Patient-Reported Outcomes (PRO) 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 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 3,125 (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 its 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 2,000 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 analog 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, by 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 one week. The self-perceived limitation question, known as the “Global Activity Limitation Indicator” (GALI), was: *For at least the last six 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 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 45.5% 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% vs. 13.4% in the 25-34 year-old group and 34.6% vs 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.

-----TABLE 1-----

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 1-----

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.

-----FIGURE 2-----

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 less than 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 to 63.16), probable psychiatric disorder (79.09 vs 63.78), strong limitations (79.65 to 48.66), a higher number of days of restriction (77.46 to 55.48) or confinement to bed (76.53 to 47.38), and with hospitalization in the previous 12 months (76.30 vs 67.14).

-----TABLE 2-----

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 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 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 0.26 (0.31 in the CHIS); and the difference between no days and more than one 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% vs 86.3% in mobility; 92.1% vs 95.9% in self-care; 86.3% vs 88.3% in usual activities; 71.7% vs 77.1% in pain/discomfort; and 83.6% vs 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 (5,473 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] vs 0.868 (95%CI 0.860 – 0.876) in the group that was 60-69 years old; 0.780 [26] vs 0.794 (95%CI 0.785 – 0.803) in the 70-79 years old group; 0.624 [26] vs 0.658 (95%CI 0.647 – 0.670) in the group aged 80-89; and 0.418 [26] vs 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.

Declarations

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.

Availability of data and materials: The Spanish National Health Survey is an official statistic included in the National Statistical Plan. The 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.

Competing interests: All authors declare that they have no competing interests.

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

Authors' 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, 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.

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

Transparency: The lead authors (GH, OG, and MF) affirm that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

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

		n (Raw%)	Weighted %	Standard error
Gender				
	Men	9412 (45.7%)	48.6%	0.412
	Women	11175 (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)	10979 (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	15779 (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	15896 (77.2%)	79.9%	0.383
Number of days with restriction, previous 2 weeks	0	18055 (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	19384 (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	18650 (90.7%)	91.3%	0.237
	Yes	1910 (9.3%)	8.7%	0.237

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. Percentages and means are adjusted by age and gender.

		% of individuals reporting problems					Health index	VAS score
Unweighted n (weighted %)		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]
	P 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]
	P 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]
	P 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]
	P 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]
	P 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]
	P 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]	0.917 [0.914 - 0.919]
	Yes	0.49%	0.10%	0.56%	4.15%	5.36%	0.81 [0.81 - 0.82]	0.814 [0.807 - 0.821]
	P value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

References

1. Mishoe, S.C., & Maclean, J.R. (2001). Assessment of health-related quality of life. *Respir Care*, 46, 1236-1257.
2. 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). *Contemp Clin Trials*, 43, 1-9.
3. Brooks, R., & The EuroQol Group. (1996). EuroQol: the current state of play. *Health Policy*, 37, 53-72.
4. 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 Health*, 9, 262-271.
5. Hinz, A., Klaiberg, A., Brahler, E., & Konig, H.H. (2006). [The Quality of Life Questionnaire EQ-5D: modelling and norm values for the general population]. *Psychother Psychosom Med Psychol*, 56, 42-48.
6. 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. *Qual Life Res*, 26, 1-23.
7. Fransen, M., & Edmonds, J. (1999). Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. *Rheumatology (Oxford)*, 38, 807-813.
8. 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. *Qual Life Res*, 16, 509-522.
9. 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 Health*, 11, 275-284.
10. 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 Qual Life Outcomes*, 13, 171-
11. 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. *Qual Life Res*, 22, 1717-1727.

12. 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. *Qual Life Res*, 21, 1065-1073.
13. 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. *Med Care*, 45, 259-263.
14. 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. *Qual Life Res*, 23, 2355-2363.
15. 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. *Qual Life Res*, 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. *Qual Life Res*, 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. *Qual Life Res*, 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*,
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]. *Med Clin (Barc)*, 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]. *Med Clin (Barc)*, 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. *Qual Life Res*, 23, 791-803.
23. Schmidt, S., & Pardo, Y. (2014). Normative Data. In Michalos, A. C. (Eds.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 4375-4379). Dordrecht, Netherlands: Springer.
24. Szende, A., Janssen, B., & Cabases, J. (2014). *Self-Reported Population Health: An International Perspective based on EQ-5D*. SpringerOpen.
25. König, H.H., Bernert, S., Angermeyer, M.C., Matschinger, H., Martinez, 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. *Qual Life Res*, 25, 1313-1321.
27. Ramos-Goñi JM, Craig BM, Oppe M, Ramallo-Fariña Y, Pinto-Prades JL, Luo N et al. (2016). How to Handle Data Quality Issues in EQ-5D-5L Valuation Studies. The Spanish Case. *Value Health*. Nov; 19(7). 19, A347-A766.
28. Martin-Fernandez, 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. *Gac Sanit*,
29. Ministerio de Sanidad, S.S.e.I., & Instituto Nacional de Estadística (2012). *Encuesta Nacional de Salud de España 2011/12. Accesible en*
<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). *Qual Life Res*, 20, 1727-1736.
31. Oppe, M., Devlin, N.J., van, H.B., Krabbe, P.F., & de, C.F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value 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. *Soc Psychiatry Psychiatr Epidemiol*, 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, O.H., Van der Heyden, J., Perenboom, R., & Jagger, C. (2006). Monitoring population disability: evaluation of a new Global Activity Limitation Indicator (GALI). *Soz Präventivmed*, 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]. *Med Clin (Barc)*, 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 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. *Qual Life Res*, 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 Oncol*, 39, 477-484.
41. Power, M., Quinn, K., & Schmidt, S. (2005). Development of the WHOQOL-old module. *Qual Life Res*, 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. *Qual Life Res*, 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. *J Clin Epidemiol*, 67, 93-99.
44. Golicki, D., & Niewada, M. (2017). EQ-5D-5L Polish population norms. *Arch Med Sci*, 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 EuroQoI 5-D in the Catalan general population: feasibility and construct validity. *Qual Life Res*, 7, 311-322.
47. Puhon, M.A., Ahuja, A., Van Natta, M.L., Ackatz, L.E., & Meinert, C. (2011). Interviewer versus self-administered health-related quality of life questionnaires - does it matter? *Health Qual Life Outcomes*, 9, 30-

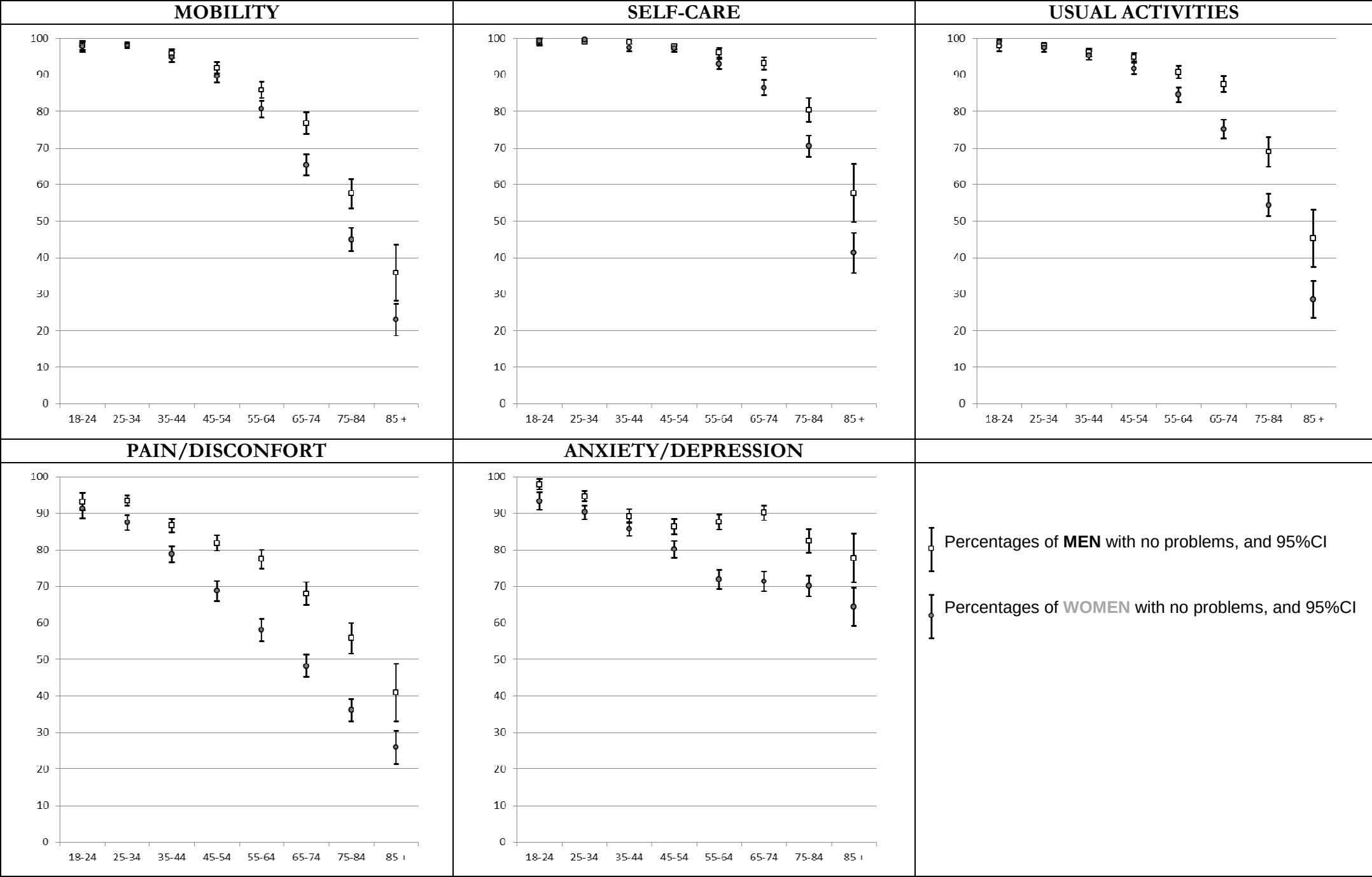
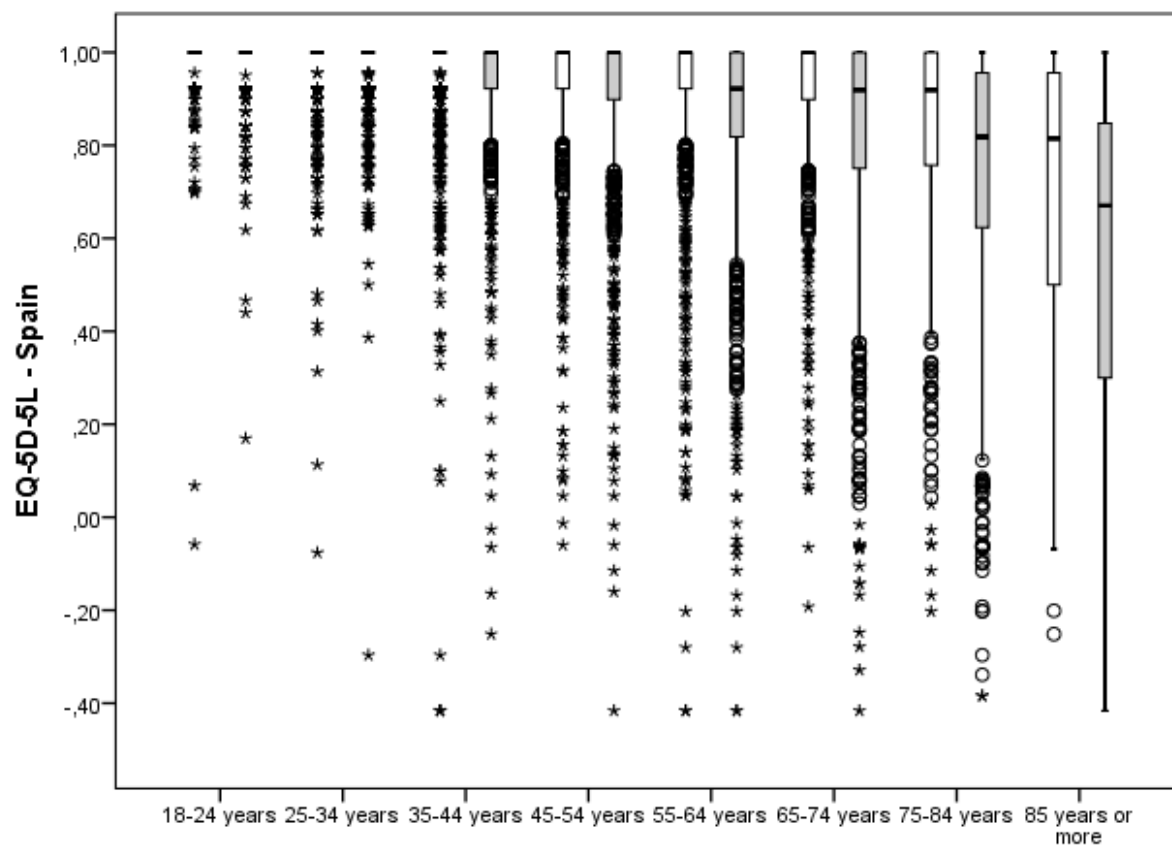


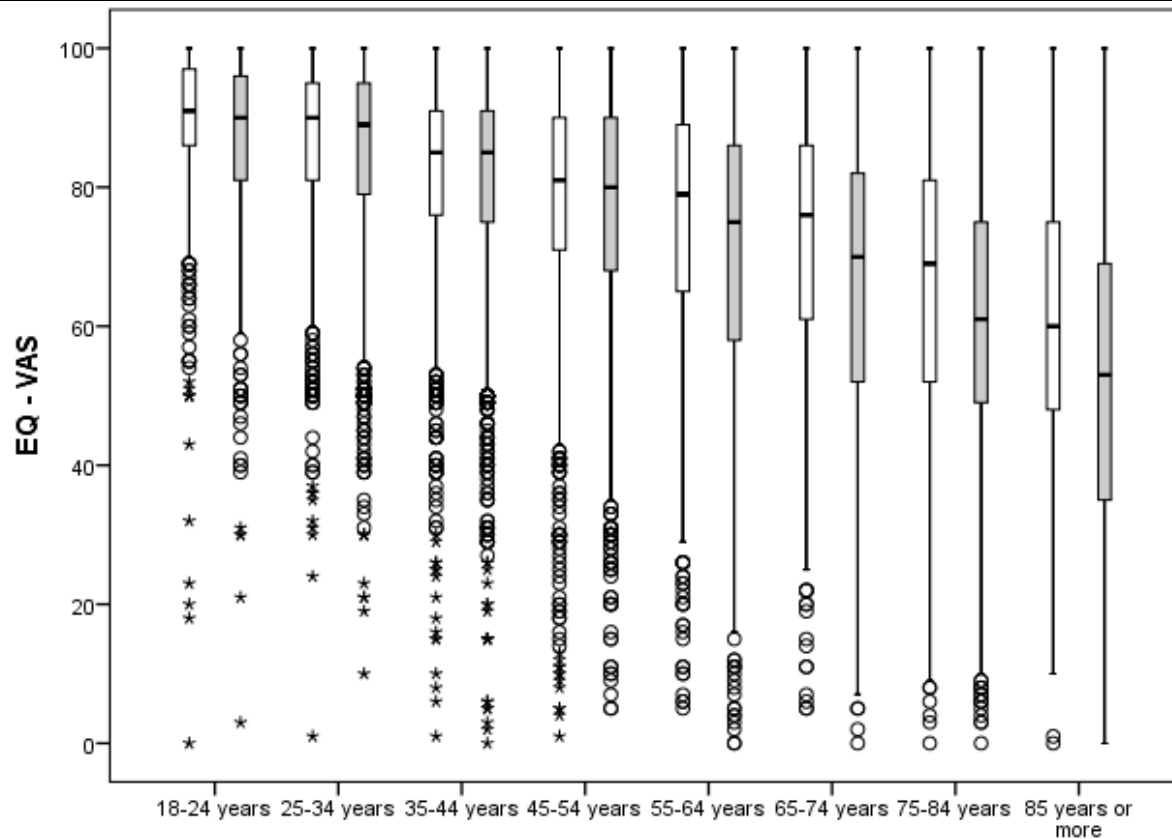
Figure 1. Percentages of individuals reporting no problems by dimension of EQ-5D-5L index regarding age and gender

EQ-5D-5L Index



Age

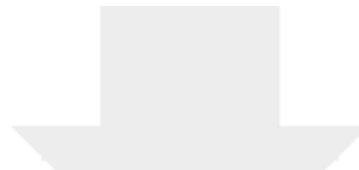
EQ-VAS



Age

Men
Women

Figure 2. EQ-5D-5L index and EQ-VAS box-plot, regarding age and gender.



[Click here to access/download](#)

Electronic supplementary material

Supplemental File_EQ5D5L_Spanish Norms_QUIRE.doc

