Journal of Patient-Reported Outcomes

RESEARCH Open Access

Psychometric properties of the EQ-5D-5L in diabetes mellitus patients in Spain

Amaia Bilbao-González[1,2,3,4*], Marta González-Sáenz de Tejada[2,5], Montse Ferrer[6,7,8], Yolanda Ramallo-Fariña[2,9], Miguel Paja-Fano[10,11], Carlos García-Forero[12], Daniela Mestre[1,2,3] and Iñigo Gorostiza-Hormaetxe[1,2,3]

Abstract

Background The EQ-5D-5L five-dimensional instrument, is one of the most widely used generic preference-based questionnaires to measure health-related quality of life and to estimate utility indices for use in economic evaluation. This study aimed to assess the psychometric properties of the Spanish EQ-5D-5L questionnaire in patients with Diabetes Mellitus (DM) assessing reliability, validity, and item-level properties such as item functioning.

Methodology We included 133 patients with DM who completed the EQ-5D-5L, the Audit on Diabetes-Dependent Quality of Life (ADDQoL), the Hospital Anxiety and Depression Scale (HADS), one question about general health and sociodemographic, and clinical data. The reliability was assessed by Cronbach's alpha, and the item functioning by the item response theory (IRT). Convergent validity was tested using the Spearman correlation coefficient between EQ-5D-5L, ADDQoL, HADS and the general health question. We examined known-groups validity by comparing the EQ-5D-5L scores between subgroups defined by age, gender, BMI, regular physical activity, disease duration, glycemic control by glycosylated blood hemoglobin (HbA1c) (%), type of DM, general health and anxiety and depression level using *t*-test, ANOVA, Wilcoxon or Kruskal-Wallis tests.

Results The reliability was supported with a Cronbach's alpha of 0.78. The IRT results supported the unidimensionality and showed adequate item functioning, except for the anxiety/depression dimension. The item with highest discriminatory power was usual activities dimension, followed by self-care and mobility dimensions. The EQ-5D-5L showed adequate convergent validity, with high correlation with the ADDQoL, HADS and general health. Older age, women, obese, no regular physical activity, ≥ 10 years of disease duration, poor glycemic control, poorer general health and higher anxiety and depression level linked with lower EQ-5D-5L scores.

Conclusions These findings support the adequate psychometric properties of the EQ-5D-5L in patients with DM, supporting its use for clinicians and researchers as an outcome measure and for use in economic evaluation studies.

Keywords Diabetes mellitus, EQ-5D-5L, Utility index, Psychometric properties, Item response theory

*Correspondence:
Amaia Bilbao-González
amaia.bilbaogonzalez@osakidetza.eus
[1]Osakidetza Basque Health Service, Research and Innovation Unit, Basurto University Hospital, Bilbao, Spain
[2]Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), Bilbao, Spain
[3]Biosistemak Institute for Health Systems Research, Bilbao, Spain
[4]Department of Medicine, Faculty of Health Sciences, University of Deusto, Bilbao, Spain
[5]Biobizkaia Health Research Institute, Barakaldo, Spain
[6]Health Services Research Group, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
[7]CIBER de Epidemiología y Salud Pública, CIBERESP, Madrid, Spain
[8]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[9]Canary Islands Health Research Institute Foundation, Tenerife, Spain
[10]Osakidetza Basque Health Service, Department of Endocrinology, Basurto University Hospital, Bilbao, Spain
[11]Faculty of Medicine, University of the Basque Country UPV/EHU, Leioa, Spain
[12]School of Medicine, Universitat Internacional de Catalunya, Sant Cugat del Vallès, Barcelona, Spain

Background

Diabetes mellitus (DM) is one of the most important public health problems worldwide due to its high prevalence, morbidity, influence on patients' health related quality of life (HRQOL) and impact on the healthcare system [1]. In Spain, the age- and sex-adjusted prevalence of DM is 13.8%, and the prevalence of undiagnosed diabetes is 6% [2]. In the long term, diabetes is associated with vascular complications, episodes of hypoglycemia, lifestyle changes and fear of long-term consequences, which may impair HRQOL [3]. This leads to a higher demand for medical and patient care and higher health care costs per patient. In Spain, the total cost of caring for patients with type 2 diabetes was around 2,500 euros per patient per year in 2010 [4].

HRQOL can be defined as a subjective and multidimensional concept that comprises the physical, psychological and social functioning [5]. In developed countries where life expectancy has been steadily increasing, HRQOL outcomes are essential for describing health. It is recommended that HRQOL be measured using specific and generic questionnaires [6]. Generic measures can be preference-based, allowing utilities to be estimated for use in cost-effectiveness studies, and they enable comparisons with the general populations [7, 8].

Cost-effectiveness evidence has become more important than ever for decision-makers at various levels [9]. In these studies, the most commonly used technique has been to estimate the incremental cost per quality-adjusted life years (QALY) [10] gained with new health technologies. The QALY combines life years and quality of life and allows a broad comparison between different treatment strategies, patient populations, and clinical settings. The parameters needed to calculate QALY are the utility for a given health state, and the amount of time spent in that state. The utilities, usually measured with generic questionnaires, represent the strength of a person's preferences for health states, with values ranging from 0 (dead) to 1 (full health), although scores can also take negative values for states worse than death [10].

As a generic preference-based measure for describing and valuing HRQOL, the EQ-5D is one of the most widely used [11–13], developed by the EuroQol Research Foundation [14]. The standard format of the EQ-5D consists of five dimensions of health, each with three response options (EQ-5D-3L). There is extensive literature to support the validity and reliability of this 3-level version in many conditions and populations [14]. Nonetheless, its restricted ability to discriminate between different levels of health [15, 16], presence of ceiling effects and skewed distribution have long been recognized [17, 18]. Seeking to overcome these shortcomings, the EQ-5D-5L was developed [12], expanding the range of responses to each dimension from three to five levels. Previous studies have indicated that the 5L version improves upon the 3L version in measurement properties with smaller ceiling effects and a better ability to discriminate between different levels of health [19–22].

Due to the extensive use of EQ-5D-5L in Spanish patients with DM for HRQoL assessing in research, clinical practice and health economics [23–25], it must be ascertained that EQ-5D-5L is a valid and reliable tool specifically in this population. In other words, it is a requirement proving that EQ-5D-5L is a useful tool that accurately and consistently measures the HRQOL in Spanish DM patients. Several studies have evaluated the psychometric properties of the EQ-5D-5L in patients with DM in other cultures and languages, which accumulates evidence about its validity in those cohorts [26–32]. Nonetheless, to the best of our knowledge, it has yet to be validated in this population in Spain. Moreover, no studies have explored its item-level properties assessing item functioning in addition to performing classical validation.

The aim of the current study was to carry out a comprehensive study of the psychometric properties of the Spanish version of the EQ-5D-5L questionnaire in patients with DM investigating its reliability and validity, as well as the item-level properties assessing item functioning.

Methods

Study population

We included patients from two hospitals in the Basque Country between July 2013 to December 2015 with an ICD-9-CM diagnosis code of 250. All patients were given a letter informing them about the study and inviting them to participate during a DM follow-up appointment. Patients were excluded if they had a terminal illness, or psychiatric and/or sensory disturbances that could prevent them from answering the questionnaires, as well as if they did not provide informed consent. The institutional review board of the hospitals approved the study (identification number PI2012156-PI12-01473) on 4rd February 2013.

Measurements

In the follow-up visit, patients who accepted participating were asked to complete the EQ-5D-5L [12], the Audit on Diabetes-Dependent Quality of Life (ADDQoL) [33], the Hospital Anxiety and Depression Scale (HADS) [34] and one question about general health from the 36-item Short Form Health Survey (SF-36) [35].

In addition, a clinician gathered other data on patients in a case report form created adhoc with information obtained for and during the visit that included: -sociodemographic characteristics, namely, age, gender, employment status, level of education; and–clinical characteristics, namely, body mass index (BMI), disease duration, smoking status, regular physical activity, type

of DM, insulin treatment, self-monitoring of blood glucose and glycosylated hemoglobin (HbA1c) (%) like a measure of the glycemic control as recommended by the American [1] and Spanish [36] Diabetes Association. Trained personnel collected comorbidities measured by the Charlson Comorbidity Index [37] from the patients' medical records.

HRQOL questionnaires

The EQ-5D-5L [12] contains items concerning five dimensions, mobility, self-care, usual activities, pain/discomfort and anxiety/depression, rated on a five-point scale from 1 (no problems) to 5 (unable to/extreme problems). The combined dimensions describe 5^5=3,125 theoretically possible states of health that can be converted into a weighted index score. For the Spanish value-set, the index score ranges from − 0.416 to 1, with a higher score indicating better HRQOL [38]. Additionally, the questionnaire includes a visual analogue scale (EQ-VAS) on which individuals rate their own health today on a scale from 0 (worst imaginable health) to 100 (best imaginable health).

The ADDQoL [33] is a specific tool for DM consisting of 21 items, 19 of which refer to specific aspects of life (such as social life and work life) scored on a 5-point scale. The impact of diabetes on each of these 19 specific aspects of life is weighted according to its importance to the patient's quality of life, yielding an average weighted impact (WI) score. This score ranges from − 9 (maximum negative impact of diabetes) to + 3 (maximum positive impact of diabetes). The other two items are summary items: one measures the impact of diabetes on quality of life (-3 indicating maximum negative impact to + 1 as maximum positive impact) and the other measures the current quality of life (-3 indicating extremely poor and + 3 excellent). We used the already validated Spanish version [39, 40].

The HADS is a self-report measure designed specifically for people with a physical illness [34], which is divided into two subscales, anxiety (seven items) and depression (seven items). Each of the 14 items is rated on a 4-point Likert scale and the total score for each subscale ranges from 0 to 21, with a higher score indicating a higher level of mood disorder. A score ≤ 7 corresponds to "no depression/anxiety", a score from 8 to 10 corresponds to "minor depression/anxiety", and a score ≥ 11 indicates "moderate to severe depression/anxiety". The Spanish version used has been validated [41, 42].

We used the question about general health "In general would you say your health is?" from the SF-36 questionnaire [35], which has five response options as follows: 1 (poor), 2 (fair), 3 (good), 4 (very good) and 5 (excellent).

Statistical analysis

Data from patients that responded to the whole EQ-5D-5L were used for the analyses. For the descriptive analyses, means and standard deviations (SDs), or frequencies and percentages were used. To describe the EQ-5D-5L scores, we also calculated the median, range and floor and ceiling effects, which should be small (< 15%) [43]. The missing data was also examined.

Reliability

We assessed internal consistency using Cronbach´s alpha [44]. A coefficient over 0.70 was considered acceptable [45]. Further, the inter-item correlation and the item-scale correlation correcting for overlap were assessed by calculating Spearman's correlation coefficient.

Item functioning

The item functioning of the EQ-5D-5L was studied using item response theory (IRT), which unlike classical test theory, focuses on item-level rather than scale-level properties [46]. A requirement of IRT is that the scales must be unidimensional. Therefore, we first tested for unidimensionality by exploratory (EFA) and confirmatory (CFA) factor analysis. Regarding EFA, an item was considered to contribute to the factor if the factor loading was ≥ 0.40 [47], and a ratio of the eigenvalues of the first and second unrotated components of at least 3:1 was taken as evidence of unidimensionality [48, 49]. CFA was also performed to confirm that the five items load on a single factor. The maximum likelihood estimator was used, and the following indices were calculated [50, 51]: the root mean square error of approximation (RMSEA), for which a value < 0.08 was considered acceptable; and the non-normed fit index (NNFI) and Comparative Fit Index (CFI), for both of which values > 0.90 were considered acceptable.

Then, having confirmed the unidimensionality, we used an IRT model, specifically, the graded response model (GRM), which is useful in the evaluation of polytomous items [52]. In the GRM, the two common item parameters are the item slope (α), which captures the ability of an item to discriminate between people with different levels of the latent trait, and the item difficulty (β), also called the threshold, which indicates the point on the scale of the latent trait where a person has 0.5 probability of responding positively to an item category [46]. Therefore, item responses are conceptualized in terms of the slope parameter (α), and a series of k–1 category thresholds, where k is the number of item response options. In this case, each item has five response options, but due to the low frequency of response options 4 and 5, they were merged. Therefore, each item is defined by α and three β thresholds. The slope α is comparable to the factor loading of CFA, with higher values indicating that an item

has greater discriminatory power while the β thresholds are indicative of the spacing of item responses along the trait dimension [49]. That is, the range of the latent trait that the estimated β parameters encompass indicates the adequacy of an item to represent low, medium, or high levels of a trait. We use a Bayesian estimator [53] in the GRM models.

This GRM was applied to the EQ-5D-5L items and the slope (α) and thresholds (β), as well as standard errors (*SEs*), were calculated [54]. These parameters were used to plot a category response curve (CRC) for each item, which represents the probability of a positive response to each item's response option as a function of the latent trait [48]. In addition, item information curves (IICs) were generated [49], which identify the position on a given trait spectrum at which the item provides the most information. In these graphs, the x-axis represents the latent trait, with a scale standardized to have a mean of 0 and *SD* of 1. The fit of the GRM model was assessed by: (a) determining the residuals of the model, by comparing the observed and expected proportions of each response category of each item; and (b) calculating a likelihood ratio χ^2 for each item, to assess the goodness-of-fit between the expected and observed frequencies [46, 48, 49].

Convergent and discriminant validity

The convergent and discriminant validity was assessed by analyzing the relationship of the EQ-5D-5L with the ADDQoL, HADS, and the general health scores, by using Spearman's correlation coefficient (ρ). We considered a coefficient $|\rho| \geq 0.5$ high; between 0.3 and 0.5, moderate; and between 0.1 and 0.3, small [55]. We hypothesized that the ADDQoL present HRQOL item would be highly correlated (convergent validity) with all EQ-5D-5L scores. Further, we hypothesized that the correlation of the EQ-5D-5L with the influence of diabetes on HRQOL and the average WI would be moderate or low because they are less directly related to different aspects of patients' HRQOL (discriminant validity). Because four out of the five dimensions of the EQ-5D-5L (mobility, self-care, usual activities and pain/discomfort) are closely related to physical aspects, we hypothesized that the correlation of these four dimensions, the EQ-5D-5L index and the EQ-VAS with the HADS domains would be moderate or low (discriminant validity), and high with the anxiety/depression dimension of the EQ-5D-5L (convergent validity). Further, we hypothesized that there would be a high correlation between the SF-36 general health question and the EQ-5D-5L index and EQ-VAS (convergent validity), and a moderate correlation with each EQ-5D-5L specific dimension (discriminant validity).

Known-groups validity

Known-groups validity of the EQ-5D-5L utility index and EQ-VAS was evaluated by comparing subgroups of patients known to differ in health status [28, 29, 31, 56, 57]: age (age ≤ 49; 50 ≤ age ≤ 59; 60 ≤ age ≤ 69; age ≥ 70 years), gender, BMI (< 30 non-obese vs. ≥ 30 obese), regular physical activity (yes/no), disease duration (< 10 vs. ≥ 10 years), glycemic control (poor HbA1c ≥ 7% vs. optimal HbA1c < 7%), type of diabetes (1 vs. 2), SF-36 general health question (poor; fair; good; very good + excellent), and HADS anxiety and depression level (≤ 7 vs. 8–10 vs. ≥11). We hypothesized that both scores would be lower in patients who were older, women, obese, and smokers, did not do regular physical activity, have a disease duration of ≥ 10 years, poor glycemic control (≥ 7%), diabetes type 2, and had a poorer SF-36 "health today rating" and higher depression or anxiety level. These comparisons were made using a t-test or analysis of variance, with Scheffe's test for multiple comparisons, or the non-parametric Wilcoxon or Kruskal-Wallis tests.

All effects were considered statistically significant at $p < 0.05$. The statistical analyses were performed with SAS 9.4 for Windows (SAS Institute, Cary, NC) and Mplus (version 6.1) [58].

Results

We recruited 145 patients with DM who met the selection criteria and agreed to participate. Of this sample, 133 patients (91.72%) completed all the questionnaires without missing information. The sociodemographic and clinical characteristics of the sample are summarized in Table 1. The mean age was 62.74 years (SD, 12.18), 47.37% were women, 45.38% were obese and 70.68% had a disease duration of ≥ 10 years.

Table 2 lists the descriptive statistics of the EQ-5D-5L. The percentage of missing values ranged from 5.52 to 8.28%. The percentage of patients who had "no problems" in all EQ-5D-5L dimensions ranged from 45.86% for pain/discomfort to 88.72% for self-care. The EQ-VAS score showed no floor effect and minimal ceiling effect (1.54%). Then EQ-5D-5L utility index showed no floor effect, but a ceiling effect of 22.56%. The distribution of the utility index is shown in Fig. 1.

Reliabililty

Cronbach's alpha was 0.78, being higher than the threshold of 0.70. The inter-item correlations ranged from 0.31 to 0.61, except for the correlation of the anxiety/depression dimension with the rest, being lower (between 0.05 and 0.27). The item-total correlations ranged from 0.21 to 0.68 (Table 3).

Table 1 Characteristics of the study participants

Variables	Diabetes mellitus (*n* = 133)
Sociodemographic variables	
Age (years), mean (SD)	62.74 (12.18)
Age categorized (years), n (%)	
≤49	17 (12.78)
50–59	32 (24.06)
60–69	45 (33.84)
≥70	39 (29.32)
Gender, women, n (%)	63 (47.37)
BMI, n (%)	
Non obese (< 30 kg/m^2)	71 (54.62)
Obese (≥ 30 kg/m^2)	59 (45.38)
Employment status, n (%)	
Active	40 (33.33)
Unemployed	13 (10.83)
Retired	67 (55.83)
Level of education, n (%)	
No studies	3 (2.61)
Primary studies	46 (40.00)
Secondary studies	22 (19.13)
High school diploma	22 (19.13)
Bachelor's degree	22 (19.13)
Clinical variables	
Disease duration (years), mean (SD)	17.78 (10.54)
Disease duration categorized, n (%)	
< 10 years	39 (29.32)
≥ 10 years	94 (70.68)
Smoking status, n (%)	
Non-smoker	60 (46.15)
Ex-smoker	51 (39.23)
Smoker	19 (14.62)
Regular physical activity, n (%)	55 (44.00)
Type of DM, n (%)	
Type 1	24 (18.05)
Type 2	107 (80.45)
Mixed DM	2 (1.50)
Insulin treatment, n (%)	81 (61.36)
Self-monitoring of blood glucose, n (%)	104 (78.79)
Glycemic control, n (%)	
Poor control (HbA1c ≥ 7%)	74 (56.06)
Optimal control (HbA1c < 7%)	58 (43.61)
Charlson Index, mean (SD)	2.38 (1.67)
Charlson Index, n (%)	
1	50 (37.88)
2	38 (28.79)
> 2	44 (33.33)
HRQOL questionnaires	
ADDQoL, mean (SD)	
Present HRQOL	0.80 (0.84)
Diabetes influence on HRQOL	-1.68 (0.88)
Average weighted impact	-2.02 (1.57)
HADS anxiety, mean (SD)	6.06 (3.85)
HADS anxiety level, n (%)	
No (≤ 7)	90 (71.43)
Minor (8–10)	17 (13.49)
Moderate to severe (≥ 11)	19 (15.08)
HADS depression, mean (SD)	4.10 (3.21)
HADS depression level, n (%)	
No (≤ 7)	104 (82.54)
Minor (8–10)	13 (10.32)
Moderate to severe (≥ 11)	9 (7.14)
SF-36 general health today, mean (SD)	2.65 (0.81)
SF-36 general health today, n (%)	
Poor	8 (6.20)
Fair	46 (35.66)
Good	60 (46.51)
Very good	13 (10.08)
Excellent	2 (1.55)

SD, Standard deviation; BMI, Body Mass Index; HbA1c, glycosylated hemoglobin; HRQOL, Health Related Quality of Life; HADS, Hospital Anxiety and Depression Scale (each subscale ranges from 0 to 21, with a higher score indicating a higher level of mood disorder); ADDQoL, Audit on Diabetes-Dependent Quality of Life (Diabetes influence on HRQOL scored from − 3 as maximum negative impact to + 1 as maximum positive impact; the present HRQOL ranges from − 3 (extremely poor) to + 3 (excellent); and the average weighted impact (WI) ranges from − 9 to + 3, the lower the value of the average WI score, the worse the aspect of life); SF-36, 36-item Short-Form Health Survey; the SF-36 general health today ranges from 1 to 5, in which the higher the score, the better general health

Item functioning

Regarding unidimensionality, the EFA (Table 3) yielded factor loadings between 0.36 and 0.91, exceeding the benchmark of 0.40, except for the anxiety/depression item. The percentage of variance explained by the factor was 64.82%. The ratio of the eigenvalues of the first and second unrotated factors was 12.67:1, exceeding the benchmark of 3:1. The results of the CFA showed adequate fit indices (Table 3): (a) the RMSEA was 0.064, less than 0.08; and (b) the CFI and NNFI were 0.989 and 0.973, respectively. Factor loadings ranged from 0.40 to 0.80 and were all significant ($p < 0.001$).

Table 3 lists the slope (α) and threshold (β) parameters estimated from the GRM. Regarding α, all values were acceptable and large, except for the anxiety/depression item. The highest α was obtained for item 3 (usual activities), with a value exceeding 2, followed by items 2 (self-care) and 1 (mobility), with values somewhat lower, but still high (between 1.5 and 2). Item 4 (pain/discomfort) showed a slightly lower discriminatory power ($\alpha = 0.93$), but not negligible. Finally, item 5 (anxiety/depression) had negligible discriminatory power ($\alpha = 0.37$). Regarding β, item 3 (usual activities) covers the largest range of the trait, with values of around 1 *SD* from the mean in β_1 to around 4.5 *SD* in β_3, indicating that this item is more apparent at higher levels of severity. Item 2 (self-care) also covers quite a large range of the trait, but is most apparent at higher levels of severity, with values from 2.5

Table 2 Descriptive statistics of the EQ-5D-5L dimensions ($n = 133$)

EQ-5D-5L	Completion*	EQ-5D-5L dimensions's response options				
	***n* (%) of missing**	**No problems *n* (%)**	**Slight problems *n* (%)**	**Moderate problems *n* (%)**	**Severe problems *n* (%)**	**Unable to / extreme problems *n* (%)**
Mobility	8 (5.5)	73 (54.89)	30 (22.56)	21 (15.79)	9 (6.77)	0 (0)
Self-care	10 (6.9)	118 (88.72)	8 (6.02)	5 (3.76)	2 (1.50)	0 (0)
Usual activities	10 (6.9)	93 (69.92)	20 (15.04)	15 (11.28)	3 (2.26)	2 (1.50)
Pain/discomfort	9 (6.2)	61 (45.86)	42 (31.58)	19 (14.29)	10 (7.52)	1 (0.75)
Anxiety/depression	12 (8.3)	84 (63.16)	31 (23.31)	14 (10.53)	4 (3.01)	0 (0)
		EQ-5D-5L scores				
		Mean (SD)	**Median (IQR)**	**Range (min, max)**	**Floor, %**	**Ceiling, %**
EQ-5D-5L utility index	12 (8.3)	0.823 (0.191)	0.872 (0.756 – 0.922)	(-0.09, 1)	0	22.56
EQ-VAS	7 (4.8)	69.56 (19.12)	70 (60 – 85)	(10, 100)	0	1.54

*The completion n (%) of missing are calculated with the total of 145 patients

SD: Standard deviation; IQR: interquartile range

EQ-5D-5L utility index score range from – 0.416 to 1, with higher score indicating better HRQOL; EQ-VAS scale range from 0 (worst imaginable health) to 100 (best imaginable health)

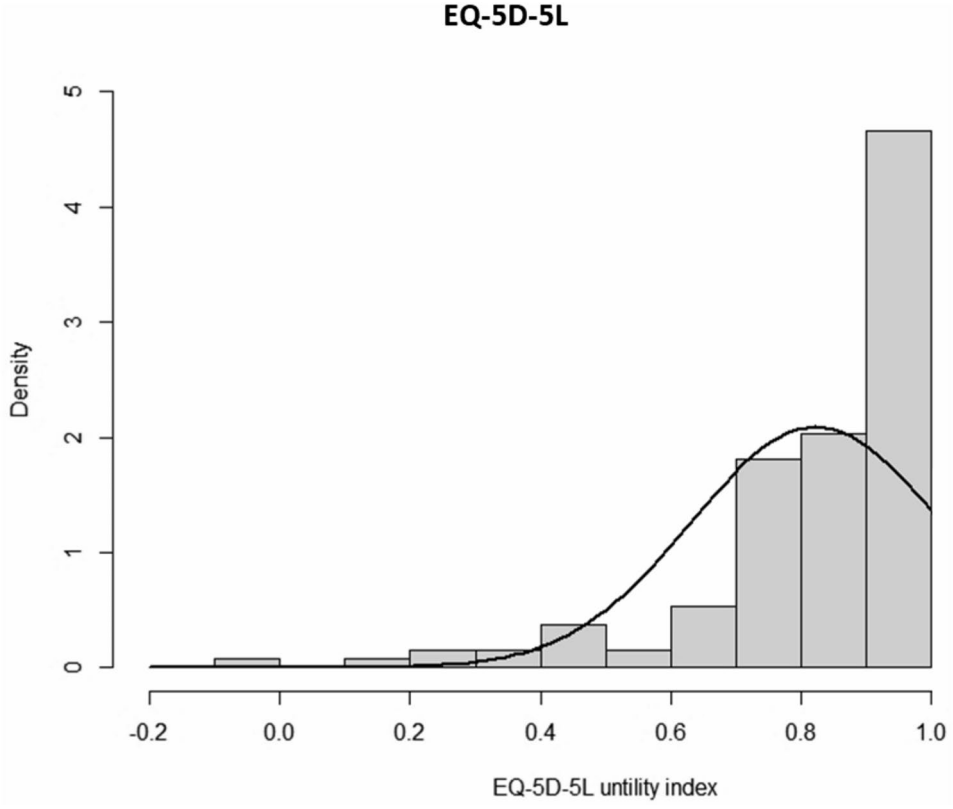

Fig. 1 Distribution of the EQ-5D-5L utility index in the cohort of study. Density is represented in the histogram's bars, while the black line indicates the normal distribution. The EQ-5D-5L utility index ranges from – 0.416 (worse than dead) to 1 (full health)

SD from the mean in β_1 to 5 *SD* in β_3. Item 1 (mobility) covers a similar range of the trait but is most apparent at medium levels of severity. Finally, items 4 (pain/discomfort) and 5 (anxiety/depression) cover a narrower range of the trait, with values of around – 0.25 *SD* to 0.25 *SD* from the mean in β_1, respectively, to around 2 *SD* in β_3, being more apparent at lower levels of severity.

The CRCs for all items showed a shift to higher trait levels as the level of the response increased, but not very notable for item 5 (anxiety/depression) (Fig. 2). Regarding

Table 3 Exploratory and confirmatory factor analysis, item response theory parameter estimates from graded response model for the EQ-5D-5L questionnaire, and item-scale correlation ($n = 133$)

Items	Exploratory factor analysis	Confirmatory factor analysis	Graded response model				Item-scale correlation*
	Factor loading	Factor loading	α	β_1	β_2	β_3	ρ
EQ1. Mobility	0.84	0.79	1.49 (0.24)	0.18 (0.17)	1.28 (0.22)	2.78 (0.37)	0.52
EQ2. Self-care	0.91	0.70	1.78 (0.36)	2.37 (0.37)	3.35 (0.47)	5.04 (0.96)	0.43
EQ3.Usual activities	0.90	0.80	2.27 (0.41)	1.23 (0.34)	2.60 (0.49)	4.67 (0.84)	0.68
EQ4. Pain/discomfort	0.68	0.63	0.93 (0.16)	-0.17 (0.14)	1.03 (0.17)	1.98 (0.21)	0.48
EQ5. Anxiety/depression	0.36	0.40	0.37 (0.13)	0.32 (0.12)	1.15 (0.15)	2.04 (0.23)	0.21
% variance explained	64.82%						
χ^2 (df)		6.15 (4)					
RMSEA (90% CI)		0.064 (0–0.157)					
CFI		0.989					
NNFI		0.973					

df = degrees of freedom; RMSEA = root mean square error of approximation; CI: confidence interval; CFI = comparative fit index; NNFI = Non-normed fit index; α = slope parameter estimate from the graded response model; β_1, β_2, and β_3 = threshold parameters from the graded response model. Standard error estimates for each parameter estimate are listed in parentheses

*Item-scale correlation correcting for overlap calculated by Spearman correlation coefficient

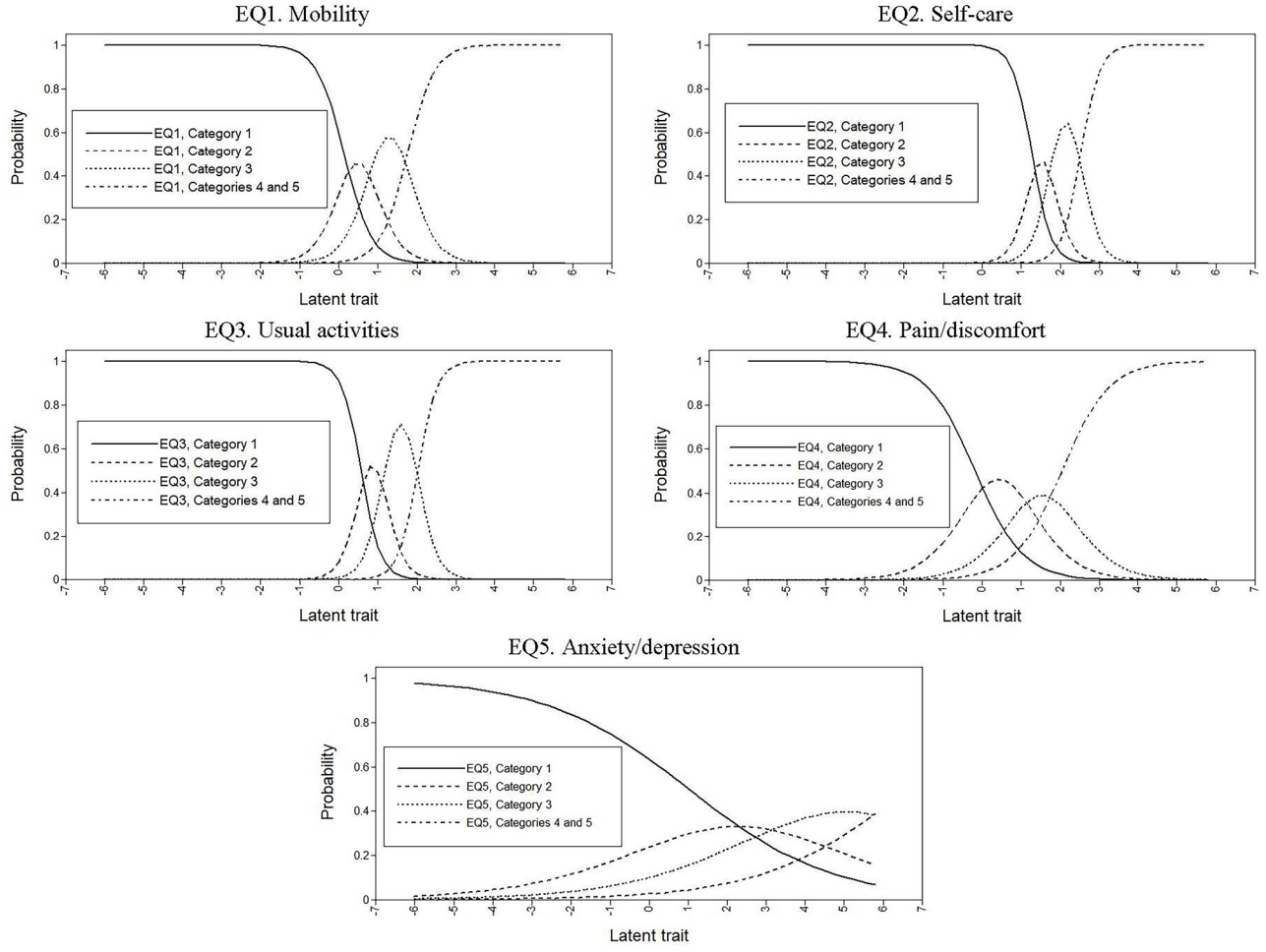

Fig. 2 Category response curves for the EQ-5D-5L items. The x-axis represents the latent trait and the y-axis represents the probability of a positive response to each item's response option. Category 1 refers to "No problems", category 2 to "Slight problems", category 3 to "Moderate problems", category 4 to "Severe problems" and category 5 to "Unable to/Extreme problems"

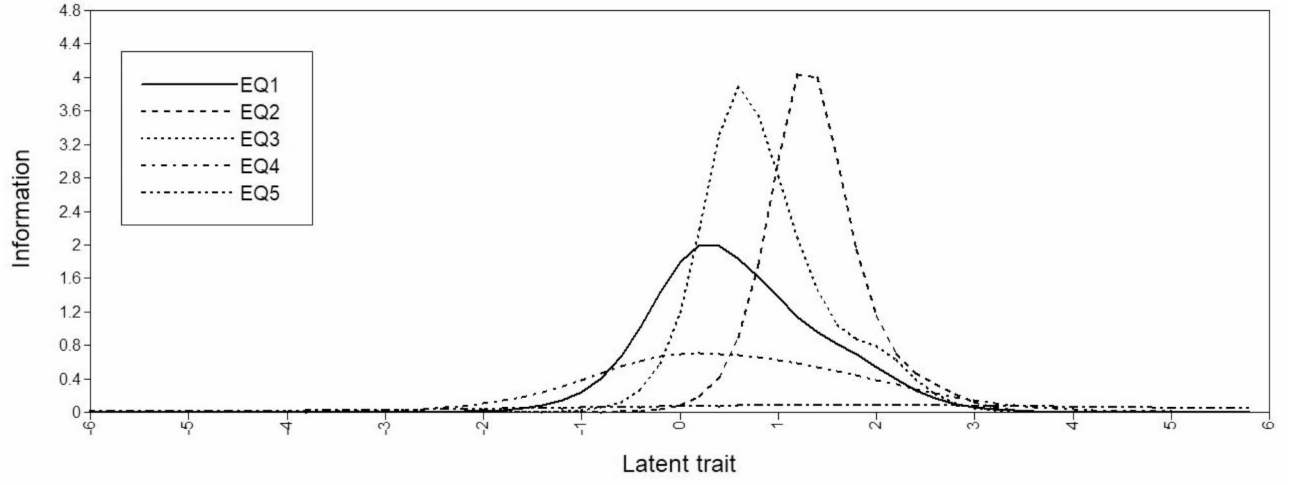

Fig. 3 Q

Table 4 Convergent and divergent validity ($n = 133$)

	Mobility	Self-care	Usual activities	Pain / discomfort	Anxiety / depression	EQ-5D-5L utility index	EQ-VAS
ADDQoL							
Present HRQOL	-0.50	-0.42	-0.58	-0.54	-0.27	0.62	0.54
Diabetes influence on HRQOL	0.08	-0.01	0.01	-0.04	0.08	-0.04	-0.14
Average weighted impact	-0.23	-0.07	-0.14	-0.18	-0.13	0.27	0.10
HADS							
Anxiety	0.04	0.15	0.23	0.34	0.51	-0.37	-0.26
Depression	0.35	0.33	0.47	0.48	0.50	-0.61	-0.46
SF-36 general health today	-0.49	-0.31	-0.40	-0.58	-0.19	0.59	0.58

HADS = Hospital Anxiety and Depression Scale; ADDQoL = Audit on Diabetes-Dependent Quality of Life

Data are given as Spearman's correlation coefficient

The scores for the ADDQoL subscales range from −3 to 3 in present HRQOL item, with higher scores indicating better health status; from −3 to 1 in the diabetes influence on HRQOL item, with higher scores indicating more positive impact; and from −9 to 3, with higher scores indicating more positive impact in diabetes. The scores for the HADS ranges from 0 to 21, with higher score indicating a higher level of mood disorder. Items concerning the five dimensions in EQ-5D-5L (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) are rated from 1 (no problems) to 5 (unable to/extreme problems). The scores for the EQ-5D-5L utility index ranges from −0.416 to 1, with higher score indicating better HRQOL, and EQ-VAS scale range from 0 (worst imaginable health) to 100 (best imaginable health); SF-36, 36-item Short-Form Health Survey; the SF-36 general health today ranges from 1 to 5, in which the higher the score, the better general health

the IICs (Fig. 3), we can see that values for items 2 (self-care) and 3 (usual activities) were relatively high across the trait spectrum compared with the rest, and in particular compared with item 5 (anxiety/depression). Regarding the model fit, the residuals of all categories of all items were not significantly different from 0, and there were no significant differences between the observed and expected likelihoods in any of the five items, demonstrating the good fit.

Convergent and discriminant validity

The ADDQoL present HRQOL dimension showed high correlations with all EQ-5D-5L dimensions, except in the cases of self-care (ρ=-0.42) and anxiety/depression (ρ=-0.27); and a small correlation with the influence of diabetes on HRQOL item and the average WI dimension. Further, we found a high correlation between the two HADS dimensions and the EQ-5D-5L anxiety/depression domain, and between HADS depression dimension and the EQ-5D-5L utility index, and a lower correlation with the other four dimensions and the EQ-VAS. The general health item presented a high correlation with the pain/discomfort dimension, EQ-5D-5L index and EQ-VAS (ρ=-0.58, 0.59 and 0.58, respectively), small correlation with anxiety/depression (ρ=-0.19), and moderate correlation with the rest of dimensions (Table 4).

Known-groups validity

There were significant differences in the EQ-5D-5L index between subgroups defined by age, gender, BMI, regular physical activity, disease duration, HbA1c, general health today rating, and the HADS anxiety and depression level. EQ-5D-5L utility index scores were significantly lower in patients who were older, women, obese, did not do regular physical activity, with disease duration of ≥ 10 years, poor glycemic control, poorer general health status, and

higher anxiety or depression level. In EQ-VAS, there were significant differences only between subgroups defined by age, sex, regular physical activity, disease duration, general health today, and HADS anxiety and depression level (Table 5).

Discussion

The results of this prospective study with a cohort of DM patients seen in two hospitals in the Basque Country (Spain), combining both classical psychometric and IRT approaches, support the validity and reliability of the EQ-5D-5L at item and scale levels. To the best of our knowledge, this is the first study assessing the psychometric properties of the Spanish EQ-5D-5L in patients with DM.

Regarding completion of the EQ-5D-5L, the missing data rate of between 5.52% and 8.28% was quite high [27, 30]. This is likely attributable to the combined length of the questionnaires used (EQ-5D-5L, ADDQoL, and HADS), and hence, completion of the questionnaires may have been a considerable burden for participants.

As stated, the EQ-5D-5L was developed to address the main limitations of the EQ-5D-3L, in particular, the ceiling effect of the utility index [18, 27]. However, in some studies the ceiling effect has been found to remain high when the EQ-5D-5L is used in DM, with ceiling effects greater than 15% [27, 30–32, 59], as in our study (22.56%). This is probably because our respondents were likely to perceive themselves as healthy, consistent with their EQ-VAS median score of 70. Nevertheless, other studies in DM patients have shown small ceiling effects in the 5L utility index [26, 29]. Further, in our sample, the ceiling effect of the EQ-5D-5L at the item level (those responding "no problems") was also high. In any case, considering that items are categorical, despite having five categories instead of three, it is easy for floor or ceiling effects to occur. Similar results have been found in other studies [26–32, 59–63], being the highest ceiling effect in the self-care dimension.

The Cronbach's alpha exceeded the threshold of 0.70 [45], indicating that 5L version maintained good internal consistency. This result agrees with those found by Abedini et al. [28].

The item functioning of the EQ-5D-5L was studied using IRT, and more specifically, the graded response model (GRM), which is a two-parameters IRT model [52]. Given that the factor loading of the anxiety/depression item is much lower than the others and therefore contributes less to the trait, we need to apply a two-parameter IRT model allowing different discriminatory ability to each EQ-5D-5L item. The results of the factor analyses support the unidimensionality and adequacy of applying IRT. The results of the IRT indicated that the data were well fitted by the GRM. The item with the highest discriminatory power was usual activities, followed by

Table 5 Known-groups validity of the EQ-5D-5L

	EQ-5D-5L utility index		EQ-VAS	
	N	Mean (SD)	*N*	Mean (SD)
Age (years)				
≤49[a]	17	0.839 (0.261)	17	74.41 (17.93)
50–59[b]	32	0.901 (0.104)[d]	32	75.13 (15.10)
60–69[c]	45	0.816 (0.172)	43	69.05 (22.92)
≥ 70[d]	39	0.759 (0.213)[b]	38	63.29 (16.41)
p-value		0.018		0.046
Gender				
Men	70	0.880 (0.134)	70	73.63 (16.65)
Women	63	0.759 (0.222)	63	64.82 (20.79)
p-value		0.0002		0.0137
BMI				
Obese (≥ 30 kg/m^2)	59	0.775 (0.227)	57	67.02 (20.59)
Non obese (< 30 kg/m^2)	71	0.870 (0.137)	70	72.11 (17.31)
p-value		0.004		0.132
Regular physical activity				
Yes	55	0.906 (0.102)	53	78.28 (14.31)
No	70	0.761 (0.223)	69	62.68 (20.32)
p-value		< 0.001		< 0.001
Disease duration (years)				
< 10	39	0.896 (0.100)	39	76.67 (16.20)
≥ 10	94	0.792 (0.211)	91	66.52 (19.54)
p-value		0.005		0.004
Glycemic control				
Poor control (HbA1c ≥ 7%)	74	0.781 (0.227)	72	67.28 (19.61)
Optimal control (HbA1c < 7%)	58	0.875 (0.119)	57	72.79 (18.17)
p-value		0.005		0.104
Type of DM				
Type 1	24	0.854 (0.225)	24	73.71 (20.03)
Type 2	107	0.816 (0.185)	104	68.79 (19.02)
p-value		0.150		0.179
Health today (SF-36)				
Poor[a]	8	0.430 (0.353)[bcd]	8	32.5 (14.64)[bcd]
Fair[b]	46	0.752 (0.167)[acd]	45	63.11 (16.66)[acd]
Good[c]	60	0.899 (0.089)[ab]	59	75.24 (12.81)[abd]
Very good + Excellent[d]	15	0.934 (0.120)[ab]	14	87.86 (17.18)[abc]
p-value		< 0.001		< 0.001
HADS anxiety level				
No (≤ 7)[a]	90	0.872 (0.138)[bc]	89	72.74 (17.99)
Minor (8–10)[b]	17	0.755 (0.180)[a]	16	63.75 (20.29)
Moderate to severe (≥ 11)[c]	19	0.684 (0.243)[a]	19	62.58 (19.17)
p-value		< 0.001		0.0376
HADS depression level				
No (≤ 7)[a]	104	0.870 (0.128)[bc]	103	73.05 (16.80)[c]
Minor (8–10)[b]	13	0.614 (0.301)[a]	12	60.83 (25.30)

Table 5 (continued)

	EQ-5D-5L utility index		EQ-VAS	
	N	**Mean (SD)**	***N***	**Mean (SD)**
Moderate to severe (≥ 11)[c]	9	0.580 (0.200)[a]	9	46.11 (11.40)[a]
p-value		< 0.001		< 0.001

SD: Standard deviation; BMI, Body Mass Index; HbA1c, Glycosylated hemoglobin; DM: Diabetes; SF-36, 36-item Short-Form Health Survey; HADS = Hospital Anxiety and Depression Scale

The score for the EQ-5D-5L utility index range from −0.416 to 1, and the score for the EQ-VAS range from 0 to 100, with higher scores indicating better health status

[a, b,c, d,e]Superscript letters indicated differences among the subgroups by Scheffe's test for multiple comparisons at $p < 0.05$

self-care dimension, and mobility dimension. Likewise, α parameters of the GRM are consistent with the EFA and CFA factor loadings, the anxiety/depression dimension being the item with the lowest discriminatory power and usual activities with the highest. Moreover, regarding the range of the trait, the usual activities dimension covers the largest range, and it is more apparent at higher levels of severity in these patients. On the other hand, the pain/discomfort and anxiety/depression items cover a narrower range of the trait and are more apparent at lower levels of severity. These results make sense in that these items have shown not to be particularly useful in this disease in which patients do not have high levels of severity.

At scale-level, the convergent and discriminant validity was also demonstrated confirming our hypotheses. Nonetheless, the correlation of EQ-5D-5L self-care and anxiety/depression dimension with the ADDQoL present HRQOL dimension was lower than expected. This could be due to these EQ-5D-5L dimensions reflect more psychological aspects of HRQOL, while perhaps, ADDQoL present HRQOL dimension reflects more physical aspects. Further, we found a slightly higher correlation between the SF-36 general health and pain/discomfort dimension, perhaps because pain is often closely related to self-perceive HRQOL. Similarly good convergent validity has also been reported in other studies of DM. Jankowska et al. [29] also confirmed the convergent validity of the EQ-5D-5L in respondents with self-reported diabetes from a representative general population survey, examining the association between this instrument and the SF-12. Pattanaphesaj et al. [30] also found satisfactory results for convergent validity of the EQ-5D-5L dimensions studying the correlations with the SF-36.

The EQ-5D-5L also showed excellent known-groups validity, with statistically significant differences according to different characteristics. As other previous studies [28, 29, 31, 57], we have also found statistically significant lower EQ-5D-5L utility index scores in patients who were older, women, obese, did not do regular physical activity, have a higher disease duration, poor glycemic control (HbA1c ≥ 7%), had a poorer SF-36 "health today rating", and higher level of anxiety and depression. Therefore, the 5L version seems to have the ability to discriminate between different levels of health. This validity has also been shown in the general population [63–66], as well as in specific diseases other than DM [21, 22, 27, 67].

In addition to those mentioned above, this study has various limitations. First, the recruitment period is almost 10 years ago. However, it is not usual to update the psychometric properties of instruments every few years, because analysis regarding their underlying structure, reliability, relationship with other instruments or the known-groups validity are valid regardless to the recruitment timing. Then, our findings are of a timeless nature, also, useful and reproducible in the modern-day context. Second, we were unable to explore differences between patients who agreed to participate in the study and those who were invited but did not accept. Thus, we were unable to assess the representativeness of the sample. Nonetheless, we believe that the patients included in this study cover the different levels of severity commonly seen in daily practice. To complete the validity study, the questionnaire's responsiveness would need to be analyzed, and although we studied the internal consistency, the reliability study should be complemented with a test-retest study. Further, the sample size is somewhat small for IRT analysis, and although we have used a Bayesian estimator to minimize this limitation, the results should be treated with caution. As the sample was not large enough to allow IRT analysis by groups, it was not possible to investigate differential item functioning, which occurs when different groups within the sample respond in different ways to an individual item. Another limitation could be the use of patient-generated outcome measures, such as the ADDQoL questionnaire, since they do not provide a form of standardization needed for comparison of results. However, we are using the ADDQoL only to study the convergent validity of the EQ-5D-5L, and despite this limitation, these measures are adequate in the development of other instruments [68]. Finally, although serum Hb1Ac is still the gold standard measure for the glycemic control, advances over the last decade have brought to light its problematics regarding the diabetes monitoring and the association with the risk of complications in certain clinical instances [69, 70]. Besides, Hb1Ac has been reported not to be surrogate for HRQOL when measured by other specific questionnaires [71]. However, our results are consistent with others demonstrating the relationship with the EQ-5D-5L [28, 57]. We recommend including additional diabetes monitoring variables in future researches so evaluation of HRQOL in DM can be adapted to the new management procedures.

Conclusions

Despite not all measurement properties have been tested in the current study, the results provide evidence of the validity of the EQ-5D-5L in patients with DM in Spain. The study provides evidence for clinicians and researchers concerning the appropriateness use of this generic preference-based measure for assessing HRQOL in patients with DM, and for use in economic evaluation. In conclusion, our prospective study of the psychometric properties of the Spanish version of the EQ-5D-5L in patients with DM, combining both classical test theory and IRT, provides evidence of its adequate item functioning, reliability and validity in this population.

Abbreviations

ADDQoL	Audit on diabetes-dependent quality of life
BMI	Body mass index
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CRC	Category response curve
DM	Diabetes mellitus
EFA	Exploratory factor analysis
EQ-VAS	EQ-visual analogue scale
GRM	Graded response model
HADS	Hospital anxiety and depression scale
HbA1c	Glycosylated blood hemoglobin
HRQOL	Health related quality of life
IIC	Item information curve
IRT	Item response theory
NNFI	Non-normed fit index
QALY	Quality-adjusted life years
RMSEA	Root mean square error of approximation
SD	Standard deviation
WI	Weighted impact

Acknowledgements

We thank the staff members of the various services, research and quality units, and medical records sections of the participating hospitals, and the patients who participated in the study, for their contributions.

Author contributions

AB performed all statistical analyses, was involved in the interpretation of the results and drafting the manuscript. MG and DM were involved in interpretation of the results and drafting the manuscript. IG conceived the study, participated in its design and coordination, and help drafting the manuscript. MF, YR, MP, and CG were involved in the conception and design of the study. All authors read and approved the final manuscript.

Funding

This study was supported by grants from the Carlos III Health Institute (PI12/01473, RD21/0016/0011 and RD24/0005/0019) and co-funded by European Regional Development Fund.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval

The institutional review board "Basque Research Ethics Committee on Drug Research (CEIm-E)" approved the study (Exp: PI2012156) and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Consent to participate

All persons gave their informed consent prior to their inclusion in the study.

Consent for publication

Not applicable.

Competing interests

The authors have no relevant financial or non-financial interests to disclosure.

Received: 3 October 2024 / Accepted: 28 March 2025
Published online: 29 May 2025

References

1. American Diabetes Association Professional Practice Committee, ElSayed NA, Aleppo G, Bannuru RR, Bruemmer D, Collins BS, Gabbay RA (2024) 2. Diagnosis and classification of diabetes: Standards of care in Diabetes—2024. Diabetes Care 47(Supplement1):S20–S42. https://doi.org/10.2337/dc24-S002
2. Soriguer F, Goday A, Bosch-Comas A, Bordiú E, Calle-Pascual A, Carmena R, Vendrell J (2012) Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Di@bet.es study. Diabetologia 55(1):88–93. https://doi.org/10.1007/s00125-011-2336-9
3. Jing X, Chen J, Dong Y, Han D, Zhao H, Wang X, Ma J (2018) Related factors of quality of life of type 2 diabetes patients: a systematic review and meta-analysis. Health Qual Life Outcomes 16(1):189. https://doi.org/10.1186/s12955-018-1021-9
4. Férriz Villanueva G, Rojas Blanc M, Riera Nadal N, Riera Nadal C, Fernández Martínez FJ, Aguado Jodar A (2011) ¿Qué Gasto Farmacológico genera Un diabético Bien Controlado? Atención Primaria 43(4):169–174. https://doi.org/10.1016/j.aprim.2010.01.025
5. Karimi M, Brazier J (2016) Health, Health-Related quality of life, and quality of life: what is the difference? PharmacoEconomics 34(7):645–649. https://doi.org/10.1007/s40273-016-0389-9
6. Guyatt GH, Veldhuyzen Van Zanten SJ, Feeny DH, Patrick DL (1989) Measuring quality of life in clinical trials: a taxonomy and review. CMAJ: Can Med Association J = J De l'Association Medicale Canadienne 140(12):1441–1448
7. Wang Y, Tan N-C, Tay E-G, Thumboo J, Luo N (2015) Cross-cultural measurement equivalence of the 5-level EQ-5D (EQ-5D-5L) in patients with type 2 diabetes mellitus in Singapore. Health Qual Life Outcomes 13(1):103. https://doi.org/10.1186/s12955-015-0297-2
8. Redekop WK, Koopmanschap MA, Stolk RP, Rutten GEHM, Wolffenbuttel BHR, Niessen LW (2002) Health-Related quality of life and treatment satisfaction in Dutch patients with type 2 diabetes. Diabetes Care 25(3):458–463. https://doi.org/10.2337/diacare.25.3.458
9. Guide to the methods of technology appraisal 2013. NICE process and methods. (2013) National Institute for Health and Care Excellence. Retrieved from https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781
10. Guidelines for the economic evaluation of health technologies: Canada. (1997) (3rd ed.). Ottawa, Ont.: Canadian Agency for Drugs and Technologies in Health
11. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL (2005) Methods for the economic evaluation of health care programme. Third edition. Oxford University Press, United Kingdom
12. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Badia X (2011) Development and preliminary testing of the new five-level version of EQ-5D(EQ-5D-5L). Qual Life Res 20(10):1727–1736. https://doi.org/10.1007/s11136-011-9903-x
13. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A (2017) Measuring and valuing health benefits for economic evaluation (Second edition.). Oxford New York, NY: Oxford University Press
14. Euroqol website (n.d.). Retrieved from https://euroqol.org/
15. Macran S, Weatherly H, Kind P (2003) Measuring population health: A comparison of three generic health status measures. Med Care 41(2):218–231. https://doi.org/10.1097/01.MLR.0000044901.57067.19
16. Petrou S, Hockley C (2005) An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. Health Econ 14(11):1169–1189. https://doi.org/10.1002/hec.1006
17. Wang H, Kindig DA, Mullahy J (2005) Variation in Chinese population health related quality of life: results from a EuroQol study in Beijing, China. Qual Life Res 14(1):119–132. https://doi.org/10.1007/s11136-004-0612-6
18. Brazier J, Roberts J, Tsuchiya A, Busschbach J (2004) A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 13(9):873–884. https://doi.org/10.1002/hec.866

19. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ (2008) Comparing the standard EQ-5D Three-Level system with a Five-Level version. Value Health 11(2):275–284. https://doi.org/10.1111/j.1524-4733.2007.00230.x
20. Janssen MF, Birnie E, Bonsel GJ (2008) Quantification of the level descriptors for the standard EQ-5D three-level system and a five-level version according to two methods. Qual Life Res 17(3):463–473. https://doi.org/10.1007/s11136-008-9318-5
21. Bilbao A, Martín-Fernández J, García-Pérez L, Mendezona JI, Arrasate M, Candela R, Retolaza A (2021) Psychometric properties of the EQ-5D-5L in patients with major depression: factor analysis and Rasch analysis. J Mental Health 1–11. https://doi.org/10.1080/09638237.2021.1875422
22. Bilbao A, García-Pérez L, Arenaza JC, García I, Ariza-Cardiel G, Trujillo-Martín E, Martín-Fernández J (2018) Psychometric properties of the EQ-5D-5L in patients with hip or knee osteoarthritis: reliability, validity and responsiveness. Qual Life Res 27(11):2897–2908. https://doi.org/10.1007/s11136-018-1929-x
23. Represas-Carrera FJ, Méndez-López F, Couso-Viana S, Masluk B, Magallón-Botaya R, Clavería A (2021) [Baseline characteristics and quality of life in patients with diabetes mellitus included in the EIRA randomized clinical trial]. Revista Esp De Salud Publica 95:e202103034
24. Collado Mateo D, García Gordillo, Olivares PR, Adsuar JC (2015) Normative values of EQ-5D-5L for diabetes patients from Spain. Nutrición Hospitalaria 41595–1602. https://doi.org/10.3305/nh.2015.32.4.9605
25. Rojas G, Solís Pazmiño P, Gaona R, Mollocana F, Espín I, Nunes A (2021) Calidad de Vida En Un grupo de Pacientes Con diabetes mellitus Tipo 2: Un estudio transversal. Revista Med Vozandes 32(1). https://doi.org/10.48018/rmv.v32.i1.7
26. Arifin B, Purba FD, Herman H, Adam JMF, Atthobari J, Schuiling-Veninga CCM, Postma MJ (2020) Comparing the EQ-5D-3 L and EQ-5D-5L: studying measurement and scores in Indonesian type 2 diabetes mellitus patients. Health and Quality of Life Outcomes, 18(1):22. https://doi.org/10.1186/s12955-020-1282-y
27. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Busschbach J (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(7), 1717–1727. https://doi.org/10.1007/s11136-012-0322-4
28. Abedini MR, Bijari B, Miri Z, Shakhs Emampour F, Abbasi A (2020) The quality of life of the patients with diabetes type 2 using EQ-5D-5L in Birjand. Health Qual Life Outcomes 18(1):18. https://doi.org/10.1186/s12955-020-1277-8
29. Jankowska A, Młyńczak K, Golicki D (2021) Validity of EQ-5D-5L health-related quality of life questionnaire in self-reported diabetes: evidence from a general population survey. Health Qual Life Outcomes 19(1):138. https://doi.org/10.1186/s12955-021-01780-2
30. Pattanaphesaj J, Thavorncharoensap M (2015) Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in the Thai diabetes patients. Health Qual Life Outcomes 13(1):14. https://doi.org/10.1186/s12955-014-0203-3
31. Koh D, Abdullah AMKB, Wang P, Lin N, Luo N (2016) Validation of Brunei's Malay EQ-5D questionnaire in patients with type 2 diabetes. PLoS ONE 11(11):e0165555. https://doi.org/10.1371/journal.pone.0165555
32. Pan C-W, Sun H-P, Wang X, Ma Q, Xu Y, Luo N, Wang P (2015) The EQ-5D-5L index score is more discriminative than the EQ-5D-3L index score in diabetes patients. Qual Life Res 24(7):1767–1774. https://doi.org/10.1007/s11136-014-0902-6
33. Bradley C, Todd C, Gorton T, Symonds E, Martin A, Plowright R (1999) The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. Qual Life Res 8(1/2):79–91. https://doi.org/10.1023/A:1026485130100
34. Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica 67(6):361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x
35. Ware JE, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30(6):473–483
36. Spanish Society of Diabetes (2023), October 17 Retrieved from https://www.sediabetes.org/
37. Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40(5):373–383. https://doi.org/10.1016/0021-9681(87)90171-8
38. Ramos-Goñi JM, Craig BM, Oppe M, Ramallo-Fariña Y, Pinto-Prades JL, Luo N, Rivero-Arias O (2018) Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. Value Health 21(5):596–604. https://doi.org/10.1016/j.jval.2017.10.023
39. Perrotta C, Irazola V (2002) Validación linguistica de Un cuestionario Específico Para Medir Calidad de Vida relacionada Con La Salud (CVRS)Audit of life diabetes dependent questionnaire(ADDQOL) al Español Como se Habla En La Argentina 36(107)
40. DePablos-Velasco P, Salguero-Chaves E, Mata-Poyo J, DeRivas-Otero B, García-Sánchez R, Viguera-Ester P (2014) Calidad de Vida y satisfacción Con El Tratamiento de Sujetos Con diabetes Tipo 2: resultados En España Del estudio PANORAMA. Endocrinología Y Nutrición 61(1):18–26. https://doi.org/10.1016/j.endonu.2013.05.005
41. Herrero MJ, Blanch J, Peri JM, De Pablo J, Pintor L, Bulbena A (2003) A validation study of the hospital anxiety and depression scale (HADS) in a Spanish population. Gen Hosp Psychiatry 25(4):277–283. https://doi.org/10.1016/S0163-8343(03)00043-4
42. Quintana JM, Padierna A, Esteban C, Arostegui I, Bilbao A, Ruiz I (2003) Evaluation of the psychometric characteristics of the Spanish version of the hospital anxiety and depression scale. Acta Psychiatrica Scandinavica 107(3):216–221. https://doi.org/10.1034/j.1600-0447.2003.00062.x
43. Wyrwich KW, Tierney WM, Wolinsky FD (1999) Further evidence supporting an SEM-Based criterion for identifying meaningful Intra-Individual changes in Health-Related quality of life. J Clin Epidemiol 52(9):861–873. https://doi.org/10.1016/S0895-4356(99)00071-2
44. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16(3):297–334. https://doi.org/10.1007/BF02310555
45. Nunnally JC, Berstein I (1994) Psychometric theory, 3rd edn. McGraw-Hill, New York
46. Gomez R (2011) Item response theory analyses of adult Self-Ratings of the ADHD symptoms in the current symptoms scale. Assessment 18(4):476–486. https://doi.org/10.1177/1073191110386341
47. Staquet MJ, Hays RD, Fayers PM (eds) (1998) Quality of life assessment in clinical trials: methods and practice. Oxford University Press, Oxford; New York
48. Gomez R (2012) Parent ratings of ADHD symptoms: generalized partial credit model analysis of differential item functioning across gender. J Atten Disord 16(4):276–283. https://doi.org/10.1177/1087054710383378
49. Lewis KM, Lambert MC (2006) Measuring social change preferences in African American adolescents: development of the measure of social change for adolescents (MOSC-A). Assessment 13(4):406–416. https://doi.org/10.1177/1073191106289032
50. Mulaik SA, James LR, Van Alstine J, Bennett N, Lind S, Stilwell CD (1989) Evaluation of goodness-of-fit indices for structural equation models. Psychol Bull 105(3):430–445. https://doi.org/10.1037/0033-2909.105.3.430
51. Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equation Modeling: Multidisciplinary J 6(1):1–55. https://doi.org/10.1080/10705519909540118
52. Samejima F (1969) Estimation of latent trait ability using a response pattern of graded scores (Psychometrika monograph supplement no. 17). Psychometric Society, Richmond, VA
53. Asparouhov T, Muthen B (2010) Bayesian analysis of latent variable models using Mplus. Tecnical Report. Version 4
54. Morrissey C, Cooke D, Michie C, Hollin C, Hogue T, Lindsay WR, Taylor JL (2010) Structural, item, and test generalizability of the psychopathy Checklist—Revised to offenders with intellectual disabilities. Assessment 17(1):16–29. https://doi.org/10.1177/1073191109344052
55. Cohen J (1992) A power primer. Psychol Bull 112(1):155–159. https://doi.org/10.1037/0033-2909.112.1.155
56. Kiadaliri AA, Eliasson B, Gerdtham U-G (2015) Does the choice of EQ-5D tariff matter? A comparison of the Swedish EQ-5D-3L index score with UK, US, Germany and Denmark among type 2 diabetes patients. Health Qual Life Outcomes 13(1):145. https://doi.org/10.1186/s12955-015-0344-z
57. Ozcan S, Amiel SA, Rogers H, Choudhary P, Cox A, De Zoysa N, Forbes A (2014) Poorer glycaemic control in type 1 diabetes is associated with reduced self-management and poorer perceived health: A cross-sectional study. Diabetes Research and Clinical Practice, 106(1):35–41. https://doi.org/10.1016/j.diabres.2014.07.023
58. Muthén L, Muthén B (n.d.). Mplus User's Guide. Sixth Edition. Los Angeles, CA. 1998–2010
59. Yfantopoulos J, Chantzaras A, Kontodimas S (2017) Assessment of the psychometric properties of the EQ-5D-3L and EQ-5D-5L instruments in psoriasis. Arch Dermatol Res 309(5):357–370. https://doi.org/10.1007/s00403-017-1743-2
60. Keng MJ, Leal J, Bowman L, Armitage J, Mihaylova B, ASCEND Study Collaborative Group (2022) Decrements in health-related quality of life associated

with adverse events in people with diabetes. Diabetes Obes Metabolism 24(3):530–538. https://doi.org/10.1111/dom.14610
61. Xu RH, Cheung AWL, Wong EL-Y (2017) Examining the health-related quality of life using EQ-5D-5L in patients with four kinds of chronic diseases from specialist outpatient clinics in Hong Kong SAR, China. Patient Prefer Adherence 11:1565–1572. https://doi.org/10.2147/PPA.S143944
62. Christiansen ASJ, Møller MLS, Kronborg C, Haugan KJ, Køber L, Højberg S, Svendsen JH (2021) Comparison of the three-level and the five-level versions of the EQ-5D. Eur J Health Econ 22(4):621–628. https://doi.org/10.1007/s10198-021-01279-z
63. Kangwanrattanakul K, Parmontree P (2020) Psychometric properties comparison between EQ-5D-5L and EQ-5D-3L in the general Thai population. Qual Life Res 29(12):3407–3417. https://doi.org/10.1007/s11136-020-02595-2
64. Młyńczak K, Golicki D (2021) Validity of the EQ-5D-5L questionnaire among the general population of Poland. Qual Life Res 30(3):817–829. https://doi.org/10.1007/s11136-020-02667-3
65. Kim TH, Jo M-W, Lee S, Kim SH, Chung SM (2013) Psychometric properties of the EQ-5D-5L in the general population of South Korea. Qual Life Res 22(8):2245–2253. https://doi.org/10.1007/s11136-012-0331-3
66. Hernandez G, Garin O, Pardo Y, Vilagut G, Pont À, Suárez M, Ferrer M (2018) Validity of the EQ–5D–5L and reference norms for the Spanish population. Qual Life Res 27(9):2337–2348. https://doi.org/10.1007/s11136-018-1877-5
67. Seng JJB, Kwan YH, Fong W, Phang JK, Lui NL, Thumboo J, Leung YY (2020) Validity and reliability of EQ-5D-5L among patients with axial spondyloarthritis in Singapore. Eur J Rheumatol 7(2):71–78. https://doi.org/10.5152/eurjrheum.2020.19043
68. Patel KK, Veenstra DL, Patrick DL (2003) A review of selected Patient-Generated outcome measures and their application in clinical trials. Value Health 6(5):595–603. https://doi.org/10.1046/j.1524-4733.2003.65236.x
69. Wright LA-C, Hirsch IB (2017) Metrics beyond hemoglobin A1C in diabetes management: time in range, hypoglycemia, and other parameters. Diabetes Technol Ther 19(S2). https://doi.org/10.1089/dia.2017.0029. -16-S-26
70. Kilpatrick ES (2004) HbA_{1c} measurement. J Clin Pathol 57(4):344. https://doi.org/10.1136/jcp.2003.010918
71. Walker J, Bradley C (2002) Assessing the quality of life of adolescents with diabetes: using the SEIQoL, DQoL, patient and diabetes specialist nurse ratings. Practical Diabetes Int 19(5):141–144. https://doi.org/10.1002/pdi.348

Publisher's note