Documentos de Trabajo

Lan Gaiak

COMPARING THE PSYCHOMETRIC PROPERTIES OF THE EQ-5D-5L BETWEEN MENTAL AND SOMATIC CHRONIC PATIENTS POPULATIONS

Juan M. Cabasés
María Errea
Iñigo Hernández-Arenaz

D.T. 1308

Departamento de Economía

Ekonomia Saila

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Comparing the psychometric properties of the EQ-5D-5L between mental and somatic chronic patients populations

Juan M Cabasés, María Errea, Iñigo Hernández-Arenaz

[1]Department of Economics, Public University of Navarra, 31006 Pamplona, Spain.

jmcabases@unavarra.es; maria.errea@unavarra.es

June 2013

Abstract

The validity and reliability of the EQ-5D-5L in comparison with the standard 3L has been tested through the analysis of psychometric properties making use of different samples of patients. However, it is likely that the condition of the illness may affect the power of the 5L version with respect to the 3L one. Here we report on parallel testing of EQ-5D-5L and 3L administered to a sample of chronic patients of both somatic and mental illness. The aim of this study is to check some psychometric properties in both subsamples.

Methods: We check for the usual psychometric properties: feasibility, (in)consistency, ordinality (and transitivity), informativity, face validity and convergent validity. Also, we perform new analysis for checking transitivity and the Cronbach-α for convergent validity. Finally, we proposed a complementary way for looking at the property of informativity through three different indexes (effective, absolute and overall) based on the statistical discriminatory power.

Data: We have a total of 1002 questionnaires finally collected. 444 (46.25%) chronic mental patients, 516 (53.75%) have somatic chronic illnesses; 42 observations of unknown origin of the illness have been dropped to perform this analysis.

Results: The mean value reported in the VAS for the full sample is 60.93. Somatic patients report a mean of 64.42 points in this scale and mental patients report 56.83 points in the VAS. Analyzing the distribution of the responses to problems on each dimension we found, for all cases, a highly skewed distribution. Moreover, the distribution of responses changes significantly between subsamples, as expected.

In all dimensions, it seems that somatic patients take more advantage of the extra levels introduced by the EQ-5D-5L. This group reduces to a greater extent the missing response rate, commit less (and of lower importance) inconsistencies, get a higher correlation of the 5L scale and the VAS within the 3L levels, complements better the dimensions to get an overall score

(measured through the Cronbach's alpha), reduces in a more significant way the "no problem" response and the Informativity gain is also superior (for both the Shannon Evenness Index and our Absolute Index). This higher performance of the EQ-5D-5L on somatic patients is endorsed by a higher preference of somatic patients toward the 5L version of the questionnaire than to the 3L one.

Conclusion: Results show the suitability of the 5L version in both subsamples, but it is much more effective for somatic patients. These subsamples' differences may be of concern when aggregating and comparing different data.

1. Background

The validity and reliability of the EQ-5D-5L in comparison with the EQ-5D-3L has been proven in prior researches through the analysis of psychometric properties making use of different samples of patients located on different countries (Pikard, A.S. et al., 2007; Kim, S.H. et al., 2012; Kim, T.H. et al., 2012; Scalone, L. et al., 2012; Janssen, M.F., 2012).

However and despite the richness of the data collected up to now, the analysis has been carried out without considering possible iterations between the source of the illness and the impact of the two intermediate levels that the EQ-5D-5L incorporates with respect to its previous version of three levels.

While certain socio-demographic variables have been reported as moderators or enhancers in previous studies (for example, gender is a relevant variable when looking possible ceiling effects (Kim, T.H. et al., 2012)), it is also very likely that the source of the illness of the patient affects the power of the 5L version with respect to the 3L one in different aspects.

In order to examine this concern, we exploit a dataset composed by patients dealing with chronic diseases[1]. This specific sample will help the analysis in two ways. A first advantage of this sample in comparison with others is that it would reduce the number of individuals expected to report to be in health state 11111 and thus will generate a greater dynamism when passing from the 3L version to the 5L one. Also, focus our attention in chronic patients will help us to discern the source of the illness affecting the patient.

Given our particular sample, the clearer and less problematic division among diseases is the one given by somatic and mental source of the illness. Being this classification feasible, the

[1] **Chronic Diseases:** diseases which have one or more of the following characteristics: they are permanent, leave residual disability, are caused by non-reversible pathological alteration, require special training of the patient for rehabilitation, or may be expected to require a long period of supervision, observation, or care. Other terms used in the scientific literature are: chronic illness, chronic conditions. rehabilitation, or may be expected to require a long period of supervision, observation, or care. Other terms used in the scientific literature are: chronic illness, chronic conditions.

aim of this study is to check some psychometric properties in both subsamples and compare the results in order to check a potential different effect over both subsamples.

The paper is organized as follows. First, in section 2, data and methods for the analysis of the psychometric properties studied are presented. Section 3 shows preliminary results. Section 4 opens to discussion together with the conclusions.

2. Data and Methods

2.1. The questionnaire

In terms of the measurement instruments used in the study, patients answered a questionnaire, which contains the two versions, EQ-5D-5L and 3L, the EQ-VAS and a page of socio-demographic questions.

The order of the questionnaire is the following: it starts with the EQ-5D-5L followed by socio-demographic questions, sex, age, education level, main activity, nationality and postal code, and if the individual is a smoker or not; then, the 3L, and the VAS. Finally, at the end of the questionnaire, we include two questions in which individuals are asked to value which of the two questionnaires they found easier to answer, and which of the two versions allowed them to better express their health status, respectively.

2.2. Data Collection process

Questionnaires were delivered by doctors at consultancy and filled by the patient. Doctors provided also extra information as the *ICPC (International Classification of Primary Care)* that helps to classify individuals in subgroups according to their diagnosed illness, and resolve the doubts patients may had when filling the questionnaires.

Data were collected in several medical centers that can be classified into three different institutions: primary care centers, specialist clinics (mental health care center, rheumatology, pneumology) and a psychiatric day hospital. Doctors were responsible of delivering the questionnaire to patients following our instructions. For specialist consultations and the psychiatric day hospital, the questionnaire was administered to all patients attending the center in the period, while on the primary care centers a criterion of convenience was used by doctors: they provided the questionnaire where they were able given their time constraints. Thus, our sample contains all the population of chronic patients attending the specialist or the psychiatric day care hospital and a random sample of chronic patients attending primary care centers.

The somatic and mental care groups are generated according to the institution patients attended and to its *ICPC* and *ICD-10* (*International Statistical Classification of Diseases and Related Health Problems, 10th Revision*). In such a fashion, somatic group is composed by 326 patients from primary care and 190 from specialist consultancies while the mental group comes from primary care (3), specialist consultancies (284) and the psychiatric day care hospital (157). Data were collected in 2010 and 2011.

2.3 Methods for data analysis

In order to test the validity and reliability of the 5L descriptive system, we follow the line opened by Janssen et al. (2008) looking as the main psychometric properties: feasibility, (in)consistency, ordinality (and transitivity), informativity, face validity and convergent validity (Jansssen et al, 2008, 2012; Kim et al 2012; Scalone et al, 2012). In addition, we perform new analysis for checking transitivity and the Cronbach-α for convergent validity. Finally, we propose a complementary way for looking at the property of informativity through three different indexes (effective, absolute and overall) based on the statistical discriminatory power. Although we lack of dimension-specific VAS, the large-scale of the study as well as the convergent validity property allow us to use the EQ-VAS where needed.

We also look at other relevant information, through regression analysis and covariates between variables, in order to explain some features of our dataset.

3. Results

3.1 Descriptive Analysis

We have a total of 1002 questionnaires finally collected. Among the full sample, 444 (46.25%) of the questionnaires correspond to chronic mental patients, 516 (53.75%) have somatic chronic illnesses, and for the remaining 42 the origin of the illness is unknown and therefore are dropped from the analysis.

The mean age of the respondents in our sample is 55.93 years ranging from 15 to 94. By type of illness, we find that the population from the somatic group is statistically older ($F(1,937)=555.42$, $p<.01$) than the one from the mental one (66.75 vs 43.51).

The percentage of women is slightly higher than men (51.42 vs 48.58). We also found that the groups are statistically unbalanced on this regard ($F(1,950)=6.65$, $p<.05$) being the proportion of male patients higher on the somatic group (52.44%) than on the mental one (47.55%).

The majority of participants in this study are Spaniards, while there are a 9.8% of respondents who have a different nationality. Among those who are from a different country, we know that the mean of years they have lived in Spain is 13.25.

A big part of the sample reports to have achieved just primary studies (454, the 48.4%) while it is much lower the percentage of patients who have reached high school (277, the

29.53%). Only the 12.47% of the population has superior studies (university or similar). The 9.59% has not received any kind of education. Looking by subsamples, mental patients tend to be more educated than somatic ones (table 1).

The 28.63% (264) are employed and the 29.72% retired. The 16.16% of respondents are house-workers, and a 9.44% are unemployed. There is a non-negligible number of individuals disabled in the sample (124, the 13.23%). Only the 2.82% of the participants are students. As happened with education, the sample is highly unbalanced attending to the origin of the illness (table 1) being especially relevant the difference on retired and unemployed population.

Table 1: Distribution of subsamples regarding studies and main activity

		Somatic		Mental	
		%	Std. Err.	%	Std. Err.
Studies	Uneducated *	11.59%	0.0142	7.23%	0.0125
	Primary **	56.78%	0.0220	38.46%	0.0235
	High School **	22.79%	0.0186	37.53%	0.0234
	Superior **	8.84%	0.0126	16.78%	0.0181
Main Activity	Employed **	22.42%	0.0186	36.12%	0.0235
	Retired **	44.64%	0.0222	11.72%	0.0158
	Disable **	9.52%	0.0131	17.70%	0.0187
	House-Worker **	21.03%	0.0182	10.29%	0.0149
	Student **	0.40%	0.0028	5.74%	0.0114
	Unemployed **	1.98%	0.0062	18.42%	0.0190

* Difference significant at p<.05

** Difference significant at p<.01

The difference between the mean time respondents have symptoms of the illness and the mean time they were diagnosed is 1.67 years for the full sample. Interesting is the fact that mental patients used to be diagnosed later than somatic (2.59 vs 1.02 years) being this difference statistically significant (p<.01). Concerning the self-perception of severity of the illness we find that 205 (21.68%) perceive severity of illness as mild, 465 (50.44%) perceive it as moderate, and 257 (27.89%) as severe. For illness origin, the groups are quite balanced although somatic patients tend to claim more often that their disease has a moderate severity (p<.05).

The mean value reported in the VAS for the full sample is 60.93. Somatic patients report in mean 64.42 points in this scale. Mental patients however, report in mean a lower value, 56.83 points in the VAS. Moreover, this difference is highly significant (p<.01).

Analyzing the distribution of the responses on each dimension (table 2) we found, for all cases, a much skewed distribution. In particular, and although we are considering only chronic patients, lower levels (5L-1 and 5L-2) are systematically more often chosen than higher ones.

Moreover, and as was expected, the distribution of responses change significantly between subsamples. The closest distribution between subsamples, takes place on the usual activities one, for which we cannot reject the equality of distribution at p<.05.

Table 2: Distribution of responses for each dimension by subsample

		mo ***	sc ***	ua *	pd ***	ad ***
Somatic	1	38.4%	59.6%	45.8%	18.5%	37.0%
	2	29.1%	23.3%	29.7%	44.7%	37.8%
	3	23.8%	12.4%	16.1%	24.7%	19.1%
	4	7.9%	3.7%	5.6%	10.7%	4.9%
	5	0.8%	1.0%	2.7%	1.4%	1.2%
Mental	1	68.5%	76.1%	44.0%	38.8%	24.4%
	2	16.9%	14.4%	23.9%	27.8%	28.0%
	3	10.6%	6.3%	19.2%	19.4%	23.5%
	4	3.2%	2.7%	7.9%	10.8%	16.9%
	5	0.9%	0.5%	5.0%	3.2%	7.2%

* Difference between samples' distributions significant at p<.1
** Difference between samples' distributions significant at p<.05
*** Difference between samples' distributions significant at p<.01

3.2 Analysis of psychometric properties

Feasibility: The overall effect of the 5L questionnaire in terms of response rate, is positive as long as all the measures are significantly (p<.01) different between the 3L and 5L questionnaires, supporting the fact that the 5L achieves a higher response rate. However, this effect is mainly driven by changes in somatic patients. Although the resulting measures from mental patients also support a higher response rate for 5L than for 3L, the changes are not significant or if so are not high. Moreover, we have also to notice that the responses rates found for this subsample of patients are very high for both 3L and 5L, so there is little room for improvement.

Comparing between subsamples, we only find a significant (p<.05) higher proportion of missing responses in somatic patients for mobility and anxiety/depression when looking the 3L questionnaire. Interesting enough is the fact that this difference disappears on the 5L instrument.

Table 3: Feasibility

		5L				3L			
		% Missing	Std. Err.	[95% Conf. Interval]		% Missing	Std. Err.	[95% Conf. Interval]	
Full Sample	Profile **	0%	0	0%	0%	1.25%	0.00359	0.55%	1.95%
	Partial profiles **	0.94%	0.00311	0.33%	1.55%	2.50%	0.00504	1.51%	3.49%
	mo **	0%	0	0%	0%	1.67%	0.00413	0.86%	2.48%
	sc **	0%	0.00104	0%	0%	1.46%	0.00387	0.70%	2.22%
	ua **	0.21%	0.00147	-0.08%	0.50%	1.56%	0.004	0.78%	2.35%
	pd **	0.31%	0.00180	-0.04%	0.67%	1.35%	0.00373	0.62%	2.09%
	ad **	0.42%	0.00208	0.01%	0.82%	1.77%	0.00426	0.94%	2.61%
Somatic	Profile **	0%	0	0%	0%	1.74%	0.00577	0.61%	2.88%
	Partial profiles *	1.36%	0.00510	0.36%	2.36%	3.49%	0.00809	1.90%	5.08%
	mo **	0%	0.00000	0%	0%	2.52%	0.00691	1.16%	3.88%
	sc **	0%	0.00194	0%	1%	1.94%	0.00607	0.74%	3.13%
	ua **	0.19%	0.00194	-0.19%	0.57%	2.13%	0.00636	0.88%	3.38%
	pd *	0.39%	0.00274	-0.15%	0.93%	1.74%	0.00577	0.61%	2.88%
	ad **	0.58%	0.00335	-0.08%	1.24%	2.52%	0.00691	1.16%	3.88%
Mental	Profile	0%	0	0%	0%	0.68%	0.00389	-0.09%	1.44%
	Partial profiles *	0.45%	0.00318	-0.17%	1.08%	1.35%	0.00549	0.27%	2.43%
	Mo	0%	0	0%	0%	0.68%	0.00389	-0.09%	1.44%
	sc*	0%	0	0%	0%	0.90%	0.00449	0.02%	1.78%
	Ua	0.23%	0.00225	-0.22%	0.67%	0.90%	0.00449	0.02%	1.78%
	Pd	0.23%	0.00225	-0.22%	0.67%	0.90%	0.00449	0.02%	1.78%
	Ad	0.23%	0.00225	-0.22%	0.67%	0.90%	0.00449	0.02%	1.78%

Profile: The full profile is completely blank. Partial Profile: At least one dimension is blank. Includes also profiles completely blank

* Difference between 3L and 5L significant at $p<.05$

** Difference between 3L and 5L significant at $p<.01$

Inconsistencies: There is a non-negligible percentage of the population, 14.77%, who committed inconsistencies when reporting the health states in the 5L and in the 3L questionnaires. This 14.77% is responsible for 177 inconsistent responses (3.8% of the total amount of pairs 3L-5L). Disaggregating the sample into somatic/mental patients it is found that mental patients are significantly ($F(1,948)=3.92$; $p<.05$) more prone to commit inconsistencies than somatic.

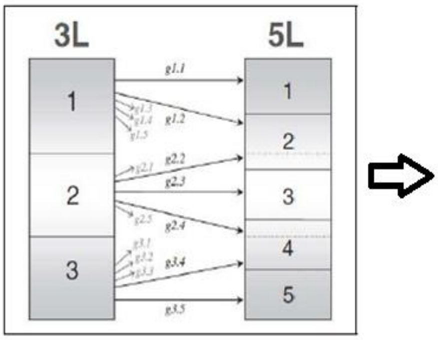
Table 4: Percentage of inconsistent patients found by groups

		N	Proportion	Std. Err.	[95% Conf.Interval]	
Full Sample	Consistent	948	85.23%	0.01153	82.97%	87.49%
	Inconsistent		14.77%	0.01153	12.51%	17.03%
Somatic	Consistent	507	87.38%	0.01476	84.48%	90.27%
	Inconsistent		12.62%	0.01476	9.73%	15.52%
Mental	Consistent	441	82.77%	0.01800	79.23%	86.30%
	Inconsistent		17.23%	0.01800	13.70%	20.77%

Making use of the proposed weights for inconsistencies by Janssen et al. (figure 1) and examining each dimension separately (figure 2), we find that, excepting for the self-care dimension in which there are only inconsistencies of weight one, mental patients commit in average higher inconsistencies. Despite this, the differences somatic/mental on this regard are only statistically significant ($p<.05$) for mobility and usual activities.

In addition, we can compute the specific patient j's average inconsistency weight (AIW) as the sum of all the weights of the inconsistencies committed by patient "j", divided by the amount of inconsistencies done by this patient. Between those who committed at least one inconsistency, the AIW are 1.46, 1.22 and 1.66 for the full sample, somatic subsample and mental subsample respectively. Moreover, the difference between somatic and mental patients is highly significant ($p<.01$) reinforcing not only the idea that mental patients do commit more inconsistencies, but also that given the existence of an inconsistency the one from mental patients is of a higher order.

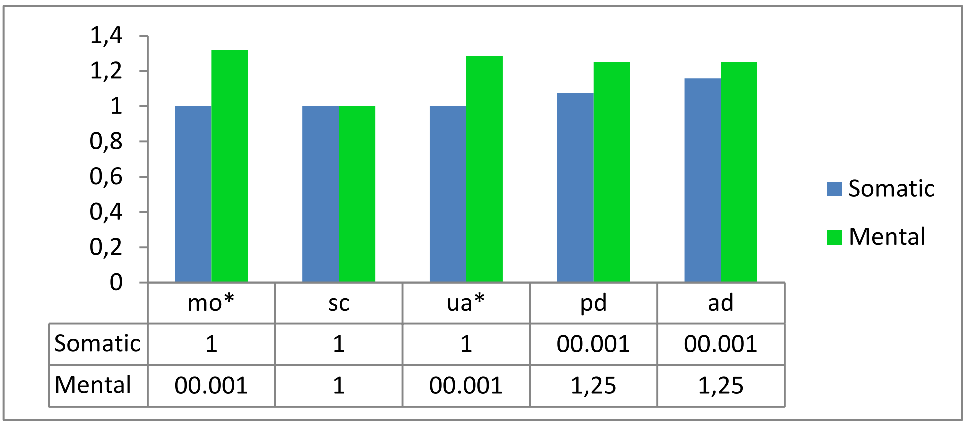
Figure 1: Diagram of redistribution between levels of the 3L and 5L (left) and identification of consistencies and inconsistencies (right). Source: Janssen *et al.* (2008)



3L 5L	Level 1	Level 2	Level 3	Level 4	Level 5
Level 1	0	0	1	2	3
Level 2	1	0	0	0	1
Level 3	3	2	1	0	0

The dark cells represent inconsistent responses (size of inconsistency is given in cells)

Figure 2. Mean inconsistency weight by dimension and by group of illness. (*) Difference significant at p<.05



Ordinality: Weak transitivity is satisfied in all dimensions for both subsamples (table 5). However the property of strict transitivity is not satisfied for the dimensions “mobility”, “usual activities” and “pain-discomfort” for mental care subsample. On the contrary, for the somatic subsample, all dimensions hold strong transitivity.

It should be noticed that we are computing this property differently to how typically is done. The difference lays in the fact that we lack of dimension specific VAS and thus, our classification is done according to the general VAS. However, we argue that given our large data set and the relationship between VAS and each dimension score (see the property of validity results), the result should be similar for both methods.

Table 5: Ordinality for each subsample

			SOMATIC							MENTAL						
	3L	VAS subgroup	N	VAS mean	VAS p50	5L mean	5L p50	WEAK	STRONG	N	VAS mean	VAS p50	5L mean	5L p50	WEAK	STRONG
MOBILITY	1	Upper	98	85.82	85	1.08	1	YES	YES	141	81.30	80	1.08	1	YES	YES
		Lower	112	62.65	65	1.16	1			160	45.73	50	1.09	1		
	2	Upper	66	78.41	78.5	2.42	2	YES	YES	40	71.25	70	2.43	2	YES	NO
		Middle	118	63.69	65	2.57	2			33	50.55	50	2.39	2		
		Lower	97	39.76	45	2.91	3			48	21.25	22.5	2.69	3		
	3	Upper	3	41.67	40	3.67	4	YES	YES	1	5.00	5	1.00	1	YES	YES
		Lower	5	14.00	10	4.60	5			5	0.00	0	2.60	2		
SELF-CARE	1	Upper	155	83.09	80	1.03	1	YES	YES	146	81.23	80	1.04	1	YES	YES
		Lower	162	59.59	62	1.14	1			174	46.53	50	1.06	1		
	2	Upper	43	74.12	71	2.19	2	YES	YES	32	70.88	70	1.94	2	YES	YES
		Middle	59	60.68	60	2.36	2			35	44.86	45	2.20	2		
		Lower	76	36.91	40	2.80	3			37	13.00	10	2.62	3		
	3	Upper	2	62.50	62.5	4.00	4	YES	YES	1	50.00	50	4.00	4	YES	YES
		Lower	6	29.17	37.5	4.17	4			2	0.00	0	4.50	4.5		
USUAL ACTIVITIES	1	Upper	110	85.55	85	1.04	1	YES	YES	1	4.00	4	1.01	1	YES	YES
		Lower	141	64.24	70	1.21	1			2	4.50	4.5	1.35	1		
	2	Upper	67	75.39	75	2.30	2	YES	YES	50	73.42	70	2.30	2	YES	YES
		Middle	73	60.84	60	2.34	2			70	53.11	50	2.59	2		
		Lower	88	41.98	45	2.75	3			63	27.40	30	2.73	3		
	3	Upper	11	47.27	40	4.45	5	YES	YES	19	36.05	35	4.21	4	YES	NO
		Lower	11	13.27	10	4.45	5			20	6.00	5	3.95	4		
PAIN –DISCOMFORT	1	Upper	53	89.08	90	1.21	1	YES	YES	82	85.71	85	1.07	1	YES	YES
		Lower	66	67.64	70	1.37	1			89	56.02	60	1.16	1		
	2	Upper	105	78.89	80	2.23	2	YES	YES	53	79.15	75	2.32	2	YES	NO
		Middle	89	67.18	67	2.39	2			50	60.68	60	2.32	2		
		Lower	150	48.16	50	2.69	3			105	37.53	40	2.64	3		
	3	Upper	17	65.24	65	3.81	4	YES	YES	24	41.04	35	3.92	4	YES	YES
		Lower	23	27.17	25	4.22	4			24	5.63	5	4.25	4		
ANXIETY/DEPRESSION	1	Upper	93	86.05	85	1.12	1	YES	YES	53	89.15	90	1.11	1	YES	YES
		Lower	117	59.32	60	1.21	1			70	61.80	65	1.29	1		
	2	Upper	69	78.57	79	2.22	2	YES	YES	74	79.16	80	2.32	2	YES	YES
		Middle	100	65.40	65	2.33	2			54	60.91	60	2.65	3		
		Lower	94	42.28	45	2.62	3			96	39.06	40	2.89	3		
	3	Upper	12	68.42	67	3.50	4	YES	YES	35	49.00	50	4.09	4	YES	YES
		Lower	14	31.50	40	3.93	4			45	13.69	10	4.44	4		

The property of Ordinality –through the analysis of transitivity–measures an existing relationship between the VAS and the health dimensions that the EQ-5D-3L is unable to capture. It restricts the sign of this relationship. For example, transitivity on the 1-U, implies that a change of VAS is accompanied by a change on the 5L (existence of relationship VAS, 5L) of contrary sign (restriction on the relationship), i.e. a decrease on the VAS would imply an increase on the 5L.

Thus, and with this thought in mind, it seems very suitable to compute the correlations on each dimension between the 5L punctuation and the VAS on each of the 3L's levels, i.e. compute

$$Corr(5L^j, VAS|3L^j=i) \text{ with } i\in\{1,2,3\} \text{ and } j\in\{mo,\ldots,ad\}$$

Performing such an analysis, it is found that all coefficients are negative on their values (table 6). However, we can only reject the null hypothesis of independency ($Corr(5L^j,VAS|3L^j=i)=0$) at $p<.1$ on 10 out of 14 of coefficients for somatic patients and on 9 out of 15 for mental ones. Decreasing the threshold to $p<.01$ for somatic patients we can still reject the independency hypothesis for 7 out of the 14 coefficients, while on the mental care the rejection rate falls down to a third of the cases (5 out of 15).

Table 6: Spearman's rank correlation between VAS and 5L within 3L groups

	3L	Somatic	Mental
mo	1	-0.12*	-0.0562
	2	-0.3058***	-0.1530*
	3	-0.706*	-0.4174
sc	1	-0.2089***	-0.1029*
	2	-0.3954***	-0.2923***
	3	-0.2906	-0.5000
ua	1	-0.3306***	-0.3638***
	2	-0.0607	-0.2693***
	3	(.)	-0.0321
pd	1	-0.2249**	-0.0566
	2	-0.3478***	-0.3401***
	3	-0.4506***	-0.3072**
ad	1	-0.0966	-0.1224
	2	-0.275***	-0.3905***
	3	-0.1907	-0.2637**

* Reject independency of distribution at p<.1
** Reject independency of distribution at p<.05
*** Reject independency of distribution at p<.01
(.) Not enough observations to perform the analysis

Validity: The correlations between 3L and 5L scales are significant ($p<.01$) and strong ($r>.7$) for all dimensions and both subsamples (table 7). Moreover, this behavior within scales is stable among samples, and not significant change takes place when moving from one subsample to another.

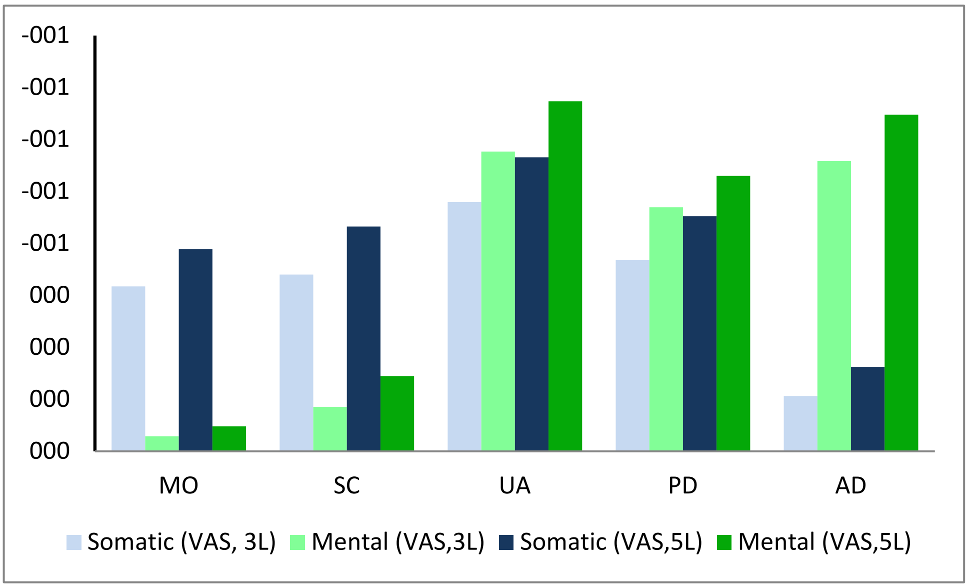
The correlations between the VAS and the dimensions are in all cases highly significant ($p<.01$) and of negative sign as was expected. While for the somatic sample passing from the 3L to the 5L generates an increase on the Spearman's rank correlation coefficient for all dimensions, on the mental care subsample this change diminishes the coefficient for mobility and self-care dimensions. Linking this with the points above, this could be an effect of the existence of more likely and of higher order inconsistencies on this group.

Replicating the analysis but restricting to consistent responses only, it is found that the somatic sample remains quite impassive to this sample change, while the subgroup of mental patients changes significantly increasing on a notable way all coefficients regarding the relationship between 3L and 5L, and improving the relationship between the VAS and the 5L. This data screening is able to explain the decreases of “mobility” and “self-care” dimensions correlations with the VAS for the full sample of mental care patients when passing from 3L to 5L.

Table 7: Spearman’s rank correlation between VAS, 5L and 3L

		SOMATIC			MENTAL		
		3L,5L	VAS,3L	VAS,5L	3L,5L	VAS,3L	VAS,5L
Full samples	MO	0.835	-0.450	-0.490	0.811	-0.337	-0.282
	SC	0.883	-0.475	-0.518	0.821	-0.354	-0.345
	UA	0.841	-0.540	-0.563	0.816	-0.584	-0.605
	PD	0.726	-0.468	-0.529	0.845	-0.527	-0.548
	AD	0.822	-0.349	-0.389	0.845	-0.578	-0.618
Consistent Responses	MO	0.872	-0.459	-0.494	0.918	-0.314	-0.324
	SC	0.924	-0.470	-0.516	0.944	-0.343	-0.372
	UA	0.894	-0.540	-0.583	0.915	-0.588	-0.637
	PD	0.759	-0.484	-0.526	0.912	-0.535	-0.565
	AD	0.866	-0.353	-0.381	0.885	-0.579	-0.624

Figure 3: Correlations between the 3L and 5L scales and the VAS, by subsample of patients using consistent responses



We find that the higher differences with regard to the correlation between dimensions and the VAS are found on "mobility" -dimension in which somatic coefficient is much more correlated - and "anxiety/depression" –where we find a notable higher correlation for mental patients. These findings point out the existing heterogeneity on the weights given to each dimension when explaining the VAS and more particularly, how the type of illness affects these weights. Also important on this aspect is the fact that the change from 3L to the 5L version has not effect on attenuating this relationship between type of illness and correlation (VAS, Dimension) but, on the contrary, it seems to enhance it.

As an auxiliary instrument to evaluate the property of Convergent Validity we calculate the Chronbach- α (table 8), finding that it is higher for the 5L scale (0.8175) than for the 3L scale (0.7789), meaning that there is a stronger dependence between the items in the EQ-5D-5L than in the 3L questionnaire. Once again, analyzing the Cronbach's alpha for each group, differences arise. The scale reliability coefficient (Cronbach's alpha) is for both questionnaires higher on the somatic subsample. Moreover, using the 5L increases the gap between these two subsamples, pointing out that the effect of the 5L instrument is deeper for somatic patients. This result comes at hand by the slight increase generated by the 5L-instrument on the mental sample with respect to the 3L instrument.

Table 8: Convergent Validity

		5L	3L
Full Sample	Average interitem covariance	0.495718	0.138048
	Number of items in the scale	5	5
	Scale reliability coefficient	0.8175	0.7789
Somatic	Average interitem covariance	0.503487	0.13088
	Number of items in the scale	5	5
	Scale reliability coefficient	0.8547	0.7936
Mental	Average interitem covariance	0.50678	0.150926
	Number of items in the scale	5	5
	Scale reliability coefficient	0.8087	0.7898

Discriminatory Power/Informativity: The number of health profiles without problems (11111) decreases from the 3L to the 5L for both subsamples as well as for the full sample, being this decrease statistically significant at p<.05 for all cases. Looking dimension by dimension, for the somatic sample it is found a highly significant (p<.01) reduction of "no problem" responses in all dimensions. In opposition, for the mental subsample it is found a significant reduction (p<.05) of "no problem" responses only for "usual activities" and "anxiety/depression".

Table 9: Proportion of "no problem" responses

	Full Sample			Somatic			Mental		
	3L	5L	Change (%)	3L	5L	Change (%)	3L	5L	Change (%)
11111	14.98	12.6	-15.89***	14	12.4	-11.43**	16.1	12.84	-20.25***
Mo	55.3	52.29	-5.44***	42.35	38.37	-9.4***	70.07	68.47	-2.28
Sc	68.82	67.26	-2.27*	63.24	59.61	-5.74***	75.23	76.13	1.20
Ua	48.99	44.99	-8.16***	50.3	45.83	-8.89***	47.5	44.02	-7.33**
Pd	31.36	27.9	-11.03***	23.47	18.48	-21.26***	40.45	38.83	-4.00
Ad	35.74	31.17	-12.79***	42.15	37.04	-12.12***	28.41	24.38	-14.19***

* Difference between 3L and 5L significant at p<.1
** Difference between 3L and 5L significant at p<.05
*** Difference between 3L and 5L significant at p<.01

Table 10: Informativity through Shannon Index (H') and Shannon Eveness Index (J')

		5L			3L			
		H'	MaxH'	J'	H'	MaxH'	J'	Change
SOMATIC	mo	1.89	2.32	0.81	1.09	1.58	0.69	15.30%
	sc	1.55	2.32	0.67	1.04	1.58	0.66	1.14%
	ua	1.84	2.32	0.79	1.21	1.58	0.77	2.97%
	pd	1.90	2.32	0.82	1.15	1.58	0.73	10.80%
	ad	1.80	2.32	0.78	1.23	1.58	0.78	-0.34%
MENTAL	mo	1.37	2.32	0.59	0.96	1.58	0.61	-3.00%
	sc	1.13	2.32	0.49	0.85	1.58	0.54	-10.83%
	ua	1.98	2.32	0.85	1.35	1.58	0.86	-0.48%
	pd	2.01	2.32	0.87	1.38	1.58	0.88	-1.15%
	ad	2.21	2.32	0.95	1.45	1.58	0.92	3.41%

We observe that for the subsample of patients with somatic illnesses the level of informativity increases by using the 5L scale in four of the five dimensions. However, for those patients with mental diseases, we observe the opposite: informativity decreases in four of the five dimensions, observing an increase in informativity only for anxiety/depression. Although we can think this is not a good result, we cannot ignore the correlations between the dimensions and the VAS, as long as loses on informativity take place on these dimensions with lower correlation. This relationship tells us that, in fact, the 5L-version is able to increase the informativity of the EQ-5D on the relevant patient-specific dimensions. This asymmetry on the informativity variation over dimensions was expected given our sample and the methodology to compute this measure. Scores on dimensions not related to the illness of the respondent (mobility for mental care patients, for example) tend to be concentrated on low values and thus, the extra levels for high punctuations will remain unused, reducing the value of J'.

However, we find this method to compute the information gain/loss by switching from the 3L to the 5L quite theoretical. From a more practical point of view, we argue that the scale more valuable in terms of informativity is the one that can split the sample into the highest number of "meaningful" groups. Obviously, the 3L-instrument split the sample into three groups on each dimension (3L-1, 3L-2 and 3L-3) while the 5L-instrument do it into five (5L-1, 5L-2, 5L-3, 5L-4 and 5L-5), but the question we pose here is: Do these groups have a meaning by themselves?. Each of these subgroups would allow us to assign statistically a particular rank of the VAS to each individual belonging to it (the p-confidence interval of the mean). However, it may well be the case that this rank for two groups is overlapped. In such a case, one can argue that these two groups are not separable and thus they should merge into one. From a mathematical point of view what we are looking is for the existence of an injective correspondence such that

$$c: L \rightrightarrows VAS \subseteq [0,100] \quad s.t. \; c(i) \cap c(j) = \emptyset, \qquad \forall i,j \in L \; \& \; i \neq j$$

with $\mathrm{L} = \{1,2,3\}$ for the 3L version and $\mathrm{L} = \{1,2,3,4,5\}$ for the 5L. For example computing the mean in mobility for the 5L-2 on the mental care sample, we get that its 95% confidence interval is equal to (46.13402, 56.5882) i.e. c(2)= (46.13402, 56.5882) but then, computing c(3) we find that c(3)=(40.45239, 55.54761). Since $c(2) \cap c(3) \neq \emptyset$, level 3 doesn't really discriminate on regard the VAS with respect to level 2 and thus it doesn't add information to the data set.

Making use of this reasoning, we suggest the use of the following index for each dimension:

$$I^k = \frac{\#\{(i,j) \in L \times L | i \neq j \; \& \; c(i) \cap c(j) = \emptyset\}}{\#\{(i,j) \in L \times L | i \neq j\}}$$

where k denote the dimension analyzed. This index is just the number of not overlapping groups (numerator) over the maximum number of not-overlapping groups (denominator). Consequently, the index reflects the proportion of useful or meaningful groups. We will call it **Effective usefulness.** This index has an upper bound of one, and would correspond to the existence of an injective correspondence from the levels of the dimension to the VAS. In this case all levels would be perfectly distinguishable between them in terms of VAS and thus all levels contain unique information. The minimum score would be zero, and corresponds to the case in which all confidence intervals overlap between them and thus the scale itself provides zero information.

However, it is easy to seethat the gain on informativity from increasing the number of levels may depend on the actual number of levels[2]. Therefore we assume this relationship to be concave reflecting the fact that the potential gain on informativity is decreasing on the number of levels[3]. For this purpose, we modify the above formula to compute an **Absolute index** (A^k) on the following way:

$$A^k = I^k \ln(Max\{L\})$$

where $Max\{L\}$ just represents the number of levels of the scale (Max{L}=5 or Max{L}=3 depending on the instrument). This absolute index accounts for the potential informativity of the scale ($\ln(Max\{L\})$) weighted by the effective usefulness of it (I^k). Notice that this index has as

[2] Notice that a comparison based only on I^k would imply that having just one group is as informative as having 100 meaningful groups.

[3] This assumption just implies that the most important splits are the first ones as they generate a clear and powerful discrimination. As illustration, consider a dataset. Clearly the first (meaningful) split i.e. from passing from one group to two, is more important to a new split that generates 1001 groups instead of 1000.

lower bound of zero, which is attained when the effective usefulness of the scale is zero or on the trivial case in which there is just one level.

In order to aggregate this information into an **Overall index** (OI) reflecting the informativity of the scale among all dimensions, we add all the A^k indexes, weighted by the respective correlations between the dimensions and the VAS (in absolute value), i.e.

$$OI = \sum_{k} |Corr\ VAS, k\ | A^k$$

Looking at the results for our dataset (table 11), we see that the overall index of informativity (OI) increases in both, mental and somatic groups, when we change the 3L version for the 5L. Looking closer, it is found that for somatic patients all dimensions gain in informativity changing from 3L to 5L. Moreover, results are consistent with the ones provided by the Shannon Evenness Index (more by coincidence than by construction), with “mobility” and “pain/discomfort” as the principal benefited dimensions, followed by “usual activities” and “self-care”. The less benefited found is in this case “anxiety depression”.

Although the ranking for dimensions on regards to informativity is the same as the one provided by the Shannon Evenness index, the magnitude is critically different. Especially striking is the change on “anxiety/depression”, dimension for which the Shannon Evenness index accounts for a loss while our absolute index for a gain. In our case, the gain is explained by the fact that the increase on the potential informativity of the scale is not completely counteracted by the decrease of the effective usefulness (I^{ad}). Although this decrease is quite important, conceptually we can understand that we can construct three groups that discriminate correctly: 5L-1,5L-2 and 5L-[3,4,5], where this last subgroup comes from the merge of the three overlapping groups. Thus we can say up to here that we have three effective subgroups on this 5L scale. However, we also have the extra information that 5L-1 and 5L-2 discriminate effectively versus 5L-3, 5L-4 and 5L-5 so in fact we have more information that the one that follows from the existence of three groups. Therefore, the information we get from the introduction of the new levels is -although of small magnitude- positive. The result for the rest of dimensions is easily explained by looking the tables. On self-care we get that only 5L-5 is problematic and thus effective usefulness of the scale does not seem very affected. Similar argument can be used to explain the result on the “usual activities” dimension.

Regarding mental care patients, things become harder to interpret. First thing we should notice is that this new indicator of informativity provides results completely different to the ones of the Shannon Evenness index. All dimensions gain on informativity except the one of mobility. The decrease on this dimension is clear when examining the table: From getting three groups completely meaningful on the 3L version, we get none on the 5L. This dramatic change is caused by the extreme behavior of 5L-5, that avoids the existence of any injective relationship, i.e. the mean VAS of the 5L-5 cannot be distinguished from any other group. To this, we should also add that the mean behavior of the VAS on the 5L-2 and 5L-3 cannot either be distinguished. This dramatic change on the effective usefulness of the 5L scale, push down the Absolute index up to get a negative change. The case of self-care dimension, is similar in essence to the case of anxiety depression for somatic patients except for the fact that the 3L scale on this case is not injective (it fails for 3L-2 and 3L-3) and thus the change on the effective usefulness index when passing from the 3L to the 5L is of lower magnitude. On usual activities and pain discomfort, only the pairs 5L-4 and 5L-5 fail to be distinguishable and for the anxiety depression dimension, the relationship between 5L and VAS remain completely injective attaining in both 3L and 5L the maximum absolute index.

Table 11: Informativity through Effective Index (I^k), Absolute Index (A^k) and Overall Index (OI)

Somatic

	mo	sc	ua	pd	Ad	OI	mo	sc	Ua	pd	ad	OI
1	73.63	71.66	73.76	79.11	71.65		73.46	71.08	73.58	77.18	71.16	
2	65.37	60	63.38	67.47	64.92		58.89	53.77	57.83	62.46	60.59	
3	56.27	49.58	53.85	57.37	55.46		24.37	37.5	30.27	43.35	48.54	
4	45.77	37.21	38.48	47.93	45.67							
5	12.5	30	34	24	44.17							
Overlap	0	2 (3,5)(4,5)	1 (4,5)	0	3 (3,4)(3,5)(4,5)		0	0	0	0	0	
Not Over	10	8	9	10	7		3	3	3	3	3	
I^k	1	0.80	0.9	1	0.7		1	1.00	1	1	1	
A^k	1.61	1.29	1.45	1.61	1.13	3.56	1.10	1.10	1.10	1.10	1.10	2.49
Correlation with VAS	-0.4918	-0.5199	-0.562	-0.527	-0.392		-0.449	-0.471	-0.5367	-0.467	-0.345	

Mental

	mo	sc	Ua	pd	Ad	OI	mo	sc	ua	pd	ad	OI
1	61.33	61.71	71.51	70.08	74.09		62.40	62.36	70.26	70.26	73.59	
2	51.36	48.28	54.02	60.03	65.72		45.77	41.53	49.81	53.70	57.58	
3	48.00	35.00	45.01	47.91	52.90		0.83	16.67	20.64	23.33	29.14	
4	17.69	26.36	28.56	28.64	37.00							
5	51.25	2.50	27.84	19.15	22.48							
Overlap	5 (2,3)(1,5)(2,5)(3,5)(4,5)	4 (2,3)(3,4)(3,5)(4,5)	1 (4,5)	1 (4,5)	0		0	1 (2,3)	0	0	0	
Not Over	5	6	9	9	10		3	2	3	3	3	
I^k	0.5	0.60	0.9	0.9	1		1	0.67	1	1	1	
A^k	0.80	0.97	1.45	1.45	1.61	3.23	1.10	0.73	1.10	1.10	1.10	2.48
Correlation with VAS	-0.2861	-0.3458	-0.603	-0.552	-0.617		-0.336	-0.354	-0.5804	-0.527	-0.577	

	Change Somatic						Change Mental					
	mo	sc	ua	pd	ad	OI	mo	sc	ua	pd	ad	OI
I^k	0,0%	-20,0%	-10,0%	0,0%	-30,0%		-50,0%	-10,0%	-10,0%	-10,0%	0,0%	
A^k	46,5%	17,2%	31,8%	46,5%	2,5%		-26,8%	31,8%	31,8%	31,8%	46,5%	
OI						43,1%						30,3%

Face Validity: We analyze here the opinion reported by patients concerning the capacity of both questionnaires to reflect individuals' health status. A first look at the data gives us the following result: the proportion of individuals who report to prefer the 5L questionnaire (24.7%) is significantly different (p<.01) than the proportion of individuals preferring the 3L questionnaire (18%). However the majority of respondents are indifferent between both questionnaires.

Table 12: Face validity by subsamples of patients

	Somatic		Mental	
	%	Std. Err.	%	Std. Err.
5L ***	28.48%	0.020307	20.43%	0.019673
3L **	15.15%	0.016132	21.38%	0.020005
Indifferent	48.89%	0.022491	50.36%	0.024397
N/A	7.47%	0.011832	7.84%	0.013115

* Difference between samples significant at p<.1

** Difference between samples significant at p<.05

*** Difference between samples significant at p<.01

Performing separate analysis for somatic and mental patients (table 12), we observe that for somatic patients we find a higher support for 5L than for the 3L (p<.01) but when looking the mental care subsample this relationship is no longer significant. Consequently with these findings, we also get that somatic patients are more likely to express their preference toward the 5L and less likely toward 3L than mental ones (p<.01 and p<.05 respectively).

Another interesting result is found when we look at preferences for 3L and 5L questionnaires distinguishing by the self-perception of the patient of his illness' severity (table 13). Examining somatic patients, it is found that those patients with severe illness' perception tend to report their preference towards the 5L significantly more frequently than those with intermediate perception (p<.05) that, at the same time, report it more frequently than those with mild perception (p<.05).

Doing the same exercise for mental care patients, not effect of self-perception of the illness on this regard it is found. Only the proportion of indifferent responses presents a marginal statistically significant effect (p<.1) when comparing those reporting a mild perception with those reporting a severe one.

Table 13: Face validity by subsamples of patients

		Somatic		Mental	
		%	Std. Err.	%	Std. Err.
Mild *	5L	16.13%	0.038346	16.49%	0.037879
	3L	11.83%	0.033669	20.62%	0.041291
	Indifferent	56.99%	0.051617	57.73%	0.050417
	N/A **	15.05%	0.037282	5.15%	0.022567
Intermediate	5L	28.35%	0.028334	22.34%	0.03046
	3L	17.72%	0.024004	20.21%	0.029367
	Indifferent	47.64%	0.0314	50.53%	0.036562
	N/A	6.30%	0.015274	6.91%	0.018553
Severe ***	5L ***	39.37%	0.043525	21.19%	0.037778
	3L **	14.17%	0.031071	24.58%	0.039803
	Indifferent	40.94%	0.043807	44.92%	0.045985
	N/A	5.51%	0.020331	9.32%	0.026879

* Difference between samples significant at p<.1

** Difference between samples significant at p<.05

*** Difference between samples significant at p<.01

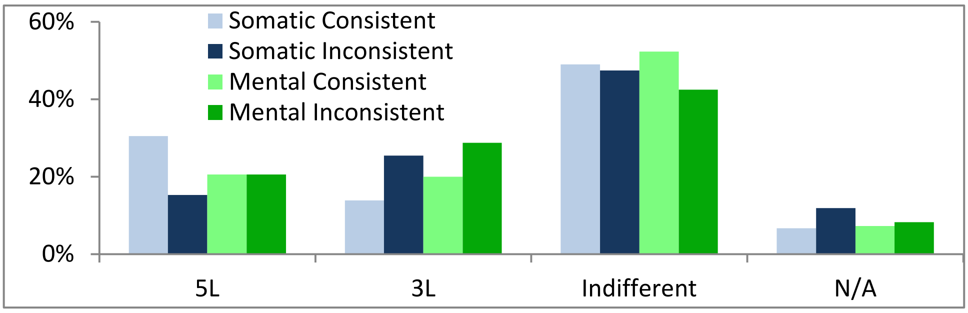
Closing this point, we can compare the patterns of both subsamples within a severity reported level by performing a simultaneous test of multiple constraints. This test will shed light about whether we can reject that somatic and mental care patients' data are dropped from the same distribution. This exercise tell us that distributions differ marginally on the mild level (p<.1)[4] and on a very strong way on the severe level (p<.01).

The last point to be considered is the effect of being inconsistent on the perceived usefulness of the survey mode (Figure 4). Analyzing this factor as a potential determinant of preference for one or another survey, we found that the proportion of patients who strictly prefer the 5L to the 3L survey is significantly higher among consistent than for inconsistent ones (p<.05). Similarly, we can say that this relationship is reversed when analyzed the strict preference between 3L and 5L survey (p<.01). Performing a simultaneous test of multiple constraints, we can reject that both distributions are equal (p<.05) pointing out the importance for the patient of being or not consistent when evaluating the usefulness of the survey. Withal, and once again, we have to talk about differences when evaluating its effect on the somatic or mental care group of patients. While for the former group, the results of the whole subsample can be extrapolated, for the second one, we cannot reject the hypothesis of equal distribution with respect to being consistent/inconsistent. More important still is the fact that the distribution of somatic-inconsistent patients and the one of mental-inconsistent ones look so similar that we

[4] This result however seems to be more an accident than a fact, since on the somatic subsample there is an unusual high proportion of "not applicable" responses.

cannot reject that these are the same, i.e. we can assume that their distribution of responses are the same. On the other hand, we can reject the equality of distributions ($p<.01$) when we look only at consistent patients.

Figure 4: Face validity results, considering inconsistencies



3.3 Other results

We can see that there is a significant difference between somatic and mental patients in how they value their health in the VAS. There is a positive difference of 7.6 points, in mean, in favor of somatic patients. A t-test assuming unequal variances for the two subsamples returns us a t-statistic of 5.2668 (Satterthwaite's degrees of freedom equals 784.791), rejecting the null hypothesis ($p<.01$) of equality on means and accepting that somatic patients tend to report higher VAS than mental ones do.

However, this can be due to differences in the sample composition. Apart from the observed differences on the descriptive data between mental and somatic patients on regard to gender composition (“gender”), level of education (“studies”) and the main economic activity (“activity”) we have an extra source of imbalance between our two samples: The institution (primary, specialist and day care hospital) in which the survey has been delivered.

Table 14: Distribution of patients attending to the institution where the questionnaire was delivered

	Primary	Specialist	Day Care	Total
Somatic	326	190	0	516
Mental	3	284	157	444
Total	329	474	157	960

These features could be responsible for a problem of omitted variable biasing the above analysis, since if a correlation between a variable and the VAS exists, the above difference may be explained not by being a mental or somatic patient but by a correlation between this classification with the variable really relevant.

Performing an ANCOVA (table 15) introducing all the relevant information to shed light over this issue, we find that there is not direct link between mental/somatic classification and the VAS. Moreover, performing the same analysis but excluding the variable "institution" it is found that the variable mental is highly significant ($p<.01$) providing evidences that what has carried out the difference between mental and somatic VAS's score is its imbalanced on regard the institution in which patients are being treated. Looking the coefficients of the variable "institution", it is found that patients being attended on the mental care centre and by the specialist, score significantly lower ($p<.01$) than those attended on a primary health care centre, being the rest of variables equal. Moreover, can also be checked that patients from the day care hospital penalize their VAS more than those coming from the specialist ($p<.01$).

Table 15: ANCOVA using the VAS as response variable

	Partial SS	Df	MS	F	Prob>F
Model	206,747.99	34	6,080.82	27.36	0
Mobility	287.69	4	71.92	0.32	0.8623
Self-Care	1,609.83	4	402.46	1.81	0.1248
Usual Activities	11,408.81	4	2,852.20	12.83	0
Pain Discomfort	7,715.32	4	1,928.83	8.68	0
Anxiety/Depression	14,557.92	4	3,639.48	16.37	0
Gender	22.99	1	22.99	0.1	0.7479
Studies	1,881.32	3	627.11	2.82	0.038
Activity	1,550.55	5	310.11	1.4	0.2238
Severity	1,145.72	2	572.86	2.58	0.0766
Institution	3,948.45	2	1,974.22	8.88	0.0002
Mental	3.99	1	3.99	0.02	0.8935
Residual	180,943.76	814	222.29		
Total	387,691.75	848	457.18		

4. Conclusion and final remarks

This paper analyzes the psychometric properties of the new EQ-5D-5L questionnaire through a parallel testing of the standard 3L and 5L versions of the EQ-5D instrument in two well established populations of chronic patients (mental/somatic).

Results show that, for both populations of chronic patients analyzed, the new EQ-5D-5L is sensitive to the introduction of two intermediate severity levels. Although this assertion is also true for each of the subsamples, the strength of the sensitiveness is mixed.

In all dimensions, it seems that somatic patients take more advantage of the extra levels introduced by the EQ-5D-5L. This group, in comparison with the mental one, reduces to a greater extent the missing response rate, commit less (and of lower importance) inconsistencies, get a higher correlation of the 5L scale and the VAS within the 3L levels, complement better the dimensions to get an overall score (measured through the Cronbach's alpha), reduce in a more significant way the "no problem" response and the informativity gain is also superior (for both the Shannon Eveness Index and our Absolute Index).

This higher performance of the EQ-5D-5L on somatic patients is endorsed by a higher preference of somatic patients toward the 5L version of the questionnaire than to the 3L one. A critical factor affecting this preference toward one or another version of the EQ-5D is whether the patient is consistent or inconsistent. Results show that mental patients do more and higher inconsistencies. In addition, this feature seems to be supported by previous research on the topic. Comparing our result with the ones of Janssen et al. (2012) and Kim et al. (2012), we find a higher proportion of inconsistent responses (3.8% versus 2.9% and 3.5% respectively). However, excluding mental patients from the analysis the proportion of inconsistent responses decreases to 2.9%, exactly the same found by Janssen et al. (2012).

Looking to the weights of inconsistencies (table 16) we get mixed effects replicating somatic patients better the results of Kim et al. (2012) and mental ones slightly better in Janssen et al.'s ones. However, we should be aware that this last result is explained by the huge difference existing on the dimension usual activities; while for the remaining four dimensions somatic patients minimize differences. Withal, when looking at this result we should remember that the dataset of Janssen et al. makes use of a non-negligible sample of mental patients (18% of the total), which according to the results presented above, may increase the average inconsistency. This point seems strengthened by noticing that at each dimension (excluding self-care) the dataset with lower average inconsistence is our somatic one or Kim et al.'s one (composed by cancer patients) followed by Janssen et al.'s and, finally, with the higher average inconsistency weight, our mental sample. Looking the average inconsistency weight for the full samples, we can notice a trend increasing on the mental care patients proportion used to generate the sample.

Table 16: Inter-research comparison of mean inconsistency weights

	Somatic	Mental	Kim (2012)	Janssen (2012)
Mo	1.00	1.32	1.09	1.146
Sc	1.00	1.00	1.21	1.186
Ua	1.00	1.29	1.08	1.197
Pd	1.08	1.25	1.06	1.131
Ad	1.16	1.25	1.14	1.185
Average Inconsistency Weight	1.05	1.222	1.12	1.17
Distance w.r.t. Kim	0.24	0.43		
Distance w.r.t. Janssen	0.31	0.30		

Therefore, our results, as well as this basic comparison with similar studies seem to confirm that mental patients commit more and higher inconsistencies.

This point, together with our finding suggesting that inconsistent patients tend to underestimate the utility of the 5L version, open the debate of whether the introduction of intermediate levels is really helping patients with mental illnesses or, on the contrary, is introducing an extra difficulty that is not corresponded with a gain on the quality of the data collected.

This research should be taken as a caveat about the importance of other factors different from the mere 3L and 5L states when evaluating the convenience and suitability of the different versions of the EQ-5D. On this paper we have analyzed the somatic/mental illness classification, but may be other relevant classificatory variables affecting the impact of moving from the 3L to the 5L. While our results show the suitability of the 5L version in both subsamples, they also show that it is much more suitable and effective for somatic patients. These subsamples' differences may be a concern when aggregating and comparing different data. While an aggregation will provide us with a general and wide picture of the effects, it would also provide us non-realistic results as long as these results are an average of well distinguishable groups that may have different reactions to the introduction of the two intermediate levels.

We have also made an effort in suggesting new ways to measure the effects of the introduction of intermediate levels, understanding that there are several complementary ways to measure it but also that the tools developed up to now may be insufficient to cover all the relevant aspects that should be analyzed on this particular context. This should not be read as a critic but as an encouragement to introduce new methodologies and properties in future researches that may shed light over the topic at hand.

References

Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., Swinburne, P. & Busschbach, J. (2012). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2012 Nov 25. [Epub ahead of print].

Janssen, M. F., Birnie, E., Haagsma, J. A., & Bonsel, G. J. (2008). Comparing the standard EQ-5D three-level system with a five-level version. *Value in Health*, *11*(2), 275-284.

Kim, S. H., Kim, H. J., Lee, S. I., & Jo, M. W. (2012). Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea. Quality of Life Research, 21(6), 1065-1073.

Kim, T. H., Jo, M. W., Lee, S. I., Kim, S. H., & Chung, S. M. (2012). Psychometric properties of the EQ-5D-5L in the general population of South Korea. Qual Life Res. 2012 Dec 6. [Epub ahead of print]

Pickard, A. S., De Leon, M. C., Kohlmann, T., Cella, D., & Rosenbloom, S. (2007). Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Medical care, 45(3), 259-263.

Scalone, L., Ciampichini, R., Fagiuoli, S., Gardini, I., Fusco, F., Gaeta, L. Del Prete, A, Magini, G, A Baldan,A and Mantovani, L. G. (2012). 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.Nov 29. [Epub ahead of print]