

Research

Open Access

Evaluation of the Functional Assessment of Cancer Therapy-General (FACT-G) Spanish Version 4 in South America: Classic Psychometric and Item Response Theory Analyses

Juan J Dapuerto*¹, Carla Francolino¹, Liliana Servente², Chih-Hung Chang³, Irene Gotta², Roberto Levin⁴ and María del Carmen Abreu¹

Address: ¹Departamento de Psicología Médica, Facultad de Medicina – Universidad de la República, Uruguay, ²Departamento de Biofísica, Facultad de Medicina – Universidad de la República, Uruguay, ³Buehler Center on Aging, Center for Healthcare Studies, Division of General Internal Medicine, and the Department of Medicine, Feinberg School of Medicine, Northwestern University, Illinois, USA and ⁴Departamento de Oncología Clínica, Facultad de Medicina – Universidad de la República, Uruguay

Email: Juan J Dapuerto* - jdapuerto@hc.edu.uy; Carla Francolino - carlaf@hc.edu.uy; Liliana Servente - lservernt@fmed.edu.uy; Chih-Hung Chang - chchang@northwestern.edu; Irene Gotta - igotta@fmed.edu.uy; Roberto Levin - bdr1@adinet.com.uy; María del Carmen Abreu - mabreu@hc.edu.uy

* Corresponding author

Published: 20 August 2003

Received: 16 May 2003

Health and Quality of Life Outcomes 2003, **1**:32

Accepted: 20 August 2003

This article is available from: <http://www.hqlo.com/content/1/1/32>

© 2003 Dapuerto et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The FACT-G has gone through many validation studies. However, little research has been conducted in South American Spanish speaking patients. The present study aimed to evaluate the FACT-G Spanish Version 4 in Uruguayan cancer patients.

Methods: The data analyzed were collected from 309 patients, with various tumor sites, at different stages of disease and receiving different treatments.

Results: Reliability was evaluated using Cronbach's coefficient alpha and showed high internal consistency for each of the subscales and its total scale (range = .78 – .91) of the FACT-G. The FACT-G total score also showed significant mean differences among known groups (performance status, in vs. outpatients) when tested by ANOVA and t-test. When the tumor stage (Local and Regional vs. Metastatic disease) was used as a clinical anchor, the FACT-G total score, the Physical Well-being (PWB), and Functional Well-being (FWB) subscale scores showed mean differences, ranging from 5 to 10 points in a scale from 0–108 (effect sizes = 0.30–0.60). Item response theory (IRT)-based evaluation using mean square fit statistics (.60–1.4) criteria showed that only two items misfit: "Estoy satisfecho(a) con mi vida sexual" (I am satisfied with my sex life) and "Estoy satisfecho(a) de cómo estoy enfrentando mi enfermedad" (I am satisfied with how I am coping with my illness).

Conclusion: The results indicated that, using both traditional and IRT approaches, the Spanish FACT-G has good reliability and validity to be used as a QOL instrument among Uruguayan cancer patients.

Background

The Functional Assessment of Cancer Therapy-General

(FACT-G) Questionnaire, designed to measure quality of life (QOL) in cancer patients [1], has gone through many

validation studies, both in its English and translated versions [2–5]. However, little research has been conducted in South American Spanish speaking patients. A previous study performed among Uruguayan cancer patients using FACT-G Spanish Version 2 [6] showed acceptable to good reliability and validity except for the emotional well-being subscale. Cella and his colleagues also recognized this potential problem and subsequently revised the items in a new Spanish Version 3 and its most recent Spanish Version 4 [2] in order to improve its reliability. Major changes include adding one previously available but not scored item to the scoring algorithm ("I worry my condition will get worse" – "Me preocupa que mi enfermedad empeore"), rephrasing some other items to improve readability and removing the 2-item Relationship with Doctor subscale. These remarks the need for additional validation studies for the Spanish-speaking cancer patient population in South America.

Another reason for further psychometric studies on FACT-G is that, up to date, its Spanish version has been validated using classic test theory (CTT) approaches like other QOL questionnaires. Traditionally, classical psychometric based procedures have dominated the health status assessment. More recently, item response theory (IRT) measurement models have entered the field and researchers are increasingly enthusiastic for the prospect of deriving better definitions of underlying constructs and the opportunity to turn attention away from static tests and scales to items and the incremental information they provide [7]. The present study aimed to evaluate the performance of the FACT-G Spanish Version 4 in Uruguayan cancer patients using both classic psychometric and item response theory based approaches.

Methods

Patients

The data were collected from cancer patients, between 18 and 75 years of age, with various tumor sites, at different stages of disease and under different forms of treatment. To be eligible patients must have been fluent in Spanish. Potential participants were identified from the daily record of office visits, treatment visits and inpatient hospitalizations. To ensure sufficient experience with treatment related side-effects, patients must have completed a minimum of two cycles of chemotherapy and/or 10 radiation therapy sessions or one month of hormone therapy. There must have been at least one month since last surgery. To ensure heterogeneity of the socioeconomic features, patients from one private (Centro de Asistencia del Sindicato Médico del Uruguay – CASMU) and three public (Hospital de Clínicas de la Universidad de la República, Instituto de Oncología del MSP, Servicio de Oncología Radioterápica del Hospital Pereira Rossell) hospitals of the city of Montevideo were recruited for the study.

Patients with ostensible cognitive deficits or serious psychiatric dysfunctions were excluded. Ability to give informed consent was required. Approval from the corresponding ethics committees was obtained.

Measures

Patients were assessed using a battery of instruments. Two physician rated QOL questionnaires, the ECOG Performance Status Rating and the Spitzer's Quality of Life-Index doctor version (QLI-*d*) were completed either by the treating physician or oncologist. Patient self-reported questionnaires included the Functional Assessment of Cancer Therapy-General (FACT-G) Spanish Version 4, the Spitzer's Quality of Life-Index patient version (QLI-*p*), the Profile of Mood States, Short Form (POMS-SF) and the Marlowe-Crowne Social Desirability Scale (MCSDS).

ECOG Performance Status Rating (ECOG PSR) is a five-point scale [8] ranging from 0 (fully ambulatory) to 4 (not being able to leave bed).

The *Quality of Life Index* (QLI) [9] is a five-item questionnaire where each one of them explores a dimension or domain of quality of life: health, activity, daily living, support and outlook. Every item has three response categories indicating different levels of functional impairment. Although it was originally developed as an observer rated scale (QLI-*d*), it can also be used as a patient rated scale (QLI-*p*). For the purpose of the study, the QL-I was translated into Spanish following a forward and backward translation procedure, carried out by an English native speaking linguist and a native Spanish-speaking English translator. The Spanish versions of the two QLI questionnaires are available from the authors upon request.

The *Functional Assessment of Cancer Therapy-General Questionnaire* (FACT-G) Spanish Version 4 [2] is a widely used QOL instrument. It comprises 27 questions that assess four primary dimensions of QOL: physical (PWB; 7 items), social and family (SFWB; 7 items), emotional (EWB; 6 items), and functional well-being (FWB; 7 items). It uses 5-point Likert-type response categories ranging from 0 = 'not at all' to 4 = 'very much'. The total FACT-G score is the summation of the 4 subscale scores and ranges from 0 to 108. Data from a previous study on the FACT-G Spanish version 2 [6] conducted among Uruguayan cancer patients suggested that its reliability was acceptable to good in all subscales except for the EWB scale. An important question raised by these results was whether these subscales showed sufficient internal consistency to justify their use across cultures and whether there was equivalence of the Spanish EWB to its English counterpart. Ever since, developers of the FACT-G revised the questionnaire into its most recent version 4 [2]. Major changes were the inclusion of an additional item ("I worry

my condition will get worse" – "Me preocupa que mi enfermedad empeore") in the EWB subscale, the removal of the two-item "Relationship with the Doctor" subscale and the rewording of 12 Spanish items to improve their readability.

The *Profile of Mood States, Short Form (POMS-SF)* [10] is a widely used scale measuring subjective mood states, such as anxiety, tension, vigor, depression, fatigue and confusion. The POMS-SF is a valid measure of affective states and psychological adjustment in cancer patients and is available in Spanish. A Total Mood Disturbance score (POMS TMD) may be obtained by summing the five scores of Tension, Depression, Anxiety, Fatigue and Confusion subscales and subtracting Vigor from these scores. Only patients with a 6th grade or higher level of reading abilities were included in the analysis, according to the instrument developers' instructions.

The *Marlowe-Crowne Social Desirability Scale (MCSDS)* [11]. The 10 items short form of the MCSDS provides a measure of the degree to which participants endorse socially desirable characteristics. A validated Spanish version of the questionnaire [12] was completed by the study participants.

Demographic, disease and treatment information was collected from patients, the treating physician and verified by the research assistants with the participants' medical record.

Statistical analysis

Classical psychometric approach for analysis of the data consisted of an examination of the reliability and validity of the FACT-G Version 4. Reliability was examined by internal consistency (Cronbach's coefficient alpha) for each subscale and the overall scale. Alpha coefficients of 0.70 or higher were considered acceptable [13]. Construct validity was assessed by comparing mean differences in FACT-G total and subscale scores according to known groups, i.e. patients performance status, in vs. out-patients and by studying the correlations between the instruments (convergent and discriminant validity).

Several methods have been described in order to establish clinical significance of QOL measures [14]. Anchor-based methods examine the relationship between scores on the instrument whose interpretation is under question (target instrument) and some independent measure (an anchor). This approach requires that 1) the anchor is easily interpretable and 2) there must be appreciable association between the target and the anchor. Differences in scores in relation to the clinical anchors can be then used to set the minimum important difference or clinically meaningful change [15] in order to evaluate outcomes in clinical tri-

als. Anchor-based clinical significance was studied in order to determine clinically significant differences in QOL assessments as measured by FACT-G total and subscale scores, using tumor stage as a definite clinical criterion commonly used in the Oncology field [16]. Differences of 5 to 10 score points in a 100 point scale are considered relevant in determining the clinical significance of QOL measures [17]. To examine the statistical magnitude of the observed differences, each mean difference score was standardized by relating this score to its standard deviation (effect size). An effect size of $d = 0.2$ was taken to indicate a small difference, $d = 0.5$ a moderate difference, and $d = 0.8$, a large one [18].

Rasch analysis

The item response data were analyzed using Andrich's [19–22] rating scale model (RSM). The RSM is an item response theory (IRT)-based measurement model and has been implemented in the WINSTEPS computer program [23]. The RSM specifies two facets (person latent trait, B_n ; item location, D_i), and the step threshold (F_j). The probability of person n responding in response category j to item i can then be expressed by the formula:

$$\ln \left[\frac{P_{nij}}{P_{ni(j-1)}} \right] = B_n - D_i - F_j$$

in which P_{nij} is the probability of person n endorsing or choosing in category j of item i , $P_{ni(j-1)}$ is the probability of person n endorsing or choosing in category $j - 1$ of item i , B_n is the latent trait measure (e.g., fatigue) of person n , and D_i is the location of item i , and F_j is the step threshold between categories $j - 1$ and j . In the present study, for example, F_1 is the transition from intensity category 1 ("not at all") to category 2 ("a little bit") and F_4 is the transition from category 4 ("quite a bit") to category 5 ("very much"). That is the point on the latent trait scale (i.e., PWB) at which two consecutive category response curves intersect.

Item fit statistics

In order to examine the fit of each item to measure a unidimensional construct (e.g., PWB), the infit and outfit mean square (MNSQ) item fit statistics provided in the WINSTEPS program were evaluated. Fit implies meeting the measurement requirements of item homogeneity and unidimensionality. It also indicates the validity of the item calibrations and person measures. Item misfit indicates that an item is not measuring the same underlying construct as other items within the same scale.

The infit MNSQ is an information-weighted fit statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's trait level. The weighting reduces the influence of less informative, low variance, off-target responses. The outfit MNSQ is an

Table 1: Socio-demographic and clinical features of the sample (n = 309)

	N	%
Gender		
Male	139	45.0
Female	170	55.0
Age		
	Mean (SD)	56.7 (12.9)
Marital Status		
Married	164	53.1
Unmarried couple	20	6.5
Divorced	27	8.7
Separated	16	5.2
Widow	37	12.0
Single	45	14.6
Family living		
Lives alone	36	11.7
Lives with family	259	83.8
Lives with others	11	3.6
Lives in institution	3	1.0
Years of education		
Primary School <3 years	36	11.7
Primary School 3 to 5 years	46	14.9
Primary School 6 years	101	32.7
High School 1 to 3 years	51	16.5
High School 4 to 6 years	32	10.4
College/University	40	12.9
Employment		
Full-time	43	13.9
Part-time	17	5.5
Retired	107	34.6
Social welfare	3	1.0
Unemployed	64	20.7
Medical leave	22	7.1
Housekeeping	40	12.9
Others	9	2.9
	n	%
Satisfaction with income		
"I can save"	10	3.2
"I get enough, no money problems"	23	7.4
"I get just fair for my needs"	100	32.4
"I don't get enough for my needs"	136	44.0
"I am in need"	38	12.3
Family monthly income in US dollars		
Means	512	
Median	300	
Range	0 – 5000	
Interquartile range	192 – 571	
Religion		
No	129	41.7
Yes	180	58.3
Tumor site		
Breast	98	31.7
Lung	44	14.2
Colorectal	35	11.3
Head & Neck	24	7.8
Uterus & ervix	20	6.5
Prostate	18	5.8
Testicle	16	5.2
Gastrointestinal	12	3.9
Urothelial	12	3.9

Table 1: Socio-demographic and clinical features of the sample (n = 309) (Continued)

Ovary	8	2.6
Lymphoma	7	2.3
Primitive unknown	7	2.3
Soft tissue	4	1.3
Melanoma	3	1.0
Brain	1	.3
Tumor stage		
In situ	2	.6
Local	64	19.9
Regional	132	41.0
Metastatic	118	37.2

outlier-sensitive fit statistic, more sensitive to unexpected behavior by persons on items far from the person's trait level. These statistics have an expected value of 1.0, and range from 0 to infinity. Values substantially below 1 indicate local dependencies in the data; values substantially above 1 indicate noise. These values are on a ratio scale, so that 1.2 indicates 20% excess noise. We set .60–1.4 for infit MNSQ and outfit MNSQ as cut-off criteria for items with good fit.

Results

A total of 361 patients were approached and asked to participate. Among them, 36 patients (10%) either refused to participate, were too sick to complete the battery of questionnaires or a family member prevented the patient from participating. Data collection for 16 other participants (4%) was not completed due to other reasons. There were no significant differences between participants and non-participants with regard to age, gender, institution or tumor stage, however more non-participants were inpatients (chi-square = 13.5, *p* = .001) and rated higher in the ECOG PSR ratings (chi-square = 7.0, *p* = .03). The study analysis was based on the remaining 309 patients. Detailed demographic and clinical information about these patients can be found in Table 1. Certain facts to be highlighted are that this is a sample of severely ill patients with 78% of them having regional and metastatic diseases; only 13.9% of patients being fully employed, 27% reached less then the sixth year of formal education and with a very unequal income distribution. The distribution of income shows a very wide range in monthly income with an interquartile range of 192 and 571 dollars/month that highlights the kind of low income and unequal distribution that prevails in South American countries. As many as 41.7% of patients in the study sample stated not to have any religion, a typical characteristic of Uruguayan culture, quite dissimilar to other South American countries.

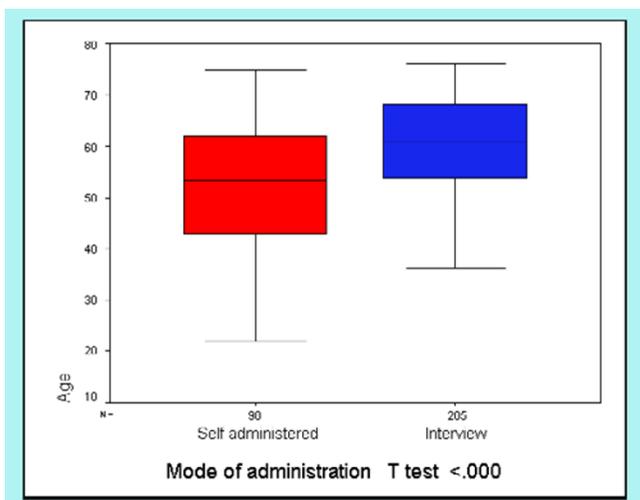


Figure 1
Differences in mode of administration with age

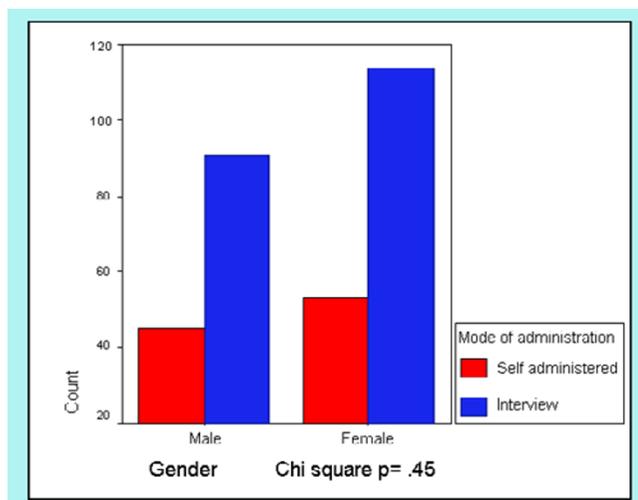


Figure 2
Differences in mode of administration with gender

Mean time since cancer diagnosis was 29 months (SD = 39.5 months; ranged 2 weeks to 24.5 years). As for treatment characteristics, 172 patients had undergone surgery (55.7%); 204 received only one treatment modality, 95 of them were treated with chemotherapy (ChT) (30.7%), 98 with radiotherapy (RT) (31.7%) and 11 with hormone therapy (HT) (3.6%), while 100 patients received a combination of ChT and RT; 4 (1.3%) were treated with RT and HT and 1 (0.3%) with ChT and HT. The timing of testing in relation to treatment varied along the sample. In 119 cases (38.5%), patients were interviewed during the week following their last treatment (ChT cycle or RT session); in 64 cases (20.7%) there had been between one week and one month since the last treatment; in 44 cases (14.2%) between 3 months and a year and 82 (26.5%) had been off treatment for more than one year.

The mode of administration of the FACT-G (self-administration vs. read in interview) was registered in 303 cases. Although the FACT-G is designed for self-administration, most patients in our sample (n = 205, 67.7%) requested some help from the interviewer to fill out the questionnaire. Figures 1,2,3,4 show differences in sociodemographic and somatic features of the two groups of patients. T-test of independent samples showed that those patients capable of self administration were younger ($p < .000$) while chi-square statistics showed that this group of patients had a higher education level ($p < .000$) and better performance status (lower ratings in ECOG scale) ($p < .000$). No differences in gender were observed.

Reliability

The reliability of the each of the FACT-G scales was evaluated with Cronbach's coefficient alpha. The alpha coefficients for the total scale and the four FACT subscales (PWB, SFWB, EWB, FWB) were quite good (ranged .78 – .91; See Table 2). Table 2 also shows alpha coefficients of FACT-G English and Spanish Version 2. No relevant differences were found in Cronbach alpha coefficients for Total FACT-G and its subscales when the group of patients who filled out the questionnaires by self administration (Cronbach's alpha = .78 – .91) and by interview (Cronbach's alpha = .79 – .89) were studied separately.

Validity

Mean score differences among known groups were tested by ANOVA or t-test. Scheffé post-hoc comparisons showed significant differences in FACT-G total and subscale scores according to ECOG PSR (Table 3), indicating that patients with worse functional status rated lower in the QOL multidimensional assessment. Due to the small number, patients rated "3" or "4" on ECOG PSR were combined with individuals rated "2". Total FACT-G scores and subscale scores, with the exception of the PWB subscale, also enabled to discriminate among groups according to patient location at the moment of filling out the questionnaire (out-patient vs. in-patient), see Table 3.

As another evidence of validity, tumor stage (Local vs. Regional vs. Metastatic disease) was examined as a clinical anchor and differences in mean scores and effect sizes were calculated (Table 4). Differences of approximately 5 to 10 % points in total FACT-G score (5.6 to 10.5 points

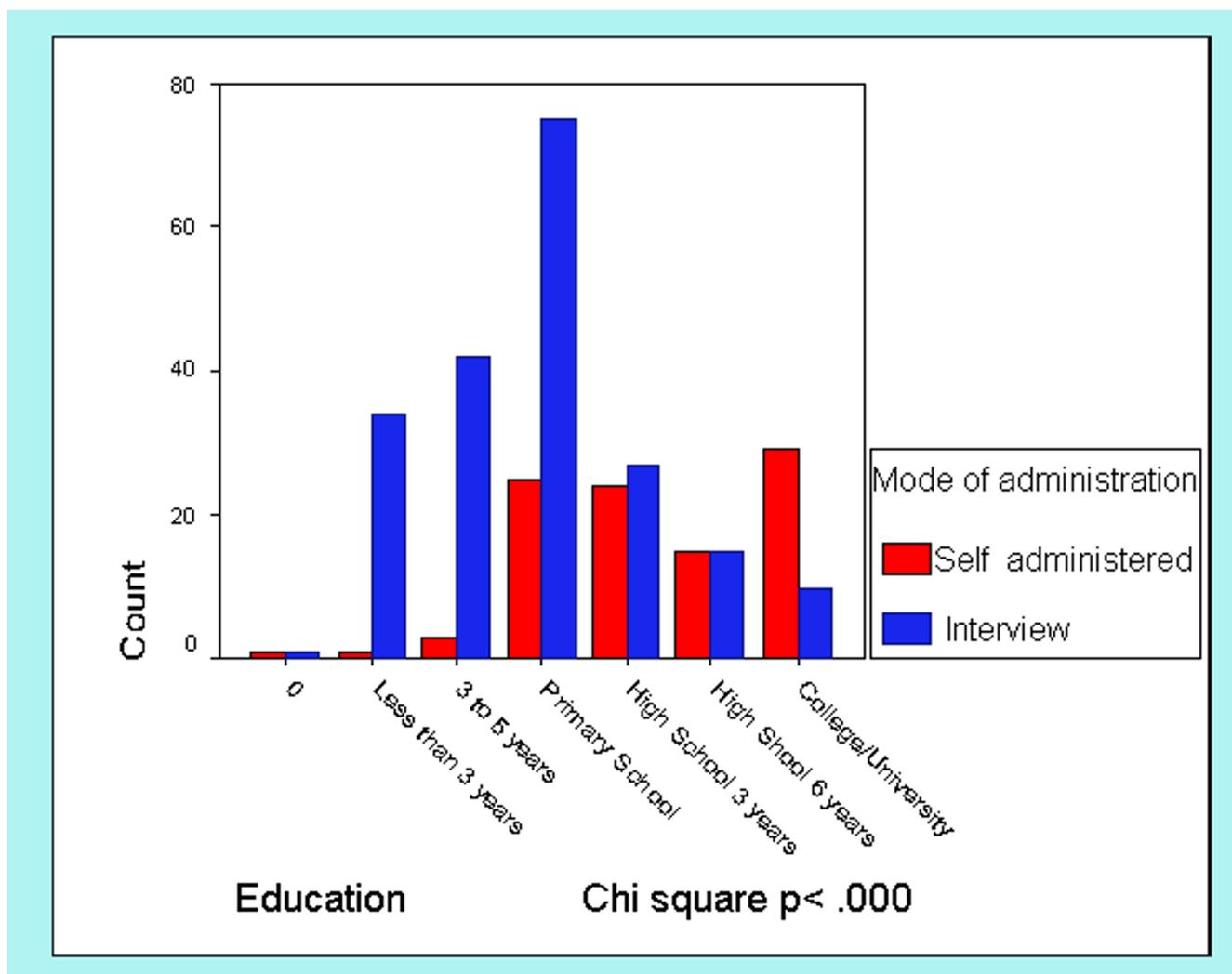


Figure 3
Differences in mode of administration with education

in a 108 point scale) were observed between local vs. metastatic and regional vs. metastatic diseases. Accordingly differences of 11 to 18% points in the PWB subscale score (3 to 5 points in a 28 point subscale) and of 9% in the FWB subscales (2.6 points in a 28 point subscale) were observed between the ratings of patients with local and regional diseases vs. those with spread forms of cancer. Calculated effect sizes of these differences ranged between 0.30–0.60.

Convergent and discriminant validity was evaluated using data from the whole set of patients with the exception of the POMS-SF questionnaire. In this case, only patients that had completed the sixth grade of primary school and higher were included according to the POMS manual's

instructions. As expected, Pearson correlation coefficients were high between the FACT-G total score and its subscale scores, ranging from $r = .53$ to $r = .80$, see Table 5. Similarly, moderate but statistically significant correlations ($r = .30$ and higher) were observed between QL-I total score in both doctor and patient versions as well as between QL-I items and the corresponding FACT-G subscales (QL-Id and QL-Ip Activity, Daily Living and Health and FWB; QL-Id and QL-Ip Support and SFW), (Table 5). Correlation between the FACT-G and the POMS-SF Total Mood Disturbance Score (POMS TMD) is also relatively high. As expected, the correlation coefficients between the FACT-G total and subscale scores and the MCSDS-10 are low.

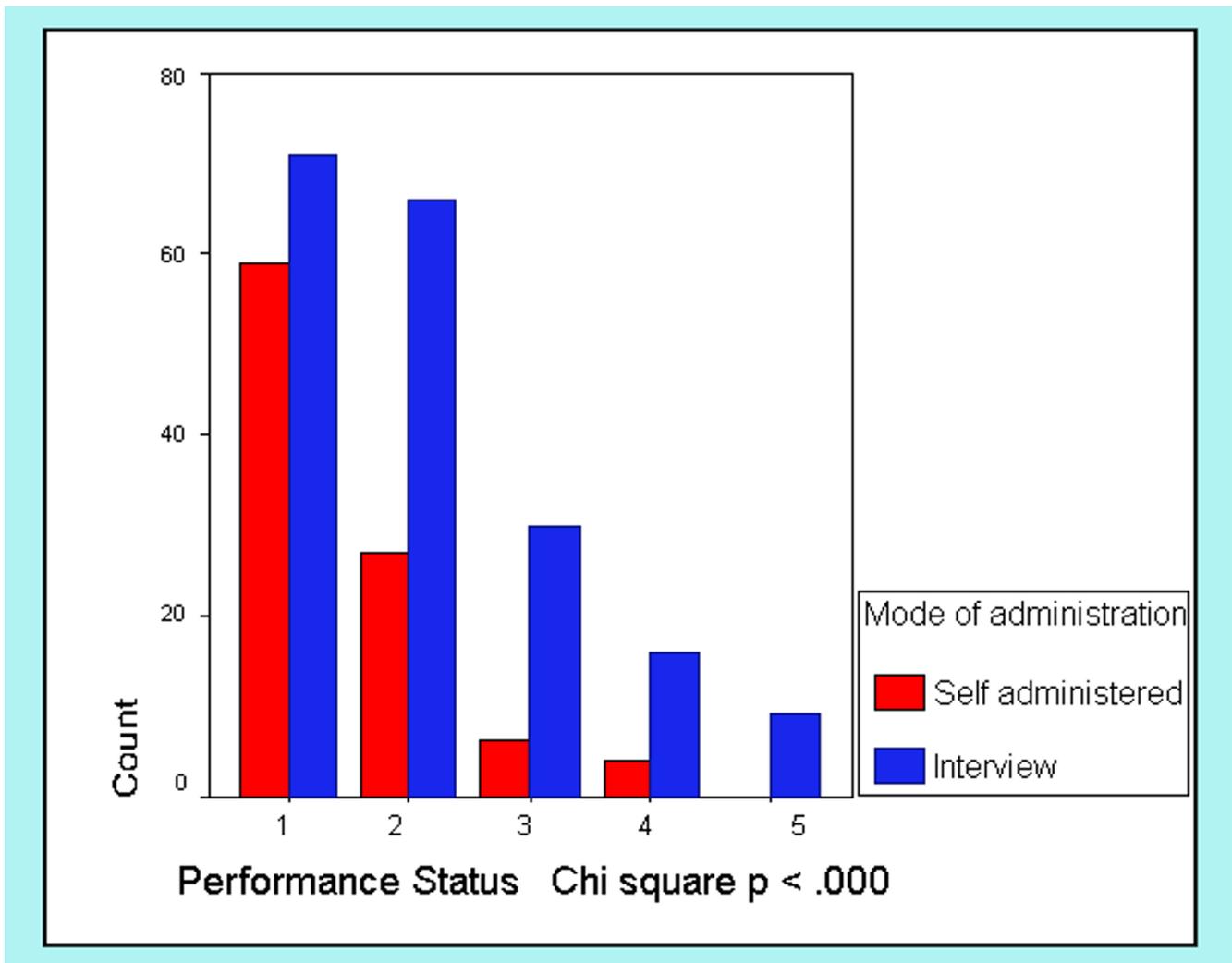


Figure 4
Differences in mode of administration with ECOG-PSR

Table 2: Internal consistency of FACT-G and subscales

Subscale	English Version 2 [1]		Spanish Version 2 [6]		Spanish Version 4	
	Mean ± SD	Alpha coefficients	Mean ± SD	Alpha coefficients	Mean ± SD	Alpha coefficients
PWB (7 items)	20.5 ± 5.5	.82	19.0 ± 5.9	.79	15.7 ± 8.9	.91
SFWB (6 items)	----	----	----	----	16.6 ± 5.1	.82
SFWB (7 items)*	21.9 ± 4.8	.69	18.0 ± 5.7	.65	18.6 ± 5.9	.81
EWB (5 items)	14.8 ± 5.0	.74	14.2 ± 3.7	.60	----	----
EWB (6 items)	---	----	----	----	13.5 ± 6.3	.78
FWB (6 items)	18.0 ± 6.1	.80	12.9 ± 5.6	.72	16.0 ± 6.0	.81
Relationship with Doctor (2 items)	6.9 ± 1.5	.65	6.8 ± 1.2	.66	----	----
Total FACT-G	82.1 ± 15.9	.89	71.9 ± 15.7	.83	61.9 ± 18.4	.89
Total FACT-G (*)					63.7 ± 19.0	.89

PWB = Physical Well-being; FWB = Functional Well-being; SFWB = social and Family Well-being; EWB = Emotional Well-being (*) Includes "satisfaction with sexual life".

Table 3: Differences in FACT-G scores according to criteria groups

Performance Status	n	FACT Total	Physical	Functional	Social	Emotional
		Means (SD)	Means (SD)	Means (SD)	Means (SD)	Means (SD)
0	133	70.5 (19.7)	17.8 (10.2)	19.7 (5.6)	19.5 (5.7)	14.9 (6.7)
1	94	60.7 (15.8)	14.1 (7.5)	15.0 (4.8)	19.0 (5.4)	12.6 (5.7)
2 – 4	66	57.0 (16.6)	15.3 (8.7)	12.0 (5.5)	17.0 (5.8)	12.7 (6.2)
P		<.000	<.00	<.000	<.05	<.00
Subgroup Difference*		0>1,2–4	0>1,2–4	0>1>2–4	0>2–4	0>2–4

Mean score differences by patient location						
Patient Location	n	FACT Total	Physical	Functional	Social	Emotional
		Means (SD)	Means (SD)	Means (SD)	Means (SD)	Means (SD)
Out-patients	223	66.5 (18.5)	16.0 (9.4)	17.2 (5.8)	19.4 (5.3)	13.8 (6.3)
In-patients	86	57.0 (17.0)	14.9 (7.4)	12.8 (5.0)	16.6 (6.4)	12.8 (6.4)
p**		<.000	NS	<.000	.001	<.000

* Scheffé Comparisons; the symbol > separates groups with scores significantly higher than those with lower scores. ** T test for independent samples. Significant p < 0.01

Table 4: Differences in FACT-G and subscale scores with tumor stage and effect size

Tumor stage	n	FACT – G Total		PWB		FWB	
		Means (SD)	95% confidence interval	Means (SD)	95% confidence interval	Means (SD)	95% confidence interval
Local	61	69.8 (20.2)	64.7–75.0	18.3(9.0)	16.0–20.6	17.5(6.4)	15.9–19.2
Regional	129	65.1 (18.9)	57.5 – 64.0	16.6(9.3)	15.0–18.2	16.3(5.4)	15.4–17.2
Metastatic	115	59.3 (16.2)	53.0 – 61.1	13.3(7.7)	11.9–14.7	14.9(6.2)	13.8–16.8
		Score difference	Effect size	Score difference	Effect size	Score difference	Effect size
Local – Metastatic		10.5	0.51	5.0	0.55	2.6	0.4
Regional – Metastatic		5.6	0.30	3.3	0.35	---	---

Table 5: Correlation between FACT-G, QLI doctor and patient version, POMS-SF, MCSDS-10

Subscale/Scale	PWB	SFWB	EWB	FWB	FACT – G
PWB	1	.08	.76**	.14*	.80**
SFWB		1	.07	.48**	.53**
EWB			1	.17**	.78**
FWB				1	.60**
FACT – G					1
QL-Index doctor version					
QLId Activity	.20**	.18**	.19**	.45**	.35**
QLId Daily Living	.11	.13*	.08	.36**	.23**
QLId Health	.11	.12	.14*	.37**	.25**
QLId Support	.02	.30**	.02	.20**	.17*
QLId Outlook	.05	.05	.04	.22**	.12
QLId Total	.16*	.26**	.15*	.48**	.35**
QL-Index patient version					
QLIp Activity	.07	.22**	.10	.51**	.30**
QLIp Daily Living	.05	.04	.00	.30**	.13*
QLIp Health	.02	.14*	.05	.38**	.19**
QLIp Support	-.06	.37**	-.11	.22**	.12*

Table 5: Correlation between FACT-G, QLI doctor and patient version, POMS-SF, MCSDS-10 (Continued)

QLIp Outlook	.03	.11	.13	.28**	.19**
QLIp Total	.04	.26**	.06	.56**	.30**
POMS-SF					
Tension	-.15*	-.03	-.20**	-.32**	-.25**
Depression	-.10	-.20**	-.15*	-.50**	-.32**
Anxiety	-.12	-.19**	-.14	-.32**	-.27**
Vigor	.10	.36**	.19**	.60**	.41**
Fatigue	-.10	-.04	-.19**	-.33**	-.23**
Confusion	-.23	-.20**	-.27**	-.45**	-.41**
POMS TMD	-.17	-.24**	-.25**	-.57**	-.42**
MCSDS-10					
Total MCSDS-10	-.11	.03	-.11	.08	-.06

** Correlation is significant at the 0.01 level. * Correlation is significant at the 0.05 level.

Table 6: IRT based analysis of FACT-G Spanish Version 4.

Physical Well-Being					
MEASURE	INFIT MNSQ	OUTFIT MNSQ	ITEM		
.07	.84	.79	GP7	I am forced to spend time in bed	
.06	.88	1.15	GP1	I have a lack of energy	
.04	1.21	1.13	GP3	Because of my physical condition, I have trouble meeting the needs of my family	
.02	1.05	1.15	GP4	I have pain.	
.01	1.23	1.03	GP5	I am bothered by side effects of treatment	
-.01	.75	.78	GP6	I feel sick	
-.19	1.08	.87	GP2	I have nausea	
Social/ Family Well-Being					
.80	1.58	1.94	GS7	I am satisfied with my sex life	
.23	.89	.88	GS1	I feel close to my friends	
.10	1.10	1.08	GS3	I get support from my friends	
-.05	.85	1.00	GS4	My family has accepted my illness	
-.28	.78	.82	GS5	I am satisfied with family communication about my illness	
-.37	1.28	1.24	GS6	I feel close to my partner (or the person who is my main support)	
-.42	.89	.86	GS2	I get emotional support from my family	
Emotional Well-Being					
.36	1.22	1.17	GE6	I worry that my condition will get worse	
.08	.64	.61	GE1	I feel sad	
.08	.95	.82	GE5	I worry about dying	
.06	.67	.68	GE4	I feel nervous	
-.14	1.01	.86	GE3	I am losing hope in the fight against my illness	
-.43	1.57	1.93	GE2	I am satisfied with how I'm coping with my illness	
Functional Well-Being					
.67	1.11	1.24	GF1	I am able to work (include work at home)	
.21	1.03	1.04	GF7	I am content with the quality of my life right now	
-.04	1.16	1.21	GF5	I am sleeping well	
-.10	1.04	.93	GF2	My work (include work in home) is fulfilling	
-.11	.84	.84	GF6	I am enjoying the things I usually do for fun	
-.14	.77	.74	GF3	I am able to enjoy life	
-.49	1.07	1.08	GF4	I have accepted my illness	

The results of Rasch analyses confirmed the unidimensionality of each subscale of the FACT-G with two exceptions (see Table 6). Items in each table were listed in the order of item difficulty. GS7 (I am satisfied with my sex life) of the SFWB subscale had both MNSQ fit statistics greater 1.40, indicating it did not perform well with the other items in the same scale. GE2 (I am satisfied with how I'm coping with my illness) of the EWB also had large MNSQ fit statistics, indicating its poor fit to the unidimensionality model with the other items on the same scale.

Discussion

Since most instruments designed to evaluate quality of life have been developed in the United States or in Western Europe, it is necessary to adapt them to be used in other cultural settings. Thus, it is important to produce a culturally equivalent measure that can be used to accurately evaluate different groups of people. The final step in the complex process of cross-cultural adaptation is to validate the instrument through the study of the psychometric properties of the measure.

A validation study of the FACT-G Spanish Version 4 was conducted in a sample of Uruguayan cancer patients. A large variability in the biological and sociodemographic features of the sample was ensured to study the general performance of the questionnaire. The frequency of tumor sites represents very closely the incidence of solid tumors in the Uruguayan general population [24].

Reliability analysis showed high internal consistency as indicated by Cronbach alpha coefficients ranging from (.78 - .91). The comparison of these data with the results obtained from the FACT-G English and Spanish Version 2 (Table 2) points out a remarkable improvement in FACT-G Spanish Version 4 total scale as well as subscale reliability coefficients. Based on these results it is safe to conclude that the FACT-G Spanish Version 4 shows sufficient internal consistency to justify its use across cultures.

As an evidence of construct validity, the FACT-G questionnaire appeared to be capable of discriminating among groups of patients according to their level of performance status, showing differences in the total scale and subscale scores with the criterion groups. The FACT-G total and the Functional Well-being subscale showed the best discriminative ability. Differences among known groups were also observed in the FACT-G total and subscale scores in relation to in-patients vs. out-patients. As expected, the FACT-G scores are higher (better QOL) among outpatients and differences were statistically significant on the Functional, Social and Emotional Well-being subscales.

Quality of life researchers have been concerned about the clinical significance of measures and have pursued the

objective to find practical and comprehensive criteria that could be interpretable by clinicians when conveying research results. Clinical status and disease characteristics were considered as clinical anchors. In our study, differences ranging from 5 to 10 points in the overall scale and of approximately 3 points in the physical and functional subscales can be considered clinically relevant since they can discriminate between patients with loco regional and metastatic diseases, a clear-cut criterion commonly used by oncologists. Accordingly, effect size calculations showed moderate values ranging from 0.30 to 0.60. These findings are consistent with those found in a longitudinal study using the FACT-G to assess treatment outcomes in a sample of advanced lung cancer patients [15].

In CTT, the most common form of construct validation of HRQL measures has been the study of convergent and discriminant validity [25]. We included it in our study, along with the more recent IRT approach, because it provides relevant information on the relationship of the questionnaire with other measures of QOL and related constructs, i.e., performance status or psychological distress after a priori hypothesis were made about the magnitude and direction of the correlations. As an evidence of convergent validity, moderate but significant, correlations were found between the FACT-G and a set of instruments (ECOG PSR, QL-I and POMS-SF) that are expected a priori to be related to QOL assessments while no correlation was found between the FACT-G and the MCSDS-10, supporting divergent validity.

An important issue to be considered is the technical equivalence of the FACT-G when used in a sample of cancer patients of a South American country [26]. As mentioned earlier, most patients in the Uruguayan sample preferred the questionnaire to be read out loud by an interviewer instead of filling it out by themselves. This is not a common finding in studies with patients from the United States. These may raise the issue as to whether this difference in the method of assessment is comparable in each culture with respect to the data that it yields. In our study, the internal consistency of the FACT-G did not vary when studied separately, by means of the Cronbach alpha coefficients for the FACT-G total and subscale scores for two groups of patients. In a study of the impact of socio-cultural and clinical factors on Hispanic and African American cancer patients' quality of life, Wan et al [27] also found no significant effect of the mode of administration of the FACT-G on the reporting of overall QOL. Recently, audio-visual computerized based assessments of QOL provide an innovative way for gathering and using self-report data and may be feasible for individuals with limited literacy skills (Hahn et al. unpublished data).

Based on the CTT approach, we may conclude that Spanish FACT-G Version 4 is a psychometrically sound instrument to assess QOL in the population being studied. However, in the present study, we moved on to introduce IRT based analyses of the data considering the significant advantages shown by this method when used to evaluate health outcomes measures. Despite the long history of CTT, there remain major limitations in some areas that have been summarized by Hambleton [28]. First of all, the CTT-based statistics that describe test performance are sample dependent. IRT is more useful since it provides more robust item statistics that are independent and invariant over sample populations that vary in the trait measured by the test. Another limitation of CTT is that the scores that are commonly used as a measure of the examinees' ability are test dependent. Potential advantages of using IRT in health outcome assessments are: more comprehensive and accurate evaluation of item characteristics, assessment of group differences in item and scale functioning, evaluation of scales containing items with different response formats, improvement of existing measures, computer adaptive testing (CAT) applications, and evaluation of person fit [29]. Thus, IRT can facilitate the development of new items and scales to improve existing measures. It may raise attention on redundant items or the location along the trait continuum (in our case, quality of life) where the scale provides little information and needs to be improved.

An IRT analysis was included in the evaluation of FACT-G Spanish Version 4 because it provides additional information on the reliability of the scale than that provided by the CTT approach. A rating scale model (RSM) [20] was used which assumes that the logit-transformed measures of the item scores within each subscale vary along the latent trait level (quality of life) and are aligned according to the difficulty (or location) the patients had to endorse each item, with negative values representing those items that are easier to endorse and positive values those that are more difficult. In the present study, item fit statistics confirmed the unidimensionality of each subscale with two exceptions. In the case of GS7 (I am satisfied with my sex life), many patients were reluctant to give information about their sexual life and this may be a cause of inconsistency in their responses. Another reason for discrepancy may be related to some problem in translation or comprehension. In both items (GS7 and GE2) the word "satisfaction" was translated into Spanish as "satisfacción" which implies in Spanish a degree of fulfillment that patients may be not prone to express when answering such question. Another possible explanation is that these items are the only ones phrased positively in their respective subscales while the rest of the items refer to negative conditions for quality of life.

Other IRT models could have been used for the analysis of the data. For instance, the graded response model (GRM), an extension of the two parameter logistic model [30] is also appropriate to use when item responses can be characterized as ordered categorical responses. However, scores obtained using several models were highly correlated showing that these approaches yield comparable results [31].

IRT analyses demands large sample sizes in order to obtain stable and invariant item and latent trait estimates. However, several studies using this procedure to assess the psychometric properties of QOL measures addressed rather small sample of patients (range: 100 to 400 patients) [32–37].

Cella and Chang [7] warned of the possible limitations of using IRT methods in the evaluation of health measures since they were originally developed for and used with a fairly homogeneous educational assessment population. When we apply these methods to more heterogeneous clinical populations there may be limitations to obtain item-free estimates of sample latent traits. They remark that the context, selection and sequence of questions, considering both item diversity and clinical diversity, may produce sample-dependent item difficulty estimates and therefore unreliable item-dependent estimates of patient ability. The continuous monitoring of item calibrations involved in the process of item banking will help to solve these uncertainties.

The present study is the first one in South America reporting results on item functioning on a health related quality of life measure. Future studies with larger sample of patients could lead to a better understanding of differences in item functioning across different South American countries and cultures and move forward to item banking and CAT technology suitable for developing countries.

Conclusions

We conclude that the FACT-G Spanish Version 4 showed, using classic psychometric and IRT approaches, good reliability and validity and is a valid instrument to set clinical significant differences in longitudinal studies of cancer treatment. Thus, the FACT – G Spanish language version, as reported here, provides sufficient assurance of equivalence to its original English version to be used in future research on quality of life among South American Spanish speaking patients.

Authors' contributions

JJD conceived of the study, participated in its design and coordination, statistical analyses, preparation of drafts and the final manuscript. CF participated in the development of protocol, subject enrollment and revision of

drafts. LS and IG performed statistical analyses. CHC performed IRT based analysis and interpreted results, assisted in the preparation and revision of the manuscript. RL conceived of the study and revision. MCA participated in subject enrollment. All authors read and approved the final manuscript.

Acknowledgements

This study was partially funded by the Comisión Honoraria de Lucha Contra el Cáncer, Uruguay. Dr. Dapuelto's research at the Center on Outcomes, Research and Education, Evanston Northwestern Healthcare was supported by an International Union Against Cancer ICRETT Fellowship. We gratefully acknowledge Professors Enrique Barrios M.D., Ricardo Bernardi, MD., David Cella, Ph.D., and Ignacio Muse, M.D. for their cooperation and support.

References

1. Cella D, Tulskey D, Gray G, Sarafian B, Linn E and Bonomi A *et al.*: **The Functional Assessment of Cancer Therapy Scale: development and validation of the general measure.** *J Clin Oncol* 1993, **11**:570-579.
2. Cella D, Hernandez L, Bonomi AE, Corona M, Vaquero M, Shiomoto G and Baez L: **Spanish language translation and initial validation of the functional assessment of cancer therapy quality of life instrument.** *Med Care* 1998, **36**:1407-18.
3. Bonomi AE, Cella DF, Hahn EA, Bjordal K, Sperner-Unterweger B, Gangeri L, Bergman B, Willems-Groot J, Hanquet P and Zittoun R: **Multilingual translation of the Functional Assessment of Cancer Therapy (FACT) quality of life measurement system.** *Qual Life Res* 1996, **5**:309-320.
4. Winstead-Fry P and Schultz A: **Psychometric analysis of the Functional Assessment of Cancer Therapy-General (FACT-G) scale in a rural sample.** *Cancer* 1997, **15**:2446-52.
5. Yu CL, Fielding R, Chan CL, Tse VK, Choi PH, Lau WH, Choy DT, O SK, Lee AW and Sham JS: **Measuring quality of life of Chinese cancer patients: A validation of the Chinese version of the Functional Assessment of Cancer Therapy-General (FACT-G) scale.** *Cancer* 2000, **88**:1715-27.
6. Dapuelto JJ, Francolino C, Gotta I, Levin R, Barrios E, Alonso I, Afonso Y and Cambiasso S: **Evaluation of the Functional Assessment of Cancer Therapy – General Questionnaire (FACT-G) Version 2 in a Spanish Speaking Population.** *Psycho-Oncology* 2001, **10**:88-92.
7. Cella D and Chang CH: **A discussion of Item Response Theory and its applications in health status assessment.** *Med Care* 2000, **Suppl II**:66-72.
8. Zubrod CG, Schneiderman M and Frei E *et al.*: **Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and triethylene thiophosphoramide.** *J Chron Dis* 1960, **11**:7-33.
9. Spitzer W, Dobson A, Hall J, Chesterman E and Levi J *et al.*: **Measuring the quality of life of cancer patients. A concise QL-Index for use by physicians.** *J Chron Dis* 1981, **34**:585-597.
10. McNair M, Lorr M and Droppleman L: **Manual of the Profile of Mood States.** San Diego: EDITS Educational and Industrial Testing Service 1992.
11. Crowne D: **A new scale of social desirability independent of psychopathology.** *Journal of Consulting Psychology* 1960, **4**:349-354.
12. Lara-Cantú M and Suzan-Reed M: **La escala de Deseabilidad Social de Marlowe y Crowne: un estudio psicométrico.** *Salud Mental* 1988, **11**(3):25-29.
13. DeVellis RF: **Scale and Applications.** Newbury Park: Sage Publications 1991.
14. Guyatt GH, Osoba D, Wu AW, Wyrwich KW and Norman GR: **Clinical significance consensus meeting group. Methods to explain the clinical significance of health status measures.** *Mayo Clin Proc* 2002, **77**:371-83.
15. Cella D, Eton D, Fairclough DL, Bonomi P, Heyes AE, Silberman C, Wolf MK and Johnson DH: **What is a clinically meaningful change on the Functional Assessment of Cancer Therapy – Lung (FACT-L) Questionnaire? Results form Eastern Cooperative Oncology Group (ECOG) Study 5592.** *J Clin Epidemiol* 2002, **55**:285-295.
16. Sobin LH, Wittekind Ch and Eds: **TNM Classification of Malignant Tumours.** New York: Wiley-Liss Inc 1997.
17. Osoba D, Rodrigues G, Myles J, Zee B and Pater J: **Interpreting the significance of changes in health-related quality-of-life scores.** *J Clin* 1998, **16**:139-44.
18. Cohen J: *Statistical Power Analysis for the Behavioral Sciences* Secondth edition. Edited by: Hillsdale NJ. Lawrence Erlbaum and Associates; 1988.
19. Andrich D: **A binomial latent trait model for the study of Likert-style attitude questionnaires.** *British Journal of Mathematical and Statistical Psychology* 1978, **31**:84-98.
20. Andrich D: **A rating formulation for ordered response categories.** *Psychometrika* 1978, **43**:561-573.
21. Andrich D: **Scaling attitude items constructed and scored in the Likert tradition.** *Educational and Psychological Measurement* 1978, **38**:665-680.
22. Andrich D: **Application of a psychometric rating model to ordered categories which are scored with successive integers.** *Applied Psychological Measurement* 1978, **2**:581-594.
23. WINSTEP – MINISTEP Rasch Model Computer Programs Chicago: John M. Linacre; 2002.
24. Vasallo JA, Barrios E, De Stefani E and Ronco A: **II Atlas de Incidencia del Cáncer en el Uruguay.** Montevideo: Comisión Honoraria de Lucha contra el Cáncer 2001.
25. Tulskey D: **An Introduction to Test Theory.** *Oncology* 1990, **4**:43-48.
26. Flaherty JA, Gaviria FM and Pathak D *et al.*: **Developing instruments for cross-cultural psychiatry research.** *J Nerv Mental Dis* 1988, **176**:257-263.
27. Wan G, Counte M and Cella D: **The Influence of Personal Expectations on Cancer Patients' Reports of Health Related Quality of Life.** *Psycho-Oncology* 1997, **6**:1-11.
28. Hambleton RK: **Emergence of Item Response Modeling in instrument development and data analysis.** *Med Care* 2000, **Suppl II**:60-64.
29. Hays RD, Morales LS and Reise SP: **Item Response Theory and health outcomes measurement in the 21st Century.** *Med Care* 2000, **Suppl II**:28-41.
30. Samejima F: **Graded Response Model.** In *Handbook of modern item response theory* Edited by: Van der Linden WJ, Hambleton RK. New York: Springer; 1997.
31. Chang CH and Cella D: **One- versus two-parameter item response theory (IRT) measurement models applied to MOS PF-10 scores: How much does it really matter? [abstract].** *Qual Life Res* 2001, **10**:302.
32. Young MA, Blodgett C and Reardon A: **Measuring seasonality: psychometric properties of the Seasonal Variation.** *Psychiatric Res* 2003, **117**:75-83.
33. Orlando M and Marshall GN: **Differential item functioning in a Spanish translation of the PTSD checklist: detection and evaluation of impact.** *Psychol Assess* 2002, **14**:50-9.
34. Rysler L, Wright BD, Aeschlimann A, Mariacher-Gehler S and Stuchi G: **A new look at the Western Ontario and Mc Master Universities Osteoarthritis Index using Rasch analysis.** *Arthritis Care Res* 1999, **12**:331-5.
35. Leplege A, Ruede N, Ecosse E, Ceinos R, Dohin E and Pouchot J: **Measuring Quality of life from view of HIV-positive subjects the HIV-QL31.** *Qual Life Res* 1997, **6**:585-94.
36. Gibbons RD, Clark DC and Kupfer DJ: **Exactly what does the Hamilton Depression Rating Scale measure? J Psychiatr Res 1993, **27**:259-73.**
37. Chang CH and Gehlert S: **The Washington Psychosocial Seizure Inventory (WPSI): psychometric evaluation and future applications.** *Seizure* 2003, **12**:261-267.