
A new statistical approach for quality of life questionnaires in the assessment of non-small-cell lung cancer Cuban patients

Carmen Viada¹, Javier Ballesteros^{2,3}, Martha Fors⁴, Patricia Luaces¹, Liset Sánchez¹, Bárbara Wilkinson¹, Aymara Fernández¹, Camilo Rodríguez¹, Tania Crombet¹

¹Center of Molecular Immunology (CIM), Havana, Cuba

²University of the Basque Country UPV/EHU, Leioa, Spain

³Centre for Biomedical Network Research on Mental Health (CIBERSAM), Spain

⁴National Coordinating Center for Clinical Trials (CENCEC), Havana, Cuba

Email address:

carmen@cim.sld.cu (C. Viada), javier.ballesteros@ehu.es (J. Ballesteros), martaf@cencec.sld.cu (M. Fors), patricial@cim.sld.cu (P. Luaces), lsanchez@cim.sld.cu (L. Sánchez), wilkinson@cim.sld.cu (B. Wilkinson), aymaraf@cim.sld.cu (A. Fernández), camilo@cim.sld.cu (C. Rodríguez), taniac@cim.sld.cu (T. Crombet)

To cite this article:

Carmen Viada, Javier Ballesteros, Martha Fors, Patricia Luaces, Liset Sánchez, Bárbara Wilkinson, Aymara Fernández, Camilo Rodríguez, Tania Crombet. A New Statistical Approach for Quality of Life Questionnaires in the Assessment of Non-Small-Cell Lung Cancer Cuban Patients. *Cancer Research Journal*. Vol. 2, No. 1, 2014, pp. 1-8. doi: 10.11648/j.crj.20140201.11

Abstract: Objectives: To evaluate the dimensionality and item characteristics of the European Organization for the Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30) and the lung cancer module (QLQ-LC13) and explore the possibility of reduction of the scales. Methods: We analyzed the answers recorded for the QLQ-C30 and QLQ-LC13 in patients diagnosed with non-small-cell lung cancer (NSCLC) participating in 4 Cuban multicenter clinical trials. We assessed the dimensionality underlying both scales with a Mokken nonparametric item response analysis. We used the parametric Samejima's graded response model to assess the item characteristics; we also conducted a confirmatory factor analysis (CFA) to test the dimensionality of both scales. Taking into account the previous results we compared different reduced scales using the Receiver Operator Curves (ROC Analysis). Results: 873 patients with NSCLC that completed the EORTC QLQ-C30 and 840 patients that completed the QLQ-LC13 were included. Mokken analysis of both scales resulted in 1-dimensional scales. All items showed scalability indices over 0.30. The overall scalability for the QLQ-C30 was 0.43, defining a medium scale according to Mokken's criteria, while the overall scalability of the QLQ-LC13 was 0.44. Unconstrained Samejima's graded response models showed appropriate fit, with most items of both scales presenting pertinent difficulty and discrimination parameters. The results of the CFA supported an underlying 1-dimensional latent structure for perceived quality of life (QLQ-C30 comparative fit index [CFI]=0.98; root-mean-square error of approximation [RMSEA]=0.05; QLQ-LC13 CFI=0.99 and RMSEA=0.04). All factor loadings were above 0.30. Conclusions: The QLQ-C30 and the QLQ-LC13 represent in patients with lung cancer a 1-dimensional structure of patient-perceived quality of life. All the reduced scales had similar performance compared with both original scales.

Keywords: Quality of Life, Cancer, Confirmatory Factor Analysis, Item Response Theory, Mokken Analysis, Samejima's Graded Response Model, Receiver Operator Curves

1. Introduction

Regulatory approval of anti-cancer treatments requires of randomized clinical trials with appropriate primary outcome measures. Secondary outcomes as quality of life (QoL) assessments are currently regarded as desirable, but not mandatory, by regulatory agencies around the world.

For cancer, an often-used specific questionnaire is the Quality of Life Core Questionnaire of the European Organization for the Research and Treatment of Cancer (EORTC QLQ-C30) [2]. The EORTC QLQ-C30 was designed as an integrated system for assessing the health-related quality of life of cancer patients participating in international clinical trials. This core questionnaire is completed with specific questionnaires or modules for

specific types of cancer, as the lung cancer module (QLQ-LC13) that is administered in parallel with the core QLQ-C30 and it is designed for use among patients receiving treatment with chemotherapy and/or radiotherapy.

Classical psychometric properties of the QLQ-C30 and the QLQ-LC13 support their use as general measures of patient perceived QoL [1-3] and had shown responsiveness to clinical changes [4-5]. Their current scoring and interpretation assume (i) a multi-dimensionality of the latent constructs that underlie the items measurement with major dimensions for functionality and symptoms severity, and (ii) a similar item discrimination within each dimension that is reflected by the unweighted approach to estimate the appropriate total score as reported in the EORTC manual [6]. We have tested these assumptions by analyzing the item responses and conducting confirmatory factor analysis (CFA) of both QoL scales in lung cancer patients participating in several trials with a therapeutic anticancer vaccine (Cimavax EGF) manufactured and licensed in Cuba for use in adult patients with stage IIIB/IV non-small-cell lung cancer (NSCLC), and designed to boost the individual's immune response against the epidermal growth factor (EGF) [7-10].

2. Main Body

2.1. Study Design and Population

This study includes a validation sample of 873 patients with NSCLC that completed the EORTC QLQ-C30 and 840 patients that completed the QLQ-LC13. The patients participated in 4 Cuban multicenter clinical trials (1 phase II, 2 phase III, and 1 phase IV) conducted since June 2002 till January 2013 [Cuban public clinical trials registration number: RPCEC 00000161]. All trials enrolled patients older than 18 years with proven histology or cytology for NSCLC with IIIB and IV stages in phase II or III and all stages in phase IV at the moment of enrollment. Patients randomized to the immunotherapy arm received vaccination at 4 sites (2 deltoids & 2 gluteus), equivalent to 2.4 mg of the antigen, distributed in the 4 anatomic sites, corresponding to 0.6 mg of EGF in 1.2 mL water in oil emulsion per site plus best supportive care. Patients randomized to the control arm received only best supportive care. Trial protocols were approved by the Cuban Regulatory Agency and the appropriate local Ethical and Research Committees. All patients gave informed consent to participate in the trials.

2.2. Data Collection Procedures and Measures

Patients completed at trial inception and during follow-up the Spanish versions of the EORTC QLQ-C30 (version 3.0), and the lung cancer module QLQ-LC13. The EORTC QLQ-C30 is a 30-item self-report scale that incorporates 5 functional scales (physical, role, cognitive, emotional, and social), 3 symptom scales (fatigue, pain, and nausea and vomiting), and 2 single items, one of global health status and the other of global QoL perception. The timeframe for the

QoL assessment is the past week. QLQ-C30 has 28 items coded in a 4-point Likert scale (not at all [1] to very much [4]) and 2 general items (item #29: how would you rate your overall health during the past week? and item #30: how would you rate your overall quality of life during the past week?) coded with a 7-point analogue scale that ranges from very poor (1) to excellent (7). The QLQ-LC13 has 12 questions assessing lung cancer associated symptoms and treatment related side effects. It is coded as the QLQ-C30 with a 4-point Likert scale. Additionally it includes a 2-part item regarding pain medication (item #13, did you take any medicine for pain? and if the answer is yes, how much did it help?). To assess the item responses to the 28 Likert scale items of the QLQ-C30 and the 12 Likert scale items of the QLQ-LC13 we used the pre-treatment (baseline) measurements.

2.3. Statistical Analysis

We followed a 3-step approach to assess the psychometric characteristics of the QLQ-C30 and the QLQ-LC13. First, the item responses were analyzed by a non-parametric item response theory (IRT) analysis (Mokken analysis) to estimate the number of probable subscales underlying the latent dimensions. All items linked to Mokken scales were retained and the scales interpreted if they had at least three items attached to them. Interpretation of Mokken scales followed usual rules of thumb regarding the scalability indexes of the Loevinger's H coefficient: weak scales for $0.3 \leq H < 0.4$, medium scale $0.4 \leq H < 0.5$ and strong scale $H \geq 0.5$. The total scale should have at least a scalability coefficient $H \geq 0.30$. We also checked with the Mokken analysis the assumptions associated with the selection of the parametric IRT model to use for the second step in the analyses. We fitted 2 Mokken models to test if the overall score of the scales was a valid tool to order and classify the subjects according to the degree of the latent dimension exhibited, and whether an order existed among items to rate the corresponding dimension that is independent of the selected sample: the monotone homogeneity model (MHM) and the double monotonicity model (DMM). The MHM assumes different item discrimination; items difficulty is not the same for all subjects and thus item invariant ordering (IIO) is not possible. On the contrary the DMM assumes similar item discrimination; items difficulty are the same for all subjects, and thus IIO is possible. Whereas the MHM would suggest the fitting of a 2-parameter logistic model as the Samejima's Graded Response Model (GRM), the DMM would suggest the fitting of 1-parameter logistic model as the extended Rasch model.

In the second step of the analyses we fitted the appropriate IRT parametric model suggested by the Mokken analysis to obtain estimates of the relationship between the latent dimension and the item characteristics. We estimated the item response characteristic curve parameters (ICC) and the item information function, an analogous to the reliability in the classical test theory. An appropriate ICC should present

an ordered shape discriminating among the category thresholds, each one within item should present a different probability of being selected more than any other category for a specific difficulty.

We assessed the dimensionality of the scales by CFA using robust weighted least squares on the sample variance-covariance matrix. Goodness of fit for CFA was evaluated by the comparative fit index (CFI), and the root mean square error approximation (RMSEA). A model was considered acceptable if CFI>0.95 and RMSEA < 0.08.

The discriminative validity of the two scales and their short scales was assessed by comparing receiver operating characteristic (ROC) curve analysis and the areas under the ROC curves (AUC).

We run all the analyses with the R package (v2.15.3) and the libraries *mokken* (non-parametric IRT analyses), *ltm* (parametric IRT analyses), and *lavaan* (CFA)[11-15].

3. Results and Discussion

QLQ-C30 and QLQ-LC13 answers were obtained for 873 and 840 patients respectively. Patients were predominantly males (65.2%) with a mean age of 60.2 years (SD=9.9). For those patients whose clinical information was available, the main clinical characteristics

were as follows: 219 patients (29.2%) presented a performance status (PS) 0, 329 patients (43.9%) presented PS=1, and 201 patients (26.8%) presented PS equal or greater than 2; 270 cancers were classified as adenocarcinoma (36.7%); and regarding cancer staging, 411 patients (54.8%) presented stage IIIB, 285 patients (38.0%) presented stage IV (38.0%), and 54 patients (7.2%) were classified as others.

3.1. Mokken Analysis: Dimensionality of the QLQ-C30 and the QLQ-LC13

The Mokken analysis of the QLQ-C30 and QLQ-LC13 singled out one latent dimension for each questionnaire (see Tables 1 and 2). All 28 QLQ-C30 items on study were retained in one scale, but for the QLQ-LC13 two items were not ascribed to any scale (item #1: cough, and item #5: short of breath when climbing stairs). The general scalability index for QLQ-C30 was 0.43 with all items showing scalability indices over 0.30 (Table 1). The general scalability index for QLQ-LC13 was 0.44 with all items showing scalability indices over 0.30 (Table 2). When items #1 and #5 were removed, the QLQ-LC13 scalability increased to 0.52 (a strong scale according to Loewinger's criteria).

Table 1. Endorsement frequencies and Mokken's scalability (Loewinger's H_i coefficients) for the items of the QLQ-C30 (N=873)

Item number and description	Mean (SD)	Endorsement frequencies (%)				H_i
		1	2	3	4	
1. Trouble carrying a heavy shopping bag	2.35 (1.13)	275 (32)	201 (23)	217 (25)	180 (21)	0.44
2. Trouble taking a long walk	2.41 (1.13)	252 (29)	203 (23)	223 (26)	195 (22)	0.47
3. Trouble taking a short walk	1.68 (0.97)	527 (60)	161 (18)	120 (14)	65 (7)	0.46
4. Stay in bed/chair during day	1.67 (0.96)	532 (61)	164 (19)	114 (13)	63 (7)	0.47
5. Help with eating, dressing, washing or using the toilet	1.34 (0.75)	695 (80)	84 (10)	67 (8)	27 (3)	0.50
6. Limited doing daily activities	1.88 (1.07)	450 (52)	187 (21)	126 (14)	110 (13)	0.49
7. Limited in leisure time activities	1.73 (1.01)	516 (59)	160 (18)	118 (14)	79 (9)	0.47
8. Short of breath	1.86 (0.96)	411 (47)	236 (27)	167 (19)	59 (7)	0.37
9. Had pain	2.08 (1.06)	342 (39)	233 (27)	181 (21)	117 (13)	0.41
10. Need to rest	2.05 (1.01)	333 (38)	253 (29)	197 (23)	90 (10)	0.51
11. Trouble sleeping	1.80 (0.98)	459 (53)	198 (23)	149 (17)	67 (8)	0.38
12. Felt weak	1.99 (1.05)	381 (44)	222 (25)	164 (19)	106 (12)	0.51
13. Lacking appetite	1.93 (1.09)	440 (50)	166 (19)	154 (18)	113 (13)	0.43
14. Felt nauseated	1.53 (0.90)	606 (69)	120 (14)	98 (11)	49 (6)	0.42
15. Vomiting	1.32 (0.72)	709 (81)	76 (9)	65 (7)	23 (3)	0.42
16. Constipation	1.69 (1.00)	541 (62)	139 (16)	119 (14)	74 (8)	0.31
17. Diarrhea	1.19 (0.56)	761 (87)	65 (7)	37 (4)	10 (1)	0.30
18. Tiredness	2.05 (0.99)	316 (36)	281 (32)	189 (22)	87 (10)	0.52
19. Pain interferes with daily activities	1.77 (1.02)	494 (57)	164 (19)	135 (15)	80 (9)	0.47
20. Difficulty in concentrating on things	1.42 (0.79)	646 (74)	117 (13)	81 (9)	29 (3)	0.44
21. Feel tense	1.95 (1.00)	370 (42)	261 (30)	155 (18)	87 (10)	0.44
22. Worried	2.21 (1.02)	255 (29)	302 (35)	190 (22)	126 (14)	0.43
23. Irritation	1.76 (0.99)	476 (55)	211 (24)	107 (12)	79 (9)	0.38
24. Depression	2.02 (1.04)	356 (41)	258 (30)	148 (17)	111 (13)	0.45
25. Difficulties remembering things	1.52 (0.84)	585 (67)	154 (18)	102 (12)	32 (4)	0.36
26. Interference with family life	1.68 (0.98)	530 (61)	156 (18)	120 (14)	67 (8)	0.41
27. Interference with social activities	1.84 (1.04)	467 (53)	173 (20)	142 (16)	91 (10)	0.41
28. Financial difficulties	1.88 (1.10)	461 (53)	172 (20)	121 (14)	119 (14)	0.32

Endorsement frequencies. 1: not at all; 2: a little; 3: quite a bit; 4: very much

Table 2. Endorsement frequencies and Mokken's scalability (Loevinger's H_i coefficients) for the items of the QLQ-LC13 ($N=840$)

Item number and description	Mean (SD)	Endorsement frequencies (%)				H_i
		1	2	3	4	
1. Cough	2.02 (1.11)	366 (44)	238 (28)	89 (11)	147 (18)	0.16
2. Cough up blood	2.48 (1.46)	394 (47)	41 (5)	13 (2)	392 (47)	0.58
3. Short of breath when resting	2.22 (1.36)	419 (50)	104 (12)	31 (4)	286 (34)	0.49
4. Short of breath when walking	2.02 (1.17)	390 (46)	213 (25)	65 (8)	172 (20)	0.32
5. Short of breath when climbing stairs	2.13 (1.15)	335 (40)	239 (28)	91 (11)	175 (21)	0.25
6. Sore mouth/tongue	2.51 (1.48)	400 (48)	21 (2)	8 (1)	411 (49)	0.59
7. Trouble swallowing	2.42 (1.44)	401 (48)	52 (6)	22 (3)	365 (43)	0.55
8. Tingling hands/feet	2.32 (1.33)	360 (43)	137 (16)	54 (6)	289 (34)	0.37
9. Hair loss	2.53 (1.41)	349 (42)	77 (9)	33 (4)	381 (45)	0.52
10. Pain in chest	2.25 (1.34)	391 (47)	133 (16)	35 (4)	281 (33)	0.48
11. Pain in arm/shoulder	2.22 (1.31)	389 (46)	140 (17)	48 (6)	263 (31)	0.41
12. Pain in other body parts	2.20 (1.32)	400 (48)	130 (15)	51 (6)	259 (31)	0.39

Endorsement frequencies. 1: not at all; 2: a little; 3: quite a bit; 4: very much

The reliability of both the QLQ-C30 and the QLQ-LC13 was good according to Mokken ($MS=0.95$ and 0.91 respectively) and Cronbach's criteria ($\alpha=0.94$ and 0.89 respectively). Both questionnaires presented a good fit to MHM but not to DMM, leading then to a 2-parameter logistic model to further study the item characteristics.

Table 3. Mokken's scalability (Loevinger's H_i coefficients), Samejima's graded response model parameters, and standardized loadings from confirmatory factor analysis for the QLQ-C30 ($N=873$)

Item number and description	H_i	a	b1	b2	b3	Information	CFA loadings
1. Trouble carrying a heavy shopping bag	0.44	1.58	-0.77	0.10	1.19	3.26	0.62
2. Trouble taking a long walk	0.47	1.77	-0.82	0.03	1.05	3.28	0.66
3. Trouble taking a short walk	0.46	2.11	0.29	1.00	1.93	3.16	0.69
4. Stay in bed/chair during day	0.47	2.25	0.30	1.02	1.88	3.17	0.70
5. Help with eating, dressing, washing or using the toilet	0.50	2.42	0.98	1.50	2.39	2.94	0.64
6. Limited doing daily activities	0.49	2.44	-0.01	0.71	1.42	3.11	0.74
7. Limited in leisure time activities	0.47	2.29	0.24	0.91	1.70	3.09	0.71
8. Short of breath	0.37	1.28	-0.20	0.99	2.53	3.48	0.55
9. Had pain	0.41	1.31	-0.52	0.61	1.83	3.34	0.59
10. Need to rest	0.51	2.50	-0.43	0.47	1.55	3.57	0.76
11. Trouble sleeping	0.38	1.30	0.04	1.06	2.39	3.30	0.55
12. Felt weak	0.51	2.50	-0.24	0.57	1.44	3.34	0.76
13. Lacking appetite	0.43	1.65	-0.03	0.69	1.64	3.07	0.63
14. Felt nauseated	0.42	1.50	0.71	1.41	2.48	3.01	0.58
15. Vomiting	0.42	1.39	1.35	2.01	3.26	2.90	0.51
16. Constipation	0.31	0.93	0.57	1.55	2.92	2.95	0.44
17. Diarrhea	0.30	0.77	2.70	4.00	6.17	2.79	0.31
18. Tiredness	0.52	2.48	-0.50	0.52	1.58	3.63	0.76
19. Pain interferes with daily activities	0.47	1.96	0.16	0.87	1.79	3.15	0.69
20. Difficulty in concentrating on things	0.44	1.66	0.87	1.63	2.77	3.11	0.60
21. Feel tense	0.44	1.52	-0.34	0.84	1.95	3.44	0.63
22. Worried	0.43	1.36	-0.92	0.53	1.71	3.54	0.61
23. Irritation	0.38	1.28	0.15	1.29	1.26	3.15	0.56
24. Depression	0.45	1.63	-0.39	0.73	1.66	3.35	0.66
25. Difficulties remembering things	0.36	1.13	0.72	1.80	3.40	3.23	0.50
26. Interference with family life	0.41	1.44	0.34	1.15	2.25	3.11	0.59
27. Interference with social activities	0.41	1.54	0.07	0.88	1.89	3.12	0.61
28. Financial difficulties	0.32	0.93	0.10	1.18	2.27	2.90	0.46

Notes. a: discrimination parameter; b's: threshold parameters; CFA: confirmatory factor analysis

Table 4. Mokken's scalability (Loevinger's H_i coefficients), Samejima's graded response model parameters, and standardized loadings from confirmatory factor analysis for the QLQ-LC13 (N=840)

Item number and description	H_i	a	b1	b2	b3	Information	CFA loadings
1. Cough	NE	NE	NE	NE	NE	NE	NE
2. Cough up blood	0.58	3.83	-0.28	-0.20	-0.15	4.51	0.88
3. Short of breath when resting	0.49	2.31	-0.16	0.20	0.30	3.10	0.69
4. Short of breath when walking	0.32	1.31	-0.26	0.34	0.62	1.78	0.37
5. Short of breath when climbing stairs	NE	NE	NE	NE	NE	NE	NE
6. Sore mouth/tongue	0.59	4.62	-0.23	-0.18	-0.14	5.21	0.92
7. Trouble swallowing	0.55	4.12	-0.19	-0.09	-0.04	4.94	0.85
8. Tingling hands/feet	0.37	1.70	-0.30	0.43	0.59	2.52	0.56
9. Hair loss	0.52	3.79	-0.32	-0.17	-0.12	4.77	0.78
10. Pain in chest	0.48	2.62	-0.21	0.35	0.44	4.03	0.69
11. Pain in arm/shoulder	0.41	1.98	-0.20	0.50	0.64	3.02	0.59
12. Pain in other body parts	0.39	1.73	-0.14	0.51	0.67	2.50	0.55

Notes. a: discrimination parameter; b's: threshold parameters; CFA: confirmatory factor analysis; NE: not estimated

3.2. Item characteristics: Samejima's Graded Response Analysis

Both the QLQ-C30 and the QLQ-LC13 presented a good fit to the unconstrained Samejima's GRM that assumes different discrimination parameters by items. The likelihood ratio test comparing the constrained and unconstrained models was 534.97 on 27 df ($p < 0.001$) for QLQ-C30 and 2630.7 on 9 df ($p < 0.001$) for QLQ-LC13. The value for intercepts, thresholds and information across items is recorded in Table 3 for the QLQ-C30 and in Table 4 for the QLQ-LC13. Supplemental figures show ICCs (S1 for QLQ-C30, S2 for QLQ-LC13). As seen, most items for the QLQ-C30 presented a shape and category threshold compatible with appropriate difficulty and discrimination parameters. The main exceptions were items #15, #16, #17 and #25. On the contrary the ICC for the QLQ-LC13 items strongly suggest that 4 Likert categories could be too many for this questionnaire to elicit appropriate responses in the patients enrolled in the Cuban immunological trials.

3.3. Confirming Unidimensionality: CFA

The QLQ-C30 and QLQ-LC13 CFA results fit to a 1-dimensional latent structure (CFI= 0.98 and 0.99 respectively; RMSEA=0.052 and 0.042 respectively). Inspection of standardized residuals and modification indexes indicated no localized points of ill fit in the final solution apart from including correlated measurement errors between most items of QLQ-C30 (what strongly points to important redundancy in the measurement of the latent dimension and thus might support a reduction in the number of items composing the scale), and between items #3 and #4, and #10 and # 11 for the QLQ-LC13. All freely

estimated unstandardized parameters were statistically significant ($p < 0.001$). Tables 3 and 4 display the completely standardized parameter estimates (factor loadings), showing loadings above 0.30 for all items in both scales.

3.4. Validation of the Reduced Versions by using ROC Analysis

The selection of items with standardized factor loadings > 0.70 lead to a 6-item QLQ that showed good discriminative validity. The first reduced version (0 to 100 with inverse items) QLQ-C6 had 3 positives items (#4, #6 & #7, related with the physical and role function) and 3 negatives items (#10, #12 & #18, related with the fatigue dimension).

The selection of items with standardized factor loadings > 0.60 lead to a 17-item QLQ that showed good discriminative validity. The Second reduced version (0 to 100 with inverse items) QLQ-C17 had 12 positives items #1 to #7 (physical & role function), #20 to #22 (cognitive & emotional function), #24 (emotional function) and #27 (social function) and 5 negatives items: # 10, #12, #13, #18 & #19 (fatigue, appetite loss & pain).

The QLQ-C30 version of 28 items (0 to 100 with inverse items) had 15 positive items and 13 negative items.

We compare the reduced scales and the original version and we obtained the following results: QLQ-C6 (AUC=0.760 CI95% [0.7240, 0.7962]), QLQ-C17 (AUC=0.735 CI95% [0.6974, 0.7739]) and QLQ-C28 (AUC=0.700 CI95% [0.6607, 0.7399])

In case of QLQ-LC13, the selection of items with standardized factor loadings > 0.70 lead to a 4-item QLQ that showed good discriminative validity. The first reduced version (0 to 100 with inverse items) QLQ-LC4 had 4 negatives items: #2 (cough up blood), #6 (sore mouth), #7

(trouble swallowing) and #9 (hair loss).

The selection of items with standardized factor loadings > 0.60 lead to a 6-item QLQ that showed good discriminative validity. The Second reduced version (0 to 100 with inverse items) QLQ-LC6 had 6 items: #2, #3 (short of breath when resting), #6, #7, #9 and #10 (pain in chest).

We also compared the reduced scales and the original version and we obtained the following results: QLQ-LC4 (AUC=0.761 CI95% [0.7244, 0.7962]), QLQ-LC6 (AUC=0.757 CI95% [0.6975, 0.7739]) and QLQ-LC12 (AUC=0.704 CI95% [0.6677, 0.7399])

4. Conclusions

Our aims with this study were twofold, to test the dimensionality of the QLQ-C30 and the QLQ-LC13, and to test the assumption of equal item discrimination for both questionnaires. The findings suggest that one underlying latent dimension of perceived QoL is measured by both QLQ-C30 and QLQ-LC13. In both cases the 1-dimensional structure first suggested by Mokken analysis was further supported by CFA. These results agree with recent research focusing on CFA of higher order models for the QLQ-C30 [20]. In that article the researchers, using data from different clinical trials collated by the EORTC, suggest retaining and interpreting a 2-dimensional model including dimensions for both physical and mental health. However the most parsimonious 1-dimensional model also presented a good fit to the data with similar indexes to ours (CFI=0.90, RMSEA=0.058 for the quoted study [16], CFI=0.98, RMSEA=0.052 in our study). The preference to retain 2 factors instead of 1 seems to be based more on the possibility to compare QLQ-C30 results with other QoL results than on substantive theoretical grounds. We think the parsimony attained with the 1-dimensional model presents a major advantage since it leads to the analysis of a single outcome measure and thus prevent the use of adjusting for multiple testing in clinical trials when using multidimensional QoL outcomes.

Results for ICC are mixed. For the QLQ-C30 most items presented appropriate difficulty and threshold discrimination parameters, however the CFA results of modification indices pointed to an important overlapping of information among items that suggests a shorter scale could be as informative as the current 28 items scored with the 4-points Likert scale. To inform the scale reduction it might be of interest to take into account (i) the information function obtained by IRT, (ii) the ordering of the CFA loadings, and (iii) the pattern of inter-item covariances. For the QLQ-LC13 most items but items #1 (cough) and #5 (short of breath when climbing stairs) present poor difficulty discrimination with main endorsements at the extreme values of response and very low response frequencies at the middle values. In fact items #1 and #5 behave so differently from the rest that Mokken analysis did not ascribed them to any scale. On the other hand, both items present good ICC and in fact might be considered as

representative items of perceived QoL specific symptoms in lung cancer patients. We do not know of any other study reporting specifically the ICC of QLQ-C30 or QLQ-LC13, however we know the EORTC is developing a computerized adaptive testing of the QLQ-C30 that surely is based on an ICC questions bank.

A possible limitation of the current study refers to the direct use of the Spanish version of the questionnaires without any linguistic adaptation to local uses of Spanish in Cuba. Others have validated Spanish versions of the EORTC QLQ-C30 and other cancer specific questionnaires (colorectal and prostate cancer) concluding that these scales appeared as reliable and valid instruments when applied to persons with Spanish as mother tongue [17-18]. Also a previous study has shown an overall good linguistic equivalence for most of the EORTC QLQ-C30 items [19]. However it also showed strongly discrepant results for some translations for several scales, with effects that could be large enough to impact on the results of clinical trials [19]. This differential item functioning if present in this study could make our results not comparable to others since the measurement of the QoL depends, as a latent variable, on the understanding and perception of individuals to the items presented in a self-administered questionnaire. Another limitation of this study, that could make comparisons with others difficult, concerns the use of most but not all the items of the questionnaires on our analyses. For the QLQ-C30 we did not include in the analyses the 2 items related with the overall perception of health and QoL that are not coded with a 4-point Likert scale but with a 7-point analogue scale. For the QLQ-LC13 we did not include in the analyses the complex item with 2-linked sentences regarding the use of drugs to alleviate pain. In both cases, the information or range of possible responses conflicted with the information and range of the other items and thus could exert an improper influence in the results. Also the responses to the general items of QLQ-C30, due to their position at the end of the questionnaire, are somehow anchored to the answers elicited to the previous 28 items. In that sense we did consider them as second order factors instead of a first order factor representing the underlying QoL dimension.

In agreement with others [16] we have found that 1-dimensional structure for perceived QoL with QLQ-C30 (and also with QLQ-LC13) fits the observed data and makes an economical explanation of the underlying latent dimension because of its parsimony. CFA and ICC analyses would support a possible reduction of the QLQ-C30 on different grounds that those previously reported by the EORTC Quality of Life Group [20-21], and a restructuring of the QLQ-LC13 to obtain scales with similar item discrimination. We consider these jointly with the analysis of sensitivity to change and mapping to generic preference-based instruments to elicit utility values at the individual level as research lines worth to pursue [22-23].

All the short scales showed similar performance on both sensitivity to change and discriminative power to define the quality of life of these patients compared with original

scales (QLOC30 and QLQLC13), nevertheless, there is still need to make further validation of these results with other groups of lung cancer patients.

5. CGT-EGF Principal Investigators

The CGT-EGF principal investigators are: Gisela González, Beatriz García, Agustín Lage (Centro de Inmunología Molecular, La Habana, Cuba), and Elia Neninger (Hospital Hermanos Ameijeiras, La Habana, Cuba).

Conflicts of Interest

None related with this paper. Cimavax EGF is a therapeutic anticancer vaccine produced by the CIM and licensed in Cuba.

Acknowledgements

We gratefully acknowledge the EORTC Headquarters for permission to use the questionnaires QLQ-C30 and QLQ-LC13 in the Cuban clinical trials of safety, efficacy, and effectiveness of Cimavax EGF.

References

- [1] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993; 85(5):365-76.
- [2] Niezgodna HE, Pater JL. A validation study of the domains of the core EORTC quality of life questionnaire. *Qual Life Res* 1993; 2(5): 319-25.
- [3] Nicklasson M, Bergman B. Validity, reliability and clinical relevance of EORTC QLQ-C30 and LC13 in patients with chest malignancies in a palliative setting. *Qual Life Res* 2007; 16(6):1019-28.
- [4] Osoba D, Zee B, Pater J, Warr D, Kaizer L, Latreille J. Psychometric properties and responsiveness of the EORTC Quality of Life Questionnaire (QLQ-C30) in patients with breast, ovarian and lung cancer. *Qual Life Res* 1994; 3(5):353-64.
- [5] Uwer L, Rotonda C, Guillemín F, Miny J, Kaminsky MC, Mercier M, Tournier-Rangéard L, Leonard I, Montcuquet P, Rauch P, Conroy T. Responsiveness of EORTC QLQ-C30, QLQ-CR38 and FACT-C quality of life questionnaires in patients with colorectal cancer. *Health Qual Life Outcomes* 2011; 9:70 [URL <http://www.hqol.com/content/9/1/70>].
- [6] Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). European Organisation for Research and Treatment of Cancer, Brussels 2001.
- [7] Rodríguez PC, Neninger E, García B, Popa X, Viada C, Luaces P, González G, Lage A, Montero E, Crombet T. Safety, immunogenicity and preliminary efficacy of multiple-site vaccination with an Epidermal Growth Factor (EGF) based cancer vaccine in advanced non small cell lung cancer (NSCLC) patients. *J Immune Based Ther Vaccines* 2011, 9:7 (doi: 10.1186/1476-8518-9-7).
- [8] Rodríguez PC, Rodríguez G, González G, Lage A: Clinical development and Perspectives of CIMAvaxEGF, Cuban vaccine for non-small-cell lung cancer therapy. *MEDICC Rev* 2010; 12(1):17-23.
- [9] González G, Crombet T, Catalá M, Mirabal V, Hernández JC, González Y, Marinello P, Guillén G, Lage A: A novel cancer vaccine composed of human-recombinant epidermal growth factor linked to a carrier protein: report of a pilot clinical trial. *Ann Oncol* 1998; 9(4):431-5.
- [10] González G, Crombet T, Torres F, Catalá M, Alfonso L, Osorio M, Neninger E, García B, Mulet A, Pérez R, Lage R: Epidermal growth factor-based cancer vaccine for non-small-cell lung cancer therapy. *Ann Oncol* 2003, 14(3):461-6.
- [11] R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2013. [URL <http://www.R-project.org/>]
- [12] van der Ark. Mokken scale analysis in R. *Journal of Statistical Software* 2007; 20(11):1-19. [URL <http://www.jstatsoft.org/v20/i11/>]
- [13] van der Ark. New developments in Mokken scale analysis in R. *Journal of Statistical Software* 2012; 48(5):1-27. [URL <http://www.jstatsoft.org/v48/i05/>]
- [14] Rizopoulos D. ltm: an R package for latent variable modeling and item response theory analysis. *Journal of Statistical Software* 2006; 17(5):1-25. [URL <http://www.jstatsoft.org/v17/i05/>]
- [15] Rosseel Y. lavaan: an R package for structural equation modeling. *Journal of Statistical Software* 2012; 48(2):1-36. [URL <http://www.jstatsoft.org/v48/i02/>]
- [16] Gundy CM, Fayers PM, Groenvold M, Petersen MA, Scott NW, Sprangers MAG, Velikova G, Aaronson NK. Comparing higher order models for the EORTC QLQ-C30. *Qual Life Res* 2012; 21:1607-17. [DOI 10.1007/s11136-011-0082-6]
- [17] Arraras JI, Suárez J, Arias de la Vega F, Vera R, Asín G, Arrazubi V, Rico M, Teijeira L, Azparren J. The EORTC Quality of Life questionnaire for patients with colorectal cancer: EORTC QLQ-CR29 validation study for Spanish patients. *Clin Transl Oncol.* 2011; 13(1):50-6. [DOI 10.1007/s12094-011-0616-y].
- [18] Arraras Urdaniz JI, Villafranca Iturre E, Arias de la Vega F, Domínguez Domínguez MA, Lainez Milagro N, Manterola Burgaleta A, Martínez Lopez E, Romero Rojano P, Martínez Aguillo M. The EORTC quality of life questionnaire QLQ-C30 (version 3.0). Validation study for Spanish prostate cancer patients. *Arch Esp Urol.* 2008; 61(8):949-54.
- [19] Scott NW, Fayers PM, Bottomley A, Aaronson NK, de Graeff A, Groenvold M, Koller M, Petersen MA, Sprangers MA; EORTC and the Quality of Life Cross-Cultural Meta-Analysis Group. Comparing translations of the EORTC QLQ-C30 using differential item functioning analyses. *Qual Life Res* 2006; 15(6):1103-15.

- [20] Bjorner JB, Petersen MA, Groenvold M, Aaronson N, Ahlner-Elmqvist M, Arraras JI, Brédart A, Fayers P, Jordhoy M, Sprangers M, Watson M, Young T; European Organisation for Research and Treatment of Cancer Quality of Life Group. Use of item response theory to develop a shortened version of the EORTC QLQ-C30 emotional functioning scale. *Qual Life Res* 2004; 13(10):1683-97.
- [21] Petersen MA, Groenvold M, Aaronson N, Blazeby J, Brandberg Y, de Graeff A, Fayers P, Hammerlid E, Sprangers M, Velikova G, Bjorner JB; European Organisation for Research and Treatment of Cancer Quality of Life Group. Item response theory was used to shorten EORTC QLQ-C30 scales for use in palliative care. *J Clin Epidemiol* 2006; 59(1):36-44.
- [22] Teckle P, Peacock S, McTaggart-Cowan H, van der Hoek K, Chia S, Melosky B, Gelmon K. The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities. *Health Qual Life Outcomes* 2011; 9:106. [URL <http://www.hqlo.com/content/9/1/106>]
- [23] Kim SH, Jo M-W, Kim H-J, Ahn J-H. Mapping EORTC QLQ-C30 onto EQ-5D for the assessment of cancer patients. *Health Qual Life Outcomes* 2012; 10:151. [URL <http://www.hqlo.com/contents/10/1/151>]