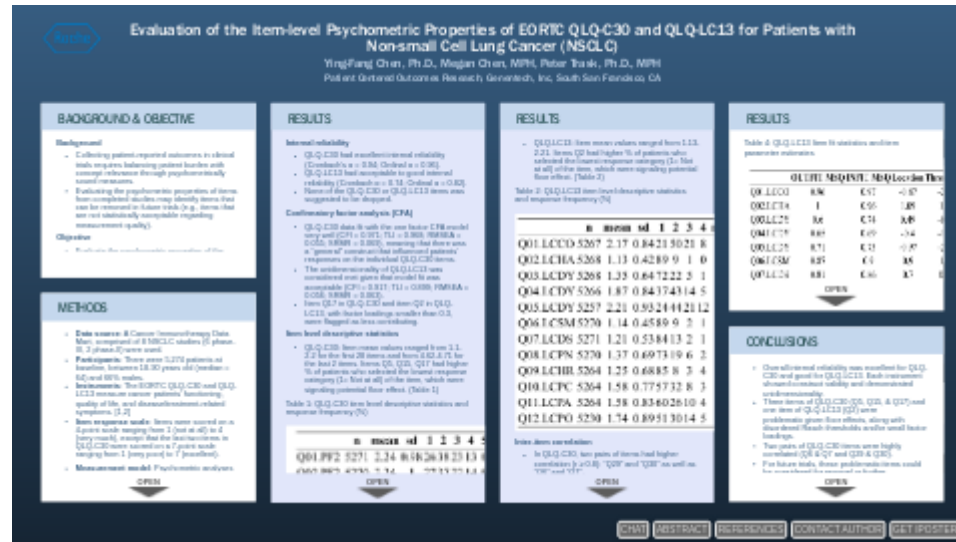# Evaluation of the Item-level Psychometric Properties of EORTC QLQ-C30 and QLQ-LC13 for Patients with Non-small Cell Lung Cancer (NSCLC)

Ying-Fang Chen, Ph.D., Megan Chen, MPH, Peter Trask, Ph.D., MPH

Patient Centered Outcomes Research, Genentech, Inc, South San Francisco, CA

BACKGROUND & OBJECTIVE

Background

- Collecting patient-reported outcomes in clinical trials requires balancing patient burden with concept relevance through psychometrically sound measures.

- Evaluating the psychometric properties of items from completed studies may identify items that can be removed in future trials (e.g., items that are not statistically acceptable regarding measurement quality).

Objective

- Evaluate the psychometric properties of the item-level performance of the EORTC QLQ-C30 and QLQ-LC13 in NSCLC patients.

METHODS

- **Data source**: A Cancer Immunotherapy Data Mart, comprised of 8 NSCLC studies (6 phase-III, 2 phase-II) were used.

- **Participants**: There were 5,274 patients at baseline, between 18-90 years old (median = 64) and 66% males.

- **Instruments**: The EORTC QLQ-C30 and QLQ-LC13 measure cancer patients' functioning, quality of life, and disease/treatment-related symptoms. [1,2]

- **Item response scale**: Items were scored on a 4-point scale ranging from 1 (not at all) to 4 (very much), except that the last two items in QLQ-C30 were scored on a 7-point scale ranging from 1 (very poor) to 7 (excellent).

- **Measurement model**: Psychometric analyses were conducted for each instrument using classical test theory (CTT) and item response theory (IRT)—Rasch Partial Credit model (PCM). [3]

- **Psychometric analysis**: Included item level descriptive statistics and response frequency, confirmatory factor analysis (CFA), internal reliability [4,5], inter-item correlation, item-total correlation, item fit, and PCM item parameter estimates.

RESULTS

Internal reliability

- QLQ-C30 had excellent internal reliability (Cronbach's α = 0.94; Ordinal α = 0.96).

- QLQ-LC13 had acceptable to good internal reliability (Cronbach α = 0.74; Ordinal α = 0.82).

- None of the QLQ-C30 or QLQ-LC13 items was suggested to be dropped.

Confirmatory factor analysis (CFA)

- QLQ-C30 data fit with the one factor CFA model very well (CFI = 0.971; TLI = 0.968; RMSEA = 0.055; SRMR = 0.069), meaning that there was a "general" construct that influenced patients' responses on the individual QLQ-C30 items.

- The unidimensionality of QLQ-LC13 was considered met given that model fit was acceptable (CFI = 0.917; TLI = 0.899; RMSEA = 0.056; SRMR = 0.063).

- Item Q17 in QLQ-C30 and item Q2 in QLQ-LC13, with factor loadings smaller than 0.3, were flagged as less contributing.

Item level descriptive statistics

- QLQ-C30: Item mean values ranged from 1.1-2.2 for the first 28 items and from 4.62-4.71 for the last 2 items. Items Q5, Q15, Q17 had higher % of patients who selected the lowest response category (1= Not at all) of the item, which were signaling potential floor effect. (Table 1)

Table 1: QLQ-C30 item level descriptive statistics and response frequency (%)

	n	mean	sd	1	2	3	4	5	6	7	miss
Q01.PF2	5271	2.24	0.98	26	38	23	13	0	0	0	0
Q02.PF2	5270	2.24	1	27	37	22	14	0	0	0	0
Q03.PF2	5267	1.52	0.78	62	26	9	3	0	0	0	0
Q04.PF2	5272	1.65	0.83	54	30	12	4	0	0	0	0
Q05.PF2	5270	1.13	0.45	90	7	2	1	0	0	0	0
Q06.RF2	5269	1.92	0.94	41	35	16	8	0	0	0	0
Q07.RF2	5266	1.89	0.97	45	30	16	9	0	0	0	0
Q08.DY	5271	1.98	0.9	34	42	16	8	0	0	0	0
Q09.PA	5269	2	0.94	36	37	19	8	0	0	0	0
Q10.FA	5272	2.14	0.85	23	47	23	7	0	0	0	0
Q11.SL	5266	1.92	0.93	40	35	17	8	0	0	0	0
Q12.FA	5269	2.04	0.88	29	44	19	8	0	0	0	0
Q13.AP	5271	1.74	0.91	51	29	13	6	0	0	0	0
Q14.NV	5272	1.32	0.65	76	17	5	2	0	0	0	0
Q15.NV	5270	1.14	0.46	90	8	2	1	0	0	0	0
Q16.CO	5269	1.52	0.81	64	24	8	4	0	0	0	0
Q17.DI	5268	1.19	0.48	84	13	2	1	0	0	0	0
Q18.FA	5264	2.09	0.83	24	50	20	6	0	0	0	0
Q19.PA	5267	1.76	0.93	51	29	13	7	0	0	0	0
Q20.CF	5267	1.42	0.7	68	23	7	2	0	0	0	0
Q21.EF	5266	1.75	0.79	44	41	12	3	0	0	0	0
Q22.EF	5265	2.03	0.87	29	46	17	7	0	0	0	0
Q23.EF	5265	1.65	0.77	50	37	10	3	0	0	0	0
Q24.EF	5260	1.67	0.8	50	36	10	3	0	0	0	0
Q25.CF	5267	1.5	0.69	60	32	7	2	0	0	0	0
Q26.SF	5266	1.65	0.83	53	32	11	4	0	0	0	0
Q27.SF	5268	1.8	0.92	47	32	14	7	0	0	0	0
Q28.FI	5266	1.63	0.87	58	26	11	5	0	0	0	0
Q29.QL2	5266	4.62	1.32	2	4	13	26	29	20	7	0
Q30.QL2	5265	4.71	1.41	2	5	13	23	27	22	10	0

RESULTS

- QLQ-LC13: Item mean values ranged from 1.13-2.21. Items Q2 had higher % of patients who selected the lowest response category (1= Not at all) of the item, which were signaling potential floor effect. (Table 2)

Table 2: QLQ-LC13 item level descriptive statistics and response frequency (%)

	n	mean	sd	1	2	3	4	miss
Q01.LCCO	5267	2.17	0.84	21	50	21	8	0
Q02.LCHA	5268	1.13	0.42	89	9	1	0	0
Q03.LCDY	5268	1.35	0.64	72	22	5	1	0
Q04.LCDY	5266	1.87	0.84	37	43	14	5	0
Q05.LCDY	5257	2.21	0.93	24	44	21	12	0
Q06.LCSM	5270	1.14	0.45	89	9	2	1	0
Q07.LCDS	5271	1.21	0.53	84	13	2	1	0
Q08.LCPN	5270	1.37	0.69	73	19	6	2	0
Q09.LCHR	5264	1.25	0.68	85	8	3	4	0
Q10.LCPC	5264	1.58	0.77	57	32	8	3	0
Q11.LCPA	5264	1.58	0.83	60	26	10	4	0
Q12.LCPO	5230	1.74	0.89	51	30	14	5	1

Inter-item correlation

- In QLQ-C30, two pairs of items had higher correlation ($r \geq 0.8$): "Q29" and "Q30" as well as "Q6" and "Q7".
- In QLQ-LC13, no high correlation between items was identified.

Item-total correlation

- Each of the individual items in QLQ-C30 and QLQ-LC13 had good item discrimination ($r \geq 0.3$), evaluated via item-total correlation coefficients.

Item fit statistics

- Overall data fit PCM very well for the QLQ-C30 (mean of INFIT = 0.98, mean of OUTFIT = 0.99; Table 3) and QLQ-LC13 (mean of INFIT = 0.92, mean of OUTFIT = 0.94; Table 4)
- None of the individual items from QLQ-C30 or QLQ-LC13 were identified as misfitting to the measurement. (Tables 3 & 4)

Table 3: QLQ-C30 Item fit statistics and item parameter estimates

	OUTFIT MSQ	INFIT MSQ	Location	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Threshold 5	Threshold 6
Q01.PF2	0.87	0.87	-0.42	-1.77	-0.09	0.61			
Q02.PF2	0.9	0.89	-0.43	-1.71	-0.07	0.49			
Q03.PF2	0.7	0.81	0.91	0.08	1.09	1.57			
Q04.PF2	0.87	0.87	0.69	-0.3	0.76	1.61			
Q05.PF2	0.7	0.95	1.81	2.08	1.88	1.46			
Q06.RF2	0.67	0.71	0.14	-0.96	0.43	0.95			
Q07.RF2	0.7	0.75	0.16	-0.69	0.3	0.86			
Q08.DY	1.06	1.09	0.05	-1.39	0.57	0.96			
Q09.PA	1.07	1.04	0.03	-1.23	0.28	1.03			
Q10.FA	0.69	0.7	-0.18	-2.07	0.23	1.3			
Q11.SL	1.29	1.18	0.16	-0.97	0.37	1.08			
Q12.FA	0.7	0.71	-0.03	-1.66	0.43	1.13			
Q13.AP	1.01	1.02	0.42	-0.4	0.59	1.07			
Q14.NV	0.96	1.02	1.31	0.86	1.48	1.6			
Q15.NV	0.96	1.03	1.79	2.01	1.93	1.43			
Q16.CO	1.47	1.28	0.81	0.18	1.12	1.13			
Q17.DI	1.79	1.22	1.86	1.36	2.16	2.07			
Q18.FA	0.7	0.71	-0.09	-2.07	0.46	1.35			
Q19.PA	0.73	0.82	0.38	-0.41	0.54	1.02			
Q20.CF	0.85	0.88	1.18	0.36	1.31	1.86			
Q21.EF	1.08	1.04	0.59	-0.96	1.06	1.66			
Q22.EF	1.15	1.11	-0.02	-1.71	0.56	1.1			
Q23.EF	1.09	1.05	0.79	-0.61	1.21	1.76			
Q24.EF	0.98	0.99	0.69	-0.6	1.14	1.53			
Q25.CF	1.27	1.18	1.15	-0.17	1.6	2.03			
Q26.SF	0.86	0.92	0.67	-0.39	0.94	1.46			
Q27.SF	0.78	0.82	0.34	-0.63	0.58	1.07			
Q28.FI	1.81	1.45	0.63	-0.08	0.78	1.19			
Q29.QL2_r	0.96	0.97	-0.27	-3.08	-1.68	-0.54	0.59	1.49	1.61
Q30.QL2_r	1.03	1.04	-0.17	-2.77	-1.42	-0.45	0.55	1.32	1.73

RESULTS

Table 4: QLQ-LC13 Item fit statistics and item parameter estimates

	OUTFIT MSQ	INFIT MSQ	Location	Threshold 1	Threshold 2	Threshold 3
Q01.LCCO	0.96	0.97	-0.87	-2.56	-0.27	0.22
Q02.LCHA	1	0.96	1.09	1.16	1.19	0.92
Q03.LCDY	0.6	0.74	0.49	-0.03	0.78	0.73
Q04.LCDY	0.65	0.69	-0.4	-1.64	0.14	0.3
Q05.LCDY	0.71	0.73	-0.97	-2.27	-0.42	-0.21
Q06.LCSM	0.87	0.9	0.9	1.21	1	0.49
Q07.LCDS	0.81	0.86	0.7	0.72	1.06	0.31
Q08.LCPN	1.11	1.04	0.36	0.07	0.43	0.58
Q09.LCHR	1.59	1.2	0.28	1.16	0.31	-0.61
Q10.LCPC	0.91	0.9	0.07	-0.76	0.44	0.52
Q11.LCPA	1	0.99	-0.01	-0.5	0.01	0.45
Q12.LCPO	1.13	1.07	-0.25	-0.88	-0.23	0.37

PCM item parameter estimates

- Item difficulties ranged from -0.43 to 1.86 logits for QLQ-C30 (Table 3) and from -0.97 to 1.09 logits for QLQ-LC13 (Table 4).

- Items Q5 and Q15 in QLQ-C30 (Table 3) and the item Q2 in QLQ-LC13 (Table 4) had disordered item option estimates.

CONCLUSIONS

- Overall internal reliability was excellent for QLQ-C30 and good for QLQ-LC13. Each instrument showed construct validity and demonstrated unidimensionality.

- Three items of QLQ-C30 (Q5, Q15, & Q17) and one item of QLQ-LC13 (Q2) were problematic given floor effects, along with disordered Rasch thresholds and/or small factor loadings.

- Two pairs of QLQ-C30 items were highly correlated (Q6 & Q7 and Q29 & Q30).

- For future trials, these problematic items could be considered for removal or further qualitative review.

- The two highly correlated pairs constitute the role functioning and GHS/QoL scales, and while highly correlated, measured different aspects of the same constructs and should be retained.

- All remaining individual items in QLQ-C30 and QLQ-LC13 performed well based on multiple psychometric indications.

ABSTRACT

OBJECTIVES: Collecting patient-reported outcomes in clinical trials requires balancing patient burden with concept relevance through psychometrically sound measures. Evaluating psychometric properties of items from completed studies may identify items that can be removed in future trials (e.g., items that are not statistically acceptable regarding measurement quality). This study evaluated this idea using item-level psychometric analyses of the EORTC QLQ-C30 and QLQ-LC13 in patients with NSCLC.

METHODS: This research used data from a Cancer Immunotherapy Data Mart, comprising eight NSCLC studies (six phase-III, two phase-II). There were 5,274 patients at baseline, between 18-90 years old (median = 64) and 66% males. The QLQ-C30 and QLQ-LC13 measure cancer patients' functioning, quality of life, and disease/treatment-related symptoms. Psychometric analyses were conducted for each instrument using classical test theory (CTT) and item response theory (IRT)—Rasch Partial Credit model (PCM).

RESULTS: Internal reliability was excellent for QLQ-C30 and good for QLQ-LC13. Each instrument showed construct validity and demonstrated unidimensionality; all individual items had statistically significant positive factor loadings. Data fit with PCM well. All individual items in each instrument provided promising item fit and displayed adequate item discrimination. Item difficulties ranged from -0.43 to 1.86 logits for QLQ-C30 and from -0.97 to 1.09 logits for QLQ-LC13. Three items of QLQ-C30 (Q5, Q15, & Q17) and one item of QLQ-LC13 (Q2) were flagged as potentially less productive to the measurement, given floor effects (i.e., the majority of patients were not experiencing the item's description), along with disordered Rasch thresholds and/or small factor loadings. Two pairs of QLQ-C30 items were considered redundant (Q6 & Q7 and Q29 & Q30). All remaining individual items in QLQ-C30 or QLQ-LC13 performed well based on multiple psychometric indications.

CONCLUSIONS: Overall, QLQ-C30 and QLQ-LC13 are reliable, valid, and psychometrically sound. Few items were identified as potential candidates for removal in future assessments of NSCLC patients.

REFERENCES

1. Aaronson, N. K., Ahmedzai, S., Bergman, B, & et al. (1993). 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. 85:365-76.

2. Bergman, B., Aaronson, N. K., Ahmedzai, S., Kaasa, S., & Sullivan, M. (1994). Thea modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. Eur J Cancer, 1994; 30A: 635-642

3. Masters,G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

4. Cronbach, L. J (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 45, 99-105.

5. Zumbo, B. D., A. M. Gadermann, and C. Zeisser. (2007). Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales. Journal of Modern Applied Statistical Methods 6: 21–29.