original article

Annals of Oncology 19: 2053–2060, 2008
doi:10.1093/annonc/mdn417
Published online 24 July 2008

Validation of the simplified Chinese version of EORTC QLQ-C30 from the measurements of five types of inpatients with cancer

C. Wan[1]*, Q. Meng[1], Z. Yang[1], X. Tu[2], C. Feng[2], X. Tang[3] & C. Zhang[3]

[1]*School of Public Health, Kunming Medical University, Kunming, Yunnan, China;* [2]*Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, New York, USA;* [3]*Yunnan Tumor Hospital, Kunming, Yunnan, China*

Received 20 June 2007; revised 8 November 2007; revised 26 April 2008; accepted 10 June 2008

Background: European Organization for Research and Treatment quality of life (QOL) questionnaire (QLQ-C30) has been used frequently and many language versions have been developed, including the simplified Chinese version. It is important to study psychometric properties of the simplified Chinese version from the clinical standpoint.
Patients and methods: The simplified Chinese version of the QLQ-C30 was used in a longitudinal study of 600 patients with five types of cancer: lung, breast, head and neck, colorectal, and stomach. The psychometric properties of the scale were evaluated by indicators of validity and reliability coefficients such as Cronbach's α and Pearson's correlation coefficient *r*, standardized response mean (SRM), correlational analysis, *t*-tests, and structural equation models.
Results: Correlation and structural equation model analyses confirmed good construct validity with root mean square error of approximation 0.054, standardized root mean square residual 0.037, non-normed fit index 0.972, and comparative fit index 0.980. The α coefficients for all domains are >0.7 except for cognitive functioning (0.49). The test–retest reliability coefficients for most domains are >0.80 except for appetite loss (0.77) and diarrhea (0.75). The QOL score changes after treatments were of statistical significance with higher or moderate SRM in most domains.
Conclusion: The simplified Chinese version of QLQ-C30 has good validity, reliability, and responsiveness and can be used to measure QOL for Chinese cancer patients.
Key words: quality of life, QLICP-GM, structural equation model, standardized response mean, psychometric properties

introduction

In the past three decades, research on quality of life (QOL) has become an international phenomenon. Research on QOL for cancer patients has taken center stage in the medical field, with thousands of articles on QOL published every year. QOL has also been valued as one of the indispensable outcomes in clinical medicine. This trend makes it all the more necessary to have a clear methodology for the development and application of QOL instruments. Many QOL instruments, such as the quality of life questionnaire (QLQ) series from European Organization for Research and Treatment (EORTC) [1, 2], the Functional Assessment of Cancer Therapy (FACT) series from Center on Outcomes, Research and Education in the United States [3, 4], and the functional living index-cancer [5] have been developed and widely used in cancer clinical research.

Of the instrument systems mentioned above, the QLQs from EORTC were developed to assess the health-related QOL of cancer patients participating in international clinical trials. The core questionnaire of the system, QLQ-C30, is used to measure QOL for all cancer patients. By adding specific modules into QLQ-C30, EORTC developed several cancer-specific questionnaires, such as the QLQ-C30 and QLQ-BR23 for breast cancer [6] and QLQ-C30 and QLQ-LC13 for lung cancer [7]. The current QLQ-C30 (V3.0) is a 30-item cancer-specific scale addressing various aspects of QOL classified into 15 domains including five functional subscales (physical, role, emotional, cognitive, and social), three multi-item symptom subscales (fatigue, pain, and nausea/vomiting), a global health/QOL subscale, and six single items addressing various symptoms and perceived financial impact. All items use a four-point Likert scale, namely, not at all, a little, quite a bit, and very much, except for the global health status (QL)/QOL (Q29 and Q30), in which a seven-point scale is used. Standardized scores for all 15 domains ranged from 0 to 100, with a higher score indicating better QOL for the functional

Correspondence to: Prof. C. Wan, Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA. Tel: +01-585-275-5294; Fax: +01-585-273-1031; E-mail: wanchh@hotmail.com

and global health domains and worse QOL (a greater degree of symptoms) for the symptom domains.

Many research studies [8–10] have confirmed that QLQ-C30 is an excellent QOL instrument with good psychometric properties. For example, Kemmler et al. [9] reported Cronbach's α coefficients ranging from 0.562 to 0.941, with 0.769 for physical functioning (PF), 0.759 for social functioning (SF), 0.797 for emotional functioning (EF), 0.864 for role functioning (RF), 0.661 for cognitive functioning (CF), and 0.941 for QL. Luo et al. [10] studied its validation in Singapore and the results showed that Spearman's correlations between the QLQ-C30 and SF-36 scales assessing similar dimensions of health related quality of life ranged from 0.35 to 0.67, subjects with mild symptoms had better scores than those with severe symptoms for all six QLQ-C30 QOL scales ($P < 0.05$ for five scales), and Cronbach's α ranged from 0.19 for the cognitive functioning scale to 0.91 for the global QOL scale.

As the QLQ-C30 is of greatest importance (since it serves as the core of the system), it has been translated into >54 language versions including Chinese. In particular, EORTC developed two versions of QLQ-C30 (the simplified Chinese version for Mainland China and the traditional Chinese version for Taiwan and Hong Kong) according to strict translation procedures including forward translation, back translation, and cultural adaptation [11–14]. The simplified Chinese version was developed using simplified Chinese characters that are used in Mainland China. Chie et al. [11–13] reported the validation of the traditional Chinese version of QLQ-C30 (V3.0) from data in patients with lung cancer, breast cancer, and nasopharyngeal carcinoma, respectively. Zhao and Kanda [14] presented the translation procedures and validation of the simplified Chinese version of QLQ-C30 (V2.0). Psychometrics of the current simplified Chinese version of QLQ-C30 (V3.0) must be confirmed in field tests before it can be used broadly in China. The translation procedures can be seen in other papers [14]; in this paper, we study psychometrics of the current simplified Chinese version of QLQ-C30 (V3.0) using data from a cancer trial consisting of patients with five types of cancer conducted in Mainland China.

materials and methods

patients

The study population consisted of inpatients with one of the five types of cancers—lung cancer, breast cancer, colorectal cancer, head and neck cancer, and stomach cancer—in Yunnan Tumor Hospital, at any age, any clinical stage, and any treatments, who can read and understand the questionnaires. The two exclusion criteria were an inability to read the questionnaires because of illiteracy and an inability to fill out the questionnaires because of deteriorated disease. We limited to these five cancers for they have higher prevalence and thus are good representatives of all cancer patients in China.

instruments

The study subjects were assessed by the simplified Chinese version of QLQ-C30 (V3.0) and the quality of life instruments for cancer patients-general module (QLICP-GM). The latter was developed by our research group for use with Chinese cancer patients and was used here to compare the properties of the two instruments.

The QLICP-GM was developed on the basis of QOL as defined by the World Health Organization and on programmed decision procedures, including multiple nominal and focus group discussions and pilot and field tests [15]. It includes 32 items classified into four domains and nine facets, with each item scored on the basis of a five-point Likert scale. The scoring method of this instrument is similar to that of FACT-G [3, 4] (see Table 1 for details). Regarding psychometrics, correlational and structural equation model analyses confirmed good construct validity [15]. The internal consistency α for the domains of physical (PHD), psychological (PSD), social, and common symptom/side-effects, and the overall scale were 0.77, 0.85, 0.61, 0.71, and 0.88, respectively. The test–retest reliability for all domains and the overall scale are >0.85. QOL score changes after treatment were of statistical significance on the three domains of physical, psychological, and common symptoms and side-effects, and the overall instrument with standardized response mean (SRM) ranged from 0.16 to 0.67. Overall, the QLICP-GM has good validity, reliability, and responsiveness and can be used to measure QOL for patients with cancer in China.

collection of data

The simplified Chinese versions of QLQ-C30 and QLICP-GM were administered to a sample of 600 cancer inpatients with five types of cancers

Table 1. Scoring method of the quality of life instrument QLICP–GM

Domains/facets	Number of items	Range of scores	Scoring method (sum of item's scores)
Physical domain (PHD)	7	7–35	BPF + SXF + IDF
Basic physiologic function (BPF)	2	2–10	GPH1 + GPH2
Sexual function (SXF)	1	1–5	GPH3
Independence function (IDF)	4	4–20	GPH4 + GPH5 + GPH6 + GPH7
Psychological domain (PSD)	12	12–60	EMO + REC
Emotion (EMO)	9	9–45	GPS1 + … + GPS8 + GPS11
Recognition (REC)	3	3–15	GPS9 + GPS10 + GPS12
Social domain (SOD)	6	6–30	SSS + ELE
Social support and safety (SSS)	4	4–20	GSO1 + GSO2 + GSO3 + GSO4
Effects on life and economics (ELE)	2	2–10	GSO5 + GSO6
Common symptom and side-effect domain (SSD)	7	7–35	SEF + CST
Side-effect (SEF)	4	4–20	GSS1 + GSS2 + GSS3 + GSS6
Common symptom (CST)	3	3–15	GSS4 + GSS5 + GSS7
Total (TOT)	32	32–160	PHD + PSD + SOD + SSD

GPH, general physical function; GPS, general psychological function; GSO, general social function; GSS, general symptom and side effects.

in Yunnan Tumor Hospital. After obtaining appropriate institutional review board approvals, the investigators explained the trial and the scale to the patients and obtained informed consent from patients who agreed to participate in the study and met the inclusion and exclusion criteria. Each participant was asked to fill in the questionnaires by himself/herself at the time of admission to the hospital. To calculate the test–retest reliability, we assessed 554 inpatients a second time 1–2 days after hospitalization. A subsample of patients (335) were measured a third time at discharge (after ~4 weeks of treatment) to evaluate the responsiveness. To ensure good quality, investigators were on hand each time to review the questionnaires. If missing values were found, the questionnaire would be returned to the patient for completion.

analysis method

The Chinese version of QLQ-C30 was evaluated for reliability, validity, and responsiveness. Construct validity was evaluated by the Pearson's correlation coefficient, *r*, among items and domains and by the confirmatory factor analysis using structural equation models. Multi-trait scaling analysis [16] was employed to test item convergent and discriminant validity, with two criteria: (i) convergent validity is supported when an item–domain correlation is ≥0.40 and (ii) discriminant validity is revealed when item–domain correlation is higher than that with other domains. Also, Pearson's correlation coefficients between the QLQ-C30 and the QLICP-GM were calculated to evaluate the criterion-related validity due to the lack of a gold standard. It was hypothesized that conceptually related domains would correlate substantially high with each other. The known-groups method was used to assess the clinical validity [2]. Internal consistency reliability was evaluated using Cronbach's α coefficient for each domain. Test–retest reliability was evaluated by the Pearson's correlation coefficient between the first and the second assessment, and intraclass correlation (ICC) under the two-way mixed model [17, 18]. Responsiveness was assessed by comparing the mean difference between the first and the third assessments (pre-treatment and post-treatment) using SRM as the computed effect size [19, 20]. It was on the basis of our experience and expectation that QOL scores would change after treatments and good instruments would reflect these changes. All statistical analyses were carried out using SPSS (V13.0) and LISREL 8.54 on a Windows XP platform.

results

sociodemographic and clinical characteristics

The sample included 600 cancer patients: lung cancer 85, breast cancer 186, colorectal cancer 110, head and neck cancer 133, stomach cancer 86, with age ranging from 11 to 87 years [median = 52.0 and mean (standard deviation) = 52.4 (12.6)]. Among them, 303 (50.5%) were male; 145 (24.1%) patients finished primary school, 343 (47.2%) completed high school, and 109 (18.2%) had a college degree (see Table 2 for details).

construct validity

Pearson's correlation and the structural equation model were used to study the construct validity of the instrument.

The correlation coefficients among items and multi-item domains showed strong correlations between items and their own domains (all correlation coefficients *r* are >0.70 with a range of 0.71–0.94). On the other hand, there were weak correlations between items and other domains (see Table 3). These also demonstrated item convergent and discriminant validity.

Table 2. Sociodemographic and clinical characteristics of the sample (n = 600)

Characteristics	*n*	%
Gender		
Male	303	50.5
Female	297	49.5
Ethnic groups		
Han	512	85.3
Others	86	14.4
Missing	2	0.3
Age		
<30	19	3.2
30–39	78	13.0
40–49	158	26.3
50–59	159	26.5
≥60	186	31.0
Income[a]		
Poor	253	42.2
Fair	320	53.3
High	10	1.7
Missing	17	2.8
Occupation		
Worker	165	27.5
Farmer	70	11.7
Teacher	79	13.2
Cadre	125	20.8
Others	156	26.0
Missing	5	0.8
Marital status		
Married	582	97.0
Others	18	3.0
Clinical stage		
Early	33	5.5
Medium	114	19.0
Late	106	17.7
Missing	347	57.8
Education		
Primary school	145	24.1
Middle school or high school	231	38.5
Professional secondary school	112	18.7
College	109	18.2
Missing	3	0.5
Medical insurance		
Self-paid	157	26.2
Partly public insurance	242	40.3
Public insurance	183	30.5
Missing	18	3.0
Treatments		
Radiotherapy	17	2.8
Chemotherapy	73	12.2
Surgery	80	13.3
Chemotherapy + radiotherapy	73	12.2
Surgery + radiotherapy	40	6.7
Chemotherapy + surgery	177	29.5
Chemotherapy + radiotherapy + surgery	37	6.2
Others	26	4.3
Missing	77	12.8

[a]These categories are reported by patients on the basis of their perceptions with no real objective numbers.

Table 3. Pearson's correlation coefficients among items and multi-item domains of the simplified Chinese version of quality of life questionnaire (QLQ-C30) ($n = 600$)

Item	PF	RF	EF	CF	SF	QL	FA	NV	PA
Q1	0.81	0.36	0.26	0.27	0.30	0.22	−0.40	−0.13	−0.33
Q2	0.84	0.41	0.26	0.29	0.33	0.24	−0.46	−0.09	−0.32
Q3	0.76	0.36	0.18	0.32	0.28	0.25	−0.39	−0.13	−0.33
Q4	0.79	0.43	0.22	0.30	0.33	0.21	−0.43	−0.13	−0.38
Q5	0.73	0.41	0.19	0.25	0.28	0.22	−0.42	−0.18	−0.35
Q6	0.49	0.94	0.36	0.40	0.56	0.27	−0.59	−0.23	−0.47
Q7	0.45	0.94	0.39	0.39	0.54	0.32	−0.56	−0.26	−0.47
Q8	−0.25	−0.25	−0.29	−0.35	−0.25	−0.23	0.40	0.25	0.42
Q9	−0.32	−0.35	−0.39	−0.41	−0.31	−0.28	0.55	0.32	0.88
Q19	−0.45	−0.53	−0.46	−0.50	−0.47	−0.35	0.62	0.30	0.90
Q10	−0.46	−0.51	−0.34	−0.38	−0.41	−0.33	0.83	0.23	0.57
Q12	−0.43	−0.49	−0.42	−0.47	−0.45	−0.29	0.84	0.35	0.55
Q18	−0.44	−0.53	−0.46	−0.47	−0.46	−0.29	0.83	0.30	0.54
Q11	−0.33	−0.38	−0.38	−0.45	−0.38	−0.28	0.49	0.30	0.45
Q13	−0.33	−0.37	−0.40	−0.44	−0.36	−0.29	0.55	0.48	0.48
Q14	−0.16	−0.27	−0.32	−0.31	−0.27	−0.19	0.37	0.92	0.34
Q15	−0.13	−0.22	−0.26	−0.27	−0.22	−0.15	0.28	0.92	0.29
Q16	−0.11	−0.11	−0.20	−0.19	−0.13	−0.10	0.25	0.27	0.27
Q17	−0.04	−0.12	−0.11	−0.23	−0.08	−0.12	0.23	0.23	0.21
Q20	0.37	0.40	0.47	0.84	0.35	0.30	−0.50	−0.30	−0.52
Q25	0.19	0.26	0.34	0.75	0.35	0.20	−0.32	−0.19	−0.30
Q21	0.23	0.32	0.79	0.38	0.37	0.19	−0.34	−0.26	−0.36
Q22	0.25	0.37	0.84	0.40	0.49	0.23	−0.38	−0.26	−0.39
Q23	0.21	0.24	0.71	0.40	0.29	0.22	−0.38	−0.24	−0.37
Q24	0.17	0.29	0.76	0.40	0.39	0.23	−0.39	−0.22	−0.36
Q26	0.36	0.48	0.44	0.38	0.90	0.21	−0.43	−0.23	−0.37
Q27	0.33	0.58	0.45	0.41	0.87	0.22	−0.50	−0.25	−0.44
Q28	−0.23	−0.33	−0.36	−0.22	−0.51	−0.13	0.30	0.21	0.23
Q29	0.25	0.32	0.29	0.28	0.25	0.94	−0.39	−0.19	−0.37
Q30	0.27	0.27	0.24	0.30	0.21	0.94	−0.29	−0.15	−0.27

PF, physical functioning; RF, role functioning; EF, emotional functioning; CF, cognitive functioning; SF, social functioning; QL, global health status; FA: fatigue; NV, nausea and vomiting; PA, pain.

Structural equation model analysis showed that the structure of the QLQ-C30 can be grouped into 15 domains (see Table 4), with goodness of fit chi-square $\chi^2 = 762.28$ ($P < 0.0001$), root mean square error of approximation (RMSEA) = 0.054, 90% confidence interval = (0.049–0.059), non-normed fit index = 0.972, comparative fit index (CFI) = 0.980, and standardized root mean square residual (SRMR) = 0.037. These indicators demonstrated excellent fitting of data [21, 22].

criteria-related validity

The QLICP-GM developed by our research group was used as criteria due to the lack of a gold standard. The Pearson's correlation coefficients of scores among domains of two instruments (QLQ-C30 and QLICP-GM) are presented in Table 5. It can be seen from Table 5 that the between-instrument correlations are higher for the same and similar domains than across different and nonsimilar domains. For example, the coefficient between PF of QLQ-C30 and PHD of QLICP-GM is 0.41, higher than that between PF of QLQ-C30 and any other domains of QLICP-GM. Also the correlation between EF and PSD is 0.67, higher than that between EF and any other domains of QLICP-GM.

clinical validity (known-groups validity)

It is well known that a patient's clinical stage can affect QOL. Generally speaking, patients at an earlier stage can be treated by surgery and those at a late stage are often treated by other methods such as chemotherapy. It is expected that the patients who had surgery would have different QOL scores as compared with those who received chemotherapy. Thus, we selected two subgroups of inpatients when they were hospitalized, a surgery group (80 cases) and a chemotherapy group (78 cases), and compared the mean QOL scores between the two treatment groups by *t*-tests. As shown in Table 5, there were significant differences for nine out of 15 domains, including almost all of the symptom domains (exception of insomnia and diarrhea). The surgery group had higher QOL than the chemotherapy counterpart on the whole because it had a higher mean score for functioning domains and a lower score for symptom domains. In other words, surgical patients

have better QOL in role, social, and cognitive functioning, and fewer symptoms in fatigue, pain, nausea/vomiting, constipation, and dyspnea domains (a lower score means fewer symptoms and thus higher QOL for these symptom domains).

Table 4. Results of the structure of the simplified Chinese version of quality of life questionnaire (QLQ-C30) revealed by structural equation model ($n = 600$)

Domains (subscales/single items)	Items/standardized path coefficients (error)
Physical functioning	Q1/0.77 (0.41), Q2/0.83 (0.32), Q3/0.72 (0.48), Q4/0.73 (0.46), Q5/0.67 (0.55)
Role functioning	Q6/0.89 (0.21), Q7/0.87 (0.25)
Emotional functioning	Q21/0.72 (0.48), Q22/0.81 (0.35), Q23/0.58 (0.57), Q24/0.68 (0.54)
Cognitive functioning	Q20/0.66 (0.57), Q25/0.45 (0.80)
Social functioning	Q26/0.78 (0.39), Q27/0.78 (0.39)
Global health status/quality of life	Q29/0.98 (0.04), Q30/0.82 (0.32)
Fatigue	Q10/0.71 (0.49), Q12/0.77 (0.41), Q18/0.77 (0.41)
Nausea and vomiting	Q14/0.94 (0.11), Q15/0.78 (0.39)
Pain	Q9/0.71 (0.49), Q19/0.83 (0.31)
Dyspnea	Q8/1.0 (0.00)
Insomnia	Q11/1.0 (0.00)
Appetite loss	Q13/1.0 (0.00)
Constipation	Q16/1.0 (0.00)
Diarrhea	Q17/1.0 (0.00)
Financial difficulties	Q28/1.0 (0.00)

reliability

The Cronbach's α and test–retest reliability coefficients (correlation coefficients *r* and ICC) of all domains are presented in Table 6. As can be seen in Table 6, the Cronbach's α coefficients of all domains are >0.7 except for CF (0.49). The test–retest reliability coefficients for most domains are >0.80 except for AP (0.77) and DI (0.75). The paired *t*-tests indicated no statistically significant change of domain scores between the first and the second measurements ($P > 0.05$). ICC is very similar to correlation coefficient *r*, indicating no significant drift in the mean response for all domains.

responsiveness

The third assessment (post-treatment) was completed by 335 patients (breast cancer 94, lung cancer 61, colorectal cancer 54, head and neck cancer 60, and stomach cancer 66). The mean duration of time between pre- and post-treatment assessments was 26.6 ± 18.2 days (median = 21.0). Paired *t*-tests were used to examine changes of mean scores from pre- to post-treatment for each domain of the QLQ-C30.

Patients were divided into two groups to analyze responsiveness because the direction of change of QOL after treatment differed across the cancer types; the change was negative for the group of breast and lung cancer, but positive for the group of colorectal, stomach, and head and neck cancer. The results are reported in Tables 7 and 8, respectively, with the SRM being the difference (absolute value) divided by its standard deviation.

It can be seen from Table 7 that for the breast and lung cancer groups, QOL score changes after treatment were of statistical significance on all domains except for QL, with SRM being >0.50 for all domains except for QL (0.10) and FI (0.28). Table 8 shows that for the colorectal, stomach, and head

Table 5. Criterion-related validity and clinical validity of the simplified Chinese version of quality of life questionnaire (QLQ-C30)

QLQ-C30 domains	Criterion-related validity, correlations with QLICP-GM				Clinical validity, comparisons of two treatments					
	PHD	PSD	SOD	SSD	Surgery		Chemotherapy		*t*	*P*
					Mean	SD	Mean	SD		
Physical functioning	**0.41**	0.22	0.16	0.23	78.35	19.22	75.43	19.31	0.94	0.350
Role functioning	0.39	0.29	0.25	0.37	64.32	30.36	53.65	25.34	2.33	0.021
Emotional functioning	0.26	**0.67**	0.37	0.39	69.94	21.26	65.26	22.72	1.31	0.193
Cognitive functioning	0.31	**0.44**	0.32	0.44	75.21	18.56	68.04	23.53	2.09	0.039
Social functioning	0.27	0.40	**0.38**	0.32	65.19	24.05	55.48	25.62	2.41	0.017
Global health status	0.35	0.23	0.20	0.28	56.52	23.67	53.36	26.60	0.77	0.443
Fatigue	−0.43	−0.37	−0.22	**−0.55**	33.19	21.83	44.14	19.13	−3.28	0.001
Nausea and vomiting	−0.14	−0.28	−0.21	**−0.55**	17.30	22.08	28.70	26.42	−2.89	0.004
Pain	−0.36	−0.38	−0.22	**−0.53**	29.11	23.03	39.50	24.92	−2.67	0.008
Dyspnea	−0.22	−0.27	−0.16	**−0.35**	17.08	22.50	28.31	27.03	−2.80	0.006
Insomnia	−0.36	−0.32	−0.25	**−0.39**	30.38	26.79	36.07	30.30	−1.23	0.221
Appetite loss	−0.32	−0.30	−0.17	**−0.47**	30.38	24.57	42.92	26.92	−3.00	0.003
Constipation	−0.07[a]	−0.14	−0.12	**−0.31**	23.21	23.48	31.46	27.54	−1.98	0.050
Diarrhea	−0.08[a]	−0.12	0.00[a]	−0.27	16.03	23.18	15.28	22.33	0.20	0.839
Financial difficulties	−0.14	−0.33	−0.46	−0.23	56.54	33.06	58.90	32.16	−0.45	0.656

The numbers in bold are the correlation coefficients between the same/similar domians of the two instruments.

[a]Correlations: no statistical significance ($P > 0.05$) and others significant ($P < 0.05$).

PHD, physical domain; PSD, psychological domain; SOD, social domain; SSD, common symptom/side-effects domain; QLICP-GM, quality of life instruments for cancer patients-general module; SD, standard deviation.

Table 6. Reliability of the simplified Chinese version of quality of life questionnaire (QLQ-C30) ($n = 600$ for α, $n = 554$ for r)

Domain (subscales/items)	Internal consistency (Cronbach's coefficient)	Test–retest reliability[a] (correlation coefficient)	Intraclass correlations (95% CI)
Physical functioning	0.85	0.89	0.89 (0.88–0.91)
Role functioning	0.86	0.83	0.83 (0.81–0.86)
Emotional functioning	0.79	0.84	0.84 (0.82–0.87)
Cognitive functioning	0.49	0.83	0.82 (0.80–0.85)
Social functioning	0.75	0.80	0.80 (0.77–0.83)
Global health status	0.87	0.89	0.89 (0.87–0.90)
Fatigue	0.78	0.84	0.84 (0.81–0.86)
Nausea and vomiting	0.82	0.86	0.86 (0.84–0.88)
Pain	0.72	0.87	0.87 (0.85–0.89)
Dyspnea	–	0.87	0.87 (0.85–0.90)
Insomnia	–	0.85	0.85 (0.83–0.87)
Appetite loss	–	0.77	0.77 (0.74–0.80)
Constipation	–	0.80	0.80 (0.76–0.83)
Diarrhea	–	0.75	0.75 (0.71–0.78)
Financial difficulties	–	0.84	0.84 (0.81–0.86)

[a]All correlation coefficients are statistically significant.
–, represent missing, because the domain consisted of one item.
CI, confidence interval.

Table 7. Responsiveness of the simplified Chinese version of quality of life questionnaire (QLQ-C30) for measuring quality of life of patients with breast and lung cancer ($n = 155$)

Domains (subscales/items)	Pre-treatment		Post-treatment		Differences		SRM	t	P
	Mean	SD	Mean	SD	Mean	SD			
Physical functioning	70.38	20.98	56.98	26.53	13.41	19.57	0.69	8.48	<0.0001
Role functioning	54.82	27.44	40.57	29.01	14.25	25.40	0.56	6.92	<0.0001
Emotional functioning	64.11	18.93	50.18	23.48	13.93	22.19	0.63	7.76	<0.0001
Cognitive functioning	67.88	18.96	52.32	23.42	15.56	22.08	0.70	8.66	<0.0001
Social functioning	54.14	27.26	39.11	29.58	15.03	27.69	0.54	6.72	<0.0001
Global health status	54.65	26.47	56.91	24.61	−2.26	22.72	0.10	−1.13	0.26
Fatigue	42.34	20.47	56.46	23.42	−14.12	22.37	0.63	−7.81	<0.0001
Nausea and vomiting	27.96	24.37	46.42	27.30	−18.46	27.00	0.68	−8.35	<0.0001
Pain	39.11	22.57	53.92	26.41	−14.81	24.37	0.61	−7.52	<0.0001
Dyspnea	24.89	23.55	38.67	26.22	−13.78	24.48	0.56	−6.89	<0.0001
Insomnia	34.91	26.18	51.35	26.76	−16.44	27.08	0.61	−7.39	<0.0001
Appetite loss	38.00	26.20	53.11	27.61	−15.11	28.01	0.54	−6.61	<0.0001
Constipation	28.86	25.89	43.40	31.17	−14.54	28.30	0.51	−6.27	<0.0001
Diarrhoea	17.33	21.75	32.44	28.89	−15.11	28.53	0.53	−6.49	<0.0001
Financial difficulties	63.38	33.83	71.27	30.46	−7.89	28.38	0.28	−3.43	0.001

SD, standard deviation; SRM, standardized response mean.

and neck cancer groups, all domain scores have statistically significant changes after treatment except for EF, CF, SF, AP, DI, and FI, with SRM being <0.50 for all domains except PF (0.73).

discussion

QLQ-C30 has been used in hundreds of research studies and its psychometric properties have been well documented [2, 10–14, 23–25], including its traditional Chinese version used in Taiwan and Hong Kong [11–13]. This paper focuses on psychometric properties of the simplified Chinese version used in the Mainland of China.

Instrument validity indicates how an instrument can capture what it purports to measure. We used correlational analysis and structural equation models to confirm the construct and criterion-related validity of the QLQ-C30. Multi-trait scaling analysis showed that all item–domain correlation coefficients met the standards of item convergent and discriminant validity. It can be seen from Table 3 that overall the correlations between the same and similar domains of QLQ-C30 and QLICP-GM are higher than those between different and nonsimilar

Table 8. Responsiveness of simplified Chinese version of quality of life questionnaire (QLQ-C30) for measuring quality of life of patients with colorectal, stomach, and head and neck cancer ($n = 170$)

Domains (subscales/items)	Pre-treatment		Post-treatment		Differences		SRM	t	P
	Mean	SD	Mean	SD	Mean	SD			
Physical functioning	71.23	23.43	83.87	17.07	−12.64	17.32	0.73	−9.77	<0.0001
Role functioning	52.04	29.72	62.59	26.70	−10.56	26.48	0.40	−5.35	<0.0001
Emotional functioning	64.48	21.70	66.68	18.08	−2.20	16.68	0.13	−1.77	0.08
Cognitive functioning	67.23	23.70	68.25	23.08	−1.02	18.54	0.06	−0.74	0.46
Social functioning	56.02	28.45	53.89	27.75	2.13	24.22	0.09	1.18	0.24
Global health status	53.42	24.41	60.44	26.10	−7.02	23.10	0.30	−4.06	<0.0001
Fatigue	45.50	25.44	41.31	21.62	4.19	22.36	0.19	2.51	0.01
Nausea and vomiting	24.58	24.82	18.62	22.71	5.96	24.26	0.25	3.29	0.001
Pain	36.39	26.22	32.69	22.72	3.70	23.54	0.16	2.11	0.04
Dyspnea	24.81	27.32	14.26	22.29	10.56	25.28	0.42	5.60	<0.0001
Insomnia	37.27	29.06	26.97	23.44	10.30	23.50	0.44	5.85	<0.0001
Appetite loss	41.11	29.08	36.85	27.41	4.26	29.70	0.14	1.92	0.06
Constipation	26.26	26.20	21.42	23.05	4.84	24.51	0.20	2.64	0.01
Diarrhoea	21.79	24.03	23.28	25.69	−1.49	25.44	0.06	−0.78	0.43
Financial difficulties	58.29	31.97	62.20	63.83	−3.91	62.04	0.06	−0.84	0.40

SD, standard deviation; SRM, standardized response mean.

domains. These correlations confirm the criterion-related validity and also demonstrate the convergent and divergent validity of the domains to a reasonable degree. For structural equation modeling, RMSEA, SRMR, CFI, and TLI are recommended as sensitive indices to model misspecification [21]. For RMSEA, Browne and Cudeck [22] indicated that a value not greater than 0.05 would indicate a 'close fit', a value between 0.05 and 0.08 would indicate a 'reasonable fit', and a value >0.10 would indicate 'unacceptable fit'. The CFI and TLI with values close to 0.95 and SRMR <0.08 reflect good fit of model to the data [21]. On the basis of these criteria, it can be seen from Table 4 that the constructs of the instrument are consistent with the original concept, with excellent model fit.

Reliability refers to the reproducibility or consistency of scores from one assessment to another. Test–retest reliability (r) and internal consistency (Cronbach's α) are the most frequently used indicators. It is well recognized that internal consistency (α) should be at least 0.70 and reliability (r) should be >0.80 in a test–retest situation [26]. Our results (Table 6) show that the simplified Chinese version of the QLQ-C30 has good reliability, with the exception of a few domains such as CF, AP, and DI.

The most important characteristic of a QOL scale is that it can detect change of QOL after treatments or interventions. Generally speaking, the assessment of responsiveness can be divided into two categories: internal and external [19, 20]. Internal responsiveness characterizes the ability of a measure to change in response to an effective intervention. One widely used method of assessing internal responsiveness is to evaluate the change in a measure within the context of a randomized clinical trial involving a treatment that has previously been shown to be efficacious [27, 28]. External responsiveness reflects the extent to which changes in a measure over a specified time frame relate to corresponding changes in a reference measure of health status. In this paper, we focused on internal responsiveness. The paired t-test was used to compare the change of mean responses between the first and the third assessments (pre-treatment and post-treatment), with the hypothesis that QOL scores would change after treatment. The responsiveness indicator, SRM, was also computed, with values of 0.20, 0.50, and 0.80 proposed to represent small, moderate, and large responsiveness, respectively [19, 20]. Our results showed that QOL score changes after treatment were of statistical significance on most domains for not only the breast and lung cancer groups (Table 7) but also the colorectal, stomach, and head and neck cancer groups (Table 8). Some possible reasons for the domains with no significant changes are (i) the observation period (∼4 weeks) may be too short to find significant changes in these domains and (2) some domains have stable scores that simply do not change after treatment, for example, social function (SF) would not be expected to change during hospitalization. Therefore, it can be inferred that this instrument has reasonably good responsiveness and is more sensitive for the first group than for the second group.

From our findings and those in the literature, it can be seen that the psychometric properties of this instrument are very similar to those of the original English version and other language versions [2, 10–14, 23–25, 29, 30]. For example, the Korean version of QLQ-C30 (V3.0) [29] showed that Cronbach's α coefficients for eight multiple-item scales were >0.70, with the exception of cognitive functioning. All interscale correlations were statistically significant in the expected direction ($P < 0.01$). Multi-trait scaling analyses demonstrated that all scales met multidimensional conceptualization criteria, in terms of convergence and discrimination validity. Silpakit et al. [30] reported that Cronbach's α coefficients of the six scales of the Thai version of QLQ-C30 (V3.0) were >0.7, except for cognitive and social function scales. All test–retest reliability coefficients were high. Multi-trait scaling analysis showed that all item–scale correlation coefficients met the standards of convergent and

discriminant validity. Most subscales and items could discriminate between subgroups of patients with different clinical status assessed with the Eastern Cooperative Oncology Group scale.

It is worth noting that the study patients were divided into two groups to analyze responsiveness because the directions of change of QOL scores were not the same after treatment. According to our experience with QOL instruments for these five types of cancer patients as well as empirical evidence from descriptive statistics, QOL scores of many domains decreased after treatment for the patients with breast and lung cancer, while increasing for those with colorectal, stomach, and head and neck cancer. Further investigation is needed to study the underlying mechanisms for such different patterns of change across the different cancer groups.

funding

National Natural Science Foundation of China (30360092); Natural Sciences Funds of Yunnan Province (99C0016G).

acknowledgements

In carrying out this research project, we received substantial help from Karen West, Linda Dewolf, Jane Blazeby, Neil Aaronson, and Mirjam Sprangers at EORTC and many staff at the Yunnan Tumor Hospital. We also thank Ms Cheryl Bliss-Clark in the Department of Biostatistics and Computational Biology at the University of Rochester for her careful review of the paper to improve its presentation. This research was supported in part by National Institutes of Health grants 1 U19 AI056390-01, 1 UL1 RR024160-01, and N01 AI50029 from the United States. There is no conflict of interest.

references

1. Sprangers MA, Cull A, Groenvold M et al. The European Organization for Research and Treatment of Cancer approach to developing questionnaire modules: an update and overview. Qual Life Res 1998; 7(4): 291–300.
2. Aaronson NK, Ahmedzai S, Bergman B 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: 365–376.
3. Cella DF, Tulsky DS, Gray G et al. The functional assessment of cancer therapy scale: development and validation of the general measure. J Clin Oncol 1993; 11(3): 570–579.
4. Overcash J, Extermann M, Parr J et al. Validity and reliability of the FACT-G scale for use in the older person with cancer. Am J Clin Oncol 2001; 24(6): 591–596.
5. Schipper H, Clinch J, McMurray A, Levitt M. Measuring the quality of life of cancer patients: the functional living index-cancer: development and validation. J Clin Oncol 1984; 2(5): 472–483.
6. Sprangers MA. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. J Clin Oncol 1996; 14(10): 2756–2768.
7. Bergman B, Aaronson NK, Ahmedzai S et al. The EORTC QLQ-LC13: a 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(5): 635–642.
8. McLachlan S-A, Devins GM, Godwin PJ. Validation of the European Organisation for Research and Treatment of Cancer quality of life questionaire (QLQ-C30) as a measure of psychosocial function in breast cancer patients. Eur J Cancer 1998; 34(4): 510–517.
9. Kemmler G, Holzner B, Kopp M et al. Comparison of two quality of life instruments for cancer patients: the functional assessment of cancer therapy-general and the European Organization for Research and Treatment of Cancer quality of life questionnaire-C30. J Clin Oncol 1999; 17(9): 2932–2940.
10. Luo N, Fones CS, Lim SE et al. The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30): validation of English version in Singapore. Qual Life Res 2005; 14(4): 1181–1186.
11. Chie WC, Chang KJ, Huang CS, Kuo WH. Quality of life of breast cancer patients in Taiwan: validation of the Taiwan Chinese version of the EORTC QLQ-C30 and EORTC QLQ-BR23. Psychooncology 2003; 12(7): 729–735.
12. Chie WC, Yang CH, Hsu C, Yang PC. Quality of life of lung cancer patients: validation of the Taiwan Chinese version of the EORTC QLQ-C30 and QLQ-LC13. Qual Life Res 2004; 13(1): 257–262.
13. Chie WC, Hong RL, Lai CC et al. Quality of life in patients of nasopharyngeal carcinoma: validation of the Taiwan Chinese version of the EORTC QLQ-C30 and the EORTC QLQ-H&N35. Qual Life Res 2003; 12(1): 93–98.
14. Zhao H, Kanda K. Translation and validation of the standard Chinese version of the EORTC QLQ-C30. Qual Life Res 2000; 9(2): 129–137.
15. Wan C, Yang Z, Meng Q et al. Development and Validation of the General Module of the System of Quality of Life Instruments for Cancer Patients (QLICP-GM). Int J Cancer 2008; 122(1): 190–196.
16. Hays RD, Hayashi T. Beyond internal consistency reliability: rationale and user's guide for multitrait analysis program on the microcomputer. Behav Res Methods Instrum Comput 1990; 22: 167–175.
17. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996; 1(1): 30–46.
18. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? Qual Life Res 2004; 13: 571–586.
19. Terwee CB, Dekker FW, Wiersinga WM et al. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 2003; 12(4): 349–363.
20. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000; 53(5): 459–468.
21. Hu LT, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equation Model 1999; 6: 1–55.
22. Browne MW, Cudeck R. Alternative ways of assessing model fit. In Bollen KA, Long JS (eds): Testing Structural Equation Models. Beverly Hills, CA: Sage 1992; 136–162.
23. Groenvold M, Klee MC, Sprangers MA, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. J Clin Epidemiol 1997; 50(4): 441–450.
24. Kyriaki M, Eleni T, Efi P et al. The EORTC core quality of life questionnaire (QLQ-C30, version 3.0) in terminally ill cancer patients under palliative care: validity and reliability in a Hellenic sample. Int J Cancer 2001; 94(1): 135–139.
25. Holzner B, Kemmler G, Sperner-Unterweger B et al. Quality of life measurement in oncology-a matter of the assessment instrument? Eur J Cancer 2001; 37(18): 2349–2356.
26. Van der Steeg AFW, De Vries J, Roukema JA. Quality of life and health status in breast carcinoma. Eur J Surg Oncol 2004; 30: 1051–1057.
27. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 1997; 50: 869–879.
28. Anderson JJ, Chernoff MC. Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. J Rheumatol 1993; 20: 535–537.
29. Yun YH, Park YS, Lee ES et al. Validation of the Korean version of the EORTC QLQ-C30. Qual Life Res 2004; 13(4): 863–868.
30. Silpakit C, Sirilerttrakul S, Jirajarus M et al. The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30): validation study of the Thai version. Qual Life Res 2006; 15(1): 167–172.