

Development and Validation of the Chronic Gastritis Scale Under the System of Quality of Life Instruments for Chronic Diseases QLICD-CG Based on Classical Test Theory and Generalizability Theory

Chonghua Wan, PhD,* Ying Chen, MM,† Li Gao, MM,‡
Qingqing Zhang, MM,§ Wu Li, BM,|| and Peng Quan, PhD*

Background and Aim: Quality of life (QOL) for patients with chronic gastritis (CG) is of interest worldwide and disease-specific instruments are needed for clinical research and practice. This paper focused on the development and validation of the CG scale under the system of Quality of Life Instruments for Chronic Diseases (QLICD-CG) by the modular approach and both classical test theory and generalizability theory.

Methods: The QLICD-CG was developed based on programmed decision procedures including multiple nominal and focus group discussions, in-depth interviews and quantitative statistical procedures. Based on the data measuring QOL 3 times before and after treatments from 142 inpatients with CG, the psychometric properties of the scale were evaluated with respect to validity, reliability and responsiveness employing correlation analysis, multi-trait scaling analysis, factor analyses, *t* tests and also G studies and D studies of generalizability theory analysis.

Results: Correlation, multi-trait scaling and factor analyses confirmed good construct validity and criterion-related validity when using SF-36 as a criterion. The internal consistency α for all domains were higher than 0.70 except for the social domain (0.62). Test-retest reliability coefficients (Pearson *r* and intraclass correlations) for the overall score and all domains were higher than 0.80 except for the social domain (0.77), while they were ranging between 0.72 to 0.94 at facets level; The overall score and scores for all domains/facets had statistically significant changes ($P < 0.01$) after treatments except for facets of social effects and sexual function with standardized response mean ranging from 0.04 to 1.03, but from 0.34 to 1.03 for the domain level scores. G-coefficients

and index of dependability (Φ coefficients) confirmed the reliability of the scale further with more exact variance components, and decision information on number of items changing.

Conclusions: The QLICD-CG could be used as a useful instrument in assessing QoL for patients with CG, with good psychometric properties including validity, reliability and responsiveness and also several advantages.

Key Words: quality of life, standardized response mean, psychometric properties, intra-class correlations, multi-trait scaling analysis, generalizability theory

(*J Clin Gastroenterol* 2022;56:e137–e144)

BACKGROUND

Chronic gastritis (CG) is a common disease which causes a significant risk of gastric cancer¹ with the prevalence of 60% in China,² and the cause of cancer-related death being the third most common in China.^{3,4} Consequently, CG has received increased attention within medical practice, for its long-term inflammation of the gastric mucosa can also significantly impair the health-related quality of life (HRQOL) of the patients.⁵ The HRQOL is a patient-focused concept, referring to an impairment of functional status (physical or mental) and the sense of well-being. At present, the researches on quality of life (QoL) of CG patients were rare. Most studies on HRQOL adopted generic instruments such as 36-item Short-Form general health survey (SF-36), sickness impact profile, Nottingham health profile, quality well-being index, etc.⁶ These measures place a heavy emphasis on overall life satisfaction and general health, making them fail to capture the specific impact of specific diseases. A generic measure based on the overall life satisfaction conceptualization of QOL would arguably be inappropriate for use with the patients with specific diseases. It fails in exploring some specific relevant information in CG patients. For example, CG patients often report worse QOL than the general population of poor gastric function, upper abdominal pain, indigestion, bloating, nausea, vomiting, belching, loss of appetite, weight loss.⁷ A few of studies also used Nepean Dyspepsia Index (NDI),⁸ digestive health status instruments,^{9,10} the QOLRAD (Quality Of Life in Reflux and Dyspepsia),^{11,12} the FDDQL (Functional Digestive Disorder Quality of Life Questionnaire),¹³ and the GSRS (Gastrointestinal Symptom Rating Scale),^{14,15} the PAGI-SYM (patient assessment of upper gastrointestinal disorders-symptom severity Index)^{16,17} and the PAGI-QOL (patient assessment of upper gastrointestinal disorders-quality of life index)^{17,18} to investigate HRQOL of this patients. However, these instruments were not developed based on the popular modular approach-a general/core module plus

Received for publication May 4, 2020; accepted January 25, 2021.

From the *School of Humanities and Management, Research Center for Quality of Life and Applied Psychology, Key Laboratory for Quality of Life and Psychological Assessment and Intervention, Guangdong Medical University, Dongguan; †School of Public Health, Kunming Medical University; ‡The First Affiliated Hospital of Kunming Medical University, Kunming; §Taian city central hospital, Taian; and §Department of Statistics and Medical Record Management, Shandong Provincial Hospital Affiliated to Shandong University, Jinan, China.

C.W. and Y.C. are first co-author with the same contributions.

The paper is supported by the National Natural Science Foundation of China (71373058, 30860248), Science and Technological Planning Program of Guangdong Province (2013B021800074).

The authors declare that they have nothing to disclose.

Address correspondence to: Chonghua Wan, PhD, Xincheng Road, Songshan Lake Science and Technology Industry Park, Dongguan 523808, Guangdong, P.R. China (e-mail: wanchh@hotmail.com).

Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/MCG.0000000000001511

the specific modules.^{19,20} Since diseases within the same disease class such as digestive diseases share many characteristics such as symptoms and side effects in common, an approach widely adopted in recent years to develop QOL instruments for diseases is to combine a general module for the entire class of diseases with the specific module for each individual disease. This approach can substantially reduce the amount of time and effort in developing new instruments, and the quality of life questionnaires from the EORTC and the Functional Assessment of Cancer Therapy have been developed based on this modular principle.^{19,20} Furthermore, these are not specific scales of CG in fact. To our knowledge, no instrument for CG has been developed based on the modular approach, let alone combination of classical test theory (CTT) and generalizability theory (GT). Therefore, there is an urgent need for a disease-specific QOL for CG patients. We have developed a system of Quality of Life Instruments for Chronic Diseases (QLICD) by modular approach.^{21–24} This system includes a general module (QLICD-GM) which can be used with all types of chronic diseases, and specific modules for different diseases with each module being used for only the relevant disease.^{21–24} For example, the irritable bowel syndrome instrument (QLICD-IBS) is constructed by combining QLICD-GM with the specific module for irritable bowel syndrome.²⁴ At present, the QLICD (V1.0) includes a 30-items general module QLICD-GM (3 domains and 10 facets) and 9 specific modules with the numbers of items ranging from 14 to 21, and thus forming 9 specific scales of hypertension (QLICD-HY),²² coronary heart disease (QLICD-CHD),²³ QLICD-IBS,²⁴ chronic gastritis disease (QLICD-CG), etc.

The purpose of the current study was to describe the development and validation process for this QLICD-CG instrument.

METHODS

Development of the QLICD-CG

The instrument QLICD-CG is constructed by combining QLICD-GM with the specific module for CG as the modular approach. The process of developing and validating the QLICD-GM has been described in detail in previous published work.²¹ QLICD-CG was formed with 30 items that consisted of 3 domains and 10 facets from programmed decision method. It included focus group discussions, in-depth interviews and pretesting and 4 quantitative item screening methods. This included variation procedure, correlation procedure, and factor analysis procedure, etc.

In the general module, 27 items reflecting the symptoms, side effects, and mental health conditions specific to CG were selected to form the initial item bank of the specific module by means of the literature review and nominal/focus group discussions. The members from the 2 groups evaluated the importance of each item by ranking each item independently and then discussed those rankings. Nine items with the lowest ranking was being eliminated. The remain 18 items formed a preliminary questionnaire, which was used to carry out a pilot test and adopted on 32 patients with CG, and 14 medical doctors with extensive experience in the care of patients with CG. Patients' perspective was an important factor when determining the acceptability of an intervention and associated compliance. According to the pilot data, a development process similar to the general module (statistical procedures and focus group discussions) was used to rescreen the items to obtain the final module consisting of 14 items, coded CG1-CG14 (see Table 1 in detail), classified into 5 facets.

Validation of the QLICD-CG

Patients

The study participant was selected to CG inpatients at any stage and with any treatments who follows 2 inclusion criteria: (1) be able to provide written informed consent; (2) be able to read and write words with assistance. The CG was diagnosed primarily through endoscopy and gastric biopsy, which was consistent with the classification criteria of CG proposed by Chinese society of digestive endoscopy.⁵ Gastroscopy and histopathologic examination of gastric biopsy tissues is the main diagnostic and differential diagnostic methods. Upper gastrointestinal endoscopy is the main method to diagnose CG. Furthermore, it plays a vital role in evaluating the severity and excluding other causes.

Data Collection and Scoring

The survey was carried out at the First Affiliated Hospital of Kunming Medical University with the ethics committee of Kunming Medical University approval. The investigators who were doctors and medical postgraduate students sought the inform consent of each respondent. Each respondent was required to answer the questionnaires at the time of admission to the hospital. In order to evaluate the test-retest reliability, a random sub-sample was selected to participate in a second assessment of the following day or the second day after hospitalization. Considering that responsiveness defined on the change over time, all subjects available at the third scheduled assessment time-point would complete the measures at discharge to evaluate responsiveness of the instrument.

The Chinese version of SF-36²⁵ was also used to provide data for assessing the criterion-related validity, convergent, and discriminant validity of the QLICD-CG. Owing to the lack of a consensus based on the gold standard for assessing QOL of CG. The baseline socio-demographic characteristics including age, sex, education, marital status, etc., clinical history, and treatments were recorded from the hospital medical record.

Answers were checked immediately each time by the investigators in order to ensure its integrality and quality.

The raw scores of items, domains/facets and overall scale were calculated based on the data collected. As each item in a 5-point Likert format (not at all, a little bit, somewhat, quite a bit, and very much), the positively stated items are directly scored from 1 to 5, while the negatively stated items are reverse scored. The domains/facets and overall scale score were obtained by adding together the relevant item scores, and all scores were linearly converted to standardized score with a 0 to 100 scale, indicating a higher score on the QLICD-CG represent a better level of QoL for both raw scores and standardized scores.

Psychometric Analysis

The validity, reliability, and responsiveness of the QLICD-CG were evaluated. For validity, the content validity, construct validity, and criterion-related validity were often evaluated. In this paper, construct validity was evaluated by the Pearson correlation coefficient r among items and domains and by factor analysis. Criterion-related validity was evaluated by correlating corresponding domains of the QLICD-CG and SF-36. Multi-trait scaling analysis²⁶ was employed to test item convergent validity and discriminant validity, with the 2 criteria: (1) convergent validity is supported when an item-domain correlation is 0.40 or greater; (2) discriminant validity is revealed when item-domain correlation is higher than that with other domains.

TABLE 1. Correlation Coefficients *r* Among Items and Domains of QLICD-CG (n = 142)

Code	Items Brief Description in English*	Physical	Psychological	Social	Specific
PH1	Take care of daily life (eg, eating)?	0.54	0.17	0.37	0.03
PH2	Felt easily fatigued?	0.71	0.34	0.24	0.31
PH3	Have trouble walking 800 m or more?	0.72	0.23	0.23	0.08
PH4	Have trouble going up and down stairs?	0.73	0.37	0.26	0.13
PH5	Need to take medication?	0.61	0.19	0.14	0.15
PH6	A good appetite?	0.55	0.22	0.30	0.21
PH7	Satisfied with your sleep?	0.49	0.18	0.19	0.07
PH8	Felt pain or uncomfortable?	0.61	0.31	0.29	0.31
PS1	Memory and concentration affected?	0.32	0.59	0.41	0.29
PS2	Felt mentally miserable?	0.51	0.68	0.42	0.34
PS3	Felt lonely and helpless?	0.30	0.65	0.37	0.35
PS4	Felt pessimism and despair?	0.31	0.75	0.33	0.30
PS5	Worried about disease?	0.20	0.73	0.39	0.41
PS6	Felt fretful or irritable?	0.27	0.71	0.40	0.48
PS7	Felt nervous and anxious?	0.34	0.78	0.42	0.46
PS8	Stop medication because of side effects?	0.02	0.40	0.10	0.35
PS9	To be a burden to the family?	0.33	0.64	0.42	0.28
PS10	Felt self-abasement because of disease?	0.23	0.71	0.41	0.28
PS11	Hidden emotions but could not forget?	0.14	0.69	0.33	0.37
SO1	Interfered with work/housework?	0.13	0.43	0.43	0.35
SO2	Family roles?	0.18	0.06	0.55	0.08
SO3	Decreased caring and attention to family?	0.39	0.35	0.57	0.21
SO4	Good relations with family?	0.06	0.19	0.49	0.02
SO5	Help and support from family?	0.01	0.04	0.35	0.01
SO6	Affected participating in leisure activities?	0.14	0.34	0.45	0.22
SO7	Treat illness positively and optimistically?	0.23	0.26	0.59	0.13
SO8	Treatments received good for curing?	0.22	0.06	0.40	0.02
SO9	Economic problems caused by illness?	0.16	0.49	0.50	0.45
SO10	Support from friends and relatives?	0.28	0.16	0.55	0.07
SO11	Affected sexual activities?	0.30	0.43	0.36	0.24
CG1	Frowsty/discomfort in upper abdomen?	0.14	0.11	0.14	0.42
CG2	Pain in upper abdomen?	0.14	0.25	0.30	0.41
CG3	Feel full bilge when eating?	0.21	0.16	0.09	0.56
CG4	Feel full bilge/digest slowly after a meal?	0.32	0.35	0.18	0.59
CG5	Have any belch (burps)?	0.20	0.34	0.15	0.54
CG6	Have acid regurgitation?	0.01	0.11	0.06	0.44
CG7	Have nausea?	0.23	0.19	0.23	0.45
CG8	Feel heartburn?	0.19	0.17	0.20	0.29
CG9	Worried about causing severe disease?	0.03	0.43	0.20	0.56
CG10	Upset/distress for gastroscopy inspection?	0.17	0.28	0.16	0.37
CG11	Vexed for food limit?	0.00	0.31	0.13	0.45
CG12	Vexed for often taking stomach medications?	0.10	0.42	0.24	0.60
CG13	Troubled/limit by dine at fix time?	0.02	0.25	0.16	0.51
CG14	Worried inducing gastritis by irregular diet?	0.02	0.21	0.02	0.56

Correlations between each item and its designated scale are in bold type.

*The scale is developed for Chinese people and written in Chinese, and here display just a rough meaning in English.

In terms of reliability, for each domain/facet and the overall scale, the internal consistency was assessed by the Cronbach α coefficient using the first measurements data (at admission) because of larger sample size, and the test-retest reliability was evaluated by the Pearson correlation coefficients and intraclass correlation^{27,28} between the first and second assessments.

Responsiveness (sensitivity to detect change) was evaluated by comparing the mean score change between the 2 assessments before and after treatments using paired *t* tests and also the standardized response mean (SRM).²⁹⁻³¹

GT Analysis

GT was applied to investigate the score dependability of the QLICD-CG apart from the CTT analysis. GT is a modern test theory developed on the combination of the CTT, research design and variance analysis to refine the designs of measurement procedures in an attempt to yield

reliable data.³²⁻³⁶ To control the measurement errors, GT introduces the independent variables or factors that interferes the test scores into the measurement model, such as differences among the research objects, item difficulty, scoring criteria, and interaction of these factors. The effects of these variables or factors on test scores were subsequently evaluated by variance analysis, in which the variance components are used as indexes. GT has 2 type studies including G study and D study, with the G study quantifying the amount of variance associated with the different facets (factors) that are being examined, the D study providing information about which protocols are optimal for a particular measurement situation by generating generalizability (G) coefficients. This can be interpreted as reliability coefficients across various facets of the study.

In this study, G-Studies and D-Studies were performed to estimate the variance components and dependability

coefficients in one-facet crossed design [person-by-item ($p \times i$) design] for the estimate the variance components and dependability coefficients. We defined the QOL of patients as the target of measurement and items as 1 facet of measurement error. For the G-Study, a universe of admissible observations, which consisted of the object of measurement and the measurement error facets, was defined and the variance components were estimated. For the D study, we defined a universe of admissible generalizations to represent the measurement conditions based on the object of measurement and on the item facets. Simultaneously, the variance components of generalizability coefficients and dependability indexes in each facet, as well as their interactions were calculated.

RESULTS

Sociodemographic Characteristics of the Sample

A total of 142 patients with CG varied in age from 17 to 78, with median age of 41.0 and mean age 43.5 ± 15.2 . Among them, 70 cases (49.3%) were male and 126 (88.7%) were of Han ethnicity. Regarding the education level, 27 cases (19.0%) finished primary school, while 69 (48.6%) completed high school, and 46 (32.4%) had a college or postgraduate degree. For occupation, workers accounted for 26.8% (38 cases), cadre 17.6% (25), farmer 13.4% (19), teacher 7.0% (10), and others 35.2% (50). For perceived income, 19.0% (27 cases) were classified as poor while fair 71.1% (101) and high 9.9% (14).

Construct Validity

The construct validity was evaluated by item-domain Pearson correlation coefficient r and by factor analysis. Correlation analyses from data measuring at admission to hospital showed that there were strong correlations between items and their own domains (most correlation coefficients are higher than 0.40), but weak correlations between items and other domains (Table 1). For example, correlation coefficients between PHD and items of PH1 to PH8 ranging from 0.49 to 0.73 (the first column in bold) were higher than those between PHD and other items. Similarly, correlation coefficients between PSD and items of PS1 to PS11 ranged from 0.40 to 0.75 (the second column in bold) were higher than those between PSD and other items.

The factor analysis was carried out for the general module and the specific module, respectively. According to extract criteria of eigenvalues > 1 , there were 10 principal components abstracted from 10 facets of the general module (QLICD-GM), accounting for 69.8% of the cumulative variance. By using the Varimax rotation method, 10 principal components reflected

10 different facets under 3 domains of the general module with the first, sixth, and tenth principal components mainly representing the psychological domain with higher loadings (> 0.50) on PS1-PS11. The second, seventh and ninth principal components largely reflected the physical domain with higher loadings (> 0.50) on PH1 to PH8; the third, fourth, and fifth principal components generally depicting the social domain with higher loadings (> 0.50) on SO1 to SO11. Similarly, the principal component factor analysis extracted 5 principal components from the 14 items of the specific module with the cumulative variance of 63.9%, happen to reflecting 5 facets.

Criterion-related Validity

Correlation coefficients among the domain scores of the QLICD-CG and SF-36 were presented in Table 2, showing that the correlations between the same and similar domains (figures in bold in the table) were generally higher than those between different and nonsimilar domains. For example, the coefficient between the physical domain of QLICD-CG and physical function of SF-36 was 0.62, higher than any other coefficients in this row. Similarly, the coefficient between the psychological domain of QLICD-CG and mental-health of SF-36 was 0.64, higher than any other coefficients in this row.

Reliability

The reliability of the scale was presented in Table 3 in 3 procedures: internal consistency (the Cronbach α), test-retest and ICC. In Table 3, the Cronbach α for these 4 domains and the overall were higher than 0.70, with that for facets being ranged from 0.43 to 0.81.

During the second assessment (2-d follow-up), 45 patients completed the questionnaires for test-retest reliability analysis. The test-retest correlation coefficients (r) for the 4 domains and the overall were larger than 0.90, while they were ranging between 0.83 and 0.98 at facets level. The results from ICC computed based on the definition of absolute agreement were very similar to Pearson correlation coefficients (r).

Results From GT

G-Studies were performed to estimate the variance components for 4 domains of the QLICD-CG (Table 4), in which 142 patients completed the QoL the instrument with 44 items.

Table 4 of the 4 domains of physical, psychological, social and the specific, the largest source of variation were accounted to person-by-item interactions ranging from 47.53% to 76.65%. The variances accounted for by person were the second for 3 domains of psychological, social and the specific ranged from

TABLE 2. Correlation Coefficients Among Domain Scores of QLICD-CG and SF-36 (n = 142)

QLICD-CG	SF-36									
	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
PHD	0.62	0.32	0.43	0.46	0.56	0.31	0.26	0.38	0.68	0.50
PSD	0.28	0.26	0.24	0.39	0.51	0.48	0.43	0.64	0.42	0.66
SOD	0.35	0.31	0.33	0.33	0.46	0.52	0.34	0.43	0.45	0.56
SPD	0.07*	0.16*	0.23	0.30	0.37	0.12*	0.29	0.41	0.25	0.40
TOT	0.39	0.33	0.39	0.48	0.62	0.44	0.44	0.65	0.56	0.70

* $P > 0.05$, and $P < 0.05$ for all other coefficients.

BP indicates bodily pain; GH, general health; MCS, Mental Component Summary; MH, mental-health; PCS, Physical Component Summary; PF, physical function; PHD, physical domain; PSD, psychological domain; RE, role-emotional; RP, role-physical; SF, social function; SOD, social domain; SPD, specific domain; TOT, total score; VT, vitality.

TABLE 3. Reliability of the Quality of Life Instrument QLICD-CG (n = 142 for α , n = 45 for r and ICC)

Domains/Facets (Number of Items)	Internal	Test-retest	Test-retest ICC (95% CI)
	Consistency Coefficient α	Coefficient r	
Physical Function (PHD) (8)	0.77	0.95	0.95 (0.91-0.97)
Independence (3)	0.74	0.97	0.97 (0.95-0.98)
Appetite and sleep (2)	0.43	0.83	0.83 (0.72-0.91)
Physical symptoms (3)	0.64	0.88	0.88 (0.79-0.93)
Psychological function (PSD) (11)	0.88	0.99	0.99 (0.98-0.99)
Cognition (2)	0.64	0.92	0.91 (0.85-0.95)
Anxiety (3)	0.81	0.98	0.98 (0.96-0.99)
Depression (3)	0.75	0.97	0.97 (0.95-0.98)
Self-consciousness (3)	0.63	0.97	0.97 (0.95-0.98)
Social Function (SOD) (11)	0.71	0.96	0.95 (0.91-0.97)
Social support/ security (6)	0.66	0.95	0.93 (0.88-0.96)
Social effects (4)	0.64	0.97	0.97 (0.94-0.98)
Sexual function (1)	—	0.91	0.90 (0.83-0.95)
Specific domain (SPD) (14)	0.74	0.94	0.94 (0.89-0.97)
Unwell of upper abdomen (5)	0.72	0.86	0.86 (0.76-0.92)
Acid regurgitation (1)	—	0.96	0.96 (0.93-0.98)
Nausea (1)	—	0.94	0.94 (0.89-0.97)
Heart burn (1)	—	0.98	0.98 (0.97-0.99)
Effect of mental and life (6)	0.72	0.95	0.95 (0.90-0.97)
Total (TOT) (44)	0.89	0.98	0.98 (0.96-0.99)

— indicates not acceptable/suitable; CI, confidence interval; ICC, intra-class correlations; PHD, physical domain; PSD, psychological domain; SOD, social domain; SPD, specific domain; TOT, total score.

15.87% to 42.15% exception of physical domain (the second is 31.79% by item).

The D-Studies were performed to estimate the generalizability coefficient (G-coefficient) and index of dependability (Φ coefficient) for 4 domains of the QLICD-CG for the $p \times i$ current design (physical domain includes 8 items, psychological domain includes 11 items, social domain includes 11 items, and specific domain includes

14 items), as well as the alternative designs with varied numbers of items (Table 5).

Responsiveness

These data from 120 patients who completed the questionnaires after treatments were used for evaluating the responsiveness. A classical paired t test with responsiveness indicator, SRM, was used to examine changes of mean scores from each domain/facet of the QLICD-CG before and after treatment in Table 6. There were significant changes occurred for all domains/facets and the overall scale ($P < 0.01$) except for facet of sexual function with SRMs ranging from 0.24 to 1.35, but from 0.67 to 1.35 for the domain level scores implying excellent responsiveness.

DISCUSSION

This paper focused on the development and validation of the QLICD-CG, a specific QOL instrument for CG. An approach widely adopted to develop QOL instruments for diseases is to combine a general module with specific modules for individual diseases to capture both common features and differences among diseases.¹⁹⁻²¹ We employed this modular approach to systematically and more efficiently develop a system of new instruments for chronic diseases called QLICDs with the general module QLICD-GM being used for all kinds of chronic diseases. And the QLICD-CG is a specific scale of this system for only CG. This modular approach unifies all disease-specific instruments of QLICDs using the same general module with similar constructs. Therefore, the QLICD-CG has several advantages over existing instruments.²¹⁻²⁴ First, it could compare QOL across diseases by the general module and also capture the symptoms and side effects by the specific module, implying both generic and specific properties. For instance, we can use QLICD-GM to capture general QOL in patients with different diseases, say CG and peptic ulcer disease, and then employ QLICD-CG and QLICD-PU to capture differences in QOL between the 2 diseases. Second, the mean scores could be computed not only at the domain (4 domains) and the overall levels but also at the different facet levels (15 facets) to detect changes in detail for it consists of a moderate number of items with a clear hierarchical structure (items \rightarrow facets \rightarrow domains \rightarrow the overall). Users can select either 1 or both levels for a study at hand. The third important observation was the strong Chinese cultural background. For example, the Chinese culture pay more attention to family relationship and kinship, dietary, temperament and high spirit, which are all captured in the QLICD-CG by items focusing on appetite (PH6), sleep (PH7), energy (PH2) and family support (SO4, SO5, etc.).

TABLE 4. The Estimated Variance Components and Percentage of Variance Accounted for by Effects (Percent) for $p \times i$ Design in G-study for 4 Domains of Quality of Life Instrument QLICD-CG ($n'_p = 142$)

Domain	p (Person)		i (Item)		$p \times i$ (Person \times Item)	
	Variance Component	Percent (%)	Variance Component	Percent (%)	Variance Component	Percent (%)
PHD ($n'_i = 8$)	0.357	20.05	0.566	31.79	0.857	48.14
PSD ($n'_i = 11$)	0.752	42.15	0.184	10.31	0.848	47.53
SOD ($n'_i = 11$)	0.273	18.16	0.117	7.78	1.113	74.05
SPD ($n'_i = 14$)	0.273	15.87	0.132	7.67	1.315	76.65

i indicates item effect; p, person effect; PHD, physical domain; $p \times i$, person-by-item interaction effect; PSD, psychological domain; SOD, social domain; SPD, specific domain.

TABLE 5. G-coefficients and Φ -Coefficients for Different Numbers of Items for $p \times 1$ Design in D-study for 4 Domains of Quality of Life Instrument QLICD-CG

Domain	Number of Items	$\sigma^2(P)$	$\sigma^2(I)$	$\sigma^2(PI)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$	$\sigma^2(X_{PI})$	Ep^2	Φ
Physical domain	6	0.357	0.094	0.143	0.143	0.237	0.098	0.714	0.601
	8	0.357	0.071	0.107	0.107	0.178	0.074	0.769	0.668
	10	0.357	0.057	0.086	0.086	0.142	0.060	0.807	0.715
	16	0.357	0.035	0.054	0.054	0.089	0.038	0.870	0.801
Psychological domain	8	0.752	0.023	0.106	0.106	0.129	0.032	0.876	0.854
	9	0.752	0.020	0.094	0.094	0.115	0.029	0.889	0.868
	10	0.752	0.018	0.085	0.085	0.103	0.027	0.899	0.879
	11	0.752	0.017	0.077	0.077	0.094	0.025	0.907	0.889
Social domain	9	0.273	0.013	0.124	0.124	0.137	0.016	0.688	0.666
	11	0.273	0.011	0.101	0.101	0.112	0.013	0.729	0.709
	12	0.273	0.010	0.093	0.093	0.103	0.012	0.746	0.727
	18	0.273	0.007	0.062	0.062	0.068	0.009	0.815	0.800
Specific domain	20	0.273	0.006	0.056	0.056	0.062	0.008	0.831	0.816
	11	0.273	0.011	0.101	0.101	0.112	0.013	0.729	0.709
	14	0.273	0.008	0.080	0.080	0.088	0.011	0.774	0.756
	20	0.273	0.006	0.056	0.056	0.062	0.008	0.831	0.816
	21	0.273	0.006	0.053	0.053	0.059	0.008	0.837	0.823

$\sigma^2(\delta)$ is the variance components of relative error.
 $\sigma^2(\Delta)$ is the variance components of absolute error.
 $\sigma^2(X_{PI})$ is the variance components of error when estimating the universe score by using sample mean.
 Ep^2 is the generalizability coefficient.
 Φ is the index of dependability.

It was worthy to note that the scale was developed for Chinese people and written in Chinese. Other foreign language versions could be developed by rigorous translation procedures and cultural adaptation (especially regard to above items PH6, PH2, SO4, SO5) if needed.

Validity is the extent to which an instrument can capture what it purports to measure, with content, construct and criterion-related validity being evaluated as usual. By following World Health Organization's definition of QOL³⁷ and systematic development procedures, the QLICD-CG

was developed by using focus group discussions, in-depth interviews and pretesting to effectively reduce the number of items in the final version to 30 from an initial 73 item bank for the general module²¹ and to 14 from an initial bank of 27 items for the specific module, ensuring good content validity and sound conceptual structure of this instrument. Correlation and factor analysis were used to confirm the construct. Correlation analyses among items-dimensions showed strong association between items and their own domains/facets but weak correlations between items and

TABLE 6. Responsiveness of the Quality of Life Instrument QLICD-CG (n = 120)

Domains/Facets (Number of Items)	Before Treatment		After Treatment		Differences		t	P	SRM
	Mean	SD	Mean	SD	Mean	SD			
Physical function (8)	61.20	17.61	74.97	13.65	-13.78	13.14	-11.48	<0.001	1.05
Independence (3)	80.14	22.46	89.03	15.39	-8.89	14.44	-6.74	<0.001	0.62
Appetite and sleep (2)	40.94	23.14	57.92	19.44	-16.98	22.15	-8.40	<0.001	0.77
Physical symptoms (3)	55.76	21.84	72.29	18.21	-16.53	19.29	-9.38	<0.001	0.86
Psychological function (11)	69.87	19.71	82.29	15.33	-12.42	14.21	-9.58	<0.001	0.87
Cognition (2)	63.33	26.07	79.90	17.85	-16.56	22.41	-8.10	<0.001	0.74
Anxiety (3)	61.32	25.86	83.96	18.56	-22.64	23.57	-10.52	<0.001	0.96
Depression (3)	75.90	22.50	84.51	18.09	-8.61	15.35	-6.14	<0.001	0.56
Self-consciousness (3)	76.74	22.37	80.00	19.28	-3.26	11.75	-3.04	<0.001	0.28
Social function (11)	67.29	13.98	73.35	13.19	-6.06	9.06	-7.32	<0.001	0.67
Social support/security (6)	67.85	16.44	76.60	14.03	-8.75	12.06	-7.95	<0.001	0.73
Social effects (4)	65.31	21.98	68.07	20.97	-2.76	11.54	-2.62	0.010	0.24
Sexual function (1)	71.88	27.23	75.00	24.04	-3.13	19.05	-1.80	.075	0.16
Sub-total (30)	66.61	13.82	77.06	12.08	-10.45	9.60	-11.92	<0.001	1.09
Specific domain (14)	59.64	15.45	77.98	12.90	-18.33	13.55	-14.83	<0.001	1.35
Unwell of upper abdomen (5)	51.08	20.86	79.58	15.82	-28.50	21.13	-14.78	<0.001	1.35
Acid regurgitation (1)	74.17	28.98	93.13	14.84	-18.96	26.53	-7.83	<0.001	0.71
Nausea (1)	71.46	30.97	91.88	18.66	-20.42	30.91	-7.23	<0.001	0.66
Heart burn (1)	64.17	32.04	85.83	19.10	-21.67	30.56	-7.77	<0.001	0.71
Effect of mental and life (6)	61.63	21.57	70.49	18.96	-8.85	12.91	-7.51	<0.001	0.69
Total (44)	64.39	12.53	77.35	11.14	-12.96	9.70	-14.63	<0.001	1.34

SRM indicates standardized response mean.

other domains/facets. Factor analysis revealed that the components extracted from the data coincide with the theoretical construct framework of the instrument. These results confirmed evidence supporting good construct validity. Strong correlations were observed between same/similar domains of the QLICD-CG and the SF-36, while weak correlations were observed between nonsame/similar domains, thus confirming the criterion-related validity and also the construct validity (the convergent and divergent validity) to a reasonable degree.

In terms of reliability, internal consistency reliability (Cronbach α), test-retest reliability (Pearson r), and ICC were applied in the current study. All domains and the overall score of the QLICD-CG demonstrated internal consistency, with Cronbach α coefficients ranging from 0.71 to 0.89. The test-retest reliability coefficient (Pearson r and ICC) for the overall score was 0.98 while those for the QLICD-CG domains were >0.90 ranging from 0.94 to 0.99. Thus this instrument has excellent reliability considering that internal consistency coefficients above 0.70 and test-retest reliability coefficient above 0.80 are generally accepted as satisfactory.

Responsiveness (sensitivity to detect change) is the most desirable characteristic of a QOL scale for clinical applications, with the assessment methods being divided into 2 categories: internal and external.^{29–31} In this paper we focused on internal responsiveness using the paired t test to compare the mean response before and after treatments, and a responsiveness indicator SRM, with values of 0.20, 0.50, and 0.80 representing small, moderate and large responsiveness.^{29–31} The QOL scores had significant changes after treatments for all domains and the overall score ($P < 0.001$) with SRM ranging from 0.67 to 1.35 for the domain level scores, suggesting QLICD-CG has excellent responsiveness.

In this research, GT was also applied both in G-study and D-study, with the index of dependability being typically lower than G-coefficients because they consider the main error effects in addition to the interaction effects that are used for G-coefficients. This research presented both G-coefficients and Φ , and also their changes when items assumed to be changing, with the G and Φ coefficients being increased with the number of items increasing. D studies were performed to estimate the G-coefficients and Φ coefficients for the $p \times i$ design, as well as the alternative design with different numbers of items in the 5 domains.

Under the current designs, the G and Φ coefficients were >0.70 in all 4 domains, except for Φ coefficient of the physical domain is close to 0.70 (0.668), demonstrating good reliability of the QLICD-CG and enough items for different domains. Specifically, for the physical and social domains, we estimated a G-coefficient and an index of dependability of 0.769, 0.668; 0.729, 0.709, respectively, for the current design, which can be considered to meet a standard of 0.70 with desirable items. For psychological domain, a G-coefficient of 0.907 and an index of dependability of 0.889 were estimated for the current design, which was higher than the acceptable level of 0.70, implying the domain's items can be reduced. The G-coefficient estimated to be 0.876 and the index of dependability 0.854 for an alternative design with 8 items, which will meet acceptable dependability and suitable items. For the specific domain, a G-coefficient of 0.774 and an index of dependability of 0.756 were estimated for the current design, which was higher than the acceptable level of 0.70, implying the domain's items can be reduced. If the numbers of items reduced to 11 items, the G-coefficient

estimated to be 0.729 and the index of dependability 0.709, which will just meet acceptable dependability. Therefore, these analyses suggest that the number of items of psychological and the specific domains can be reduced to 8, 11 items if need.

The study has several limitations that should be noted. It would be useful to correlate objective physiological indicators and the QLICD-CG scores but we were unable to perform these analyses in this study for the objective indicators were not a criterion for study inclusion. Moreover, the subjects in this study were selected from the inpatient at hospitals. Additional studies are needed to assess the generalizability of the instrument to other settings and populations such as outpatients at a local clinic.

In conclusion, the psychometric characteristics of the QLICD-CG were found to be excellent, demonstrating good validity, reliability and responsiveness. In summary, the QLICD-CG can be used to assess disease-specific health-related quality life in Chinese patients with CG, with excellent psychometric properties and several advantages. The analysis from GT not only confirmed the reliability of the scale further, but also presented more information on items change than CTT.

ACKNOWLEDGMENTS

In carrying out this research project, the authors have received substantial assistance from staffs of the first affiliated hospital of Kunming Medical University and the affiliated hospital of Guangdong Medical University, and also prof Wendy Wong at School of Chinese Medicine, Chinese University of Hong Kong. The authors sincerely acknowledge all the support.

REFERENCES

- De Vries AC, Van Grieken NC, Looman CW, et al. Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. *Gastroenterology*. 2008;134:945–952.
- Wang XZ. Diagnosis and treatment status of chronic gastritis. *Chin Foreign Health Magazine*. 2011;8:139–140.
- Yang L. Incidence and mortality of gastric cancer in China. *World J Gastroenterol*. 2006;12:17–20.
- Chooi EY, Chen HM, Miao Q, et al. Chronic atrophic gastritis is a progressive disease: analysis of medical reports from Shanghai (1985–2009). *Singapore Med J*. 2012;53:318–324.
- Du Y, Bai Y, Xie P, et al. Chronic gastritis in China: a national multi-center survey. *BMC Gastroenterol*. 2014;14:21.
- Borgaonkar MR, Irvine EJ. Quality of life measurement in gastrointestinal and liver disorders. *Gut*. 2000;47:444–454.
- Ashorn M, Mäki M, Ruuska T, et al. Upper gastrointestinal endoscopy in recurrent abdominal pain of childhood. *J Pediatr Gastroenterol Nutr*. 1993;16:273–277.
- Talley NJ, Verlinden M, Jones M. Validity of a new quality of life scale for functional dyspepsia: a United States multicenter trial of the Nepean Dyspepsia Index. *Am J Gastroenterol*. 1999;94:2390–2397.
- Shaw MJ, Beebe TJ, Adlis SA, et al. Reliability and validity of the digestive health status instrument in samples of community, primary care, and gastroenterology patients. *Aliment Pharmacol Ther*. 2001;15:981–987.
- Shaw M, Talley NJ, Adlis S, et al. Development of a digestive health status instrument: tests of scaling assumptions, structure and reliability in a primary care population. *Aliment Pharmacol Ther*. 1998;12:1067–1078.
- Wiklund IK, Junghard O, Grace E, et al. Quality of Life in Reflux and Dyspepsia patients. Psychometric documentation of a new disease-specific questionnaire (QOLRAD). *Eur J Surg Suppl*. 1998;583:41–49.

12. Crawley J, Frank L, Joshua-Gotlib S, et al. Measuring change in quality of life in response to *Helicobacter pylori* eradication in peptic ulcer disease: the QOLRAD. *Dig Dis Sci*. 2001;46:571–580.
13. Chassany O, Marquis P, Scherrer B, et al. Validation of a specific quality of life questionnaire for functional digestive disorders. *Gut*. 1999;44:527–533.
14. Svedlund J, Sjodin I, Dotevall G. GSRS-A clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci*. 1988;33:129–134.
15. Kulich KR, Madisch A, Pacini F, et al. Reliability and validity of the Gastrointestinal Symptom Rating Scale (GSRS) and Quality of Life in Reflux and Dyspepsia (QOLRAD) questionnaire in dyspepsia: a six-country study. *Health Qual Life Outcomes*. 2008;6:12.
16. Rentz AM, Kahrilas P, Stanghellini V, et al. Development and psychometric evaluation of the patient assessment of upper gastrointestinal symptom severity index (PAGI-SYM) inpatients with upper gastrointestinal disorders. *Qual Life Res*. 2004;13:1737–1749.
17. Wyrwich KW, Mody R, Larsen LM, et al. Validation of the PAGI-SYM and PAGI-QOL among healing and maintenance of erosive esophagitis clinical trial participants. *Qual Life Res*. 2010;19:551–564.
18. de la Loge C, Trudeau E, Marquis P, et al. Cross-cultural development and validation of a patient self-administered questionnaire to assess quality of life in upper gastrointestinal disorders: the PAGI-QOL. *Qual Life Res*. 2004;13:1751–1762.
19. 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. EORTC Quality of Life Study Group. *Qual Life Res*. 1998;7:291–300.
20. Cella D, Nowinski CJ. Measuring quality of life in chronic illness: the functional assessment of chronic illness therapy measurement system. *Arch Phys Med Rehabil*. 2002;83(suppl 2):S10–S17.
21. Wan CH, Tu XM, Messing S, et al. Development and validation of the general module (QLICD-GM) of the system of quality of life instruments for chronic diseases and comparison with SF-36. *J Pain Symptom Manage*. 2011;42:93–104.
22. Wan CH, Jiang RS, Tu XM, et al. The hypertension scale of the system of quality of life instruments for chronic diseases QLICD-HY: development and validation study. *Int J Nurs Stud*. 2012;49:465–480.
23. Wan C, Li H, Fan X, et al. Development and validation of the coronary heart disease scale under the system of quality of life instruments for chronic diseases QLICD-CHD: combinations of classical test theory and generalizability theory. *Health Qual Life Outcomes*. 2014;12:82.
24. Lei P, Lei G, Tian J, et al. Development and validation of the irritable bowel syndrome scale under the system of quality of life instruments for chronic diseases QLICD-IBS: combinations of classical test theory and generalizability theory. *Int J Colorectal Dis*. 2014;29:1245–1255.
25. Yang Z, Li W, Tu XM, et al. Validation and psychometric properties of Chinese version of SF-36 in patients with hypertension, coronary heart diseases, chronic gastritis and peptic ulcer. *Int J Clin Pract*. 2012;66:991–998.
26. Hays RD, Hayashi T. Beyond internal consistency reliability: rationale and use's guide for multi-trait analysis program on the microcomputer. *Behav Res Methods Instrum Compu*. 1990;22:167–175.
27. Ren S, Yang S, Lai S. Intraclass correlation coefficients and bootstrap methods of hierarchical binary outcomes. *Stat Med*. 2006;25:3576–3588.
28. 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.
29. 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;12:349–363.
30. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61:102–109.
31. Lehman LA, Velozo CA. Ability to detect change in patient function: responsiveness designs and methods of calculation. *J Hand Ther*. 2010;23:361–370.
32. Winterstein BP, Willse JT, Kwapi TR, et al. Assessment of score dependability of the Wisconsin Schizotypy scales using generalizability analysis. *Psychopathol Behav Assess*. 2010;32:575–585.
33. Stora B, Hagtvet KA, Heyerdahl S. Reliability of observers' subjective impressions of families: a generalizability theory approach. *Psychother Res*. 2013;23:448–463.
34. Crits-Christoph P, Johnson J, Gallop R, et al. A generalizability theory analysis of group process ratings in the treatment of cocaine dependence. *Psychother Res*. 2011;21:252–266.
35. Heitman RJ, Kovaleski JE, Pugh SF. Application of generalizability theory in estimating the reliability of ankle-complex laxity measurement. *J Athl Train*. 2009;44:48–52.
36. Chavez LM, Garcia P, Ortiz N, et al. Applying generalizability theory methods to assess continuity and change on the Adolescent Quality of Life-Mental Health Scale (AQOL-MHS). *Qual Life Res*. 2016;25:3191–3196.
37. The WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): development and psychometric properties. *Soc Sci Med*. 1998;46:1569–1585.