Development and preliminary validation of the chronic obstructive pulmonary disease scale quality of life instruments for chronic diseases-chronic obstructive pulmonary disease based on classical test theory and generalizability theory

Chonghua Wan[1],[#] ⓘ, Zheng Yang[2],[#], Zhihuan Zhao[3], Peng Quan[1], Bin Wu[4] and Yunbin Yang[1]

Abstract

Quality of life (QOL) in patients with Chronic obstructive pulmonary disease (COPD) is a major global concern in respiratory care with the specific instruments used rarely being developed using a modular approach. This paper is aimed to develop the COPD scale of the system of QOL Instruments for Chronic Diseases (QLICD-COPD) by the modular approach based on Classical Test Theory and Generalizability Theory (GT). 114 inpatients with COPD were used to provide the data measuring QOL three times before and after treatments. The psychometric properties of the scale were evaluated with respect to validity, reliability and responsiveness employing correlation analysis, factor analyses, multi-trait scaling analysis, and also GT analysis. The Results showed that Multi-trait scaling analysis, correlation and factor analyses confirmed good construct validity and criterion-related validity with almost all correlation coefficients or factor loadings being above 0.40. The internal consistency α and test-retest reliability coefficients (Pearson r and Intra-class correlations ICC) for all domains except for the social domain were larger than 0.70, ranging between 0.70–0.86 with r = 0.85 for the overall. The overall score and scores for physical and the specific domains had statistically significant changes after treatments with moderate effect size SRM (standardized response mean) ranging from 0.32 to 0.44. All G-coefficients and index of dependability were all greater than 0.80 exception of social domain (0.546 and 0.500 respectively), confirming the reliability of the scale further. It concluded that the QLICD-COPD has good validity, reliability, and moderate responsiveness, and can be used as the QOL instrument for patients with COPD.

[1]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, China
[2]School of Public Health, Guangdong Medical University, Dongguan, China
[3]The First Affiliated Hospital of Kunming Medical University, Kunming, China
[4]The Affiliated Hospital of Guangdong Medical University, Zhanjiang, China

[#]Chonghua Wan and Zheng Yang are as the first co-author with the same contributions.

Corresponding authors:
Chonghua Wan, 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, 1 Xincheng Road, Dongguan 523808, China.
Email: wanchh@hotmail.com

YunBin Yang, 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 523808, China.
Email: yyb08@126.com

Introduction

Chronic obstructive pulmonary disease (COPD) is a slowly developing disease that causes a huge economic and social burden.[1,2] The disease commonly causes people to experience symptoms such as difficulty breathing, cough, chest tightness and fatigue, etc. Since most patients with COPD are still incurable, one of the main goals of care is to improve health-related quality of life (HRQOL), which is subjective in nature and involves patients' self-assessment of multiple health dimensions. As a result, many studies have included HRQOL measurements in assessing the impact and progress of the disease.[3-5] However, many studies used generic questionnaires which have a relatively wide application and can be used for comparison among different diseases, including EuroQol 5 dimension,[5-8] the MOS Short-Form 36 (SF-36) and Short-Form 12,[8-11] and the World Health Organization Quality of Life-BREF.[12] One of main limitations of generic questionnaires is that they may not be sensitive to small changes associated with specific disease. On the other hand, disease-specific questionnaires can detect minor clinical changes. For this reason, many disease-specific questionnaires have been developed, which included the St George's Respiratory Questionnaire (SGRQ), COPD Assessment Test (CAT), and COPD Clinical Questionnaire (CCQ).[13-15] Weldam et al.[16] reviewed 77 studies on QOL in patients with COPD and found that 13 disease-specific questionnaires and 10 generic questionnaires have ever been used.

The SGRQ, CCQ and CAT etc. are very well validated questionnaires and widely used in research and clinical practice. However, these specific scales are developed neither for COPD (but for respiratory diseases) nor for QOL (but for health and function status to the most extent), demonstrating some limitations used for QOL in COPD.[16-18] For example, most questions/items (80%) in SGRQ have two points (yes/no) answers. As a result, the scale/domain scores may be less responsive to changes, as two-point scales are less reliable than those with more categories.[19] In addition, they are not developed through modular method - generic/core module plus specific module to overcome the shortcomings of generic and specific instrument use alone.[20-22] Since diseases in the same class such respiratory diseases have much in common, the modular approach is helpful to capture common and also unique characteristics. The general module can be used to compare QOL across different diseases, while the specific module can be used to depicture symptoms and treatments in detail. Therefore, the modular approach for developing instruments has its significant benefits. Consequently, both the QLQs (Quality of Life Questionnaires) from EORTC (European Organization for Research and Treatment) and the FACITs (Functional Assessment of Chronic Illness Therapy) from CORE (Center on Outcomes, Research and Education) in USA were developed based on this modular principle.[20,21]

Consider combining generic and disease specificity in the questionnaire and responding directly to the need for COPD in HRQOL instruments, we have developed a system of Quality of Life Instruments for Chronic Diseases (QLICD).[23,24] 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.[23,24] Among this system, the QLICD-COPD is constructed by combining QLICD-GM with the specific module of COPD.

Considering no COPD instrument developed by the modular approach, and both classical test theory (CTT) and Generalizability theory (GT) have their advantages,[24] especially GT can gives more sophisticated results than CTT, this article is aimed to describe the development and validation of this instrument based on the modular approach by both CTT and GT. To use for reference, other instruments can be developed and validated efficiently in similar ways.

Methods

Establishment of the general module (QLICD-GM)

The study population was limited to patients with chronic diseases who were able to read and understand the questionnaires at any stages and treatments. And the study protocol and the informed consent form were approved by the IRB (institutional review board) of Kunming Medical University.

The QLICD-GM was developed followed the WHO's definition of QOL.[25] First, we applied a programmatic decision-making procedure that effectively reduced the item number from 73 item pool to 30. Qualitative analysis including data transcription and thematic coding based on in-depth interview, focus group discussions etc., and quantitative analysis procedures including importance rating, variation analysis, correlation analysis and factor analysis etc. were used for item selection. The whole process of developing QLICD-GM was described in detail in other

papers.[23] The final QLICD-GM included 30 items (3 domains and 10 facets). There are 8 items in physical domain (coded PH1-PH8), 11 items in psychological domain (PS1-PS11), and 8 items in social domain (SO1-SO8) (see Table 1 for details). The results from 620 patients of seven diseases of hypertension, coronary heart disease, chronic gastritis, peptic ulcer, irritable bowel syndrome, COPD, and chronic pulmonary heart disease showed the QLICD-GM has good psychometric properties.[23]

Establishment of the specific module

The programmed decision method was also used in item generation and selection.[23] First, the focus group discussed and confirmed the structure of the module, which included five facets. Based on literature review, reference of domestic and foreign mature scales, and the clinical experience of COPD, the nominal group proposed some possible items under each of the facet within the module. As a result, 25 items were presented to form an item pool of the specific module that reflects the symptoms of COPD such as cough, white sputum, and side effects of treatments on COPD such as nausea, vomit and dry mouth. Then, a pre-test and in-depth interview was administered to 20 patients and 20 experts of COPD (physicians/nurses), respectively. The following statistical procedures were used to re-screen the items:

(1) Importance rating by patients. Each patient was asked to score the importance of each item on a 0-100 score system (0 = extremely unimportant and 100 = extremely important). Those items with a low importance rating (<65) were deleted.

(2) Importance rating by experts. Each expert was asked to score the importance of each item same as (1). Those items with a low importance rating (<65) were deleted.

(3) Variation procedure. The standard deviation (SD) of the scores for each item was calculated and the items with smaller SD (<1.10) were deleted.

(4) Correlation procedure. The Pearson's correlation coefficients between the score of each item and the sum score of its own domain were computed and the items with smaller correlation coefficients (<0.50) were deleted.

To determine the final items, the following selection rules were used: (1) retain those items that were selected by at least three of the above procedures; and (2) retain those items selected by at least two procedures with incorporation of professional knowledge (focus group).

In the end, the specific module consists of 15 items, coded COPD1-COPD15, classified into 5 facets (see Tables 1 and 3 for details).

Validation of the QLICD-COPD

Data collection and scoring. The final QLICD-COPD consists of the above general and specific modules and is used in a field survey at the First Affiliated Hospital of Kunming Medical University to study their psychometric properties. According to the guidelines,[26,27] COPD was diagnosed by the presence of chronic cough, sputum production and/or dyspnea, and also other clinical indicators such as Forced Expiratory Volume in the first second (FEV 1), Forced Vital Capacity (FVC).

The study population was limited to COPD inpatients who were able to read and understand the questionnaires at any stages and treatments. The investigators explained the aims of the tests and the instrument to the patients and obtained written informed consent from those patients who agreed to participate in the study and met the inclusion criteria. Participants were asked to answer the questionnaire at the time of admission by themselves. Each patient was assessed a second time on the second day after hospitalization to assess the reliability of the test-retest. The Data were obtained on the third scheduled time to complete the discharge measures to assess the responsiveness of the questionnaire, after about 1 week of treatment. The Chinese version of SF-36[28] was also used simultaneously to provide data for evaluating the criterion-related validity, and also the convergent and the discriminant validity of the QLICD-COPD.

Participants took 15–30 min to complete the questionnaires. Each investigator immediately checked the answers each time to ensure their integrity. If missing values were found, the questionnaire would be returned to the patients to fill in the missing item.

Each item of QLICD-COPD is rated as a 5-point Likert rating system. Positive statement items are rated directly from 1 to 5, while negative statement items are scored in reverse.

Psychometrics analysis

There are several types of validity that can be distinguished.[29,30] The construct validity is evaluated by the Pearson correlation coefficient r between the item and the domain and the factor analysis using the Varimax rotation. We used correlating corresponding domains of the QLICD-COPD and SF-36 to evaluate criterion-related validity. Relatively high correlations among conceptually related domains and relatively low correlation among conceptually distinct domains would suggest high criterion-related validity. Multi-trait scaling analysis[31] was used to test

Table 1. Correlation coefficients r among items and domains of QLICD-COPD (*n* = 114).

Code	Items brief description	Physical	Psychological	Social	Specific
PH1	Take care of daily life (e.g. eating)?	**0.42**	0.03	0.13	0.29
PH2	Felt easily fatigued?	**0.73**	0.52	0.25	0.57
PH3	Have trouble walking 800m or more?	**0.80**	0.32	0.12	0.65
PH4	Have trouble going up and down stairs?	**0.82**	0.43	0.18	0.62
PH5	Need to take medication?	**0.80**	0.59	0.24	0.74
PH6	A good appetite?	**0.57**	0.26	0.30	0.45
PH7	Satisfied with your sleep?	**0.53**	0.14	0.20	0.36
PH8	Felt pain or uncomfortable?	**0.68**	0.62	0.39	0.58
PS1	Memory and concentration affected?	0.39	**0.65**	0.56	0.44
PS2	Felt mentally miserable?	0.59	**0.82**	0.60	0.53
PS3	Felt lonely and helpless?	0.20	**0.68**	0.51	0.38
PS4	Felt pessimism and despair?	0.36	**0.71**	0.51	0.45
PS5	Worried about disease?	0.46	**0.66**	0.27	0.43
PS6	Felt fretful or irritable?	0.59	**0.52**	0.26	0.58
PS7	Felt nervous and anxious?	0.57	**0.60**	0.19	0.47
PS8	Stop medication because of side effects?	0.20	**0.44**	0.05	0.17
PS9	To be a burden to the family?	0.09	**0.73**	0.43	0.23
PS10	Felt self-abasement because of disease?	0.25	**0.76**	0.39	0.38
PS11	Hidden emotions but could not forget?	0.12	**0.65**	0.56	0.17
SO1	Interfered with work/housework?	0.72	0.55	**0.44**	0.56
SO2	Family roles?	0.11	0.10	**0.44**	0.02
SO3	Decreased caring and attention to family?	0.49	0.51	**0.60**	0.50
SO4	Good relations with family?	0.21	0.21	**0.50**	0.09
SO5	Help and support from family?	0.22	0.25	**0.52**	0.07
SO6	Affected participating in leisure activities?	0.50	0.22	**0.25**	0.44
SO7	Treat illness positively and optimistically?	0.24	0.48	**0.72**	0.28
SO8	Treatments received good for curing?	0.20	0.08	**0.42**	0.21
SO9	Economic problems caused by illness?	0.11	0.03	**0.50**	0.18
SO10	Support from friends and relatives?	0.30	0.09	**0.33**	0.19
SO11	Affected sexual activities?	0.20	0.50	**0.46**	0.38
COPD1	Cough?	0.42	0.25	0.16	**0.64**
COPD2	Cough in morning more than daytime?	0.41	0.28	0.02	**0.54**
COPD3	Paroxysmal cough before sleeping?	0.31	0.51	0.15	**0.58**
COPD4	Yellow and purulence sputum?	0.29	0.52	0.59	**0.53**
COPD5	White sputum?	0.49	0.38	0.26	**0.53**
COPD6	Much more sputum?	0.56	0.27	0.21	**0.71**
COPD7	More sputum in morning?	0.48	0.42	0.33	**0.68**
COPD8	Expectoration before sleeping?	0.53	0.55	0.48	**0.68**
COPD9	Dyspnea in resting state?	0.54	0.19	0.24	**0.63**
COPD10	Dyspnea after walking?	0.73	0.32	0.14	**0.75**
COPD11	Dyspnea when climbing or working?	0.64	0.28	0.02	**0.65**
COPD12	Chest tightness?	0.42	0.13	0.07	**0.54**
COPD13	Sleep disorder?	0.62	0.36	0.48	**0.71**
COPD14	Need oxygen therapy?	0.43	0.25	0.12	**0.55**
COPD15	Restriction to go out?	0.37	0.61	0.53	**0.49**

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

the convergence validity and discriminant validity of the scale. There are two criteria: (1) when an item-domain correlation is 0.40 or greater, the convergence validity is supported; when item-domain correlation is higher than that with other domains, it indicates discriminant validity.

Reliability refers to the degree to which the instrument is not affected by random errors and is evaluated by internal consistency and repeatability. Internal consistency is evaluated by Cronbach's alpha coefficient of domain/facet. A high internal consistency (above 0.7) indicates that the scale

is measuring a single structure. Repeatability (test-retest reliability) is the instrument stability over time in a stable population.[30] It was evaluated by the Pearson's correlation coefficients between the first and second assessments, and intra-class correlation (ICC) with definition of absolute for single measure under the two-way mixed model.[32,33]

Responsiveness is the instrument's ability to detect clinically important change over time. It was measured by comparing the mean score change between the two assessments before and after treatments using paired t-tests as well as the effect size SRM (standardized response mean), with values of 0.20, 0.50 and 0.80 having been proposed to represent small, moderate and large responsiveness, respectively.[34,35]

GT analysis

In addition to the CTT, in the current study, we also applied GT (G theory) to analyze the score dependability. G theory let us improve the design of the measurement procedure by techniques of experiment design and analysis of variance (-ANOVA), and try to produce reliable data through two types of study: G studies and D studies.[36-39] The G study quantifies the amount of variance associated with the different facets (factors) that are being examined. The D study provides information about which protocols are optimal for a particular measurement situations by G-coefficients that can be interpreted as reliability coefficients across various facets of the study.

In our research, a random crossed design (person-by-item ($p \times i$) design) was conducted in G and D studies to estimate variance components and dependability coefficients. We define a patient's QOL as a measurement target, and use the item as a facet of measurement error. It is equivalent to all participants being asked to reply to all items, the design is a one-facet crossover design.[36-39] Therefore, the corresponding relative errors or absolute errors in the G study, as well as generalizability and reliability coefficients in the D study in each potential factor were

Results

Socio-demographic characteristics of the sample

At first measurement at admission to hospital, 114 patients with COPD were enrolled (mean age 67.4 ± 10.9, 71.9% were male). Of them, 104 subjects participated in the test-retest phase of the study, and 106 subjects completed an assessment of sensitivity to change at discharge.

Construct validity

Correlation analysis of data measurements at admission shows that there is a strong correlation between the item and

its own domain, and all correlation coefficients above 0.40 except of SO6 and SO10, but weak relationship between items and other domains (see Table 1 for detail).

Factor analyses led to the extraction of 8 principal components from the 30 items of the general module QLICD-GM based on the eigenvalues >1 criterion, accounting for 75.91% of the cumulative variance. Three domains of the general module included eight main components reflect the different facets. The physical domain(represented by the first and seventh principal components) with higher loadings on PH2(0.78), PH3(0.64), PH4(0.72), PH5(0.73), PH6(0.73), PH7(0.79), PH8 (0.78); The second and eighth major components mainly reflect the social domain with higher loadings on SO4(0.84), SO5(0.88),SO7(0.62), SO8(0.68) and SO10(0.80); Other major components usually describe a psychological domain with a higher loadings on PS1(0.82), PS3(0.65), PS6(0.80), PS7(0.67), PS8(0.86), and PS11(0.68).

Similarly, the principal component factor analysis extracted 4 principal components from the 15 items of the specific module with the cumulative variance of 70.78%, reflecting 4 facets (cough and phlegm, short breath, pulmonary encephalopathy, and effect of mental and life) with high loadings ranging from 0.61 to 0.83.

From the results described above, the theoretical construction is usually confirmed by data analysis and shows good construct validity.

Criterion-related validity

The correlation coefficient is given in Table 2, which shows the domain score of the QLICD-COPD and SF-36. It shows that the correlation between the same and similar domains is usually higher than the correlation between the different and non-similar domains. For example, the coefficient between the physical domain of QLICD-COPD and physical function (PF) of SF-36 was 0.59, higher than any other coefficients in this row (PCS and MCS are not mutually exclusive domain of SF-36).

These confirmed the criterion-related validity and also demonstrated the convergent and divergent validity to some extent.

Reliability

We evaluated the reliability of the scale (see Table 3 for details). The Cronbach's α for all domains and facets were computed using the measurements data at admission because of larger sample size. The Cronbach's α for these four domains were higher than 0.70 except for SOD (0.55).

Table 2. Correlation Coefficients among domain scores of QLICD-COPD and SF-36 (*n* = 114)*.

QLICD-COPD	SF-36									
	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
PHD	0.59	0.58	0.54	0.10	0.30	0.46	0.44	0.44	0.63	0.51
PSD	0.17	0.26	0.66	0.02	0.43	0.60	0.12	0.42	0.29	0.55
SOD	0.02	0.11	0.69	0.21	0.41	0.53	0.04	0.43	0.05	0.52
SPD	0.54	0.55	0.59	0.03	0.25	0.45	0.41	0.40	0.55	0.47

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

Table 3. Reliability of the quality of life instrument QLICD-COPD (*n* = 114 for α, *n* = 104 for *r* and ICC).

Domains/Facets	Internal consistency coefficient α	Test-retest coefficient *r*	Test-retest ICC (95% CI)
Physical domain (PHD)	0.82	0.86	0.86 (0.80–0.90)
Independence	0.68	0.83	0.83 (0.76–0.88)
Appetite and sleep	0.61	0.85	0.85 (0.78–0.89)
Physical symptoms	0.84	0.81	0.81 (0.73–0.86)
Psychological domain (PSD)	0.87	0.70	0.69 (0.58–0.78)
Cognition	0.80	0.79	0.78 (0.70–0.85)
Anxiety	0.75	0.63	0.63 (0.49–0.73)
Depression	0.74	0.69	0.67 (0.56–0.76)
Self-consciousness	0.70	0.48	0.48 (0.31–0.61)
Social domain (SOD)	0.55	0.73	0.73 (0.63–0.81)
Social support/Security	0.60	0.76	0.76 (0.67–0.83)
Social effects	0.71	0.80	0.80 (0.71–0.86)
Sexual function	—	0.84	0.84 (0.78–0.89)
Specific domain (SPD)	0.88	0.83	0.83 (0.76–0.88)
Cough and phlegm	0.83	0.79	0.79 (0.71–0.85)
Short breath	0.88	0.80	0.79 (0.71–0.85)
Pulmonary encephalopathy	—	0.73	0.72 (0.61–0.80)
Oxygen treatment	—	0.76	0.76 (0.67–0.83)
Effect of mental and life	—	0.83	0.82 (0.75–0.88)
Total (TOT)	—	0.85	0.85 (0.79–0.90)

- not acceptable/suitable.

The test-retest correlation coefficients (r) for the 4 domains and the overall QLICD-COPD were larger than 0.70, ranging between 0.70–0.86 with r = 0.85 for the overall scale. The results from ICC and their 95% confidence intervals computed based on the definition of absolute agreement were very similar to Pearson's correlation coefficients (r).

Responsiveness

The paired t-tests and responsiveness indicator, SRM, were used to examine changes of mean scores from each domain/facet of the QLICD-COPD before and after treatments, with the results presented in Table 4. There were significant changes occurred for domains of physical function and the disease-specific, and also the subtotal (QLICD-GM) and overall scale (Ps < 0.01) with effect size SRM ranging from 0.32 to 0.44. At the facets level, five of eight facets of these two domains were of statistical significance with effect size SRM ranging from 0.23 to 0.52.

Results from GT

We conduct a G-study to estimate as many as possible the variance of the four domain of QLICP-COPD, as shown in

Table 4. Responsiveness of the quality of life instrument QLICD-COPD (n = 106).

Domains/facets (number of items)	Before treatment		After treatment		Differences				
	Mean	SD	Mean	SD	Mean	SD	t	p	SRM
Physical function (8)	57.02	17.92	61.88	21.45	−4.86	15.39	−3.25	0.002	0.32
Independence (3)	49.21	21.07	56.45	23.04	−7.23	18.15	−4.10	0.000	0.40
Appetite and sleep (2)	45.99	22.20	50.83	23.92	−4.83	21.10	−2.36	0.020	0.23
Physical symptoms (3)	72.17	22.85	74.69	24.96	−2.52	18.04	−1.44	0.154	0.14
Psychological function (11)	83.88	12.59	85.42	13.39	−1.54	12.91	−1.23	0.221	0.12
Cognition (2)	82.43	19.19	82.43	20.54	0.00	15.91	0.00	1.000	0.00
Anxiety (3)	76.57	16.47	80.42	17.78	−3.85	17.49	−2.27	0.025	0.22
Depression (3)	91.04	12.95	90.96	13.17	0.08	13.24	0.06	0.951	0.01
Self-consciousness (3)	84.98	15.28	86.87	12.80	−1.89	15.82	−1.23	0.222	0.12
Social function (11)	62.18	11.80	62.59	11.90	−0.41	10.18	−0.41	0.681	0.04
Social support/Security (6)	62.46	15.23	62.46	14.19	0.00	12.47	0.00	1.000	0.00
Social effects (4)	61.14	20.88	61.26	19.87	−0.12	16.95	−0.07	0.943	0.01
Sexual function (1)	64.62	31.60	68.63	29.69	−4.01	24.91	−1.66	0.101	0.16
Sub-total (30)	68.76	11.16	70.77	12.76	−2.01	10.17	−2.04	0.044	0.20
Specific domain(15)	65.33	14.96	71.84	18.90	−6.51	14.91	−4.49	0.000	0.44
Cough and phlegm(8)	66.57	15.03	73.61	18.89	−7.05	17.18	−4.22	0.000	0.41
Short breath(4)	61.08	21.84	70.75	23.07	−9.67	18.72	−5.32	0.000	0.52
Pulmonary encephalopathy(1)	76.18	25.91	81.13	27.56	−4.95	19.95	−2.56	0.012	0.25
Oxygen treatment(1)	63.44	23.70	62.26	25.18	1.18	17.04	0.71	0.478	0.07
Effect of mental and life(1)	63.44	36.22	62.26	36.39	1.18	18.70	0.65	0.518	0.06
Total (45)	67.62	11.62	71.13	14.20	−3.51	10.86	−3.33	0.001	0.32

Table 5. The estimated variance components and percentage of variances for p×i design in G-study for four domains of quality of life instrument QLICP-COPD ($n'_p = 114$).

Domain	P (person)		i (item)		p* i (person*item)	
	Variance component	Percent (%)	Variance component	Percent (%)	Variance component	Percent (%)
PHD ($n'_i = 8$)	0.42	28.77	0.30	20.55	0.74	50.68
PSD ($n'_i = 11$)	0.22	32.84	0.07	10.45	0.38	56.72
SOD ($n'_i = 11$)	0.12	8.16	0.23	15.65	1.12	76.19
SPD ($n'_i = 15$)	0.31	28.18	0.13	11.82	0.66	60.00

p: person effect, i: item effect, p × i: person-by-item interaction effect.
PHD: physical domain, PSD: psychological domain, SOD: social domain, SPD: specific domain.

Table 5, 114 patients filled out the instrument with 45 items. For physical domain, the variances accounted for 28.77%, 20.55% and 50.68% by person, item and person-by-item interactions respectively. The largest source of variation came from person-by-item interactions in other domains, while the variances by person were in the second place (except for social domain by item).

The D-Studies were performed to estimate the G-coefficient and Φ coefficient (Table 6). It can be seen the G-coefficient and Φ coefficient were estimated for the current design and alternative designs with varied numbers of items.

Discussions

The article focuses on the main steps in development and validation of a specific QOL instrument for COPD (the QLICD-COPD) by combining the general module of the entire disease category with the specific module. As far as we know, although a number of instruments have been widely used for studying COPD impacts on patients' HRQOL, none of them combine brevity, comprehensive coverage of all dimensions of HRQOL and COPD specificity together. Moreover, there was none of them developed directly for COPD by the modular approach

Table 6. G-coefficients and Φ-coefficients for different numbers of items for p ×I design in D-study for four domains of quality of life instrument QLICP-COPD.

Domain	Number of items	$\sigma^2(P)$	$\sigma^2(I)$	$\sigma^2(PI)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$	$\sigma^2(X_{PI})$	$E\rho^2$	Φ
Physical domain	6	0.424	0.051	0.123	0.123	0.174	0.056	0.775	0.709
	8	0.424	0.038	0.092	0.092	0.130	0.043	0.821	0.765
	10	0.424	0.030	0.074	0.074	0.104	0.035	0.852	0.803
	12	0.424	0.025	0.062	0.062	0.087	0.030	0.873	0.830
	14	0.424	0.022	0.053	0.053	0.075	0.026	0.889	0.851
Psychological domain	9	0.220	0.008	0.042	0.042	0.050	0.010	0.840	0.815
	11	0.220	0.007	0.034	0.034	0.041	0.009	0.865	0.843
	13	0.220	0.006	0.029	0.029	0.035	0.008	0.883	0.864
	15	0.220	0.005	0.025	0.025	0.030	0.007	0.897	0.880
	17	0.220	0.004	0.022	0.022	0.026	0.006	0.908	0.893
Social domain	9	0.122	0.025	0.124	0.124	0.149	0.027	0.496	0.450
	11	0.122	0.021	0.101	0.101	0.122	0.023	0.546	0.500
	13	0.122	0.017	0.086	0.086	0.103	0.019	0.587	0.541
	15	0.122	0.015	0.074	0.074	0.089	0.017	0.621	0.577
	17	0.122	0.013	0.066	0.066	0.079	0.015	0.650	0.607
Specific domain	13	0.310	0.010	0.051	0.051	0.061	0.013	0.858	0.836
	15	0.310	0.009	0.044	0.044	0.053	0.012	0.875	0.854
	17	0.310	0.008	0.039	0.039	0.047	0.011	0.888	0.869
	19	0.310	0.007	0.035	0.035	0.042	0.010	0.899	0.881
	21	0.310	0.006	0.032	0.032	0.038	0.009	0.907	0.891

$\sigma^2(\delta)$ 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.
$E\rho^2$ is the Generalizability coefficient.
Φ is the index of dependability.

although the recent McGill COPD QOL questionnaire was established by combining both generic and disease-specific properties.[22,40] In addition, the QLICD-COPD has several advantages[23,24] over existing instruments: (1) it can compare HRQOL for various diseases through a generic module and capture symptoms and side effects through the specific module, showing general and specific attributes; (2) it consists of a medium number of items with a clear hierarchy (items→ facets→ domains→ overall) so that mean scores can be computed not only at the domain (four domains) and the overall levels but also at the different facet levels (15 facets) to detect changes in detail; (3) Based on the scale, the foreign language versions can be developed with strict translation and back-translation procedures, and it can be used in future research, testing its use in other languages in other countries.

On items generation and selection, both qualitative (in-depth interview, focus group discussion) and quantitative methods (importance rating, variation and correlation procedure) were used. Qualitative methods were just used for essential tools and steps to identify content for the instrument, to establish relevant domains and rate items of relevance (item content and clarity). The data were transcribed, managed, and thematic coding and analysis were done by NVivo software package. This is a big and comprehensive process for item selection and content validity of the instrument and will be reported in other paper.

In terms of quantitative methods, the thresholds for item selection were based on the statistical standards (e.g. correlation coefficients <0.50) and scale development experience or item numbers prefer to retain (e.g. importance rating <65, SD<1.10), which were subjective to some extent.

With regard to psychometrics, practical QOL instruments must be validated in at least three ways: validity, reliability and responsiveness. We used correlation and factor analysis to confirm the construct and criterion-related validity of QLICP-COPD. Correlation analysis showed a strong correlation between the items and its own domain/facet, but weak correlation between the items and other domains/facet, indicating good criterion-related validity and construct validity.

Our results indicated that the instrument has good reliability given coefficients above 0.70. For the exception of the social domain, many studies presented small Cronbach's α for relatively higher heterogeneity of this domain (social support, social relations, social security, etc.).[12-15] Accordingly, we suggested that it is acceptable for Cronbach's α 0.60 for social domain.

For responsiveness, we focus on internal responsiveness, assuming that sensitive instruments should detect changes after detection. As shown in Table 4, QOL scores had significant changes after treatments for physical and the specific domains as well as the overall score with SRM ranging from 0.32 to 0.44 at domains levels, and also 5 out of 8 facets reached statistical significance. Some possible reasons for these non-significant domains and facets are: (1) The observation period (about 1 week) was not long enough to detect significant changes; (2) They may not change over time in nature, such as the social function in hospital. In other words, the instrument revealed the changes of domain/facets scores which are expected to change. Therefore, it can be inferred that this instrument has good (moderate) responsiveness.

Traditionally, the scale is assessed by classical test theory analysis. Generalizability Theory was also applied both in G-study and D-study in this research. The index of dependability is 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 as items changing.

For social domain, we estimated a G-coefficient of 0.546 and an index of dependability of 0.500 for the current design, which was below the acceptable level of 0.70. For an alternative design with 17 items, the G-coefficient estimated to be 0.650 and the index of dependability 0.607, which will meet acceptable criteria for social domain. Theoretically speaking, it will be better to increase the numbers of items of social domain from 11 to 17 in order to reach an acceptable dependability 0.70. However, this is very difficult in practice. Therefore, we preferred to coefficient 0.60 being as acceptable dependability as many researches.[12-15] In terms of other domains, G-coefficients and index of dependability were all greater than 0.80 for the current design, and changed a little as items changing. It can be considered that current items are reasonable and acceptable for all domains.

In this paper, both CTT and GT (focusing on reliability of the scale) were used to develop and validate the QLICD-COPD to avoid one's weaknesses. The analysis from GT confirmed the reliability of the scale further, and presented much more information on items change. The numbers of items for social domain can be increased in order to obtain better reliability although 0.6 was acceptable here. If not so, the items' quality, not quantity, should be addressed for this domain in new version of the scale.

Finally, it is worth noting that there are some limitations in this study. First, the sample size of the study is not very large, which may also affect the results related to factor analysis (114 cases and 30 variables) and responsiveness. Second, the subjects in this study were only selected from hospital inpatients. Third, it takes 15–30 min to use, which possibly makes some patients feel burden. Moreover, some patients (8 cases) were not assessed at discharge because they were not in the wards when the study investigators visited the hospital at the scheduled assessment time (earlier discharge, transferred to other departments, etc.). The missing cases may affect responsiveness to some extent although it is reasonable to infer that these events all happened by chance. Further large-scale research is needed to assess the universality of the instrument for other settings and populations, such as outpatients in local clinics.

In summary, the QLICD-COPD can be used as a useful instrument with some strengths in measuring and assessing quality of life for patients with COPD who speak Chinese. The questionnaire is free to use on request by other researchers. It also needs further large-scale studies to confirm psychometric properties in different settings (outpatients, community etc.).

Acknowledgements

Author Contributions

WCH, ZY and YYB designed the study. ZZH, QP, WB performed the data collection and WCH,
QP performed data analyses, and all authors contributed to interpreting the data. WCH and ZY wrote the first draft, which was critically revised by all others. All authors have read and approved the final manuscript.

Declaration of conflicting interests

Funding

Ethics approval and consent to participate

The study protocol and the informed consent form were approved by the IRB (institutional review board) of Kunming Medical University (30860248, Kunming Medical University, 01/17/2009).

ORCID iD

Chonghua Wan https://orcid.org/0000-0002-7323-6837

Supplemental material

Supplemental material for this article is available online.

References

1. Pauwels RA and Rabe KF. Burden and clinical features of chronic obstructive pulmonary disease (COPD). *Lancet* 2004; 364: 613–620.
2. Rabe KF, Hurd S, Anzueto A, et al. Global initiative for chronic obstructive lung disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; 176(6): 532–555.
3. Peruzza S, Sergi G, Vianello A, et al. Chronic obstructive pulmonary disease (COPD) in elderly subjects: Impact on functional status and quality of life. *Respir Med* 2003; 97: 612–617.
4. Curtis JR. and Patrick DL. The assessment of health status among patients with COPD. *Eur Respir J(supp)* 2003; 41: 36–45.
5. Stahl E, Lindberg A, Jansson SA, et al. Health-related quality of life is related to COPD disease severity. *Health Qual Life Outcomes* 2005; 3: 56.
6. Rutten-van Molken MP, Oostenbrink JB, Tashkin DP, et al. Does quality of life of COPD patients as measured by the generic EuroQol five-dimension questionnaire differentiate between COPD severity stages? *Chest* 2006; 130: 1117–1128.
7. Pickard AS, Wilke C, Jung E, et al. Use of a preference based measure of health (EQ-5D) in COPD and asthma. *Respir Med* 2008; 102(4): 519–536.
8. Menn P, Weber N and Holle R. Health-related quality of life in patients with severe COPD hospitalized for exacerbations - comparing EQ-5D, SF-12 and SGRQ. *Health Qual Life Outcomes* 2010; 8: 39.
9. Garrido P, Diez J, Gutierrez J, et al. Negative impact of chronic obstructive pulmonary disease on the health-related quality of life of patients. Results of the EPIDEPOC study. *Health Qual Life Outcomes* 2006; 4: 31.
10. Cedano S, Belasco AG, Traldi F, et al. Influence that sociodemographic variables, clinical characteristics, and level of dependence have on quality of life in COPD patients on long-term home oxygen therapy. *J Bras Pneumol* 2012; 38(3): 331–338.
11. Mahler DA and Mackowiak JI. Evaluation of the short-form 36-item questionnaire to measure health-related quality of life in patients with COPD. *Chest* 1995; 107(6): 1585–1589.
12. Liang WM, Chen JJ, Chang CH, et al. An empirical comparison of the WHOQOL-BREF and the SGRQ among patients with COPD. *Qual Life Res* 2008; 17(5): 793–800.
13. Ferrer M, Villasante C, Alonso J, et al. Interpretation of quality of life scores from the St George's Respiratory Questionnaire. *Eur Respir J* 2002; 19(3): 405–413.
14. Camelier A, Rosa FW, Salim C, et al. Using the Saint George's respiratory questionnaire to evaluate quality of life in patients with chronic obstructive pulmonary disease: validating a new version for use in Brazil. *J Bras Pneumol* 2006; 32(2): 114–122.
15. Ringbaek T, Martinez G and Lange P. A comparison of the assessment of quality of life with CAT, CCQ, and SGRQ in COPD patients participating in pulmonary rehabilitation. *COPD* 2012; 9(1): 12–15.
16. Weldam SW, Schuurmans MJ, Liu R, et al. Evaluation of quality of life instruments for use in COPD care and research: a systematic review. *Int J Nurs Stud* 2013; 50(5): 688–707.
17. Malý M and Vondra V. Generic versus disease-specific instruments in quality-of-life assessment of chronic obstructive pulmonary disease. *Methods Inf Med* 2006; 45(2): 211–215.
18. Haave E, Hyland ME and Skumlien S. The relation between measures of health status and quality of life in COPD. *Chron Respir Dis* 2006; 3(4): 195–199.
19. Preston CC and Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychol (Amst)* 2000; 104(1): 1–15.
20. Nolte S, Liegl G, Petersen MA, et al. EORTC quality of life group. General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the Unites States. *Eur J Cancer* 2019; 107: 153–163. DOI: 10.1016/j.ejca.2018.11.024
21. Webster KA, O'Connor ML, Hansen AR, et al. Development of a functional assessment of chronic illness therapy item library and primary symptom list for the assessment of patient-reported adverse events associated with immune checkpoint modulators. *J Cancer Metastasis Treat* 2020; 6: 8. DOI: 10.20517/2394-4722.2019.38
22. Pakhale S, Wood-Dauphinee S, Spahija A, et al. Validation of a new questionnaire with generic and disease-specific qualities: the McGill COPD quality of life questionnaire. *Can Respir J* 2012; 19(6): 367–372.
23. 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(1): 93–104.
24. Wan C, Chen Y, Gao L, et al. Development and validation of the peptic ulcer scale under the system of quality of life instruments for chronic diseases based on classical test theory and generalizability theory. *BMC Gastroenterol* 2020; 1420(1): 422.

25. The WHOQOL Group. The word health organization quality of life assessment (WHOQOL): Devolvement and psychometric properties. *Soc Sci Med* 1998; 46(12): 1569–1585.

26. Anzueto A and Miravitlles M. Considerations for the correct diagnosis of COPD and its management with bronchodilators. *Chest* 2018; 154(2): 242–248.

27. Murphy DE and Panos RJ. Diagnosis of COPD and clinical course in patients with unrecognized airflow limitation. *Int J Chron Obstruct Pulmon Dis* 2013; 8: 199–208.

28. 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(10): 991–998.

29. Chassany O, Sagnier P, Marquis P, et al. Patient reported outcomes: the example of health-related quality of life: a European guidance document for the improved integration of health related quality of life assessment in the drug regulatory process. *Drug Inf J* 2002; 36: 209–238.

30. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002; 11: 195–203.

31. Hays RD and 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.

32. McGraw KO and Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; 1(1): 30–46.

33. 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.

34. 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.

35. Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000; 53(5): 459–468.

36. Winterstein BP, Willse JT, Kwapil TR, et al. Assessment of score dependability of the Wisconsin schizotypy scales using generalizability analysis. *Psychopathol Behav Assess* 2010; 32: 575–585.

37. Stora B, Hagtvet KA and Heyerdahl S. Reliability of observers' subjective impressions of families: a generalizability theory approach. *Psychother Res* 2013; 23(4): 448–463.

38. 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(3): 252–266.

39. Heitman RJ, Kovaleski JE and Pugh SF. Application of generalizability theory in estimating the reliability of ankle-complex laxity measurement. *J Athl Train* 2009; 44(1): 48–52.

40. Pakhale S, Wood-Dauphinee S, Spahija J, et al. Combining both generic and disease-specific properties: development of the McGill COPD quality of life questionnaire. *COPD* 2011; 8(4): 255–263.