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

chonghua wan (wanchh@hotmail.com)
Xincheng Road, Songshan Lake Science and Technology Industry Park, Dongguan 523808, Guangdong, P.R. China https://orcid.org/0000-0002-7323-6837

Ying Chen
Kunming Medical University

Li Gao
Taian City Central Hospital

Qingqing Zhang
Shandong Provincial Hospital

Peng Quan
Guangdong Medical University

Xiaoyuan Sun
Guangdong Medical University

Research article

Keywords: Quality of life, Peptic ulcer, Standardized response mean, Psychometric properties, Generalizability Theory

Posted Date: September 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-61801/v1

Version of Record: A version of this preprint was published on December 14th, 2020. See the published version at https://doi.org/10.1186/s12876-020-01562-y.

Abstract

Background: Quality of life (QOL) for patients with Peptic ulcer disease (PUD) is of interest worldwide and disease-specific instruments are needed for clinical research and practice. This paper focus on the development and validation of the PUD scale under the system of Quality of Life Instruments for Chronic Diseases (QLICD-PU) by the modular approach and both classical test theory and Generalizability Theory.

Methods: QLICD-PU is developed based on programmatic decision-making procedures, including multiple nominal and focus group discussions, in-depth interviews, and quantitative statistical procedures. Based on the data of 153 PUD inpatients, G and D studies using correlation analysis, factor analysis, t-test, and generalized theoretical analysis were used to assess the validity, reliability, and responsiveness of the scale.

Results: When SF-36 was used as the standard, correlation and factor analysis confirmed good structural validity and standard-related validity of QLICD-PU. Except for the social domain (0.62), the internal consistency α of all domains is higher than 0.70. The overall score and the retest reliability coefficients (Pearson r and intra-class correlation ICC) in all domains are higher than 0.80 (0.77 in the social domain). After treatment, the overall scores and scores of all domains have statistically significant changes ($P < 0.01$), except for social impact and sexual function scores. The SRM of field-level scores ranges from 0.34 to 1.03. The G coefficient and reliability index (Φ coefficient) further confirm the reliability of the scale through more accurate variance components and decision-making information about changes in the number of items.

Conclusions: QLICD-PU can be used as a useful measurement to assess the quality of life of PUD patients with good psychometric characteristics and multiple advantages.

Backgrounds

Peptic ulcer disease (PUD), usually in the stomach and duodenum, is a common disease in the world, and often recur [1-5]. Its annual incidence rate is 1.1-3.3%, the prevalence is 1.7-4.7% and about 10% of the people are suffered from this disease during their lifetime in the United States [1-2]. In Europe, about 10% of the population experience PUD in their lifetime [4-5]. Patients with PUD may have various gastrointestinal symptoms, including abdominal pain, vomiting, and even upper gastrointestinal bleeding and perforation, which is related to high mortality and high morbidity [6-7]. Considering that PUD can cause pain, bleeding, and gastrointestinal discomforts, such as fullness, nausea, anorexia, and certain limitations on social and mental health, therefore, assessing patient's health-related quality of life (HRQOL) is especially important [8,9]. Several studies have shown that the HRQOL of patients with PUD is significantly lower than that of the general population and that HRQOL improvement in PUD patients plays an important role in disease treatment [9,10]. It is hoped that the use of appropriate tools can improve the understanding of the treatment and service needs of PUD patients.

There are many HRQOL instruments, which can be divided into general measures and measures against diseases. The application of general and specific quality of life measures is different. General measures must be taken to compare the results of different populations and interventions. Specific measures are more sensitive to detect and quantify subtle changes that are important to clinicians or patients [11]. For the last few decades, although the measurement of general QOL has been improved using SF-36, PGWB (Psychological General Well-Being) index, etc., clinicians and researchers still need to determine the clinical significance of any measure of patients' response to treatment. Therefore, several HRQOL measures against PU have been developed for

diseases, such as QPD (quality of life in peptic disease)[12], QLDUP (Quality of Life in Duodenal Ulcer Patients) [13]. Besides, the QOLRAD (Quality Of Life in Reflux and Dyspepsia) [14,15], the FDDQL (Functional Digestive Disorder Quality of Life Questionnaire) [16] and the GSRS (Gastrointestinal Symptom Rating Scale) [17] can also be used for patients with PUD. Among them, QLDUP contains 54 items in 15 dimensions, which were developed by combining SF-36, PGWB index, and 13 disease-specific items obtained from interviews with patients and clinicians. QOLRAD is a 25-item questionnaire, each item is rated on a 7-point Likert scale and has 5 sub-scores.

However, these instruments are not developed based on popular modular methods (general/core modules and specific modules) [18,19]. Because diseases of the same disease category (such as digestive system diseases) have many common characteristics (such as symptoms and side effects), a widely used method for developing QOL for diseases in recent years is to use a general module of the entire disease to classify each disease separately. This approach can greatly reduce the time and effort of developing new instruments. For example, the European Research and Treatment Organization and the Quality of Life Questionnaire for Cancer Therapeutic Function Evaluation were developed based on this modular principle [18,19].

To the best of our knowledge, no disease-specific PUD devices have been developed based on a modular approach, let alone a combination of classic test theory (CTT) and generalizability theory (GT).To fill the gaps in QOL research and meet the needs of such research and clinical studies, we developed a set of Quality of Life Instruments for Chronic Diseases (QLICD) through a modular approach [20-23]. The system includes a general module (QLICD-GM), which can be used for all types of chronic diseases, as well as specific modules, only for related diseases [20-23]. For example, the coronary heart disease instrument (QLICD-CHD) is constructed by combining QLICD-GM with the specific module for coronary heart disease [21]. At present, QLICD (V1.0) includes 30 general-purpose modules QLICD-GM (3 domains and 10 facets) and 9 specific modules. The number of items ranges from 14 to 21, so 9 specific proportions are formed. Coronary heart disease (QLICD-CHD) At present, QLICD (V1.0) includes a 30-items general module QLICD-GM and 9 specific modules and thus forming 9 specific scales of coronary heart disease (QLICD-CHD)[21], hypertension (QLICD-HY)[22], irritable bowel syndrome (QLICD-IBS) [23], peptic ulcer disease (QLICD-PU), etc.

In the current research, we aimed to develop and validate the QLICD-PU instrument.

Methods

Development of the QLICD-PU

QLICD-PU consists of a general module QLICD-GM and a module dedicated to PUD. The development process of QLICD-GM has been described in another paper [20]. Here, we briefly summarize the development steps and results. The programmed develop procedures which include focus group discussions, in-depth interviews, pre-testing, and four quantitative statistical analyses were used in the QLICD-GM. Finally, the QLICD-GM has 30 items which included 3 domains and 10 facets. Based on the data of 620 patients with seven kinds of chronic diseases, QLICD-GM has good psychometrics (reliability, effectiveness, responsiveness), such as coronary heart disease and hypertension. [20].

For a specific module, 29 items reflecting symptoms were selected to constitute the initial item pool. These items focus on the unique side effects and mental health of PUD. We selected these items from literature reviews and nominal / focus group discussions. Focus groups evaluate the importance of each item by ranking each item

independently and then discussing the 9 lowest ranked items that are excluded. The remaining 20 items constitute a preliminary questionnaire for conducting the pilot test and also Interviews with 29 PUD patients and 14 clinicians and researchers with extensive experience. We focus on patient opinion, which is most important for assessing the acceptability of interventions and related compliance. Based on the pilot data, the items were re-screened using a development process similar to the generic module (statistical procedure and focus group discussion). The final specific module consists of 14 items, coded PU1-PU14 (see table 1 in detail), classified into 6 facets.

Validation of the QLICD-PU

Data Collection and Scoring

In this study, we enrolled participants with PUD at any stage who were: (1) be able to provide written informed consent; (2) be able to read and write words with assistance. There were no protocol requirements regarding specific clinical treatment of patients. Physicians could treat the patients according to what they deemed clinically appropriate.

The survey was carried out at the First Affiliated Hospital of Kunming Medical University after approved by the ethics committee of Kunming Medical University. Researchers, including doctors and medical graduate students, explained the purpose of the study and obtained informed consent before the test. Each interviewee was required to answer the questionnaire upon admission. To assess the reliability of the retest, a subsample is randomly selected for the second assessment on the second or second day of hospitalization. All patients available at the scheduled third evaluation time point have completed discharge measures to assess the responsiveness of the questionnaire. Besides, there is no recognized gold standard for evaluating PUD quality of life. To evaluate the standard correlation validity, convergence validity, and discriminant validity of QLICD-PU, the Chinese version of SF-36 [24] was also used in the formal test. Baseline socio-demographic characteristics were recorded from hospital medical records, including age, gender, education level, marital status, clinical history, and treatment. Each investigator checked the answers immediately to ensure their integrity.

Since each item uses the five-point Likert format (not many at all, many, many), positively stated items will be scored directly from 1 to 5, while negatively stated items will receive the opposite score. The domain/facet and overall scale scores are obtained by adding related item scores, all of which are linearly converted to standardized scores on a scale of 0-100. The higher the score of QLICD-PU means the better the quality of life of original and standardized scores.

Psychometric Analysis

Then the effectiveness, reliability, and responsiveness of QLICD-PU were evaluated. In this study, the structural effectiveness is evaluated by the Pearson correlation coefficient r between the item and the domain. Assess the validity of the standard by correlating the corresponding fields of QLICD-PU and SF-36. Multi-feature scale analysis [25] is used to test the convergence validity and discriminant validity of the item. There are two validity criteria: (1) When the item domain correlation is 0.40 or higher, it supports convergence validity; (2) The item domain correlation is higher than the discriminant validity of other domains. item In terms of reliability, for each domain/facet and the overall scale, the internal consistency is assessed using the first measurement data (at admission) by Cronbach's alpha coefficient. Evaluation of retest reliability was by Pearson correlation coefficient

and intra-class correlation (ICC) [26-27] between the first and second assessments. The responsiveness (sensitivity to detect change) was assessed by using a paired t-test to compare the average score change between the two assessments before and after treatment and the average value of the standardized response (SRM). [28-29].

Generalizability Theory Analysis

In addition to the classical test theory analysis, to study the reliability of the QLICD-PU score, we also applied the Generalizability Theory (GT) in this study. GT is a modern test theory developed based on the combination of classical test theory and analysis of variance. It is proposed as a method to improve measurement program design in an attempt to obtain reliable data [30-33]. To control the measurement errors, GT introduces independent variables or factors that interfere with test scores into measurement models, such as differences between research objects, item difficulty, scoring criteria, and the interaction between these factors. An analysis of variance was then used to assess the impact of these variables or factors on test scores, using the variance component as an index. GT includes G study and D study. G study quantified the amount of variance related to the different facets (factors) to be examined. D study provides information about which protocol is best for a particular measurement by generating a generalizability (G) coefficient, which can be interpreted as a reliability factor for all facets of the current study.

In our research, both G study and D study are completed in one measurement model to estimate the variance component and reliability factor, and to estimate the variance component and reliability factor in the one-sided cross design. [person-by-item (p × i) design]. We define the patient's quality of life as the measurement target and the item as a facet of measurement error. For G-Study, we defined an acceptable observation range composed of measurement objects and measurement errors and estimated variance components. For D-study, we define the allowable summary based on the measurement object and the measurement facet that the researchers are willing to summarize to express the measurement conditions. At the same time, the generalized coefficients of each facet and the variance components of the reliability indicators and their interactions are calculated.

Results

Socio-demographic characteristics of the sample

153 PUD patients range in age from 16 to 79, with a mean age of 45.2±14.8. 110 cases (71.9%) were male, and 134 (87.6%) were of Han ethnicity. 27 cases (17.6%) finished primary school, while 85 (55.5%) completed high school, and 40 (26.2%) had a university or graduate degree. In terms of occupation, workers accounted for 38.6% (59 cases), farmer 15.0% (23),cadre 12.4% (19), teacher 9.2% (14), and others 24.8% (38). For perceived Income, poor accounted for 30.7% (49 cases), fair 58.8% (90), and high 9.2% (14).

Construct validity

The Pearson correlation coefficient and factor analysis of the domain are used to evaluate the effectiveness of the construction. A correlation analysis from the data measured at the time of admission shows that there is a strong correlation between the items and their domain (most above 0.40). However, the relationship between the item and other domains is weak (for details, see Table 1). For example, the correlation coefficient between PHD and PH1-PH8 is between 0.49 and 0.76 (the first column in bold) are higher than those between PHD and other

items. Similarly, correlation coefficients between PSD and items of PS1- PS11 ranging from 0.44 to 0.72 (the second column in bold) are higher than the value between PSD and other items. Factor analysis was performed on common modules and specific modules. Extraction standard was set as criteria of eigenvalues >1, there were 8 main components extracted from 30 items of the general module (QLICD-GM), accounting for 63.88% of the cumulative variance. By using the Varimax rotation method, it can be seen that these eight main components reflect eight different facets under the three domains of the general module, wherein the 1st, 4th and 5th main components mainly represent the higher load psychological domains on PS1-PS11; the 2nd and 7th principal components largely reflecting the physical domain with higher loadings on PH1-PH8; the 3rd, 6th and 8th principal components generally depicting the social domain with higher loadings on SO1-SO11. Similarly, principal component factor analysis extracted 6 principal components from 14 items in a specific module, with a cumulative variance of 65.88%

Criterion-related validity

The correlation coefficients between the QLICD-PU and SF-36 domain scores were listed in Table 2, indicating that the correlation between the same and similar domains (bold in the table) is usually higher than different and dissimilar domains. For example, the coefficient between the physical domain of QLICD-PU and the physical function of SF-36 is 0.67, which is higher than any other coefficient in this row. Similarly, the coefficient between the psychological domain of QLICD-PU and the mental health of SF-36 is 0.51, higher than any other coefficient in the row

Reliability

As shown in Table 3, the Cronbach's α for these four domains were higher than 0.70 except for SOD (0.62), while they were ranging from 0.35 to 0.81 at facets level.

In the second evaluation (two-day follow-up), data from 63 patients were used for retest reliability analysis. The test-retest correlation coefficients for the 4 domains and the overall were larger than 0.80 except for SOD (0.77), while they were ranging between 0.72-0.94 at facets level. The ICC result calculated according to the definition of absolute consistency is very similar to the Pearson correlation coefficient.

Results from Generalizability Theory

G-Studies evaluated the variance components of the four domains of QLICD-PU (see Table 4), and 153 patients completed 44 quality of life instruments. As can be seen from Table 4, for the four domains of physical, psychological, social and specific domains, the biggest source of change is the interaction between people, ranging from 55.68% to 81.89%while variances accounted for by person were the second for three domains of physical, psychological and social which range from 12.20% to 35.92%.

A D study was conducted to estimate the generalization coefficient (G coefficient) and reliability index (Φ coefficient) of the four domains of QLICD-PU currently designed by p × i. (physical domain includes 8 items, the psychological domain includes 11 items**,** the social domain includes 11 items and specific domain includes 14 items), as well as alternative designs with different numbers of items (see Table 5).

Responsiveness

The data from 135 patients who completed the questionnaire after treatment were used to assess responsiveness. The paired t-test and the response index SRM were used to check the average score change of each domain/facet of QLICD-PU before and after treatment. The results are shown in Table 6. It can be seen that except for social impact and sexual function, all domains/facets and overall scale have undergone major changes ($p < 0.01$), with SRM ranging from 0.04 to 1.03 and domain-level SRM from 0.34 to 1.03 scores.

Discussion

The focus of this study was to develop and validate a special QOL instrument QLICD-PU for peptic ulcer disease. We use a method that combines a general module with a specific module for a specific disease to capture common features within the disease category and the differences between specific diseases[18-20]. We adopted a modular approach to systematically and effectively develop new instrument systems for chronic diseases, in which the general module QLICD-GM is used for various chronic diseases, and QLCID-PU is only for a specific scale of PUD. This method uses the same general module and similar structure to unify all QLICD specific disease tools. Compared with existing instruments, QLICD-PU has several advantages [20-23]. First, it can compare the QOL of various diseases through a general module, and can also capture symptoms and side effects through a specific module. We can use QLICD-GM to capture general QOL in patients with different diseases. We can also employ QLICD-PU and QLICD-CG to capture differences in QOL in PUD and chronic gastritis patients. Secondly, it consists of a moderate number of items with a clear hierarchy (item → facet → domain → overall), so the mean can be calculated to detect detailed changes, not only at the domain level (4 domains) but also at the facet level (16 facets). Users can choose one or two levels for research at their convenience. Third, the most important value of QLICD-PU is the profound Chinese cultural background behind it. For example, Chinese culture focuses on family relations and pedigree, diet, temperament, and noble spirit, all of which are reflected in QLICD-PU through items that focus on appetite, sleep, energy, and family support.

Practical QOL instruments require excellent psychometric characteristics, including validity, reliability, and responsiveness. Validity is the degree to which the tool can capture what it claims to measure. Follow WHO's definition of quality of life [34] and systematic development procedures, we developed QLICD-PU for PUD patients through focus group discussions, in-depth interviews, and pre-tests to effectively reduce the number of items. The final version includes 30 of the original 73 items of the general module [20], and it has been reduced to 14 items from the first 29 items of specific modules. This process helps us achieve good content validity and the correct tool concept structure. Correlation and factor analysis are used to confirm the construct. Correlation analysis shows that the relationship between items and their domains/facets is strong, but the relationship between items and other domains/facets is weak. Factor analysis shows that the components extracted from the data are consistent with the theoretical structural framework of the instrument. These results confirm evidence supporting the good construct validity. The correlation coefficients between the QLICD-PU and SF-36 domain scores demonstrated that the criterion-related validity and construct validity (the convergent and divergent validity) are both high.

Reliability refers to the repeatability or consistency of item ratings in different assessments. In this study, internal consistency reliability (Cronbach's α), retest reliability (Pearson r) and ICC were applied. Our results show that, except for the social function domain (0.62), the internal consistency coefficient and the overall score of the QLICD-PU domain are both greater than 0.70. The retest reliability coefficient of the overall score is 0.89, while the retest reliability coefficient of the QLICD-PU domain is greater than 0.80 (except for social function domain)

(0.77). Taking into account that the internal consistency coefficient is greater than 0.70 and the retest reliability coefficient is greater than 0.80 is considered satisfactory, these results indicate that the instrument has good reliability overall.

Responsiveness (sensitivity to detect changes) is the most important function of the QOL scale in clinical applications. There are two types of assessment methods: internal and external [28, 29]. In this study, we used a paired t-test to focus on internal responses to compare the average response before and after treatment. We used SRM as a responsiveness indicator, with 0.20, 0.50, and 0.80 representing small, moderate, and large responsiveness [28,29]. After treatment, QOL scores in all domains were significantly changed, the overall score ($p < 0.001$), SRM except for the social function field (0.34) were greater than 0.70, indicating that QLICD-PU has good responsiveness.

In addition to classical test theory analysis, this study also applies generalization theory. The reliability indicators are usually lower than the G factor because in addition to the interaction of the G factor, they also consider the main error effects. This study not only introduced the G coefficient and Φ but also introduced their changes when the item changed. For the physical and psychological fields, we estimate that the currently designed G coefficients are 0.787 and 0.830 and the credibility index is 0.755 and 0.815, respectively. It can be considered that it meets the 0.70 standards. For the social field, the current design G-factor is estimated to be 0.622, and the reliability index is 0.605, which is lower than the acceptable 0.70. Therefore, the items of this domain need to be improved. For an alternative design with 17 items, the G coefficient is estimated to be 0.717 and the reliability index is 0.703, which will satisfy acceptable reliability. For a specific domain, the G coefficient of the current design is estimated to be 0.626, and the reliability index is 0.580, which is also lower than the acceptable 0.70. Similarly, the G-coefficient estimated to be 0.742 and the index of dependability 0.703 when an alternative design with 24 items. Therefore, these analyses suggest that the number of items of the social domain needs to be increased from 11 to 17, and the specific domain from 11 to 24 to reach acceptable dependability. However, it may not be practical to increase the length of the test in practice, as reliability is reduced if the subject is required to complete too many items at the same time. Researchers or instrument users can decide to add items or tolerate reasonable low reliability.

Conclusions

In conclusion, QLICD-PU can be used as a useful tool for assessing the quality of life of patients with PUD, with good psychometric characteristics and many advantages. The analysis from Generalizability theory not only confirmed the reliability of the scale as a whole, but it also provides more information than CTT. However, the number of social and domain-specific items should be increased to increase reliability. Besides, the quality of items in these 2 domains should also be addressed.

List Of Abbreviations

QOL:Quality of life; HRQOL:Health-Related Quality Of Life;PUD :Peptic ulcer disease;PGWB: Psychological General Well-Being; QPD:quality of life in peptic disease; QLDUP:Quality of Life in Duodenal Ulcer Patients; QOLRAD :Quality Of Life in Reflux and Dyspepsia ;FDDQL:Functional Digestive Disorder Quality of Life Questionnaire; GSRS : Gastrointestinal Symptom Rating Scale; QLICD: Quality of Life Instruments for Chronic Diseases; ICC

:Intra-Class Correlation;SRM:Standardized Response Mean; CTT: Classical Test Theory; GT: Generalizability theory.

Declarations

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).

Consent to publish

The authors understand and agree to publish.

Availability of data and materials

The data can be available by request.

Competing interests

The authors declare that they have no competing interests.

Funding

The paper is supported by the National Natural Science Foundation of China (71373058, 30860248), Science and Technological Planning Program of Guangdong Province(2013B021800074). The funding bodies provided funds to support project development. The grant recipient (Chonghua Wan) designed the study, performed the data collection and data analyses, and extensively revised the manuscript.

Authors' contributions

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

Acknowledgments

In carrying out this research item, we have received substantial assistance from Liping Duan, Hongying Li and Wu Li at the first affiliated hospital of Kunming Medical University and Bin Wu and Fuzhen Liu at the affiliated hospital of Guangdong Medical University. We sincerely acknowledge all the support.

References

1. NIH Consensus Development Panel on Helicobacter pylori in Peptic Ulcer Disease: Helicobacter pylori in peptic ulcer disease. JAMA 272:65–69, 1994
2. Sonnenberg A, Everhart JE: Health impact of peptic ulcer in the United States. Am J Gastroenterol 92:614–620, 1997

3. Baghianimoghadam MH, Mohamadi S, Baghianimoghadam M, Falahi A, Roghani HS: Survey on quality of life related factors in patients with peptic ulcer based on PRECEDE model in Yazd, Iran. J Med Life. 2011;4(4):407-11
4. Ehlin AG, Montgomery SM, Ekbom A, Pounder RE, Wakefield AJ. Prevalence of gastrointestinal diseases in two British national birth cohorts. Gut. 2003;52(8):1117-21.
5. Suadicani P, Hein HO, Gyntelberg F. Genetic and life-style determinants of peptic ulcer. A study of 3387 men aged 54 to 74 years: The Copenhagen Male Study. Scand J Gastroenterol. 1999;34(1):12-7.
6. van Leerdam ME, Vreeburg EM, Rauws EA, Geraedts AA, Tijssen JG, Reitsma JB, Tytgat GN. Acute upper GI bleeding: did anything change? Time trend analysis of incidence and outcome of acute upper GI bleeding between 1993/1994 and 2000. Am J Gastroenterol. 2003;98(7):1494-9.
7. Lau JY, Sung J, Hill C, Henderson C, Howden CW, Metz DC. Systematic review of the epidemiology of complicated peptic ulcer disease: incidence, recurrence, risk factors and mortality. Digestion. 2011;84(2):102-13.
8. Mokrowiecka A, Jurek K, Pińkowski D, Małecka-Panas E. The comparison of Health-Related Quality of Life (HRQL) in patients with GERD, peptic ulcer disease and ulcerative colitis. Adv Med Sci. 2006;51:142-7.
9. Barkun A, Leontiadis G: Systematic review of the symptom burden, quality of life impairment and costs associated with peptic ulcer disease. Am J Med. 2010;123(4):358-66.
10. Zboralski K, Florkowski A, Talarowska-Bogusz M, MacanderM, Gałecki P. Quality of life and emotional functioning in selected psychosomatic disease. Postepy Hig Med Dosw (Online). 2008;62:36-41.
11. Guyatt GH, Naylor CD,Juniper E,Heyland DK,Jaeschke R,Cook DJ. Users' guides to the medical literature, XII: how to use articles about health-related quality of life. JAMA.1997;277:1232-1237.
12. De Carli G, Irvine SH, Arpinelli F, Bamfi F, Olivieri A, Recchia G. Development and validation of QPD 32, a specific questionnaire for measuring the quality of life of patients with peptic ulcer. Minerva Gastroenterol Dietol.1995;41(4):275-82.
13. Martin C, Marquis P, Bonfils S. A 'quality of life questionnaire' adapted to duodenal ulcer therapeutic trials. Scand J Gastroenterol Suppl. 1994;206:40-3
14. Wiklund IK, Junghard O, Grace E, Talley NJ, Kamm M, Veldhuyzen van Zanten S,Paré P, Chiba N, Leddin DS, Bigard MA, Colin R, Schoenfeld P. Quality of Life in Reflux and Dyspepsia patients. Psychometric documentation of a new disease-specific questionnaire (QOLRAD). Eur J Surg Suppl. 1998;(583):41-9.
15. Crawley J, Frank L, Joshua-Gotlib S, Flynn J, Frank S, Wiklund I. Measuring change in quality of life in response to Helicobacter pylori eradication in peptic ulcer disease: the QOLRAD. Dig Dis Sci. 2001;46(3):571-80.
16. Chassany O, Marquis P, Scherrer B, Read NW, Finger T, Bergmann JF. Validation of a specific quality of life questionnaire for functional digestive disorders. *Gut* 1999;44:527–33.
17. 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;33:129-34.
18. Sprangers MA, Cull A, Groenvold M, Bjordal K, Blazeby J, Aaronson NK. 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(4):291-300.

19. 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(12 Suppl 2): S10-S17.
20. Wan CH, Tu XM, Messing S, Li XM, Yang Z, Zhao XD, Gao L, Yang YP, Pan JH, Zhou ZF. 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.
21. Wan C, Li H, Fan X, Yang R, Pan J, Chen W, Zhao R. 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. doi:10.1186/1477-7525-12-82.
22. Wan CH, Jiang RS, Tu XM, Tang W, Pan JH, Yang RX, Li XM, Yang Z, Zhang XQ. 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(4):465-480.
23. Lei P, Lei G, Tian J, Zhou Z, Zhao M, Wan C. 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(10):1245-1255. doi:10.1007/s00384-014-1976-x.
24. Yang Z, Li W , Tu XM, Tang W, Messing S, Duan LP. 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.
25. 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.
26. Ren S, Yang S, Lai S. Intraclass correlation coefficients and bootstrap methods of hierarchical binary outcomes. Stat Med. 2006;25(20):3576-3588.
27. 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.
28. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. Qual Life Res 2003;12(4):349-363.
29. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes.J Clin Epidemiol. 2008;61(2):102-109.
30. 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.
31. Stora B, Hagtvet KA, Heyerdahl S. Reliability of observers' subjective impressions of families: A generalizability theory approach. Psychother Res. 2013 ;23(4):448-463.
32. 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.
33. Heitman RJ, Kovaleski JE, Pugh SF. Application of generalizability theory in estimating the reliability of ankle-complex laxity measurement. J Athl Train. 2009;44(1):48-52.
34. The WHOQOL Group. The Word Health Organization Quality of Life assessment (WHOQOL): Devolvement and psychometric properties. Sci. Med. 1998;46(12): 1569-1585.

Tables

Table 1 Correlation coefficients r among items and domains of QLICD-PU (n=153)

Code	Items brief description	Physical	Psychological	Social	Specific
PH1	Take care of daily life (e.g. eating)?	**0.70**	0.14	0.34	0.06
PH2	Felt easily fatigued?	**0.65**	0.33	0.21	0.19
PH3	Have trouble walking 800m or more?	**0.76**	0.12	0.24	0.04
PH4	Have trouble going up and down stairs?	**0.71**	0.32	0.20	0.04
PH5	Need to take medication?	**0.65**	0.33	0.23	0.19
PH6	A good appetite?	**0.49**	0.10	0.17	0.09
PH7	Satisfied with your sleep?	**0.55**	0.20	0.14	0.15
PH8	Felt pain or uncomfortable?	**0.52**	0.26	0.24	0.43
PS1	Memory and concentration affected?	0.39	**0.46**	0.40	0.16
PS2	Felt mentally miserable?	0.35	**0.59**	0.32	0.28
PS3	Felt lonely and helpless?	0.14	**0.69**	0.42	0.19
PS4	Felt pessimism and despair?	0.18	**0.72**	0.37	0.26
PS5	Worried about disease?	0.31	**0.72**	0.27	0.39
PS6	Felt fretful or irritable?	0.16	**0.56**	0.24	0.31
PS7	Felt nervous and anxious?	0.26	**0.68**	0.20	0.31
PS8	Stop medication because of side effects?	0.10	**0.44**	0.05	0.14
PS9	To be a burden to the family?	0.36**0.55**	0.37	0.23	
PS10	Felt self-abasement because of disease?	0.03	**0.69**	0.20	0.23
PS11	Hidden emotions but could not forget?	0.11	**0.68**	0.29	0.27
SO1	Interfered with work/housework?	0.29	0.35	**0.53**	0.18
SO2	Family roles?	0.16	0.03	**0.38**	0.10
SO3	Decreased caring and attention to family?	0.28	0.36	**0.34**	0.20
SO4	Good relations with family?	0.01	0.09	**0.56**	0.04
SO5	Help and support from family?	0.06	0.03	**0.52**	0.08
SO6	Affected participating in leisure activities?	0.18	0.35	**0.41**	0.07
SO7	Treat illness positively and optimistically?	0.23	0.30	**0.63**	0.23
SO8	Treatments received good for curing?	0.12	0.22	**0.35**	0.16
SO9	Economic problems caused by illness?	0.11	0.35	**0.48**	0.27
SO10	Support from friends and relatives?	0.13	0.01	**0.51**	0.06
SO11	Affected sexual activities?	0.27	0.27	**0.35**	0.16

PU1	Have pain (sore) in epigastria?	0.28	0.11	0.11	**0.53**
PU2	Have heartburn in epigastria?	0.08	0.17	0.14	**0.49**
PU3	Have pain/discomfort at night or hungry?	0.17	0.08	0.05	**0.51**
PU4	Pain/uncomfortable relieved after dinner?	0.13	0.00	0.07	**0.33**
PU5	Have acid regurgitation?	0.06	0.17	0.05	**0.48**
PU6	Have any belch (burps)?	0.19	0.12	0.08	**0.49**
PU7	Abdominal distension?	0.03	0.07	0.15	**0.53**
PU8	Salivate (flow saliva)?	0.06	0.26	0.00	**0.38**
PU9	Move bowels normal?	0.27	0.10	0.19	**0.34**
PU10	Upset/distress for gastroscopy inspection?	0.13	0.21	0.05	**0.36**
PU11	Vexed for food limit?	0.05	0.30	0.12	**0.43**
PU12	Troubled/limit by dine at fix time?	0.14	0.18	0.14	**0.35**
PU13	Worried about causing severe disease?	0.04	0.36	0.22	**0.51**
PU14	Vexed for often taking stomach medications?	0.01	0.31	0.17	**0.55**

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

Table 2 Correlation Coefficients among domain scores of QLICD-PU and SF-36 (n=153)*

QLICD-PU	SF-36									
	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
PHD	**0.67**	0.26	0.26	0.37	0.31	0.18	0.16	0.19	0.61	0.30
PSD	0.23	0.21	0.21	0.40	0.30	0.38	0.27	**0.51**	0.38	0.51
SOD	0.37	0.20	0.31	0.47	0.33	**0.54**	0.23	0.31	0.53	0.47
SPD	0.14	0.18	**0.33**	0.29	0.12	0.24	0.20	0.23	0.31	0.25
TOT	0.49	0.30	0.39	0.56	0.37	0.39	0.31	0.49	**0.64**	**0.54**

PHD: physical domain, PSD: psychological domain, SOD: social domain, SPD: specific domain, TOT: total score, 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-PU (n=153 for *α*, n=63 for *r* and ICC)

Domains/facets (number of items)	Internal consistency coefficient *α*	Test-retest coefficient *r*	Test-retest ICC (95% CI)
Physical Function PHD (8)	**0.79**	**0.89**	**0.89(0.83-0.93)**
Independence (3)	0.81	0.88	0.88(0.82-0.93)
Appetite and Sleep (2)	0.49	0.74	0.74(0.60-0.83)
Physical Symptoms (3)	0.60	0.80	0.80(0.69-0.87)
Psychological Function PSD(11)	**0.83**	**0.91**	**0.91(0.86-0.95)**
Cognition (2)	0.53	0.83	0.82(0.72-0.89)
Anxiety (3)	0.69	0.83	0.83(0.74-0.89)
Depression (3)	0.79	0.90	0.90(0.83-0.94)
Self-Consciousness (3)	0.70	0.94	0.94(0.90-0.96)
Social Function SOD(11)	**0.62**	**0.77**	**0.77(0.65-0.86)**
Social Support/Security (6)	0.73	0.84	0.84(0.75-0.90)
Social Effects (4)	0.63	0.74	0.73(0.59-0.83)
Sexual Function (1)	-	0.84	0.84(0.74-0.90)
Specific domain SPD (14)	**0.73**	**0.80**	**0.80(0.69-0.87)**
Upper abdomen pain(4)	0.44	0.72	0.72(0.57-0.82)
Acid regurgitation/salivation(2)	0.35	0.86	0.86(0.77-0.91)
Hiccup (1)	-	0.84	0.84(0.75-0.90)
Flatulent (1)	-	0.83	0.83(0.73-0.89)
Changing in stool habit (1)	-	0.77	0.77 0.64-0.85
Effects of mental and life(5)	0.63	0.73	0.72(0.58-0.82)
Total TOT(44)	**-**	**0.89**	**0.89(0.82-0.93)**

- not acceptable/suitable

Table 4 The estimated variance components and percentage of variance accounted for by effects (percent) for p×i design in G-study for four domains of quality of life instrument QLICD-PU ()

	p(person)		*i*(item)		*p* i*(person*item)	
Domain	Variance component	Percent (%)	Variance component	Percent (%)	Variance component	Percent (%)
PHD()	0.517	27.80	0.224	12.04	1.119	60.16
PSD()	0.547	35.92	0.128	8.40	0.848	55.68
SOD()	0.190	12.20	0.092	5.91	1.275	81.89
SPD()	0.143	8.99	0.257	16.15	1.191	74.86

PHD: physical domain, PSD: psychological domain, SOD: social domain, SPD: specific domain ,

p: person effect, i: item effect, p × i: person-by-item interaction effect

Table 5 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 QLICD-PU

Domain	Number of items								
Physical	6	0.517	0.037	0.186	0.186	0.224	0.042	0.735	0.698
domain	**8**	**0.517**	**0.028**	**0.140**	**0.140**	**0.168**	**0.032**	**0.787**	**0.755**
	9	0.517	0.025	0.124	0.124	0.149	0.029	0.806	0.776
	11	0.517	0.020	0.102	0.102	0.122	0.024	0.836	0.809
Psychological domain	9	0.338	0.010	0.095	0.095	0.105	0.013	0.780	0.763
	11	**0.338**	**0.007**	**0.069**	**0.069**	**0.076**	**0.010**	**0.830**	**0.815**
	13	0.338	0.006	0.059	0.059	0.065	0.009	0.852	0.839
	15	0.338	0.005	0.051	0.051	0.056	0.008	0.869	0.858
Social	9	0.190	0.010	0.142	0.142	0.152	0.012	0.573	0.556
domain	**11**	**0.190**	**0.008**	**0.116**	**0.116**	**0.124**	**0.010**	**0.622**	**0.605**
	13	0.190	0.007	0.098	0.098	0.105	0.009	0.660	0.644
	16	0.190	0.006	0.080	0.080	0.085	0.008	0.705	0.690
	17	0.190	0.005	0.075	0.075	0.080	0.007	0.717	0.703
	27	0.190	0.003	0.047	0.047	0.051	0.005	0.801	0.790
	29	0.190	0.003	0.044	0.044	0.047	0.005	0.812	0.801
Specific	11	0.143	0.023	0.108	0.108	0.132	0.025	0.568	0.520
domain	**14**	**0.143**	**0.018**	**0.085**	**0.085**	**0.103**	**0.020**	**0.626**	**0.580**
	17	0.143	0.015	0.070	0.070	0.085	0.016	0.671	0.626
	20	0.143	0.013	0.060	0.060	0.072	0.014	0.705	0.663
	24	0.143	0.011	0.050	0.050	0.060	0.012	0.742	0.703
	34	0.143	0.008	0.035	0.035	0.043	0.009	0.803	0.770
	41	0.143	0.006	0.029	0.029	0.035	0.007	0.831	0.802

is the variance components of relative error

is the variance components of absolute error

is the variance components of error when estimating the universe score by using sample mean

is the Generalizability coefficient

is the index of dependability

Table 6 Responsiveness of the quality of life instrument QLICD-PU (n=135)

Domains/facets (number of items)	Before treatment Mean SD		After treatment Mean SD		Differences Mean SD		*t*	*p*	SRM
Physical Function (8)	**58.70**	**19.53**	**71.57**	**13.06**	**-12.87**	**17.37**	**-8.61**	**<0.001**	**0.74**
Independence (3)	71.23	27.41	82.35	16.48	-11.11	23.39	-5.52	<0.001	0.47
Appetite and Sleep (2)	43.24	25.59	58.33	21.87	-15.09	24.39	-7.19	<0.001	0.62
Physical Symptoms (3)	56.48	22.11	69.63	15.75	-13.15	21.07	-7.25	<0.001	0.62
Psychological Function(11)	**75.35**	**16.15**	**83.84**	**13.85**	**-8.48**	**11.78**	**-8.37**	**<0.001**	**0.72**
Cognition (2)	68.61	22.85	81.11	17.80	-12.50	23.76	-6.11	<0.001	0.53
Anxiety (3)	69.57	22.24	84.81	17.57	-15.25	20.09	-8.82	<0.001	0.76
Depression (3)	82.47	20.70	87.35	14.71	-4.88	14.27	-3.97	<0.001	0.34
Self-Consciousness (3)	78.52	19.09	81.17	16.51	-2.65	10.95	-2.82	0.006	0.24
Social Function (11)	**67.41**	**14.21**	**71.16**	**12.92**	**-3.75**	**10.99**	**-3.97**	**<0.001**	**0.34**
Social Support/Security (6)	69.69	19.73	76.60	15.04	-6.91	13.64	-5.89	<0.001	0.51
Social Effects (4)	64.44	22.08	63.75	21.27	0.69	18.00	0.45	0.655	0.04
Sexual Function (1)	65.56 30.67	68.15	27.90	-2.59	21.22	-1.42	0.158	0.12	
Sub-Total (QLICD-GM) (30)	**68.00**	**12.56**	**75.92**	**10.90**	**-7.92**	**10.02**	**-9.19**	**<0.001**	**0.79**
Specific domain(14)	**62.72**	**11.86**	**74.44**	**12.14**	**-11.72**	**11.43**	**-11.91**	**<0.001**	**1.03**
Upper abdomen pain(4)	53.47	17.90	68.43	11.66	-14.95	19.60	-8.86	<0.001	0.76
Acid regurgitation/salivation(2)	75.83	21.28	86.94	18.19	-11.11	19.77	-6.53	<0.001	0.56
Hiccup (1)	70.00	28.78	85.56	19.19	-15.56	26.43	-6.84	<0.001	0.59
Flatulent (1)	60.19	29.97	80.00	23.22	-19.81	27.67	-8.32	<0.001	0.72
Changing in stool habit (1)	48.89	28.95	54.63	24.66	-5.74	27.14	-2.46	0.014	0.21
Effects of mental and life(5)	66.70	18.66	74.89	19.24	-8.19	13.90	-6.84	<0.001	0.59
Total (44)	**66.32**	**10.73**	**75.45**	**10.36**	**-9.13**	**9.38**	**-11.31**	**<0.001**	**0.97**

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- EQLICDPUV1.0.doc