• Review

Literature review and analysis of the development of health outcomes assessment instruments in Chinese medicine

Feng-bin Liu[1], Zheng-kun Hou[1], Yun-ying Yang[1], Zheng-zheng Zhang[1], Di Xie[1], Nelson Xie[2], Hong Thach Nguyen[3]

1. Department of Spleen-Stomach Diseases, the First Affiliated Hospital, Guangzhou University of Chinese Medicine, Guangzhou 510405, Guangdong Province, China
2. School of Medical and Molecular Biosciences, University of Technology Sydney, Ultimo 2007, New South Wales, Australia
3. Will of East Medical Ha Noi Viet Nam, Ministry of Health of Viet Nam, Ha Noi 010000, Viet Nam

OBJECTIVE: To evaluate the development of health outcomes assessment instruments in Chinese medicine.

METHODS: A comprehensive literature search for all published articles in China National Knowledge Infrastructure Database, Chongqing VIP Database and WANFANG Data was conducted. The studies that met the inclusion and exclusion criteria were used to extract information according to a predesigned assessment instrument.

RESULTS: A total of 97 instruments for health outcome assessment in Chinese medicine were identified. Of these questionnaires, 7 were generic, 12 were condition-specific and 78 were disease-specific. All instruments were suitable for adults, children, and both men and women. These instruments aimed to evaluate the health-related quality of life, signs and symptoms as well as patient satisfaction and doctor-reported outcome. However, the descriptions were poorly constructed for some of the most basic parameters, such as the domains and items, administrative mode, response options, memory recall periods, burden evaluation, format, copyright, content validity, and other properties.

CONCLUSION: The instrument development for health outcomes assessment in Chinese medicine is increasing rapidly; however, there are many limitations in current methodologies and standards, and further studies are needed.

KEYWORDS: health outcomes; health-related quality of life; patient-reported outcome; satisfaction; questionnaire; Chinese medicine; outcome assessment (health care)

DOI: 10.3736/jintegrmed2013018

Liu FB, Hou ZK, Yang YY, Zhang ZZ, Xie D, Xie N, Nguyen HT. Literature review and analysis of the development of health outcomes assessment instruments in Chinese medicine. *J Integr Med*. 2013; 11(2): 80-89.

Received August 26, 2012; accepted December 7, 2012.

Correspondence: Feng-bin Liu, MD, Professor; Tel: +86-20-36591109; E-mail: liufb163@163.com. Feng-bin Liu and Zheng-kun Hou contributed equally to this paper.

1 Introduction

Health outcome is the measurement of a change in the health of an individual, a group of people or population, which is attributable to an intervention or series of interventions[1]. Improved health outcome is the core concern and ultimate purpose of all interventions. Traditional Chinese medicine (TCM) has typically focused more on direct reports from patients and doctors, in contrast to Western medical practice, which mainly focuses on laboratory indicators. As contemporary medical practice shifts toward a "physiological-psychological-social" model, the health outcome evaluation system has become increasingly

complicated, including not only laboratory indicators, but also patient-, doctor-, and observer-reported outcomes[2] that place great weight on subjective feelings of participants. This inherent property is consistent with the holism and individualization in diagnosis and treatment principles of TCM. Therefore, many experts believe that the introduction and developments of patient-reported outcomes (PROs), clinician-reported outcomes (ClinROs) and observer-reported outcomes (ObsROs) are the key factors to overcome the bottleneck in globalization and modernization of TCM. In 1994, PRO and instruments were first introduced in TCM[3] and have since been widely recognized and used. Currently, PRO, ClinRO, ObsRO and instrument research in China are rapidly improving, and the prerequisite of using instrument development in all clinical areas has attracted many researchers. This paper aims to show the advances in health outcomes assessment instrument developments in TCM comprehensively and analyze possible problems.

2 Methods

2.1 Concepts and terminologies

"Chinese medicine category" indicates that a research project was designed and performed under the guidance of TCM theories, or TCM experts participated in the conceptual and methodological development of the project. The limitation would be achieved by limiting research field during conducting search strategies. Health outcome is defined as above[1]. Instrument refers to means of capturing data (i.e., questionnaire) plus all the information and documentation that support its use. The questionnaire must be well validated. To gain a comprehensive understanding of the state of current research, this study also included questionnaires without property evaluation.

2.2 Search strategies

2.2.1 Sources, formulas and limits in search

Literature search was performed in: *i*) China National Knowledge Infrastructure (CNKI) with the formula of "(AB = (量 + 问 + 调查)*(表 + 卷) + 工具 + 条目) and (TI= 制 *(定 + 编 + 研 + 订) + 评 *(考 + 价 + 定) + 常模 + 测试 + 检验 + 研究 + 分析 + 应用 + 筛选 + 修订 + 建立 + 设计 + (信 + 效 + 反应)* 度)". Subject domain was limited to be within the three sub-libraries of "Chinese medicine", "Chinese pharmacology" and "integrated traditional Chinese and Western medicine"; the date range was "1915 to 2012". *ii*) Chongqing VIP Information (CQVIP) with the formula of "(T = (量 + 问 + 调查)*(表 + 卷) + 工具 + 条目)*(C=R2)"; the date range was "1989 to 2012". *iii*) WANFANG Data with the formula of "Title = (量表 or 问卷 or 调查表 or 工具 or 条目) (制定 or 编制 or 研制 or 制订 or 考评 or 评价 or 评定 or 常模 or 测试 or 检验 or 研究 or 分析 or 应用 or 筛选 or 修订 or 建立 or 设计 or 信度 or 效度 or 反应度) 分类号 : 'R2*'". The date range was "1980 to 2012". The latest search was performed on January 20th, 2012.

2.2.2 Inclusion criteria

Literature included must meet all the following conditions: (1) belonging to the "Chinese medicine research" subheading; (2) containing a health outcome assessment instrument; (3) the instrument can be used for health outcome assessment even if it was not its primary intention.

2.2.3 Exclusion criteria

Literature was excluded if the study instruments met any of the following conditions: (1) designed for specific purposes, including disease diagnosis, physical measurements, syndrome description, disease screening and epidemiology survey; (2) simple theoretical or educational research; (3) for the application and/or promotion of instruments; (4) for modifying and/or adapting from existing instruments; (5) analyses on necessity of instrument development or protocols for future instruments; (6) duplicate publications or duplicate articles from different databases; (7) full-text is unobtainable.

2.3 Data management

This study included four steps: (1) Independent searches and selection of literature according to the stated inclusion and exclusion criteria were performed by two researchers (Yun-ying Yang and Zheng-zheng Zhang). If there was dispute between the two researchers, a third-party expert (Feng-bin Liu) was consulted until there was an unanimous decision. (2) The information extraction form (IEF) was designed by Zheng-kun Hou and revised by Feng-bin Liu. (3) Abiding by the IEF, data were extracted and entered separately by two researchers (Di Xie and Hong Thach Nguyen) with the software EpiData 3.02. After correcting for errors, the data were ready for analysis. (4) Software SPSS 11.0 was used to analyze the data. The aim was to conduct a qualitative research that focuses on data description and frequency analyses.

3 Results

3.1 Literature search results

Using the above method, 5 242, 604 and 347 records were obtained from "CNKI", "CQVIP", and "WANFANG Data" databases, respectively. After a strict selection with predefined inclusion and exclusion criteria, 97 articles that included health outcome assessment instruments of TCM were identified[4-100].

3.2 Basic characteristics of TCM instruments

3.2.1 Types of instruments

Seven general instruments that aimed to evaluate the health-related quality of life (HRQL) and signs and symptoms of participants were identified[30-32,64,73,81,94]. Twelve condition-specific instruments focused on qualities and/or outcomes of sub-health conditions (a relatively

modern term describing a state of imperfect health), aging, irritability or excessive anxiety, fatigue, post-surgery condition and emotions[7,10,11,16,22,33,52,56,70,71,95,99]. A total of 78 disease-specific instruments evaluated the outcomes of diseases belonging to spleen and stomach, liver, lung, heart and kidney systems, cancer, neurological disorders, endocrine disorders, infectious disease, skin diseases, bone diseases, neuropsychiatric disorders, autoimmune diseases, psychological disorders, surgical lesions, gynecological disorders and other pathological conditions. Despite recording quality of life and signs and symptoms, Li *et al*[67,82,83] also developed patient satisfaction and doctor-reported outcome scales in TCM. Some scales primarily designed for epidemiological diagnoses and surveys can also be used for health outcome assessment[7,19,33,34,41,43,90,91,99].

3.2.2 Demographics

Among the 97 instruments, one of them was specifically designed for children[39] and eight of them for females[11,13,25,40,41,85,95,97]. The remaining instruments were not specific to any type of participant. In the development and evaluation procedures of most instruments, participants were 16 years or older. Few instruments included the living environment and only one instrument was specifically designed for inpatients[23].

3.2.3 Domains, facets and items

Most instruments adopted the "concept-domain-facet-item" conceptual model. The number of "domains" ranged from 2 to 12 and the majority had 3 to 6 domains. The researchers had achieved high agreement on physiological, psychological and social domains. The "facets" of some instruments offered related information and the number ranged from 5 to 20, but differences existed on interpreting factual contents under domains between different scales. The number of "items" ranged from 7 to 124. Some scales adopted an "*n* items plus 1 total item" system, but some scales did not offer clear information on the domains and/or items.

3.2.4 Administrative modes

All information providers were patients. Most primary administrators were study interviewers or patients. However, no study evaluated the differences between primary and alternative administrative models. All instruments used paper-based scales for the administrative tools, except for the computer used in Gastralgia PRO Questionnaire based on computer-adaptive testing[100]. No letters and telephones were involved.

3.2.5 Response options

Likert and Graded Response Model was the most commonly used response option. Other options were visual analog scale, pictorial scale or checklists. Most scales used single type; some used two or more types. The option range in Likert and Graded Response scales was always 2 to 9, though results were commonly 4 or 5. The checklists always used two-category response with "Yes" or "No". Visual analog scale scored from 0 to 10. Only a few studies mentioned how to evaluate the response options.

3.2.6 Recall times and management times

Most instruments did not record the recall times and management times. Due to the lack of original data, the times could not be extracted from the published articles. In scales with complete information, the recall times varied from one day, one week, two weeks, and one month to three months, etc. Some scales did not have recall time limitations, but some had two or more. For example, "in the past one week", "in the past two weeks" and "at present" may exist in one instrument simultaneously. Few studies mentioned how they determined the recall time. No study mentioned the instrument management time for clinical application.

3.2.7 Scores

Few studies provided detailed scoring rules, but most of them provided operational scoring methods. Some scales used a total score; some used independent domain scores; most scales used the combination of the two methods mentioned above. For most scales, higher scores indicated a better health status. Some scales contained negative items, and those scales provided details on dealing with the negative items. Only a few studies used a conversion score and provided the conversion procedure, but no research explained the rationale of the procedure. Many instruments evaluated each item and domain with equal weight, but some did not. Of the studies that did not assign weights equally, only a few provided the process, reason, and source for the discrepancies in weight assignment.

3.2.8 Burden evaluation

No instruments evaluated burdens of the administrators, such as length of questionnaire or interview, format, font size and requirement on what patients could consult to complete response. The response burden evaluation mainly concentrated on recording recovery rate, completion rate and completion time. Few studies evaluated the participants' education levels; none described specified unsuitable environments or times for the surveys.

3.2.9 Format

Only a few studies offered a full instrument. Of those, over 50% did not have an instruction manual, and thus lacked clarity in their intended goals, recording specifications, formats, etc. Of those that offered manuals, some were almost the same as their main texts, and some scales required extensive and tedious social background information. No instrument described the format for typesetting formats, and clinical usage procedures. Only a few offered a detailed operating manual and also provided contact information.

3.2.10 Copyright

A minority of researchers explicitly indicated copyright. All instruments were free for use after signing a user-agreement

contract with the instrument developers. All contact information can be found in the published articles.

3.3 Property evaluation of TCM instruments

3.3.1 Content validity

3.3.1.1 Rationality and intention

Although many studies stated the necessity for developing new scales, their reasons for doing so were not always clear. For example, a few studies described existing instruments for similar diseases and conditions, but few of them performed a proper comparison. In addition, many studies mentioned "different Chinese cultural background", but few of them elaborated on how cultural background may influence PRO. In addition, most studies described the research purpose, but lacked clarity on the intended application of their instrument.

3.3.1.2 Conceptual framework establishment

Most instruments outlined different domains, mainly containing physiological, psychological and social functions. Definitions for these domains were unclear, and open to interpretation, and thus the "facets" under these domains differed significantly among studies. The sources of conceptual framework establishment included literature reviews, expert interviews, patient interviews, clinical experiences, etc. Some studies clearly described their concept and domain structure, but in many studies it was still difficult to identify the sources and methods for the development.

3.3.1.3 Item pool development and item selection

Many studies performed literature reviews in various sources including the World Health Organization Quality of Life Questionnaire (WHOQOL-100), the World Health Organization Quality of Life Questionnaire-BREF (WHOQOL-BREF), and the 36-Item Short-Form Health Survey (SF-36). Relevant Chinese medicine theoretical books, clinical books, diagnostic criteria, disease-specific scales, etc., were also consulted during their item pool development. Data extraction methods mainly focused on expert and/or patient interviews and/or working teams, but few studies had patient participation in the interview procedure, and even fewer described the specific composition and procedures of their focus group. The people performing item selection included focus group and external experts, patients, nurses, researchers on quality of life, healthy people, etc. Ideally, participant groups should be consistent with the intended study demographic in terms of education level, gender, age, etc., but few studies performed such analyses. In studies that did such analyses, many had inadequate participant composition. Few studies mentioned how to control clinical bias in their operation manual.

3.3.2 Other properties

Most studies used traditional property evaluation methods of reliability, validity and responsiveness, but differences existed among studies. Exploratory factor analysis and correlation analysis were the most commonly used methods in structure validity analysis; however, a small proportion of studies used confirmatory factor analysis, cluster analysis, and structural equation modeling, etc. Differences also existed in correlation analyses, such as correlation between individual items, items and self-domain, items and other items in the same domain, items and total score, individual domain, and domain and total score. Some studies did not analyze the correlation coefficients or modify the framework when the data showed poor correlation. The criterion validity mainly used WHOQOL-100, WHOQOL-BREF, SF-36 and KPS; the patients' general wellbeing, disease severity and general module questionnaire scores were also taken into account. Almost all correlation coefficients were very high, but little selection rational was provided. Differentiating criteria were in dispute among studies, which involved patient gender, age, disease severity, disease subtypes and healthy people. In addition, dispersion ratio, variance ratio, effects index, etc., were introduced into related assessments. Reliability analysis showed that most internal consistency reliability and test-retest reliability performed very well, but the test-retest period varied widely among studies, ranging anywhere between 3 d, 7 d, 2 weeks, to 4 weeks**,** without specifying reasons for choosing that timeframe. Cross-interviewer reliability testing is mandatory in interviewer-administrated instruments, but they were rarely presented in a proper assessment. Management times varied in responsiveness evaluations. A few studies described the intervention measures. The description and analysis were unclear on differential sub-groups that were divided according to disease severity or age.

In addition, most studies adopted computer-adaptive testing and item response theories in their instrument development and evaluation, such as the PRO scale of spleen stomach disease[26-28], the PRO scale of myasthenia gravis[29], the functional gastrointestinal disorder scale of TCM[79], and the stomachache patient-reported computer-adaptive test outcome scale[100]. These studies mainly used validity, reliability, person separation index and differential item functioning evaluation. The results showed that these scales had adequate psychological evaluations.

4 Discussion

4.1 Characteristics of the current scale development

Instrument development in TCM has made great progress recently. In the last five years, in average more than 10 scales were developed each year. The scales covered many diseases and subjects including tumors, neurological diseases, digestive diseases, and cardiovascular diseases. An increasing number of researchers and institutions are

engaged in instrument research.

As instrument research methods become more widely-known, more researchers have mastered the core process of scale development. In recent years, with the introduction of domestic PRO and publication of instrument research guidelines[101,102], many researchers can now follow the rigorous instrument development process with appropriate methods, thus significantly improving the quality of related research in Mainland China. In addition, new techniques such as the computer-adaptive testing and item response theory have been introduced. These techniques have enhanced instrument research in China; they are expected to further improve the scientific accountability and clinical adaptability of the scales, and transform the field of instrument development.

4.2 Limitations of the current scale development

4.2.1 Conceptual and structural systems

It is commonly recognized that clinical outcome assessments should include laboratory indices, PROs, ClinROs and ObsROs[2,103]. Of these, PRO includes HRQL, effect satisfaction, and signs and symptoms. HRQL can be further divided into quality of life in psychological and social status. Effect satisfaction is a combination of medical environment satisfaction and nursing satisfaction.

PRO is strictly confined to information obtained directly from the patients, and excludes any supplementary and explanatory information from others[101]. However, the extracted information from the articles showed that some studies did not define their scale properties and intentions clearly. Few studies support the rigorous property of PRO for many scales, and even fewer studies offered PRO property assessment specifically[104]. Significant differences existed in the instruments' definitions of important concepts and terminologies. For example, in the conceptual framework development procedure, some studies used the top-down method, others used the bottom-up method, while still others did not even present enough information to know what method was used. Some studies did not present conceptual information and directly proceeded to item-establishing. Few studies presented the conceptual framework model and figure. Many scales contained physiological, psychological and social domains. There were no clear and consistent domain definitions, so significant differences existed in the fact explanations. This may lead to inconsistent data-gathering, such as items in the social domain which included social relationship, social support, social adaptation, and social competence and satisfaction.

4.2.2 Item development procedure

A major problem is that many items lacked clear goals. The entire process of item development should contain item pool establishment, a preliminary selection of items, pre-testing, pilot surveys, clinical investigations and property assessment, all of which come together to form the instrument model. The administrators, participants, analyzing tools and methods in each step should have different characteristics and purposes; each area should be clearly delineated, and no information should be omitted. The guidelines for domestic instruments have presented clear descriptions[102]; however, many scales were found to have omitted stages, such as primary selection of items, pre-test and/or pilot survey. Some questionnaires did not even have a selection process at all.

A comprehensive instrument development procedure should include not only the common item pool, concept establishment and property evaluation, but also detailed inclusion and exclusion criteria, response opinions, scores, times, administrative modes, burden assessment and statistical methods[101]. However, most of the current research only reported the item selection. A large body of information was missing or not provided, which greatly reduced the reliability and operability of the scales.

Finally, the reporting of instrument development is vital for both researchers and clinicians to determine its validity. The complete instrument development report should include the title and structured abstract under "title and abstract"; rationale and objectives under "introduction"; intention, eligibility criteria, conceptual framework, item selection, response options, scoring, timeframe, administrative modes, burden assessment, property evaluation, and statistical analysis under "methods"; participants, main results and other analyses under "results"; summary of evidence, limitations, clinical details, and conclusions under "discussion"; item pools and/or final instrument subheadings in "appendixes"; and finally the "funding" section[105]. Few studies provided complete information mentioned above, and that degrades their accountability as well as clinical applicability.

4.2.3 Property evaluation

Firstly, the components in property evaluation were inconsistent across the studies. Although there are no strict rules on how many components should be included, reliability, validity and responsiveness are common properties for scale evaluation based on classical test theories. In 2002, the *Scientific Advisory Committee of the Medical Outcomes Trust* recommended eight key attributes of instruments, including conceptual and measurement model, reliability, validity, responsiveness, interpretability, respondent and administrative burden, alternate forms, and cultural and language adaptations[106]. In current studies, many components were absent, especially the more minor or specialized components such as test-retest reliability, the observer consistency reliability, criterion-related validity, validity and responsiveness. Interpretability and burden assessments were even more rare.

Secondly, content validity assessments were too simple. Content validity is the core property of a scale which requires

qualitative assessments from conceptual establishments; the rationality and integrity of the item development process; comprehensive methods, and a sufficient number of participants. Content validity assessment is an important measure for other quantitative assessments. However, in some studies, their content validity assessments were too simple and lacked explanations for method usage and limitations. Some studies used either an importance-evaluation score or a simple questionnaire, and performed correlation analyses, normality tests or content validity ratio analyses for content evaluation. But these were not supported by any guidelines or standards.

Thirdly, there was discordance in composition and index of other validity assessments. For example, researchers used exploratory factor analyses, correlation analyses, confirmatory factor analyses, cluster analyses or structural equation models, etc., in structure validity analyses; some used correlation coefficients between items, item-domain, item-other items in the same dimension, item-total score, between dimensions or dimension-total score in correlation analysis; some used WHOQOL-100 or -BREF, SF-36, KPS, general health of the patients, disease severity and general module questionnaire score as the criteria in criteria validity analyses; some used gender, age group, disease severity and disease subtypes for group validity analysis. These discrepancies made it difficult, if not impossible, to draw comparisons among different studies.

4.2.4 Statistical methods

Three major problems exist in the statistical methods of the instrument studies. Firstly, the research methods and standards used in the instrument development process were varied and inconsistent. For example, according to different studies, the maximum score for item deletion in importance assessment ranged from 40, 50 to 60, and the significant correlation coefficients were 0.2, 0.4 or 0.6. Although internal consistency reliability, split-half reliability, and criterion validity were commonly recognized, many differences existed in test-retest reliability and construct validity, especially in responsibility analyses. For example, the administrative times and selection criteria for test-retest reliability tests were different among studies. Many studies did not offer the reasons behind exploratory factor analyses, confirmatory factor analyses and correlation analyses in their constructed validity analysis. A well-constructed study should contain data from different time points and group comparisons, but many researchers only performed the latter. Only a few studies described the exact information used in their analyses, thus in the vast majority of studies it was difficult to extrapolate the effect size.

Secondly, although item response theory and computer-adaptive testing have been introduced into TCM instrument usage, few research projects discussed the role, position, limitation, interpretation and evaluation of these new tools. Information on these theories and testing techniques is scarce, and they are not yet used widely in scientific research and clinical application.

Thirdly, qualitative research methods are fairly new ideas that are still being developed. Qualitative methods play a more important role in computer-adaptive testing-based scale development than traditional methods. The reason is that modern test theory requests fewer, but also more high-quality items to achieve the same accuracy as classical test theory, but the quantitative technical indicators cannot completely meet the demand. Adding to this problem is that few quantitative indices were introduced in qualitative studies. In order to develop high-quality, useful instruments for qualitative research evaluation, more studies and further exploration are needed.

5 Conclusions

In conclusion, research on instrument development in TCM has developed rapidly. Thus far, more than 100 scales have been developed, including generic and disease-specific instruments covering many TCM and Western medicine subjects. As a milestone for instrument recognition, a domestic instrument research standard was also published. However, some limitations in the development procedure, methods and standards still exist. Many scales are not useful in clinical practice and research, and they serve the function of “trophies” more than actual tools. This is unfortunate and a terrible waste of scientific resources, but also provides more impetus for future studies.

6 Funding

This study was supported by the National Natural Science Foundation of China (Project Grant No. 81073163).

7 Acknowledgements

We would like to express our gratitude to Bob Wong (USA) for the assistance in revising this manuscript.

8 Competing interests

The authors declare that they have no competing interests.

REFERENCES

1 Frommer M, Rubin G, Lyle D. The NSW health outcomes program. N S W Public Health Bull. 1992; 3(12): 135-137.
2 Acquadro C, Berzon R, Dubois D, Leidy NK, Marquis P, Revicki D, Rothman M; PRO Harmonization Group. Incorporating the patient’s perspective into drug development and communication: An ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group

meeting at the Food and Drug Administration, February 16, 2001. Value in Health. 2003; 6(5): 522-531.
3 Wu Y, Wang ZZ. The analysis of the relationship between quality of life and kidney deficiency of vital energy. Tong Ji Yi Ke Da Xue Xue Bao. 1994; 7(4): 231-233. Chinese with abstract in English.
4 Liu FB. Study on the quality of life of the patients with gastroenteric diseases. Guangzhou Zhong Yi Yao Da Xue Xue Bao. 1997; 14(4): 225-228. Chinese with abstract in English.
5 Zhang HL. The clinical and experimental research on electroacupuncture for cervical spondylosis. Guangzhou: Guangzhou University of Chinese Medicine. 2000. Chinese with abstract in English.
6 Zhang QH, Zhou FS, Lu MH. The analysis of expert surveys for items selection on functional stomach and intestine disease questionnaire development based on the combination of diseases and patterns. Proceedings of the 15th National Symposium on Digestive Diseases by China Association of Integrative Medicine. Chongqing, China. 2003: 79-81. Chinese.
7 He LY, Liu BY, Xie YM, Hu JQ, Li WQ, Gao Y. Evaluation of traditional Chinese medicine syndrome questionnaire for sub-health state. Zhongguo Zhong Yi Ji Chu Yi Xue Za Zhi. 2004; 10(10): 64-67. Chinese.
8 Zhang RQ. Evaluating the effects of three methods against cancer by the quality of life scale. Shanghai: Second Military Medical University. 2004. Chinese with abstract in English.
9 Hu XJ, Zhang BL, Cai GX. Developing and evaluating a quality of life instrument for patients with stroke. Zhong Yi Yao Xue Kan. 2006; 24(9): 1638-1640. Chinese.
10 Wang Z, Hu SY, Cai TS, Xia DS. Development of emotion rating scale for Ganzangxiang of traditional Chinese medicine (ERSG). Zhongguo Xing Wei Yi Xue Ke Xue. 2004; 13(1): 108-110. Chinese with abstract in English.
11 Wang H. Developing emotional scale of traditional Chinese medicine for menopausal woman based on the psychology theory of traditional Chinese medicine. Guangzhou: Guangzhou University of Chinese Medicine. 2005. Chinese with abstract in English.
12 Xiang FM. A clinical research on questionnaire of quality of life for patients with HIV/AIDS in China. Guangzhou: Guangzhou University of Chinese Medicine. 2005. Chinese with abstract in English.
13 Liu MS. Development of quality of life scale on chronic pelvic inflammatory disease. Guangzhou: Guangzhou University of Chinese Medicine. 2005. Chinese with abstract in English.
14 Wu W. Study on developing a traditional Chinese medicine questionnaire of quality of life for Parkinson disease. Guangzhou: Guangzhou University of Chinese Medicine. 2005. Chinese with abstract in English.
15 Liu XL, Deng J, Liu FB, Chen GX, Chen JF. Establishment of the scale of the quality of life for ankylosing spondylitis. Guangzhou Zhong Yi Yao Da Xue Xue Bao. 2005; 22(4): 315-319. Chinese with abstract in English.
16 Wang ZH, Chen ZQ, Wang SS, Ou AH, Mao W. Preliminary establishment and assessment of postoperative recovery scale. Guangzhou Zhong Yi Yao Da Xue Xue Bao. 2005; 24(3): 195-197. Chinese with abstract in English.
17 Wang C. Development of traditional Chinese medicine questionnaire of quality of life on psoriasis. Changsha: Hunan University of Chinese Medicine. 2005. Chinese.
18 Liu FB, She SF, Luo SJ. The development of patient-reported outcome questionnaire for liver disease in Chinese medicine. Shenyang: Proceedings of 18th National Academic Seminar on Spleen-gastric Diseases, Branch of Spleen-gastric Disease of China Association of Chinese Medicine. 2006: 443-446. Chinese.
19 Zhou JW, Zhang F, Li CY, Wang M. Development and reliability assessment of TCM syndromes scale for cervical spondylosis (radiculopathy). Sichuan Zhong Yi. 2006; 24(8): 31-33. Chinese.
20 Huang YT. The set-up of vertebroarterial type of cervical spondylopathy inventory and the evaluation of clinical curative effect. Guangzhou: Guangzhou University of Chinese Medicine. 2006. Chinese with abstract in English.
21 Ruan Y, Chen Q, Yang ZJ, Zhang JX. Study on the allergic rhinitis quality of life in adult. Liaoning Zhong Yi Za Zhi. 2006; 33(3): 292-294. Chinese with abstract in English.
22 Wang TF, Xue XL, Ji SL. Imaging for the development of fatigue self-assessment scale. Proceedings of the Conference on the Establishment of the Diagnostics Branch of China Association of Chinese Medicine, Beijing, China. 2006: 250-253. Chinese.
23 Song QL. The establishing, evaluation and application of kidney medical quality of life scale. Guangzhou: Guangzhou University of Chinese Medicine. 2006. Chinese with abstract in English.
24 Wang JK. Studies on Chinese medicine quality of life scale for evaluating knee osteoarthritis treatment by therapy of lukewarm activation. Guangzhou: Guangzhou University of Chinese Medicine. 2006. Chinese with abstract in English.
25 Fu J. Reliability and validity for the scale of Chinese medicine curative effect evaluation of menopause syndrome. Guangzhou: Guangzhou University of Chinese Medicine. 2006. Chinese with abstract in English.
26 Liu FB, Wang WQ. Establishment of the spleen-stomach patient-reported outcome scale in Chinese medicine. Proceedings of the 19th National Academic Seminar for Digestive Diseases held by Branch of Spleen-stomach Diseases, China Association of Chinese Medicine, Shijiazhuang, China. 2007: 411. Chinese.
27 Huang YD. Development and evaluation of the chronic gastritis subscale in spleen and stomach disease PRO scale (SSDPRO-CG). Guangzhou: Guangzhou University of Chinese Medicine. 2009. Chinese with abstract in English.
28 Liang BJ. Development and evaluation of the irritable bowel syndrome subscale in spleen and stomach disease PRO scale (SSDPRO-IBS). Guangzhou: Guangzhou University of Chinese Medicine. 2010. Chinese with abstract in English.
29 Guo L. Development and evaluation of the myasthenia gravis patient-reported outcome scale (MG-PRO). Guangzhou: Guangzhou University of Chinese Medicine. 2007. Chinese with abstract in English.
30 Liu FB, Lang JY, Zhao L, Fang JQ, Liang GH, Lin LZ. The development of health status scale of traditional Chinese medicine. Proceedings of the 19th National Academic Seminar on Spleen-gastric Disease, Branch of Spleen-gastric Disease of China Association of Chinese Medicine, Shijiazhuang, China. 2007: 589-593. Chinese.
31 Liu FB, Zhao L, Lang JY, Lin LZ, Liang GH, Fang JQ. Development of the Chinese quality of life instrument. Zhongguo Zu Zhi Gong Cheng Yan Jiu Yu Lin Chuang Kang Fu. 2007; 11(52): 10492-10495, 10515. Chinese with abstract in English.

32 Wang YG, Zhang QM, Wang YY, Song GL, Li M, Li YQ. Assessment on the self-evaluation scale of life quality of TCM. Shandong Zhong Yi Yao Da Xue Xue Bao. 2007; 31(3): 182-185. Chinese with abstract in English.
33 Li HF, Chen X, Jin RF. Development and reliability and validity assessment of sub-health syndrome scale. Shaanxi Zhong Yi. 2007; 28(5): 565-567. Chinese.
34 Guo Q. Preliminary development of rating scales for hyper-activity of liver-yang and damp-heat in liver and gallbladder syndromes. Changsha: Central South University. 2007. Chinese with abstract in English.
35 Zhang YH. Analyses on reliability and validity of the evaluation scale of patient-reported outcome based on patients with spastic paralysis due to stroke. Beijing: China Academy of Chinese Medical Sciences. 2007. Chinese with abstract in English.
36 Guo XX, Zhang ML, Xu LR, Su FJ, Peng QH, Hao XM. The development of quality of life questionnaire for HIV/AIDS. Proceedings of the International Academic Seminar on Prevention and Cure of AIDS of China Association of Chinese Medicine, Zhengzhou, China. 2007: 218-220. Chinese.
37 Chen RM. A study on developing a traditional Chinese medicine questionnaire of quality of life on insomnia. Guangzhou: Guangzhou University of Chinese Medicine. 2007. Chinese with abstract in English.
38 Liu QH. The development and preliminary application of evaluation scale system in Xiaoke-associated eye disease for clinical curative effect. Guangzhou: Guangzhou University of Chinese Medicine. 2007. Chinese with abstract in English.
39 Liu J. Analysis of the influence of bronchial asthma on children's quality of life. Chengdu: Chengdu University of Traditional Chinese Medicine. 2007. Chinese with abstract in English.
40 Nie GN. The development and evaluation of menopause syndrome rating scale of traditional Chinese medicine. Guangzhou: Guangzhou University of Chinese Medicine. 2007. Chinese with abstract in English.
41 Wang HJ. Research on syndrome differentiation standard and curative effect evaluation scale for liver-qi rebellion of premenstrual syndrome. Jinan: Shandong University of Traditional Chinese Medicine. 2007. Chinese with abstract in English.
42 Yang XL. Development and evaluation of the liver disease PRO scale of traditional Chinese medicine for patients with liver cirrhosis. Guangzhou: Guangzhou University of Chinese Medicine. 2008. Chinese with abstract in English.
43 Le M, Huang X, Gao YQ. Establishment of measurement scale for spleen deficiency syndrome in chronic hepatitis B. Shanghai Zhong Yi Yao Za Zhi. 2008; 42(3): 6-9. Chinese with abstract in English.
44 Xu W. The development of quality of life scale in patients with chronic hepatitis B. Nanjing: Nanjing University of Chinese Medicine. 2008. Chinese.
45 Shi PY. The development and evaluation of traditional Chinese medicine quality of life scale in patients with liver cirrhosis. Nanjing: Nanjing University of Chinese Medicine. 2008. Chinese.
46 Guo LF. The preliminary development of quality of life scale for nasopharyngeal cancer. Guangzhou: Guangzhou University of Chinese Medicine. 2008. Chinese with abstract in English.
47 Zhu T, Mao JY. Design idea of item of QOL scale with TCM characteristics to patients with CAD. Liaoning Zhong Yi Yao Da Xue Xue Bao. 2008; 10(7): 11-13. Chinese with abstract in English.
48 Lin Q, Nong YB, Wan J, Lu JJ, Wen ZH, An C. A clinical study of scale for evaluating the quality of life in patients with chronic heart failure treated with traditional Chinese and Western medicine. Zhongguo Zhong Xi Yi Jie He Ji Jiu Za Zhi. 2008; 15(3): 131-134. Chinese with abstract in English.
49 Liu AM. Study on quality of life and influential factor in patients with high myopia. Guangzhou: Guangzhou University of Chinese Medicine. 2008. Chinese with abstract in English.
50 Yang ZB, Liu J, Liu X, Sun J, Zeng BJ. Development of quality of life scale for chronic eczema patients. Zhong Yi Yao Dao Bao. 2008; 14(8): 1-5. Chinese with abstract in English.
51 Zhou MH. Study on developing a traditional Chinese medicine questionnaire of quality of life on psoriasis. Guangzhou: Guangzhou University of Chinese Medicine. 2008. Chinese with abstract in English.
52 Xin L. The development, evaluation and initial clinical application of the measuring scale of quality life of TCM for sustained hemodialysis patients. Guangzhou: Guangzhou University of Chinese Medicine. 2008. Chinese with abstract in English.
53 Feng S. The establishment of primary scale for generalized anxiety disorder treated by acupuncture method. Guangzhou: Guangzhou University of Chinese Medicine. 2008. Chinese with abstract in English.
54 Wang HR, Wang QC. Design and application of clinical syndrome questionnaire for insomnia based on liver function in Chinese medicine. Zhong Yi Wen Xian Za Zhi. 2008; (2): 32-34. Chinese.
55 Li H, Liang WX. Study and development of TCM stroke scale for quality of life measurement (1) — establishment of the scale. Liaoning Zhong Yi Za Zhi. 2008; 35(3): 376-378. Chinese with abstract in English.
56 Wei YL, Kong J, Liu GL, Huang Y, Han B, Li L, Chen HM, Kong J, Liu WH, Wei BH. Application of somatic complaints of subhealth status questionnaire (Sc-SHs-Q) in subhealth status's clinical trail screen, symptoms analysis and evaluation of intervention therapy. Huan Qiu Zhong Yi Yao. 2009; 2(5): 349-353. Chinese with abstract in English.
57 Chen LM. Establish the preliminary PRO scale of key symptoms in TCM for COPD and pulmonary heart disease. Guangzhou: Guangzhou University of Chinese Medicine. 2009. Chinese with abstract in English.
58 Wang JW, Liu N, Yang ZB, Liu X. The primary study of quality of life questionnaire for acne in Chinese medicine. Proceedings of Academic Annual Conference of Cosmetic Branch of China Association of Chinese Medicine, Tianjin, China. 2009: 17-19. Chinese.
59 Huang CJ. The development of "Main Symptom Scale of Atopic Dermatitis" and the initial evaluation of its reliability, validity and responsibility. Guangzhou: Guangzhou University of Chinese Medicine. 2009. Chinese with abstract in English.
60 Gao H. Preliminary formulating and appraising of the quality of life scale for primary osteoporosis. Guangzhou: Guangzhou University of Chinese Medicine. 2009. Chinese with abstract in English.
61 Tang XD, Wang P, Liu BY, Zhai MJ. Development and analysis of a patient-reported outcome instrument for chronic gastrointestinal diseases. Zhong Yi Za Zhi. 2009; 50(1): 27-29. Chinese with abstract in English.
62 Shi X, Cai QW, Du XD, Liu J. Reliability & validity research

on measuring scale of traditional Chinese medicine for primary osteoporosis patients. Zhong Yi Za Zhi. 2009; 50(2): 124-126. Chinese with abstract in English.
63 Wang W. The development and evaluation of Quality of Life Scale of integrated TCM and Western medicine for coronary heart disease. Guangzhou: Guangzhou University of Chinese Medicine. 2009. Chinese with abstract in English.
64 Wu DR, Lai SL, Liang WX, Guo XF, Ou AH, Wen ZH. A study of scale response for health scale of traditional Chinese medicine. J Chin Integr Med. 2009; 7(8): 717-723. Chinese with abstract in English.
65 Fang J. The initial research and evaluation in Quality of Life Scale of knee osteoarthritis. Guangzhou: Guangzhou University of Chinese Medicine. 2009. Chinese with abstract in English.
66 Zhou XG. The preliminary establishment of the QOL Scale for maintenance hemodialysis patients. Jinan: Shandong University of Traditional Chinese Medicine. 2009. Chinese with abstract in English.
67 Li JS, Yu XQ, Li SY, Sun ZK, Zhang W, Zheng SP, Jia XH, Wang HF, Wang MH. Property assessment of efficacy satisfaction scale for pneumonia. Proceedings of the 10th National Conference on the Prevention and Treatment for Respiratory Diseases with Traditional Chinese and Western Medicine, Hangzhou, China. 2009: 80-84. Chinese.
68 Li JS, Yu XQ, Li SY, Sun ZK, Zhang W, Zheng SP, Shi XZ, Jia XH, Wang MH, Wang HF. The development of patient-reported outcome questionnaire for pneumonia based on therapeutic evaluation. Proceedings of Clinical Scientific Research on Internal Medicine of Traditional Chinese Medicine in 2009, Branch of Internal Medicine of the China Association of Chinese Medicine, Hangzhou, China. 2009: 85-92. Chinese.
69 Liu SN, Liu Z, Liu HM, Tao XP, Liu BY. Development of self-evaluation questionnaire for chronic liver disease. Zhongguo Zhong Yi Yao Xin Xi Za Zhi. 2009; 16(9): 93-94. Chinese.
70 Xia PF. The development of scale for insomnia with the state of atypical dysphoria and the relative research with state anxiety inventory. Jinan: Shandong University of Traditional Chinese Medicine. 2009. Chinese with abstract in English.
71 Qi XH, Teng J, Kang XL, Wang L. Initial development of scale for insomnia with the state of over-anxiety. Zhonghua Zhong Yi Yao Xue Kan. 2009; 27(12): 2490-2493. Chinese with abstract in English.
72 Wang J, He QY, Zhu MJ, Ma CS, Yao KW, Ye Y, Lin Q, Zhong JB, Nong YB. Comprehensive efficacy assessment system and clinical rating scale on TCM syndrome for angina pectoris. Proceedings of the 10th Cardiovascular Disease Conference held by Chinese Association of the Integrated Traditional Chinese and Western Medicine and the 5th Cardiovascular Disease Conference held by Jiangxi Province Association of the Integrated Traditional Chinese and Western Medicine, Nanchang, China. 2010: 234-237. Chinese.
73 Xu X, Kong JH, Li L, Yu Y, Yang QL. "Adjusting mind for calming" self-rating inventory and its reliability and validity. Zhong Yi Za Zhi. 2010; 51(10): 887-889. Chinese with abstract in English.
74 Ren M, Sun ZT, Feng JH, Shang HC, Zhang BL. Theory framework development of traditional Chinese medicine patient-reported outcome questionnaire for patients with COPD (TCMPRO-COPD). Zhongguo Zhong Yi Yao Xian Dai Yuan Cheng Jiao Yu. 2010; 8(18): 151-153. Chinese.
75 Zhu MF, Run Q, Tan QW, Liu X, Yang ZB, Wang JW, Zhang YM. Study of life quality scale on patients with alopecia areata in TCM. Hunan Zhong Yi Yao Da Xue Xue Bao. 2010; 30(9): 149-151. Chinese with abstract in English.
76 Wang XF, Gao Y, Cao KG. Items selection methods for patient-reported outcomes instrument of stroke. Liaoning Zhong Yi Za Zhi. 2010; 37(5): 772-774. Chinese.
77 Xiang YP, Chen LM, Yang ZB, Wang C. Study of quality of life (QOL) scale on patients with vitiligo in TCM. Hunan Zhong Yi Yao Da Xue Xue Bao. 2010; 30(1): 56-59. Chinese with abstract in English.
78 Zhang W. Imagining for the establishment of chronic hepatitis B questionnaire in traditional Chinese medicine. Zhong Xi Yi Jie He Gan Bing Za Zhi. 2010; 20(1): 53-55. Chinese.
79 Zhuang KH. Development and evaluation of functional gastrointestinal disorder scale of traditional Chinese medicine (FGIDS-scale of TCM). Guangzhou: Guangzhou University of Chinese Medicine. 2010. Chinese with abstract in English.
80 She SF, Liu FB, Luo SJ. Theory framework development for clinical effect assessment questionnaire for liver disease in Chinese medicine. Zhong Yao Xin Yao Yu Lin Chuang Yao Li. 2010; 21(4): 449-450. Chinese.
81 Zhong JB. Health self-assessment questionnaire designed based on five internal organs system. Liaoning Zhong Yi Yao Da Xue Xue Bao. 2010; 12(8): 86-88. Chinese with abstract in English.
82 Li JS, Yu XQ, Li SY, Sun ZK, Zhang W, Zheng SP, Shi XZ, Jia XH, Wang MH, Wang HF. Property assessment for clinician-reported outcome questionnaire for pneumonia. Proceedings of the 11th National Conference on Prevention and Treatment of Respiratory Diseases by Traditional Chinese Medicine and Western Medicine, Beijing, China. 2010: 118-123. Chinese.
83 Li JS, Wang MH, Yu XQ, Li SY, Wang HF, Li Y. Study on the efficacy satisfaction scale for chronic obstructive pulmonary disease at stable stage. Zhongguo Zhong Yi Yao Xin Xi Za Zhi. 2010; 17(6): 8-10. Chinese with abstract in English.
84 Peng JQ, Zhao H, Chen JX, Wang SX, Li XH, Cui HZ, Ji N, Li C, Sun LL, Wang N. Validity and response of clearing heat to eliminate fire efficacy scale of Chinese herbal toothpastes. Zhonghua Zhong Yi Yao Za Zhi. 2010; 25(8): 1324-1326. Chinese with abstract in English.
85 Gao XF, Liu S, Chen HF, Que HF, Lu JG, Cao YQ, Chen WJ, Qiu JY. Development and preliminary application of the TCM scale of depression after breast cancer. Zhonghua Zhong Yi Yao Za Zhi. 2010; 25(12): 2264-2267. Chinese with abstract in English.
86 Hu XC, Zhang H, Zhou Y, Liu P. Establishing and screening entry items of rating scale for patient-reported outcomes of posthepatitic cirrhosis. Proceedings of the 20th Academic Annual Conference on Traditional Chinese Medicine Diagnosis, Yinchuan, China. 2011: 363-367. Chinese.
87 Lu GJ. Study on quality of life questionnaire for chronic urticaria in TCM. Changsha: Hunan University of Chinese Medicine. 2011. Chinese with abstract in English.
88 Ma YH. Modification and property assessment of patient-reported outcome questionnaire for posthepatitic cirrhosis. Shenyang: Liaoning University of Traditional Chinese

Medicine. 2011. Chinese with abstract in English.
89 Liu LX. Development and evaluation of the scale of stable angina (the syndrome of phlegm accumulating with blood stasis). Shenyang: Liaoning University of Chinese Medicine. 2011. Chinese with abstract in English.
90 He W, Xie YM, Wang YY. Research about re-evaluation of screening of traditional Chinese medicine symptoms item of post-marketing medicine Xuezhikang. Zhongguo Zhong Yao Za Zhi. 2011; 36(20): 2898-2900. Chinese with abstract in English.
91 Pi Y, Zhou JW, Yan P, Han JT. Development of quality of life rating scale for cervical spondylotic radiculopathy. Jiangxi Zhong Yi Yao. 2011; 42(8): 6-9. Chinese.
92 Liu W, He JC, Cai XS, Shou XX, Bian YF, Chen LL. Structure of insomnia scale of TCM and measurement analysis. Zhonghua Zhong Yi Yao Za Zhi. 2011; 26(11): 2507-2510. Chinese with abstract in English.
93 Li GX, Zhao J, Liang MX, Liu J, Zhang HZ. Study the consistency between the scale for primary osteoporosis according to traditional Chinese medicine and the quality of life scale. Shi Yong Zhong Yi Nei Ke Za Zhi. 2011; 25(12): 3-4, 7. Chinese with abstract in English.
94 Wang Q, Zhang HM, Ma FL, Liang YJ. Development of self-rating questionnaire of health status in traditional Chinese medicine and pilot evaluation on its reliability and validity. Anhui Zhong Yi Xue Yuan Xue Bao. 2011; 30(5): 18-21. Chinese with abstract in English.
95 Zhang YF, Shi ZR, Wang SY. Development and evaluation of women aging questionnaire in Chinese medicine. Shandong Zhong Yi Yao Da Xue Xue Bao. 2011; 35(5): 401-403. Chinese.
96 Zhu YB, Liu YL, Wang W, Jiang FC. Development of a patient-reported outcome scale for stable chronic obstructive pulmonary disease and its clinical applicability. J Chin Integr Med. 2011; 9(8): 857-865. Chinese with abstract in English.
97 Zhang J, Shen Y, Bian WH, Pan LQ. Research on the design of primary TCM PRO questionnaire for breast cancer. Nanjing Zhong Yi Yao Da Xue Xue Bao. 2011; 27(3): 295-297. Chinese with abstract in English.
98 Li JS, Wang MH, Yu XQ, Li SY, Wang HF, Li Y. Study and preliminary assessment of chronic obstructive pulmonary disease patient-peported circulating ending scale. Zhong Yi Xue Bao. 2011; 26(154): 270-274. Chinese with abstract in English.
99 Zhao H, Chen JX, Xiong WH, Peng YQ, Ge X, Liu Y, Ji ZM, Cui HZ, Peng JQ, Zhao X, Wang LM, Xue FF. Design of TCM syndrome questionnaire of sub-health state. Beijing Zhong Yi Yao Da Xue Xue Bao. 2011; 34(1): 38-41, 51. Chinese with abstract in English.
100 Li XY. The pilot study of development of gastralgia PRO computerized adaptive testing. Guangzhou: Guangzhou University of Chinese Medicine. 2011. Chinese with abstract in English.
101 U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Guidance for industry Patient-Reported Outcome Measures: Use in medical product development to support labeling claims. (2009-12) [2012-07-30]. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf
102 Fang JQ, Liu FB, Hao YT, Liu BY, Xie YM. Guidance for medical instrument development and application (draft). Proceedings of Research on Five-organ Relation Theory and Practice and the Development Methods and Application of Chinese Medicine Questionnaire, Guangzhou, China. 2010: 152-194. Chinese.
103 Public Workshop; White O. The Food and Drug Administration Review and Qualification of Clinical Outcome Assessments Workgroup. Measurement in clinical trials: Review and qualification of clinical outcome assessments. (2011-10-19) [2012-07-30]. http://www.fda.gov/downloads/Drugs/NewsEvents/UCM276993.pdf.
104 Li XH. Research and assessment of PRO characteristic of the Chinese quality of life scale. Guangzhou: Guangzhou University of Chinese Medicine. 2007. Chinese with abstract in English.
105 Hou ZK, Liu FB, Fang JQ, Li XY, Li LJ, Lin CH. Advising on preferred reporting items for patient-reported outcome instrument development: the PRIPROID. Chin J Integr Med. 2012; [Epub ahead of print]
106 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(3): 193-205.