

# Development and validation of the system of quality of life instruments for cancer patients: breast cancer (QLICP-BR)

Chonghua Wan · Zheng Yang · Xueliang Tang ·  
Tianning Zou · Dedian Chen · Dongmei Zhang ·  
Qiong Meng

Received: 6 April 2008 / Accepted: 25 June 2008 / Published online: 15 July 2008  
© Springer-Verlag 2008

## Abstract

**Goals** To develop and validate a quality of life (QOL) instrument for patients with breast cancer, QLICP-BR, which is one of the system of QOL instruments for cancer patients in China.

**Methods** Using the programmed decision methods of instrument development, the quality of life instrument for cancer patients-breast cancer (QLICP-BR) with considering Chinese cultural background was developed, and evaluated on the data from 186 inpatients with breast cancer. The statistical methods used in this research included statistical description, Pearson correlation, factor analysis, and paired *t* test.

**Results** The test–retest reliability for the overall scale and five domains are all above 0.75. Internal consistency  $\alpha$  for each domain is higher than 0.65 except social domain (0.58).

Most correlation coefficients between each item and its domain are above 0.60. The scores differences between pretreatment and post-treatment for overall scale, general module, physical domain, psychological domain and social domain have statistical significance.

**Conclusions** The QLICP-BR is of good validity, reliability, and reasonable responsiveness, and can be used to assess quality of life for patients with breast cancer in China.

**Keywords** Breast cancer · Standardized response mean · Psychometric properties · FACT-B · QLICP-BR

## Introduction

Quality of life (QOL) is nowadays considered to be one of the most important outcome measures in medical studies, especially in the field of cancer clinical research. This trend makes it all the more necessary to have a clear methodology for the development and use of QOL instruments. Many QOL instruments for patients with breast cancer have been developed, such as LASA [9], functional assessment of cancer therapy for breast cancer (FACT-B) [3–4], European Organization for Research and Treatment of Cancer (EORTC) quality of life questionnaire QLQ-C30 and QLQ-BR23 [1, 11]. In China, more and more clinicians and researchers are more concerned with the QOL of patients with breast cancer, but very few Chinese QOL instruments are available, which impedes research and applications of QOL in this field greatly. Although the Chinese versions of QLQ-BR53 (QLQ-C30 and QLQ-BR23) and FACT-B can be used for Chinese patients [19, 21], they are lacking in Chinese cultural backgrounds to

---

Supported by the National Natural Science Foundation of China (30360092) and the Natural Sciences Funds of Yunnan Province (99C0016G).

---

C. Wan (✉) · Z. Yang · Q. Meng  
School of Public Health, Kunming Medical University,  
Kunming 650031 Yunnan, China  
e-mail: wanchh@hotmail.com

X. Tang · T. Zou · D. Chen  
Yunnan Tumor Hospital,  
Kunming 650031, China

D. Zhang  
Department of Management, Guangdong Pharmacy College,  
Guangzhou 510240, China

some extent considering their original use in English-speaking patients. For example, the family relationship and kinship play very important roles in daily life. Taoism and traditional medicine focus on good temper and high spirit. Good appetite, sleep, and energy are highly regarded in daily life with food culture being very important. This kind of culture dependence does not reflect in most QOL instruments in other languages. It is necessary to develop Chinese specific QOL instruments. Therefore, by combining a general module and cancer-specific modules, we have developed the Chinese QOL instruments system called quality of life instruments for cancer patients (QLICP) [15–18, 20]. This system includes a general module (QLICP-GM) which can be used with all types of cancer, and specific modules for different cancer with each module being used for only the relevant disease. As an example, the breast cancer instrument of this system, quality of life instruments for cancer patients-breast cancer (QLICP-BR), was formed by combining the QLICP-GM and the specific module of the breast cancer. At present, the QLICP-GM and 12 specific modules have been developed, and thus 12 cancer-specific QOL instruments were formed, namely, lung cancer (QLICP-LU), breast cancer (QLICP-BR), head and neck cancer (QLICP-HN), colorectal cancer (QLICP-CR), liver cancer (QLICP-LI), esophageal cancer (QLICP-ES), stomach cancer (QLICP-ST), bladder cancer (QLICP-BL), prostate cancer (QLICP-PR), cervical cancer (QLICP-CE), ovarian cancer (QLICP-OV), and brain cancer (QLICP-BN).

This paper aims to report the developmental process and validation of the QLICP-BR.

## Materials and methods

### Establishment of the general module (QLICP-GM)

Two working groups, composed of physicians, nurses, medical educators, teachers, and researchers formed a nominal group of 16 persons and a focus group of ten persons. The programmed decision method was used in item selection. First, the focus group discussed and confirmed the structure of the instrument, which included four domains: physical, psychological, social, and the common symptoms/side-effects. After reviewing some well-known QOL instruments such as SF-36 [12], NHP [5], FACT-G [4], QLQ-C30 [1], and considering elements of Chinese culture, the nominal group proposed some possible items under each of the facets within the domains, resulting in a 78-item pool. Then some methods such as focus group discussion, in-depth interview, and pilot-test were used to refine and select items. Four statistical procedures (variation analysis, correlation analysis, factor analysis, and cluster analysis) were used to rescreen the items based on pretest data. Finally, 32 items were selected to form the QLICP-GM, which included four domains and nine facets (see Table 1 in detail), and this scale was confirmed to have good validation based on the data of 600 patients of lung

**Table 1** Scoring method of the quality of life instrument QLICP-BR

Domains/facets	Number of items	Range of scores	Scoring method
Physical domain (PHD)	7	7–35	<b>BPF+SXF&gt;IDF</b>
Basic physiologic function(BPF)	2	2–10	GPH1+GPH2
Sexual function (SXF)	1	1–5	GPH3
Independence function(IDF)	4	4–20	GPH4+GPH5+GPH6+GPH7
Psychological domain (PSD)	12	12–60	EMO+REC
Emotion (EMO)	9	9–45	GPS1+...+GPS8+GPS11
Recognition (REC)	3	3–15	GPS9+GPS10+GPS12
Social domain (SOD)	6	6–30	SSS+ELE
Social support and safety (SSS)	4	4–20	GSO1+GSO2+GSO3+GSO4
Effects on life and economics (ELE)	2	2–10	GSO5+GSO6
Common symptom and side effect domain (SSD)	7	7–35	SEF+CST
Side effect (SEF)	4	4–20	GSS1+GSS2+GSS3+GSS6
Common symptom (CST)	3	3–15	GSS4+GSS5+GSS7
Specific domain (SPD)	7	7–35	SBR1+SBR2+...+SBR7
Breast symptom (BRS)	3	3–15	SBR1+SBR2+SBR3
Upper body effect (UBE)	2	2–10	SBR4+SBR5
Physical and psychological effect (PPE)	2	2–10	SBR6+SBR7
Core/General Module(CGM)	32	32–160	PHD+PSD+SOD+SSD
Total (TOT)	39	39–195	PHD+PSD+SOD+SSD+SPD

cancer (85), breast cancer (186), colorectal cancer (110), head and neck cancer (133), and stomach cancer (86).

The entire process of developing the QLICP-GM can be seen in other papers [15–18, 20], the main steps are summarized as follows:

- Item pool (78 items)
  - ↓ focus and nominal group discussions
- Screened refining Items (56 items)
  - ↓ importance test (50 cases interview), analysis, focus group discussions
- Primary scale (V0.0, 40 items)
  - ↓ pretest (448 cases), analysis, focus group discussions
- Final scale (V1.0, nine facets under four domains, 32 items)
  - ↓ 600 patients
- Evaluation (validity, reliability, responsiveness)

#### Establishment of the specific module

After development of the QLICP-GM, the 14 items reflecting symptoms, side effects, and special mental health of breast cancer were selected to form the item pool of the specific module, and similar methods were used to get the final module, which has seven items and can be classified into three facets (see Table 1 in detail).

#### Evaluation of the QLICP-BR

The formal QLICP-BR (general module QLICP-GM plus specific module) was used to evaluate patients with breast cancer in a great scale in order to study its validity, reliability, and responsiveness. The study population was limited to breast cancer inpatients at any stages and treatments who were able to read and understand the questionnaires. The participating investigators were doctors, nurses, and medical postgraduate students. The investigators explained the aims of the trial and the instrument to the patients and obtained informed consent from those patients who agreed to participate in the study and met the inclusion criteria. Each respondent ( $n=186$ ) answered the questionnaires at the time of admission to the hospital. Some patients ( $n=166$ ) were entered to participate in a second assessment the following day or the second day after hospitalization to evaluate test–retest reliability, and 94 cases were sampled randomly and assessed a third time after treatment in order to evaluate responsiveness.

After investigation, the raw scores of items, domains, and overall scale were calculated. Each item of QLICP-BR

is rated in a five-level scoring system, namely, not at all, a little bit, somewhat, quite a bit, and very much. The positively stated items directly obtain scores from one to five points and the negatively stated items are reversed. Each domain score is obtained by adding its own item score together. The overall scale score is the sum of five domains score (see Table 1).

For comparison, all domains scores were linearly converted to a 0–100 scale using the formula:  $SS=(RS-Min) \times 100/R$ , where SS, RS, Min and  $R$  represent the standardized score, raw score, minimum score, and range of scores, respectively.

Chinese versions of FACT-B and QLQ-BR53 [19, 21] were used simultaneously in order to compare one with the other. Psychometric properties of the instrument were subsequently analyzed. Construct validity was evaluated by calculating the Pearson correlation coefficient,  $r$ , among items and domains as well as factor analysis. Internal consistency reliability was evaluated using Cronbach's alpha coefficient for each domain, and test–retest reliability through calculating the Pearson correlation coefficient between the first and second assessment as well as intraclass correlation (ICC) defined based on absolute agreement with single measure under the two-way mixed model [7, 10]. Responsiveness was assessed through comparing the mean difference between the pretreatment and post-treatment with effect size, standardized response mean (SRM) [6, 13].

## Results

The 186 patients with breast cancer varied in age from 16 to 78, with a median age of 48.0 and mean age  $48.5 \pm 10.1$ ; 49 (26.3%) patients finished primary school, while 95 (51.1%) completed high school, and 40 (21.5%) had a college degree; ethnics of Han is 163 cases (87.6%) while others is 22 (11.8%); distributions of occupations are worker 55 cases (29.6), farmer 24 (12.9%), teacher 25 (13.9%), cadre 37 (19.9%), others 45 (24.2%).

#### Content validity

By reviewing the literature and consulting some practical panels, it was agreed that the item pool well represented WHO's concept about QOL [14] and the specific aspects of the patients with breast cancer. This was facilitated by use of the programmed decision method for item selection, which produced a scale with good content validity.

#### Construct validity

Correlational analyses showed that there were strong correlations between items and their own domains (most correlation coefficients are higher than 0.5), but weak correlations between items and other domains (see Table 2 in detail).

**Table 2** Correlation coefficients among items and domains of QLICP-BR ( $n=186$ )

Items	PHD	PSD	SOD	SSD	SPD
GPH1	<b>0.72</b>	0.04*	0.02*	0.02*	0.19
GPH2	<b>0.60</b>	0.00*	0.03*	0.04*	0.15
GPH3	<b>0.18</b>	0.39	0.26	0.34	0.13*
GPH4	<b>0.73</b>	0.01*	0.04*	0.07*	0.01*
GPH5	<b>0.39</b>	0.07*	0.09*	0.09*	0.21
GPH6	<b>0.73</b>	0.10*	0.14*	0.01*	0.19
GPH7	<b>0.61</b>	0.11*	0.14*	0.03*	0.34
GPS1	0.24	<b>0.47</b>	0.24	0.16	0.48
GPS2	0.12*	<b>0.66</b>	0.27	0.42	0.03*
GPS3	0.05*	<b>0.63</b>	0.13*	0.31	0.07*
GPS4	0.00*	<b>0.71</b>	0.37	0.40	0.12*
GPS5	0.07*	<b>0.64</b>	0.17	0.32	0.08*
GPS6	0.01*	<b>0.74</b>	0.27	0.34	0.10*
GPS7	0.02*	<b>0.74</b>	0.30	0.37	0.15*
GPS8	0.19	<b>0.56</b>	0.13*	0.27	0.16
GPS9	0.26	<b>0.34</b>	0.22	0.12*	0.13*
GPS10	0.05*	<b>0.42</b>	0.02*	0.27	0.37
GPS11	0.07*	<b>0.56</b>	0.36	0.43	0.14*
GPS12	0.40	<b>0.26</b>	0.17	0.02*	0.42
GSO1	0.27	0.11*	<b>0.49</b>	0.04*	0.05*
GSO2	0.27	0.05*	<b>0.59</b>	0.08*	0.10*
GSO3	0.27	0.25	<b>0.66</b>	0.20	0.24
GSO4	0.01*	0.15	<b>0.66</b>	0.06*	0.15
GSO5	0.05*	0.32	<b>0.52</b>	0.25	0.15
GSO6	0.01*	0.40	<b>0.52</b>	0.27	0.17
GSS1	0.09*	0.35	0.16	<b>0.70</b>	0.32
GSS2	0.09*	0.37	0.27	<b>0.74</b>	0.07*
GSS3	0.15	0.27	0.25	<b>0.34</b>	0.51
GSS4	0.05*	0.25	0.05*	<b>0.48</b>	0.03*
GSS5	0.10*	0.11*	0.14*	<b>0.56</b>	0.20
GSS6	0.14*	0.30	0.18	<b>0.41</b>	0.29
GSS7	0.19	0.30	0.02*	<b>0.53</b>	0.07*
SBR1	0.06*	0.03*	0.11*	0.22	<b>0.53</b>
SBR2	0.11*	0.14*	0.11*	0.18	<b>0.74</b>
SBR3	0.21	0.19	0.25	0.21	<b>0.73</b>
SBR4	0.12	0.26	0.21	0.27	<b>0.71</b>
SBR5	0.27	0.34	0.19	0.33	<b>0.62</b>
SBR6	0.06*	0.09*	0.03*	0.08*	<b>0.47</b>
SBR7	0.16	0.40	0.22	0.32	<b>0.46</b>

\*No statistical significance ( $p>0.05$ ), others have statistical significance ( $p<0.05$ )

There were nine principal components (the initial Eigenvalues  $>1$ ) abstracted from 32 items of the general module (QLICP-GM) by factor analysis, accounting for 74.49% of the cumulative variance. By using the Varimax rotation method, it can be seen that the nine principal components reflected nine different facets under four domains of the general module. The first and sixth principal components mainly represented the psychological domain with higher loadings on GPS1 (0.79), GPS2 (0.80), GPS3 (0.71), GPS4 (0.74), GPS5 (0.63), GPS6 (0.83), GPS8 (0.85), and GPS12 (0.69). The second and fifth principal

components mainly represented the physical domain with higher loadings on GPH2 (0.72), GPH4 (0.72), GPH5 (0.78), GPH6 (0.82), and GPH7 (0.75). The third and seventh principal components mainly represented the social domain with higher loadings on GSO1 (0.79), GSO2 (0.71), GSO3 (0.74), and GSO4 (0.84). The fourth, eighth, and ninth principal components mainly represented the common symptom and side effect domain with higher loadings on GSS1 (0.78), GSS2 (0.64), GSS3 (0.77), and GSS5 (0.63).

Similarly, the principal component factor analysis extracted three principal components from the seven items of the specific module with the cumulative variance of 74.96%, reflecting three facets of this module. And here the first principal component represented the facet of physical and psychological effect with higher factor loadings on SBR6 (0.89) and SBR7 (0.89), the second principal component represented the facet of breast symptom with higher factor loadings on SBR1 (0.61), SBR2 (0.85), and SBR3 (0.80), the third principal component represented the facet of upper body effects with higher factor loadings on SBR4 (0.79) and SBR5 (0.89).

From the results above, theoretical construct was confirmed by data analysis, showing good construct validity.

#### Criterion-related validity

In this study, we chose Chinese versions of FACT-B and QLQ-BR53 as the criteria for assessing criterion-related validity because of the lack of an agreed-upon gold standard. Correlation coefficients of scores among the domains of the QLICP-BR and FACT-B can be seen from our other article [21] and the ones among the domains of the QLICP-BR and QLQ-BR53 were presented in Table 3, with both showing that overall the correlations between the same and similar domains are higher than those between different and nonsimilar domains. For example, the coefficient between the psychological domain of QLICP-BR and emotional functioning of QLQ-BR53 was 0.64, higher than any other coefficients in this column, e.g., physical functioning (0.09), role functioning (0.29). Also the specific domain of QLICP-BR and breast symptoms of QLQ-BR53 has the biggest correlation coefficient (−0.59) in this column. These confirmed the criterion-related validity to a reasonable degree and also demonstrated the convergent and divergent validity to some extent.

#### Reliability

The reliability of the scale was evaluated by two procedures: test–retest and internal consistency Cronbach's  $\alpha$ .

The test–retest correlation coefficients ( $r$ ) for the five domains and 12 facets of QLICP-BR ranged from 0.72 to 0.91, with the overall scale 0.88. The results from ICC and

**Table 3** Correlation coefficients of scores among subscales of QLQ-BR53 and domains of QLICP-BR

QLQ-BR53 subscales/items	QLICP-BR domains				
	PHD	PSD	SOD	SSD	SPD
<b>QLQ-C30</b>					
Physical functioning	0.16	0.09*	0.01*	0.19	0.08*
Role functioning	0.13*	0.29	0.16	0.38	0.28
Emotional functioning	0.06*	<b>0.64</b>	0.24	0.47	0.39
Cognitive functioning	0.00*	0.47	0.21	0.44	0.42
Social functioning	0.01*	0.35	0.33	0.43	0.32
Global health status/QOL	0.23	0.20	0.22	0.27	0.04
Fatigue	-0.21	-0.40	-0.08	-0.53	-0.38
Nausea and vomiting	0.04*	-0.33	-0.21	-0.54	-0.34
Pain	-0.14	-0.34	-0.09	-0.51	-0.54
Dyspnea	-0.03*	-0.40	-0.12	-0.41	-0.36
Insomnia	-0.11*	-0.36	-0.10*	-0.50	-0.27
Appetite loss	-0.05*	-0.28	-0.07*	-0.45	-0.28
Constipation	0.16	-0.13*	0.00*	-0.30	-0.18
Diarrhea	-0.06*	-0.18	0.00*	-0.24	-0.13*
Financial difficulties	-0.05*	-0.34	-0.40	-0.31	-0.21
<b>QLQ-BR23</b>					
Body image	0.16	0.49	0.34	0.46	0.38
Sexual functioning	0.23	0.15*	0.21	0.06*	-0.01*
Sexual enjoyment	0.19	0.14*	0.26	0.04*	-0.09*
Future perspective	0.08*	0.54	0.25	0.39	0.26
Systematic therapy side effects	-0.10*	-0.48	-0.25	-0.58	-0.39
Breast symptoms	-0.04*	-0.35	-0.28	-0.45	<b>-0.59</b>
Arm symptoms	-0.14*	-0.39	-0.19	-0.50	-0.53
Upset by hair loss	0.02*	-0.35	-0.35	-0.42	-0.33

\*no statistical significance ( $p > 0.05$ ), others have statistical significance ( $p < 0.05$ )

their 95% confidence intervals computed based on the definition of absolute agreement for a single measure were very similar to Pearson's correlation coefficients ( $r$ ). The score differences of these domains and facets between the first and the second measurements were not statistically significant ( $p > 0.05$ ) by paired  $t$  tests.

Cronbach's  $\alpha$  for these domains and facets ranged from 0.58 to 0.90, with that of all domains being greater than 0.60 except of social domain (0.58), see Table 4 in detail.

### Responsiveness

A classical paired  $t$ -test with responsiveness indicator, SRM, was used to examine statistically significant changes of mean scores from each domain of the QLICP-BR between the assessments before and after treatment, with the results presenting in Table 5. It can be seen that three of five domains and the overall instrument were of statistical significance except of SSD and SPD.

### Discussions

The development of a QOL instrument is a lengthy process. This paper focused on the main steps of developments and validation of the QLICP-BR.

Considering same-class diseases such as cancer often share many things in common, a popular approach in recent years has been to develop a general module for a class of diseases and then additional modules to capture individual differences in different people and diseases. Since the add-on modules are much simpler, this approach can substantially reduce the amount of time and effort in developing new instruments. The QLQs from EORTC and the FACTS from Center on Outcomes, Research and Education (CORE) have been developed based on this modular principle [1, 3, 4, 11]. By building upon existing instruments for cancer patients and Chinese culture, we employed this modular approach to systematically and more efficiently develop a system of instruments for cancer patients, QLICP, with QLICP-BR being one instrument of this system [15–18, 20].

On psychometric properties of a QOL instrument, it must be validated with respect to at least three aspects: validity, reliability, and responsiveness.

Validity is the extent to which an instrument can capture what it purports to measure. By following WHO's definition of QOL [14] and the programmed decision procedures, we developed the QLICP-BR for patients with breast cancer by use of multiple turns of focus group discussion, in-depth interview, and pretesting to effectively reduce the number of items in the final version to 32 from a 78-item pool for



**Table 4** Reliability of the quality of life instrument QLICP-BR ( $n=186$  for  $\alpha$ ,  $n=166$  for  $r$  and ICC)

Domains/Facets	Test–retest reliability Pearson's $r$	Test–retest reliability ICC (95% CI)	Internal consistency coefficient $\alpha$
Physical domain (PHD)	0.76	0.74 (0.67–0.81)	0.63
Basic physiologic function (BPF)	0.81	0.80 (0.74–0.85)	0.74
Sexual function (SXF)	0.87	0.83 (0.77–0.87)	–
Independence function (IDF)	0.72	0.71 (0.63–0.78)	0.62
Psychological domain (PSD)	0.89	0.89 (0.85–0.91)	0.87
Emotion (EMO)	0.89	0.89 (0.85–0.92)	0.90
Recognition (REC)	0.85	0.83 (0.78–0.87)	0.59
Social domain (SOD)	0.91	0.91 (0.88–0.93)	0.58
Social support and safety (SSS)	0.88	0.87 (0.83–0.91)	0.69
Effect on life & economics (ELE)	0.89	0.89 (0.86–0.92)	0.60
Common symptom and side effect domain (SSD)	0.88	0.87 (0.83–0.91)	0.72
Side effect (SEF)	0.90	0.88 (0.84–0.91)	0.68
Common symptom (CST)	0.85	0.86 (0.81–0.89)	0.51
Specific domain (SPD)	0.84	0.85 (0.80–0.89)	0.65
Breast symptom (BRS)	0.90	0.90 (0.87–0.93)	0.58
Upper body effect (UBE)	0.80	0.80 (0.73–0.85)	0.66
Physical and psychological effect (PPE)	0.77	0.70 (0.62–0.77)	0.81
Core/general module (CGM)	0.87	0.87 (0.81–0.89)	0.82
Total (TOT)	0.88	0.86 (0.82–0.90)	–

–Not acceptable/suitable

the general module, and seven from 14 for the specific module, which ensured good content validity and conceptual structure of this instrument. Besides, correlation analyses and factor analysis confirmed the good construct validity and criterion-related validity.

Reliability refers to the reproducibility or consistency of item scores from one assessment to another. Test–retest

reliability (Pearson  $r$  and ICC), internal consistency reliability were applied in the current study. The test–retest correlation coefficient (Pearson  $r$ ) only reflects the consistency in tendency between two repeated tests. Therefore, in this paper, the paired  $t$  test was also used to compare the “difference” of scores between the first and second measurements. Given that no statistical significance being

**Table 5** Responsiveness of the quality of life instrument QLICP-BR ( $n=94$ )

Domains/facets	Before treatment		After treatment		Differences		$t$	$p$ value	SRM
	Mean	SD	Mean	SD	Mean	SD			
Physical domain (PHD)	45.99	16.74	42.30	17.92	3.69	16.98	2.10	0.038	0.22
Basic physiologic function (BPF)	38.03	22.44	36.17	23.00	1.86	20.57	0.88	0.383	0.09
Sexual function (SXF)	57.89	28.35	56.58	30.10	1.32	25.79	0.44	0.658	0.05
Independence function (IDF)	47.19	21.54	42.02	22.05	5.16	20.72	2.42	0.018	0.25
Psychological domain (PSD)	63.01	16.96	60.62	17.10	2.39	11.45	2.02	0.046	0.21
Emotion (EMO)	64.45	19.14	62.91	20.26	1.54	13.51	1.11	0.271	0.11
Recognition (REC)	58.95	17.80	54.17	17.78	4.79	16.33	2.84	0.005	0.29
Social domain (SOD)	61.01	17.13	58.10	18.00	2.91	13.91	2.03	0.045	0.21
Social support and safety (SSS)	66.58	19.01	61.30	20.76	5.27	16.79	3.05	0.003	0.31
Effect on life and economics (ELE)	49.60	28.52	51.21	27.71	–1.61	19.78	–0.79	0.434	0.08
Common symptom and side effect domain (SSD)	68.99	15.95	66.45	17.43	2.54	13.58	1.81	0.073	0.19
Side effect (SEF)	72.83	19.95	70.48	20.58	2.35	16.63	1.37	0.174	0.14
Common symptom (CST)	63.92	18.40	61.48	19.63	2.44	16.06	1.47	0.144	0.15
Specific domain (SPD)	73.14	15.39	72.81	14.60	0.33	11.97	0.25	0.806	0.03
Breast symptom (BRS)	79.11	19.28	81.82	17.05	–2.71	15.84	–1.53	0.130	0.17
Upper body effect (UBE)	67.75	23.87	66.51	24.35	1.23	20.60	0.54	0.591	0.06
Physical and psychological effect (PPE)	69.06	28.96	63.44	30.67	5.63	26.97	1.87	0.066	0.21
Core/general module (CGM)	60.00	11.83	57.28	11.83	2.72	9.13	2.89	0.005	0.30
Total (TOT)	61.89	11.21	59.60	11.31	2.29	8.51	2.61	0.011	0.27

found, the correlation coefficients are meaningful. In such cases, Pearson  $r$  and ICC can be used interchangeably to provide assessment of reliability, with the results close to each other. Based on the results, it can be inferred that QLICP-BR is of good reliability.

The assessment methods on responsiveness can be divided into two categories: internal and external [6, 13]. Internal responsiveness characterizes the ability of a measure to Change over a particular prespecified time frame. One widely used method of assessing internal responsiveness is to evaluate the change in a measure within the context of a randomized clinical trial involving a treatment that has previously been shown to be efficacious [2, 8]. External responsiveness reflects the extent to which changes in a measure over a specified time frame relate to corresponding changes in a reference measure of health status. In this paper we used the definition and method of internal responsiveness. Classical paired  $t$ -test was employed in this study to make mean-comparisons between the pretreatment and post-treatment assessment, also accompanying with an important responsiveness indicator, SRM, with values of 0.20, 0.50, and 0.80 having been proposed to represent small, moderate, and large responsiveness, respectively [6, 13]. Table 5 showed QOL scores changes after treatment were of statistical significance on three domains of physical function, psychological function, social function, and the overall instrument, with not higher SRM of 0.22, 0.21, 0.21, and 0.27. There are three possible reasons for common symptom/side-effects domain and specific domain which were not statistically significant: (1) the sample size may be not big enough ( $n=94$ ), (2) the interval between the measurements before and after treatment is not long enough (about 4 weeks), and (3) the score in this domain is of no change in nature. Therefore, it can be concluded that the instrument has reasonable responsiveness.

In addition, QLICP-BR has some highlights compared to Chinese versions of FACT-B and QLQ-BR53 although they have similar psychological properties [19, 21]. First, it is of strong Chinese culture background. For example, the Chinese culture pays more attention to the family relationship and kinship, eating and food, good temper, and high spirit. It includes some items focusing on these such as appetite, sleep, energy, family support etc. Second, unlike FACT-B and QLQ-BR53, it has a very clear hierarchical structure (items→ facets→ domains→ overall). It can present mean scores not only at the domains (five domains) and the overall level but also at the facet levels (12 facets) so that it can detect changes in more details. The users can select either one level or both.

To sum up, the QLICP-BR can be used as a useful instrument in measuring and assessing quality of life for patients with breast cancer in China, with good psychological properties and some highlights.

**Acknowledgements** In carrying out this research project, we have received substantial assistance from Neil Aaronson, Mirjam Sprangers, Christiane van Pottelsberghethe, Karen West and Linda Dewolf at EORTC, from David Cella, Sonya Eremenco, Benjamin J. Arnold and Hiramatsu Toshiko at CORE, Evanston–Northwestern Healthcare, and also from staffs of the Yunnan Tumor Hospital, and Keith J Bourgeois and Xin Tu at the University of Rochester. We sincerely appreciate their kind help.

## References

- Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Fibert A, Flechtner H, Fleishman SB, de Haes JC, Kaasa S, Klee M (1993) The European Organization for Research and Treatment of Cancer QLQ-C30: a quality of life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 85:365–376
- Anderson JJ, Chernoff MC (1993) Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. *J Rheumatol* 20:535–537
- Brady MJ, Cella DF, Mo F, Bonomi AE, Tulskey DS, Lloyd SR, Deasy S, Cobleigh M, Shiomoto G (1997) Reliability and validity of the functional assessment of cancer therapy-breast quality-of-life instrument. *J Clin Oncol* 15(3):974–986
- Cella DF, Tulskey DS, Gray G, Sarafian B, Linn E, Bonomi A, Silberman M, Yellen SB, Winicour P, Brannon J (1993) The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol* 11(3):570–579
- Hunt SM, Mckznna SP, McEwen J, Williams J, Papp E (1981) The Nottingham Health Profile: subjective health status and medical consultations. *Soc Sci Med* 15:221–229
- Husted JA, Cook RJ, Farewell VT, Gladman DD (2000) Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 53:459–468
- McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1(1):30–46
- Norman GR, Stratford P, Regehr G (1997) Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 50:869–879
- Selby PJ, Champon JA, Etazadi-Amoli J, Dalley D, Boyd NF (1984) The development of a method for assessing the quality of life of cancer patients. *Br J Cancer* 50(1):13–22
- Schuck P (2004) Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? *Qual Life Res* 13:571–586
- Sprangers MA (1996) The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. *J Clin Oncol* 14(10):2756–2768
- Stewart AL, Hays RD, Ware JE Jr (1988) The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* 26:724–735
- Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM (2003) On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 12:349–363
- The WHOQOL Group (1998) The World Health Organization Quality of Life assessment (WHOQOL): development and psychometric properties. *Soc Sci Med* 46(12):1569–1585
- Wan C, Luo J, Yang Z, Meng Q, Zhang X, Lu Y, Tang X, Zhang C (2007) Development of the general module of the system of quality of life instruments for cancer patients: responsiveness analysis (in Chinese). *Chin J Cancer* 26(4):337–340

16. Wan C, Luo J, Zhang C, Tang X, Chen M, Song Y (2003) Study on the system of quality of life instruments for cancer patients (QLICP) (in Chinese). *Chin J Behav Med Sci* 12(3):341–342
17. Wan C, Meng Q, Luo J, Tang X, Zhang C, Lu Y, Yang Z, Zhang X (2007) Development of the general module of the system of quality of life instruments for cancer patients: items selection and structure of the general module (in Chinese). *Chin J Cancer* 26(2):113–117
18. Wan C, Meng Q, Tang X, Zhang C, Lu Y, Luo J, Yang Z, Zhang X (2007) Development of the general module of the system of quality of life instruments for cancer patients: reliability and validity analysis (in Chinese). *Chin J Cancer* 26(3):225–229
19. Wan C, Tang X, Tu X, Feng C, Messing S, Meng Q, Zhang X (2007) Psychometric properties of the simplified Chinese version of the EORTC QLQ-BR53 for measuring quality of life for breast cancer patients. *Breast Cancer Res Treat* 105(2):187–193
20. Wan C, Yang Z, Meng Q, Feng C, Wang H, Tang X, Zhang C, Lu Y, Luo J, Zhang X (2008) Development and validation of the general module of the system of quality of life instruments for cancer patients (QLICP-GM). *Int J Cancer* 122(1):190–196
21. Wan C, Zhang D, Yang Z, Tu X, Tang W, Feng C, Wang H, Tang X (2007) Validation of the simplified Chinese version of the FACT-B for measuring quality of life for patients with breast cancer. *Breast Cancer Res Treat* 106(3):413–418