





Development and Validation the Uterine Fibroid Symptom and Quality of Life Short-Form: Integrating Classical and Modern Test Theory

Yubo Zhang ¹, Wei Xu ², Yuxian Nie¹, Jingyu Zhang¹, Dengfeng Chen², Chenxi He², Fan Xu ³, Jin Bai¹, Qiuling Shi ^{1,2}

¹State Key Laboratory of Ultrasound in Medicine and Engineering, College of Biomedical Engineering, Chongqing Medical University, Chongqing, People's Republic of China; ²School of Public Health, Chongqing Medical University, Chongqing, People's Republic of China; ³Department of Gynecology, Shanghai First Maternity and Infant Hospital, School of Medicine, Tongji University, Shanghai, People's Republic of China

Correspondence: Jin Bai; Qiuling Shi, State Key Laboratory of Ultrasound in Medicine and Engineering, College of Biomedical Engineering, Chongqing Medical University, No. 1 Yixueyuan Road, Yuzhong District, Chongqing, 400016, People's Republic of China, Tel +86 182 9058 5397, Email sajinbai@cqmu.edu.cn; qshi@cqmu.edu.cn

Purpose: To develop a shorter version of the uterine fibroid symptom and quality of life (UFS-QoL) questionnaire and evaluate its psychometric properties, aiming to provide a more efficient instrument for symptom and quality-of-life assessment.

Patients and Methods: Items were selected using classical test theory (CTT) and item response theory (IRT), and the standardized response mean was used to identify inconsistent items and to evaluate their ability to detect change. The short-form was evaluated for dimensionality, internal consistency reliability, and criterion and known-group validity. Data were obtained from a 20-center prospective cohort of premenopausal Chinese women with symptomatic uterine fibroids who underwent hysterectomy, myomectomy, or HIFU (n = 1939; development n = 969, validation n = 970).

Results: To improve adaptability and clinical application, we removed the self-consciousness subscale and retained only one item in the sexual functioning subscale. The resulting 11-item UFS-QoL (UFS-QoL-11) reduced the administrative burden by 70% (from 37 items to 11 items) compared to the original version. It exhibited a strong correlation with UFS-QoL in each domains (all exceeding 0.7). Exploratory factor analysis accounted for 64.22% of total variance, slightly higher than the original scale (63.60%). UFS-QoL-11 demonstrated excellent internal consistency and reliability, and sensitivity in detecting varying levels of current health status. The responsiveness of UFS-QoL-11 was comparable to that of UFS-QoL, with all effect sizes > 0.45.

Conclusion: The UFS-QoL-11 demonstrated comparable psychometric performance to the original scale while substantially reducing the number of items. It offers a more efficient instrument for assessing symptom severity and quality of life and shows potential for use in clinical and research settings, although further validation in independent populations is warranted.

Keywords: classical test theory, item response theory, responsiveness, Chinese women, reliability

Introduction

Uterine fibroids (UFs), also known as uterine smooth muscle tumors or uterine myomas, are the most common benign tumors in the female reproductive system. Among women of childbearing age, the overall prevalence rate is approximately 60%-70%, and about 25% of these patients have obvious clinical symptoms that require treatment.^{1,2} Patients with uterine fibroids, particularly premenopausal women undergoing medical or uterus-preserving treatment, often have a significant burden of symptoms, including moderate to severe abdominal pain,³ low back pain, urinary frequency and urgency, pain during intercourse,² and vaginal bleeding.⁴ Notably, several of these presenting symptoms are not specific to fibroids and may overlap with malignancy.⁵ Furthermore, fibroids can distort uterine anatomy, which may potentially obscure visualization or delay the early diagnosis of aggressive cancer subtypes.⁶ Given the high prevalence and symptom burden among women, UFs impose a significant healthcare burden on women's individual health as well as



on the healthcare and social security systems as a result of work productivity loss during treatment and disease recurrence.^{7,8} Consequently, measuring UF-related symptoms and QoL status in a valid and reliable manner could support high-quality practice and comprehensive patient management.

As the only questionnaire designed to assess the full spectrum of fibroid-related symptoms and their impact on QoL, the English version of the Uterine Fibroid Symptom and Quality of Life (UFS-QoL) questionnaire, published in 2002,⁹ is applicable only to women with uterine fibroids who have retained their uterus and still have a normal menstrual cycle, and not to those who have undergone hysterectomy. The UFS-QoL questionnaire consists of 37 items, eight of which assess the severity of symptoms (single domain) and 29 assess health-related quality of life (HRQL) in six subscales (concern, activities, energy/mood, control, self-consciousness, and sexual function). All the responses were classified into five Likert-scale options. The UFS-QoL questionnaire has been used and validated in Brazilian Portuguese,^{10,11} Spanish,¹² traditional Chinese,¹³ and simplified Chinese^{14,15} as a disease-specific measure of health-related QoL.

For an instrument to be clinically useful, an obstacle to its routine use is its length. Although the 37 items of the Chinese UFS-QoL—which contains the same number of items as the original English version—assess overall patient information, the length might deter its clinical efficacy, as it is time-consuming for daily high-frequency care, especially in primary hospitals with low resource settings. Shorter questionnaires (one that achieves the same effect but with fewer items) that improve response rates and reduce participant burden are recommended as follow-up measures.^{16,17} Our group evaluated the adaptability and clinical applicability of the Chinese UFS-QoL.¹⁵ We found that the symptom severity, activity, and mood subscales of the Chinese UFS-QoL scale were valid and reliable. However, psychometric evidence revealed that the self-consciousness and sexual functioning subscales exhibited ceiling effects > 15%. Notably, the self-consciousness domain showed poor internal consistency ($\alpha = 0.56$) and demonstrated the lowest detectability, with both effect size and standardized response means being only 0.38 at 6- and 12-months post-treatment. We pointed out before applying the UFS-QoL to Chinese patients, the cross-cultural test criteria of adjusting the subscales to the Chinese culture, shortening or deleting the poor discrimination items to ensure the scale validity, and further optimizing and reevaluating the practicability of these items, should be considered. Therefore, there are both practical application challenges and psychometric evidence to pursue the construction of a shorter version of the UFS-QoL while maintaining its reliability, validity, and clinical applicability and enhancing routine clinical assessments.

In this study, we employed different psychometric techniques from both classical test theory (CTT) and modern test theory (item response theory [IRT])¹⁸ to develop a short-form version of the UFS-QoL. While CTT identifies the basic properties of the data, such as eigenvalues and factor loadings, IRT provides more detailed information at the item level. These techniques complement each other when developing short scales.^{18,19} Using the short-form version of the UFS-QoL, we designed it to reduce the respondent burden and be short enough for use in clinical practice and research. We predicted that the short-form version would have similar psychometric properties and reproduce scores that would be useful in assessing the efficacy of therapies without the loss of information compared with the original 37-item UFS-QoL.

Materials and Methods

To develop the short-form scale, we extracted data from a prospective cohort study that included 20 centers (Uterine Fibroids Multicenter Network Information System: www.hifuctr.com) for a secondary analysis, which included patients who underwent self-selected hysterectomy, myomectomy, or high-intensity focused ultrasound (HIFU) therapy after being fully informed of the treatment options (the multicenter study was approved by a China-registered clinical trial ethics committee (ChiECRCT-2011034)).²⁰ Details regarding the study design, data collection, and primary outcomes regarding the efficacy of the treatment have been published.²⁰ Prior to undergoing any study-related procedures at the clinical site, the patients completed the UFS-QoL questionnaire, the study short-form-36 (SF-36), and a brief socio-demographic questionnaire. Follow-up visits were scheduled at 6 and 12 months post-surgery, including complications, magnetic resonance imaging evaluation, overall treatment effect evaluation, the UFS-QoL questionnaire (for those who had undergone HIFU or myomectomy, because the instructions of the UFS-QoL questionnaire were based on the presence of uterine fibroids and menstrual periods), SF-36, and several health care utilization items were recorded.

Patient-Reported Outcome Measures

Uterine Fibroid Symptom and Quality of Life Questionnaire (UFS-QoL)

The UFS-QoL was developed from focus groups of women with uterine fibroids.^{9,14} The Original UFS-QoL questionnaire consists of 37 items, 8 of which assess the severity of symptoms (single domain) and 29 of which assess HRQL with six subscales (concern, activities, energy/mood, control, self-consciousness, and sexual function). All the responses were classified into five Likert-scale options. A higher score on the questionnaire's severity subscale indicates more severe symptoms, whereas a lower score on the HRQL subscales indicates poorer QoL.

Medical Outcomes of the Study Short-Form 36 (SF-36)

The Chinese SF-36 is a 36-item self-administered generic measure used to assess general health status and validated its cross-cultural application, reliability and validity.^{21–24} SF-36 consists of eight subscales: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health, as well as two composite scores: physical and mental. Individual subscale items were combined to form a subscale rating, which was then converted to a 0–100 scale.^{25,26} Higher QoL scores correspond to a four-week recall period.²⁶

Item Selection

We selected two items per UFS-QoL scale using a method that maximizes content validity.^{16,27} Due to the poor internal consistency and ability to detect self-consciousness in our previous study,¹⁵ we did not retain the three items of self-consciousness. At the same time, because sexual functioning showed poor adaptability and high correlation between the two items, we only kept one item of the two items. We used the CTT and IRT to select two items for each of the five subscales. For the inconsistent items extracted from the two methods, we used responsiveness to choose items with higher ability to detect changes in clinical therapy. Therefore, we administered the 11-item version (UFS-QoL-11) to test its factor structure, criterion-related validity, and responsiveness, following the recommendations for short-form scale development. A flow chart of item selection is shown in Figure 1.

Based on the CTT to choose shortened items, we followed a method using regression analysis that maximizes the content validity of a 2-item scale. The first item had the highest correlation with the original scale. The second item selected was the one with the highest beta weight of the remaining items in multiple regression, with the original scale score as the dependent variable and individual items as predictor variables.

Item response theory offers detailed information at the item level, making it a powerful technique for developing short-form scales.¹⁸ The responses of participants to each item were used to estimate their location on a latent trait (ie, the level of symptoms). This latent trait is estimated using the responses on all individual items, and not, as in the CTT, via the sum score of all items. Within the IRT, every single item is defined by a discrimination parameter (alpha) and one or more location parameters or threshold parameters. The threshold parameters indicate the location on the scale of the

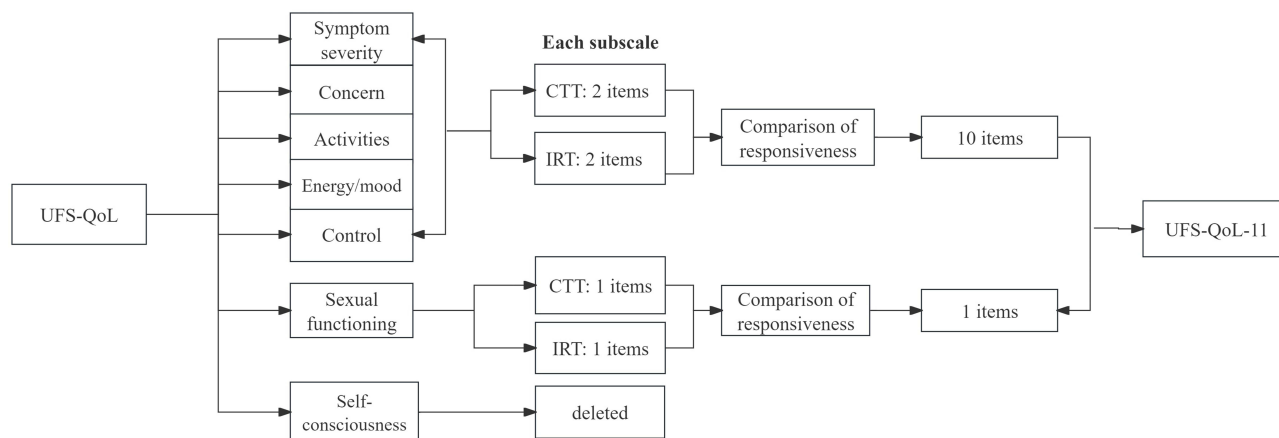


Figure 1 Flow charts of items selection.

latent continuum where the item best discriminates among individuals. The discrimination parameter reflects the true difference in theta per item and is comparable to factor loading.

To determine the candidate items for the short-form of the scale, we selected the items with two of the highest discriminatory parameters, as given by the IRT analyses. As the scale has five ordered response options (1 = “not at all”, 2 = “A little bit”, 3 = “Somewhat”, 4 = “A great deal”, 5 = “A very great deal”), we used the graded response model to estimate the item response parameters. In addition, we inspected the item information curves to select items that covered a similar range of latent traits as the full scale. Finally, we compared the effect sizes of inconsistent items to choose the final items.

Statistical Analysis

Spearman correlation coefficient was used to determine the strength of the UFS-QoL-11 correlation test with the original UFS-QoL. To examine construct validity in the UFS-QoL-11, we employed principal axis factor analysis with orthogonal rotation, which was used in conjunction with orthogonal rotation to determine the final number of factors based on their eigenvalues, congruence, and clinical significance. Each dataset was first analyzed for normality using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett test of sphericity. A KMO value > 0.5 indicated acceptable structural validity. Cronbach’s α was used to determine the internal consistency of the UFS-QoL-11. Cronbach’s α coefficient, which ranges from 0 to 1, was used to determine the degree to which items on the subscales measure were related to the same concepts. A larger value indicated a smaller measurement error, which indicated a higher level of reliability. Criterion validity was assessed using correlations between baseline scores on the UFS-QoL-11 and SF-36, which were designed to measure general health.

We examined the evidence of known-group validity so that the UFS-QoL-11 could distinguish between clinically distinct groups by testing its ability to differentiate between patients based on health status (defined as SF-36-1). The ability to detect change was evaluated by comparing the 6-month pre-treatment and post-treatment scores to the 12-month scores at 6-month intervals. The effect size (change in mean score divided by baseline standard deviation)²⁸ and standardized response mean (change in mean score divided by change standard deviation) were computed. A value of 0.2 was considered to have a “small” effect, 0.5 a “moderate” one, and ≥ 0.8 a “large” effect.

Descriptive analyses (means and standard deviations [SD]) were performed using sociodemographic and clinical characteristics. Means of differences, 95% confidence intervals (CI), and statistical significance ($P < 0.05$) were tested using independent sample t-tests. The questionnaires were scored according to developers’ instructions. Version 9.1.3 of SAS was used to conduct the analyses. All statistical tests were predetermined, and no missing data imputations were performed. All statistical tests were conducted with a fixed type I error probability of 0.05 and a two-tailed design.²⁵

Results

According to the screening criteria, we chose items based on the CTT and IRT results. All assumptions of IRT were met (ie, the scale is unidimensional, and there was no noteworthy local dependence and answer authenticity) (see [Supplements Table S1-S2](#)). As shown in [Table 1](#), the CTT and IRT parameters per item included item–total correlations ranging from 0.483 to 0.841 and IRT discrimination parameters ranging from 0.841 to 3.709, indicating adequate item discrimination ability. The control and sexual function subscales showed consistent results between the two methods. [Table 2](#) demonstrates a shortened scale with nearly identical responsiveness to original UFS-QoL. By comparing the effect size and standardized response means (SRM) in other subscales, the symptom severity and activities subscale showed higher responsiveness to CTT items. However, Concern and Energy/Mood showed a higher ability to detect changes in IRT items. As shown in [Table S3](#), all items had a distribution of $< 1\%$ at every follow-up. The final UFS-QoL-11 questionnaire is presented in [Supplementary information, Figure S1](#). The scoring method for the UFS-QoL-11, adapted from the original UFS-QoL scoring manual, is provided in the [Supplementary Method](#). In [Figure 2](#), the selected items showed strong correlations with the corresponding domains of the original UFS-QoL scale, with correlation coefficients ranging from 0.73 to 0.94 (all $P < 0.001$).

In the factor analysis, values of the KMO test > 0.7 , and the statistical significance of Bartlett’s test indicated adequate sampling.^{29,30} We used orthogonal rotation to isolate the potential UFS-QoL-11 factors ([Table 3](#)). In this study, the value

Table 1 Items Selection with CTT and IRT in Development Data Set (n =969)

Subscales		CTT		IRT [¶]				
		r _P [†]	Linear Regression Model	Beta1	Beta2	Beta3	Beta4	Dscrmn
Symptom severity	U_1	0.724*		-0.385	0.476	1.527	2.826	1.975*
	U_2	0.658	4.794*	-0.690	0.862	2.212	3.450	1.708
	U_3	0.656	3.966	0.096	1.105	2.296	3.255	1.780*
	U_4	0.614	3.35	0.229	1.488	2.596	3.765	1.454
	U_5	0.483	3.222	-0.176	2.359	4.589	7.782	0.841
	U_6	0.495	3.155	0.814	2.657	4.052	6.696	1
	U_7	0.521	3.189	0.887	2.573	4.002	5.646	1.075
	U_8	0.646	3.447	0.006	1.432	2.561	3.916	1.518
Concern	U_9	-0.592	-5.635	-0.873	0.536	2.988	4.774	0.955
	U_15	-0.841*	/	-0.704	0.062	1.089	1.691	3.409*
	U_22	-0.836	-7.609*	-0.692	-0.005	1.068	1.513	3.563*
	U_28	-0.702	-5.241	-0.585	0.514	1.939	2.724	1.485
	U_32	-0.795	-6.171	-0.595	0.354	1.490	1.981	2.404
Activities	U_10	-0.686	-3.56	0.108	0.898	2.323	3.023	1.976
	U_11	-0.765	-5.33*	-0.205	0.679	1.772	2.494	2.288
	U_13	-0.786*	/	-0.398	0.528	1.560	2.336	2.344*
	U_19	-0.708	-4.173	-0.559	0.731	2.269	3.094	1.878
	U_20	-0.751	-4.428	-0.028	0.936	2.197	2.962	2.580*
	U_27	-0.709	-3.904	-0.593	0.430	1.839	2.684	1.513
	U_29	-0.63	-3.547	-0.518	0.961	2.866	3.637	1.315
Energy/Mood	U_12	-0.753	-4.377	-0.821	0.275	1.789	2.845	1.922
	U_17	-0.706	-4.009	-1.006	0.188	2.099	3.188	1.639
	U_23	-0.754	-3.608	-0.271	0.708	1.987	2.876	2.926
	U_24	-0.814	-5.107*	-0.444	0.555	1.840	2.734	3.704*
	U_25	-0.827*	/	-0.443	0.463	1.728	2.618	3.005*
	U_31	-0.726	-3.663	-0.940	0.277	1.992	2.879	1.837
	U_35	-0.731	-4.267	-0.212	0.881	2.162	3.076	2.194
Control	U_14	-0.646	-6.098	0.244	1.357	2.705	3.401	1.920
	U_16	-0.724	-5.924	-0.416	0.662	2.356	3.704	1.664
	U_26	-0.658	-5.454	-2.347	-0.962	1.059	2.631	0.947
	U_30	-0.734	-6.789*	-0.269	0.807	2.209	2.601	2.124*
	U_34	-0.76*	/	-0.152	0.895	2.046	2.743	2.885*
Sexual function	U_36	-0.804*	/	-0.569	0.157	1.321	2.050	3.709*
	U_37	-0.658	/	-0.454	0.265	1.306	1.931	/

Notes: *Items are retained for the SF-UFS-QoL based on the CTT and IRT. [†] r_P correlation coefficient for Spearman correlation test between UFS-QoL subscale score and each item. [¶]Item response theory analysis with graded response model (item discrimination parameter).

Abbreviations: CTT, classical test theory; IRT, item response theory.

of KMO was 0.86, and the value of Bartlett's sphericity test was 2536.26 ($P < 0.001$), indicating that factor analysis was suitable for the data. The UFS-QoL-11 retains three factors: symptom severity, concern, and daily life interference. Factor loadings ranged from 0.543 to 0.897, the test variables explained 64.22% of the total variance, which was slightly higher than the original scale score of 63.60%.¹⁵

Cronbach's alpha was calculated for each UFS-QoL-11 subscale to assess internal consistency. The UFS-QoL-11 questionnaire demonstrated good internal consistency and reliability for all the subscales (> 0.7) (Table 4). Cronbach's α values were 0.81 for symptom severity, 0.83 for concern, and 0.84 for daily life interference (Table 4). There was a small increase in internal consistency for the UFS-QoL-11 compared with the full UFS-QoL. The degree of correlation between similar subscales on the UFS-QoL-11 and SF-36 was used to assess the reliability of the correlation testing for the UFS-QoL-11 (Supplements Table S4).

Table 2 Responsiveness of Women with Uterine Fibroids After Treatment Among Original Scale, CTT and IRT in Development Data Set (n=969)

Subscales	Mean Score \pm SD			Change in Score \pm SD		Effect Size [‡]		SRM [†]	
	Preoperative (n=2411)	6-Month (n=1855)	12-Month (n=1774)	6-Month vs Preoperative	12-Month vs Preoperative	6-Month	12-Month	6-Months	12-Months
Symptom severity-Parent	18.28 \pm 13.97	9.29 \pm 9.62	7.15 \pm 9.00	8.99 \pm 12.04	11.14 \pm 11.86	0.64	0.8	0.75	0.94
Symptom severity-CTT*	26.68\pm22.37	12.02\pm13.73	9.53\pm12.26	14.65\pm18.61	17.14\pm18.19	0.65	0.77	0.79	0.94
Symptom severity-IRT	23.25 \pm 22.01	10.01 \pm 13.89	7.64 \pm 12.36	13.24 \pm 18.45	15.61 \pm 17.99	0.60	0.71	0.72	0.87
Concern-Parent	67.59 \pm 21.27	78.86 \pm 16.59	82.95 \pm 15.6	-11.27 \pm 19.12	-15.36 \pm 18.77	-0.53	-0.72	-0.59	-0.82
Concern-CTT/IRT	63.03 \pm 27.42	77.35 \pm 20.71	81.63 \pm 19.56	-14.32 \pm 24.34	-18.6 \pm 23.94	-0.52	-0.68	-0.59	-0.78
Activities-Parent	75.51 \pm 17.92	83.96 \pm 13.48	86.79 \pm 12.75	-8.46 \pm 15.9	-11.28 \pm 15.66	-0.47	-0.63	-0.53	-0.72
Activities-CTT	73.83 \pm 24.27	83.36 \pm 17.4	86.22 \pm 16.27	-9.53 \pm 21.15	-12.4 \pm 20.79	-0.39	-0.51	-0.45	-0.60
Activities-IRT*	76.25\pm21.7	84.99\pm15.64	87.18\pm14.84	-8.74\pm18.95	-10.93\pm18.69	-0.40	-0.50	-0.46	-0.58
Energy/mood-Parent	72.55 \pm 18.5	81.69 \pm 14.32	84.5 \pm 13.54	-9.14 \pm 16.58	-11.95 \pm 16.32	-0.49	-0.65	-0.55	-0.73
Energy/mood-CTT/IRT	73.72 \pm 21.86	82.15 \pm 17.6	84.47 \pm 16.2	-8.43 \pm 19.87	-10.75 \pm 19.33	-0.39	-0.49	-0.42	-0.56
Control-Parent	73.38 \pm 17.42	81.66 \pm 14.04	84.63 \pm 12.96	-8.28 \pm 15.85	-11.25 \pm 15.45	-0.48	-0.65	-0.52	-0.73
Control-CTT/IRT	78.13 \pm 20.91	85.88 \pm 16.23	87.79 \pm 15.07	-7.75 \pm 18.75	-9.65 \pm 18.32	-0.37	-0.46	-0.41	-0.53
Sexual functioning-Parent	69.57 \pm 22.14	80.35 \pm 20.92	84.1 \pm 19.09	-10.78 \pm 21.55	-14.53 \pm 20.73	-0.49	-0.66	-0.5	-0.7
Sexual functioning-CTT/IRT	69.45 \pm 26.08	79.17 \pm 22.8	82.79 \pm 20.96	-9.71 \pm 24.51	-13.34 \pm 23.74	-0.37	-0.51	-0.4	-0.56

Notes: *Items are retained for the UFS-QoL-11. [‡] Calculated as change in mean score divided by the standard deviation of the baseline. [†] Calculated as change in mean score divided by the standard deviation of the change.

Abbreviations: CTT, classical test theory; IRT, item response theory.

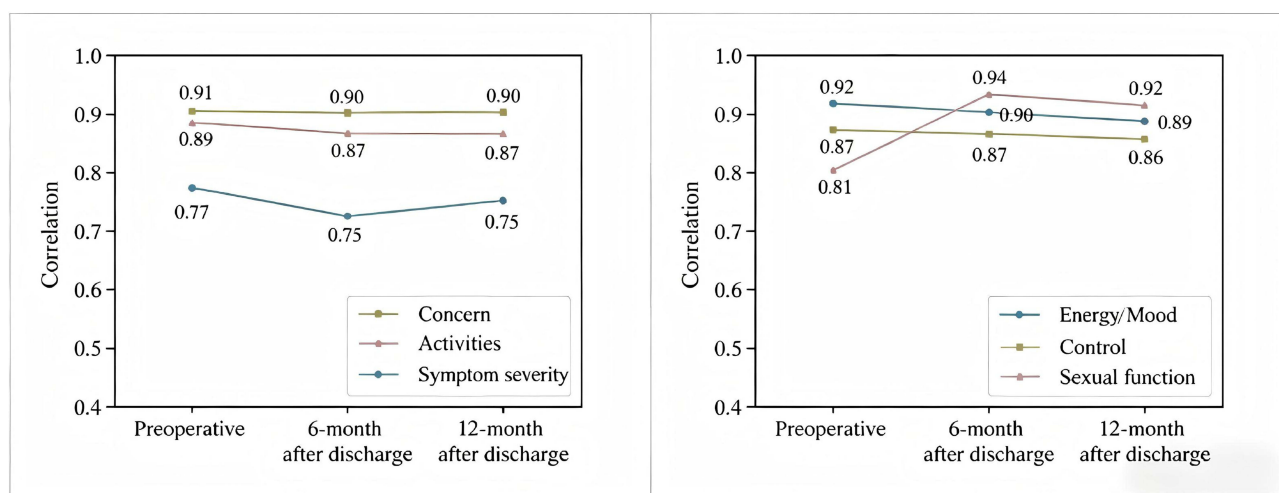


Figure 2 Correlation between selected items and original UFS-QoL in each subscale.

All domains of the SF-36 correlated positively and moderately with the “daily life interference” domain of the UFS-QoL-11 with correlation coefficients ranging from 0.35 to 0.51 (all $P < 0.001$).

The UFS-QoL-11 was sensitive enough to detect varying levels of current health status, particularly at 6 and 12 months post-surgery (Table S5). At 6 months, patients with good health had lower symptom severity scores than those

Table 3 Construct Validity of the UFS-QoL-11: Baseline Factor Loadings in Validation Data Set (n=970)

Subscales	Item	Factor 1	Factor 2	Factor 3
Daily life interference	U_34	0.792	0.026	0.072
	U_30	0.744	0.094	-0.012
	U_24	0.743	0.245	0.022
	U_25	0.727	0.257	0.057
	U_20	0.656	0.172	0.191
	U_36	0.617	0.117	0.125
	U_13	0.543	0.263	0.214
Concern	U_22	0.233	0.872	0.142
	U_15	0.231	0.872	0.164
Symptom severity	U_1	0.116	0.137	0.897
	U_2	0.111	0.14	0.893

Notes: The bold type means items are assigned to the corresponding factor through factor analysis. Kaiser–Meyer–Olkin (KMO): 0.85. Total variance explained: 64.22%.

Table 4 Internal Consistency Reliability of UFS-QoL-11 Subscales in Validation Data Set (n=970)

UFS-QOL	k	Internal Consistency (Cronbach's Alpha)		
		Baseline		
		All	HIFU	Myomectomy
Symptom severity	2	0.81	0.81	0.82
Concern	2	0.83	0.83	0.81
Daily life interference	7	0.84	0.85	0.80

Table 5 Responsiveness of Women with Uterine Fibroids After Treatment Using UFS-QoL-11 in Validation Data Set (n=970)

Subscales	Means \pm SD			Change in Score \pm SD		Effect Size*		SRM [†]	
	Baseline	6-Month	12-Month	6-Month	12-Month	6-Month	12-Month	6-Month	12-Month
Symptom severity	28.65 \pm 24.79	12.47 \pm 14.42	9.20 \pm 12.49	-16.2 (-18 to -14.3)	-19.4 (-21.2 to -17.6)	-0.65	-0.78	-0.8	-0.98
Concern	63.08 \pm 27.04	77.48 \pm 20.89	82.55 \pm 19.99	14.4 (12.2 to 16.6)	19.5 (17.3 to 21.6)	0.53	0.72	0.59	0.82
Function interference	74.79 \pm 17.49	83.1 \pm 14.16	86.42 \pm 12.76	8.3 (6.9 to 9.7)	11.6 (10.2 to 13)	0.47	0.66	0.52	0.75

Notes: * Calculated as change in mean score divided by the standard deviation of the baseline. [†] Calculated as change in mean score divided by the standard deviation of the change.

Abbreviations: UFS-QoL-11, the 11 items short form uterine fibroid symptom and quality of life; SD, standard deviation; SRM, standardized response mean.

with poor health (11.54 \pm 13.37 vs 17.07 \pm 17.97; mean difference 5.5, 95% CI 3.1–8.0), and higher scores for concern and daily life interference. Similar differences were observed at 12 months, with larger effect sizes (0.74–1.41), indicating good sensitivity to differences in health status. Similar but larger differences were observed at 12 months, with effect sizes ranging from 0.68 to 1.41 (all $P < 0.001$).

The responsiveness of UFS-QoL-11 was comparable to that of original UFS-QoL; all effect size > 0.45 , displayed a good ability to detect changes (Table 5). After treatment, there was a significant decrease in symptom severity scores and improvement in the concern and daily life interference subscale scores. The mean score change from baseline to the 12-month follow-up for symptom severity was -19.4 ($P < 0.001$), with an effect size of -0.78 and a SRM of -0.98 (Table 5). The mean change scores for the other two subscales (concern and daily life interference) ranged from 11.6 to 19.5, and the effect sizes of the two subscales were between 0.66 and 0.72. Moreover, the SRM values at 12 months after treatment were between 0.75 and 0.82.

Discussion

In this study, we used the largest cohort of patients with UFs through different psychometric techniques, CTT and IRT, to arrive at a reliable and valid short-form of UFS-QoL, which is the only questionnaire designed to assess the symptoms of UFs and their impact on QoL. Importantly, for clinical application, we used responsiveness to confirm the inconsistency between the CTT and IRT. There was a strong correlation between original UFS-QoL and its 11-item short-form, UFS-QoL-11, which supports the use of the scale in other populations as well. Factor analysis supported the hypothesized three-factor structure in 11-item, symptom severity, concern, and daily life interference. The scale and its subscales demonstrated good internal consistency, structural validity, known-group validity, and responsiveness. Throughout the treatment process, UFS-QoL-11 would be feasible for monitoring symptoms and QoL in clinical routine and long-term practice, and would enable clinicians and researchers to distinguish between symptoms and QoL changes sensitively.

The CTT and IRT are important approaches for optimizing psychological scales. According to statistics, several items were good candidates for a short-form scale to provide maximum information using two methods. An important barrier was that the statistics did not provide a clear suggestion for the selection of the best items using different approaches.¹⁹ The Food and Drug Administration (FDA) claims that a patient-reported outcome (PRO) questionnaire is an instrument not only with satisfactory psychological criteria, but should also measure what it is intended to measure, which means that the item could be sensitive enough to assess the concept of interest with each subscale in Chinese patients with UFs.³¹ Therefore, we combined these statistical parameters with clinical applicability that ability to detect change score to confirm the item. We strongly urge the use of these 11 items when assessing patients or research participants to maximize the comparability of clinical assessments.

Brief measures are favorable for clinical practice to reduce patient burden and infringement upon encountering time.¹⁷ Overall, in the present study and the original validation study, the UFS-QoL-11 performed as well as the full version in the measures of reliability, validity, and responsiveness as the full version.¹⁵ When comparing the correlations between the UFS-QoL-11 and the full scale, the correlations were nearly identical. In sum, the UFS-QoL-11 scale will provide comparable information about the latent trait of symptoms and QoL for each respondent as the full scale currently does. Simultaneously, its brevity will increase the likelihood that the assessment of symptoms and QoL will be implemented in everyday clinical practice. Therefore, the UFS-QoL-11 can be easily administered during a typical outpatient visit and

may serve as a discussion point to help clinicians understand patients' symptom experiences, especially when symptoms appear refractory to treatment or physiological improvement.

Although the results of the present study are encouraging, there are several limitations that should be highlighted so that the findings are considered with caution and in light of these deficiencies. First, although the UFS-QoL-11 demonstrated good internal consistency and concurrent validity, we could not examine other important psychometric properties such as test-retest reliability. Future studies should examine these psychometric properties for the purpose of a comprehensive psychometric evaluation of the UFS-QoL-11 among the Chinese. Second, the validation dataset ($n = 970$) was derived from the same prospective cohort as the development dataset ($n = 969$) through internal validation via sample splitting. This approach only provides preliminary psychometric evidence and cannot substitute for external validation in an independent cohort. As such, we have outlined our future research plan to conduct external validation of the UFS-QoL-11 in an independent multicenter cohort of Chinese women with symptomatic uterine fibroids, aiming to confirm its factor structure and psychometric properties across different populations. Further external validation is needed to assess the adaptability and clinical applicability of the UFS-QoL-11 in real-world settings.

Conclusion

Our findings demonstrated that the UFS-QoL-11 performs as well as the original 37-item version while reducing the administrative burden by 70%. Its brevity, if implemented, will increase the likelihood that the assessment of symptoms and QoL will be implemented in everyday clinical practice. As patients with UFs experience a high symptom burden and QoL interference, we hope that the development of UFS-QoL-11 will lead to more widespread, accurate, timely, and sensitive monitoring, in clinical and research settings.

Abbreviations

UFS-QoL, Uterine fibroid symptom and quality of life; CTT, Classical test theory; IRT, Item response theory; UFS-QoL-11, 11-item short-form version of ufs-qol; UFs, Uterine fibroids; HRQL, Health-related quality of life; UFS-QoL-SF, Short-form version of the ufs-qol; HIFU, High-intensity focused ultrasound; SF-36, Short-form-36; QoL, Quality of life; KMO, Kaiser–meyer–olkin; SD, Standard deviations; CI, Confidence intervals; SRM, Standardized response means; FDA, Food and drug administration; PRO, Patient-reported outcome.

Preprint Disclosure

This study has been previously published as a preprint on the ResearchSquare platform, with the full preprint available at: <https://www.researchsquare.com/article/rs-3027197/v1>.

Data Sharing Statement

All data in the main text and the [Supplementary Information](#) have been uploaded into a website (www.hifuctr.com). Readers have read access to the data but are not allowed to export the data. Full access to the data will be available to researchers for the purposes of research or regulatory decision-making with a signed data access agreement after approval of a proposal. All data requests will be reviewed by the research committee at Chongqing Medical University and the corresponding authors to verify whether the request is subject to any intellectual property or confidentiality obligations.

Ethics Approval and Consent to Participate

The study protocol was approved by the China Ethics Committee for Registering Clinical Trials (ChiECRCT-2011034). Because this was a multicenter study involving 20 participating medical institutions, ethical approval was obtained from this independent national ethics committee rather than from the institutional review board of a single participating center. The study was conducted in accordance with the Declaration of Helsinki, and written informed consent was obtained from all participants before enrollment.

Consent for Publication

All participants provided written informed consent for the publication of anonymized data. No identifiable images, videos, or audio recordings are included in this manuscript.

Acknowledgments

We sincerely thank all participants and clinical staff for their support in questionnaire administration, follow-up assessments, and data collection.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

The Project supported by National Natural Science Foundation of China (No. 82427901), Health Commission of Sichuan Province Medical Science and Technology Program (No. 24QNMP007), the Science, Technology Research Program of Chongqing Municipal Education Commission (No. KJQN202400408) and Chongqing Natural Science Foundation General Program (No. CSTB2022NSCQ-MSX0119).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Stewart EA, Cookson CL, Gandolfo RA, et al. Epidemiology of uterine fibroids: a systematic review. *BJOG*. 2017;124(10):1501–1512. doi:10.1111/1471-0528.14640
2. Dai Y, Chen H, Yu J, et al. Global and regional trends in the incidence and prevalence of uterine fibroids and attributable risk factors at the national level from 2010 to 2019: a worldwide database study. *Chin Med J (Engl)*. 2024;137(21):2583–2589.
3. Deal LS, Williams VS, Fehnel SE. Development of an electronic daily uterine fibroid symptom diary. *Patient*. 2011;4(1):31–44. doi:10.2165/11537290-000000000-00000
4. Egbe TO, Badjang TG, Tchounzou R, et al. Uterine fibroids in pregnancy: prevalence, clinical presentation, associated factors and outcomes at the Limbe and Buea Regional Hospitals, Cameroon: a cross-sectional study. *BMC Res Notes*. 2018;11(1):889. doi:10.1186/s13104-018-4007-0
5. Clarke MA, Long BJ, Del Mar Morillo A, et al. Association of endometrial cancer risk with postmenopausal bleeding in women: a systematic review and meta-analysis. *JAMA Intern Med*. 2018;178(9):1210–1222.
6. Doll KM, Romano SS, Marsh EE, et al. Estimated performance of transvaginal ultrasonography for evaluation of postmenopausal bleeding in a simulated cohort of black and white women in the US. *JAMA Oncol*. 2021;7(8):1158–1165. doi:10.1001/jamaoncol.2021.1700
7. Whiteman MK, Kuklina E, Jamieson DJ, et al. Inpatient hospitalization for gynecologic disorders in the United States. *Am J Obstet Gynecol*. 2010;202(6):541.e541–546. doi:10.1016/j.ajog.2009.12.013
8. Baird DD, Patchel SA, Saldana TM, et al. Uterine fibroid incidence and growth in an ultrasound-based, prospective study of young African Americans. *Am J Obstet Gynecol*. 2020;223(3):402.e401–402.e418. doi:10.1016/j.ajog.2020.02.016
9. Spies JB, Coyne K, Guaou Guaou N, et al. The UFS-QOL, a new disease-specific symptom and health-related quality of life questionnaire for leiomyomata. *Obstet Gynecol*. 2002;99(2):290–300. doi:10.1016/s0029-7844(01)01702-1
10. Silva RO, Gomes MT, Castro RA, et al. Uterine fibroid symptom–quality of life questionnaire translation and validation into Brazilian Portuguese. *Rev Bras Ginecol Obstet*. 2016;38(10):518–523.
11. Oliveira Brito LG, Malzone-Lott DA, Sandoval Fagundes MF, et al. Translation and validation of the Uterine Fibroid Symptom and Quality of Life (UFS-QOL) questionnaire for the Brazilian Portuguese language. *Sao Paulo Med J*. 2017;135(2):107–115. doi:10.1590/1516-3180.2016.0223281016
12. Calaf J, Palacios S, Cristóbal I, et al. Validation of the Spanish version of the Uterine Fibroid Symptom and Quality of Life (UFS-QoL) questionnaire in women with uterine myomatosis. *Med Clin (Barc)*. 2020;154(6):207–213. doi:10.1016/j.medcli.2019.05.027
13. Yeung SY, Kwok JWK, Law SM, et al. Uterine fibroid symptom and health-related quality of life questionnaire: a Chinese translation and validation study. *Hong Kong Med J*. 2019;25(6):453–459. doi:10.12809/hkmj198064
14. Wang XQ, Zhu L, Xu T, et al. Validation of the Chinese version of the uterine fibroid symptom and health-related quality of life. *Zhonghua Fu Chan Ke Za Zhi*. 2017;52(7):455–460. doi:10.3760/cma.j.issn.0529-567X.2017.07.005
15. Xu W, Chen W, Chen J, et al. Adaptability and clinical applicability of UFS-QoL in Chinese women with uterine fibroid. *BMC Womens Health*. 2022;22(1):372.
16. Elander J, Said O, Maratos FA, et al. Development and validation of a short-form pain medication attitudes questionnaire (PMAQ-14). *Pain*. 2017;158(3):400–407.

17. Taft TH, Guadagnoli L, Carlson DA, et al. Validation of the short-form esophageal hypervigilance and anxiety scale. *Clin Gastroenterol Hepatol.* 2022;20(2):e64–e73. doi:10.1016/j.cgh.2020.12.021
18. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res.* 2007;16(Suppl 1):5–18. doi:10.1007/s11136-007-9198-0
19. De Beurs D, Cleare S, Wetherall K, et al. Entrapment and suicide risk: the development of the 4-item Entrapment Scale Short-Form (E-SF). *Psychiatry Res.* 2020;284:112765. doi:10.1016/j.psychres.2020.112765
20. Chen J, Li Y, Wang Z, et al. Evaluation of high-intensity focused ultrasound ablation for uterine fibroids: an IDEAL prospective exploration study. *BJOG.* 2018;125(3):354–364. doi:10.1111/1471-0528.14689
21. Li J, Wang H. Evaluation on reliability and validity of SF-36 Scale (Version 2) in urban residents' quality of life in Chongqing. *Journal of the Fourth Military Medical University.* 2009;30(14):1342–1344.
22. Li L, Wang HM, Shen Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalisation. *J Epidemiol Community Health.* 2003;57(4):259–263. doi:10.1136/jech.57.4.259
23. Chang CY, Huang CK, Chang YY, et al. Cross-validation of the Taiwan version of the Moorehead-Ardelt Quality of Life Questionnaire II with WHOQOL and SF-36. *Obes Surg.* 2010;20(11):1568–1574. doi:10.1007/s11695-009-9813-y
24. Ware JE, Snow KK, Kosinski M, et al. *SF-36 Health Survey: Manual and Interpretation Guide.* Lincoln (RI): QualityMetric Inc; 1993.
25. Coyne KS, Margolis MK, Murphy J, et al. Validation of the UFS-QOL-Hysterectomy Questionnaire: modifying an existing measure for comparative effectiveness research. *Value Health.* 2012;15(5):674–679. doi:10.1016/j.jval.2012.03.1387
26. Coyne KS, Margolis MK, Bradley LD, et al. Further validation of the uterine fibroid symptom and quality-of-life questionnaire. *Value Health.* 2012;15(1):135–142.
27. Jensen MP, Keefe FJ, Lefebvre JC, et al. One- and two-item measures of pain beliefs and coping strategies. *Pain.* 2003;104(3):453–469. doi:10.1016/S0304-3959(03)00076-9
28. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.
29. Ghanbari A, Hasandoost F, Lyili EK, et al. Assessing Emergency Nurses' clinical competency: an exploratory factor analysis study. *Iran J Nurs Midwifery Res.* 2017;22(4):280–286. doi:10.4103/1735-9066.212990
30. Pourmovahed Z, Mahmoodabadi HZ, Ardekani SMY, et al. Validation of the family stability questionnaire in married couples: a confirmatory factor analysis. *Electron Physician.* 2018;10(8):7185–7195. doi:10.19082/7185
31. US Food and Drug Administration. *Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* Silver Spring (MD): US Food and Drug Administration; 2009.

Patient Preference and Adherence

Dovepress
Taylor & Francis Group

Publish your work in this journal

Patient Preference and Adherence is an international, peer-reviewed, open access journal that focusing on the growing importance of patient preference and adherence throughout the therapeutic continuum. Patient satisfaction, acceptability, quality of life, compliance, persistence and their role in developing new therapeutic modalities and compounds to optimize clinical outcomes for existing disease states are major areas of interest for the journal. This journal has been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/patient-preference-and-adherence-journal>