Qual Life Res (2016) 25:1835–1844
DOI 10.1007/s11136-015-1223-0

Quality of life in Iranian patients with bipolar disorder: a psychometric study of the Persian Brief Quality of Life in Bipolar Disorder (QoL.BD)

Amirhossein Modabbernia[1] · Mohammadhossein Yaghoubidoust[2] · Chung-Ying Lin[3] · Bengt Fridlund[4] · Erin E. Michalak[5] · Greg Murray[6] · Amir H. Pakpour[2]

Accepted: 21 December 2015 / Published online: 29 December 2015

Abstract

Purpose To assess the reliability, validity, and factor structure of the Persian Brief Quality of Life in Bipolar Disorder (QoL.BD) in Iranian patients with bipolar disorder (BD).

Methods After translation and cross-cultural adaptation of the Brief QoL.BD, we administered the questionnaire to 184 patients diagnosed with BD. To determine factor structure, we performed both exploratory and confirmatory factor analyses. To investigate the reliability, we assessed internal consistency, reproducibility and agreement. Construct validity was assessed by calculating correlations between the Brief QoL.BD and the Short Form-36 (SF-36), Positive And Negative Affect Schedule (PANAS), Hamilton Depression Rating Scale, Young Mania Rating Scale (YMRS) and Quality of Life Enjoyment and Satisfaction Questionnaire Short Form (Q-LES-Q-SF). We also investigated gender differences in interpretations of QoL.BD items.

Results The results obtained from reliability analysis confirmed internal consistency (Cronbach's alpha was 0.87 and 0.89 for two assessments) and reproducibility and agreement (the intraclass correlation coefficient ranged between 0.74 and 0.94). Validity analyses showed that the items loaded on a single-factor structure. The inter-item correlations varied from 0.31 to 0.68. Significantly lower scores on the Brief QoL.BD were observed in people diagnosed with BD I compared to BD II. Significant correlations were observed between the Brief QoL.BD and SF-36 summary measures, HAMD, YMRS, Q-LES-Q-SF and PANAS subscales. Items in the Brief QoL.BD were interpreted similarly by men and women.

Conclusions The Brief Persian QoL.BD is a psychometrically sound measure with acceptable validity and reliability and provides a rapid assessment tool for measuring QoL in patients with BD.

Keywords Bipolar disorder · Factor analysis · Quality of life · Persian · Reliability · Validity

✉ Amir H. Pakpour
Pakpour_Amir@yahoo.com

[1] Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[2] Social Determinants of Health Research Center, Qazvin University of Medical Sciences, Shahid Bahounar BLV, Qazvin 3419759811, Iran

[3] Department of Public Health, College of Medicine, National Cheng Kung Universtiy, Tainan, Taiwan

[4] School of Health Sciences, Jönköping University, Jönköping, Sweden

[5] Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

[6] Faculty of Life and Social Sciences, Swinburne University of Technology, Hawthorn, Australia

Introduction

Mental health professionals have traditionally used objective measures to assess outcomes in patients with bipolar disorders (BD). However, with an increasing emphasis on patient-centered medicine in the past several years, subjective assessment of patient experiences has increasingly drawn attention [1]. In this context, quality of life (QoL) has been recognized as a key outcome in assessment of patients with BD [2, 3]. World Health Organization (WHO) defines QoL as "an individual's perception of their position

in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" [4: p. 1405]. Based on this definition, QoL is a broad concept, which encompasses several physical, social and psychological domains of personal well-being that goes beyond the usual borders of outcome assessment in the clinical setting.

United States Food and Drug Administration [5] has emphasized the importance of using a patient-reported outcome, such as the QoL for assessing the efficacy of a certain medication. Moreover, European Medicines Agency [6] has stated that measuring QoL is important for comprehensive understanding of a patient's overall health. Therefore, it is important for mental healthcare professionals to care for the QoL of patients with BD. Several systematic reviews have demonstrated that BD places an enormous burden on QoL of patients living with the condition [7–9]. Bipolar disorder impairs QoL to a greater extent than chronic physical illness [7]. Indeed, QoL is even impaired in patients with BD who are considered to be euthymic [2]. Reliance solely on objective symptom evaluation will not, therefore, necessarily capture the entire picture. Moreover, it has been shown that patients with BD have important illness-related concerns that go beyond symptom alleviation [10]. It is therefore recommended that in patients with BD, therapeutic goals always allow for trade-offs between treatment effects and QoL [11].

Most previous studies have used generic instruments such as the 36-Item Short Form Health Survey (SF-36) or EuroQoL to assess QoL in patients with BD [9]. Generic measures are useful in that they allow for comparisons of QoL among groups with different health conditions. However, such measures often fail to capture disease-specific impairment in well-being. For example, issues related to identity, finances and religion/spirituality may be of particular concern in patients with BD, but many QoL measures fail to assess these domains of interest [12]. Moreover, compared with disease-specific measures, generic instruments are often less responsive to change and therefore are less sensitive to the effects of therapeutic interventions [13].

Given the clear rationale for a BD-specific QoL instrument, in 2010, Michalak et al. [12] developed a disease-specific QoL instrument for patients with BD (QoL.BD) with valid and reliable scores. Moreover, a brief 12-item version (contents include physical, sleep, mood, cognition, leisure, social, spirituality, finance, household, self-esteem, independence and identity) of the QoL.BD (Brief QoL.BD) showed moderate-to-high correlations with the original 56-item QoL.BD and demonstrated good test–retest reliability. Importantly, the instrument effectively works across mood states and stages of BD and is responsive to change. Furthermore, with only 12 items, the brief form can be easily administered in even the busiest clinical settings [12].

The ways in which QoL and personal well-being are perceived vary significantly from culture to culture [14]. It is essential, therefore, to study the psychometric characteristics of the Brief QoL.BD in different cultural contexts. The primary objective of the present study was to assess the reliability and validity of the Brief QoL.BD scores in Persian version using a nationwide sample of Iranian patients with BD.

Method

This study was a multicenter (six sites in the Iran) cross-sectional prospective study to examine the reliability and validity of the Brief QoL.BD score for Iranian patients. Patients were progressively recruited from clinics at the universities of Iran, Tehran, Qazvin, Zanjan, Ahvaz and Tabriz.

Study participants

This study included 184 patients who had been diagnosed with BD according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [15] confirmed by administration of Structured Clinical Interview (SCID) [16]. All study participants were selected from outpatient clinics. The inclusion criteria were (a) 18 years of age or older, (b) able to communicate in Persian with research team and (c) written informed consent. Participants with any organic brain damage or severe physical diseases were excluded from the study. Demographic characteristics and clinical information were retrieved from patient medical files. Two hundred participants were invited to participate, and 184 (92.0 %) participants provided informed consent. Sixty-six percent of the patients were female, 41 % were married, and the mean ± SD age was 41.1 ± 5.1 years. Participants' educational attainment was reported as follows: 23 % completed primary school; 56 %, secondary school and 12 %, higher education. Most patients (72.8 %) were diagnosed with BD type I (BD I). The mean age of diagnosis was 28.3 years, and the mean duration of illness was 13.0 years (Table 1). No significant differences were found between patients with BD I and those with BD type II (BD II) in terms of demographics.

Measures

Short Form-36 Questionnaire (SF-36)

The SF-36 is a generic questionnaire to assess health-related QoL in healthy and patient populations [17, 18]. It is a

Table 1 Sample characteristics for patients with bipolar disorder (BD)

Characteristic	Mean (SD) or *n* (%)			*p* value
	Whole sample ($n = 184$)	BD type I group ($n = 134$)	BD type II group ($n = 50$)	Between-group comparisons
Age (years)	41.12 (5.10)	39.72 (5.90)	42.64 (6.81)	0.39
Age of diagnosis (years)	28.31 (3.19)	26.72 (4.92)	30.16 (4.51)	0.28
Time since diagnosis (years)	12.99 (2.17)	11.82 (3.7)	13.73 (2.94)	0.05
Number of bipolar episodes	3.61 (1.02)	3.39 (1.24)	3.81 (1.39)	0.62
Number of psychiatric admissions	2.91 (1.11)	2.83 (1.15)	2.98 (1.22)	0.34
Marital status				0.15
Married	76 (41.3 %)	56 (41.8 %)	20 (40.0 %)	
Single	91 (49.5 %)	69 (51.5 %)	22 (44.0 %)	
Divorced	17 (9.2 %)	9 (6.7 %)	8 (16.0 %)	
Gender				0.22
Male	63 (34.2 %)	42 (31.3 %)	21 (42.0 %)	
Female	121 (65.8 %)	92 (68.7 %)	29 (58.0 %)	
Education				0.37
Unlettered	16 (8.7 %)	11 (8.2 %)	5 (10.0 %)	
Primary school	43 (23.4 %)	34 (25.4 %)	9 (18.0 %)	
Secondary school	103 (56.0 %)	76 (56.7 %)	27 (54.0 %)	
College school or above	22 (11.9 %)	13 (9.7 %)	9 (18.0 %)	
Family income (US$)				0.93
<800	40 (21.7 %)	29 (21.6 %)	11 (22.0 %)	
800–1500	123 (66.8 %)	89 (66.4 %)	34 (68.0 %)	
>1500	21 (11.5 %)	16 (11.9 %)	5 (10.0 %)	
Bipolar disorder type				–
I	134 (72.8 %)	–	–	
II	50 (21.2 %)	–	–	

self-administered questionnaire which has 36 items covering eight dimensions: physical functioning (PF), role limitations due to physical health (RP), bodily pain (BP), general health perception (GH), social functioning (SF), role limitations due to emotional problems (RE), vitality (VT) and mental health (MH). These scales are commonly aggregated into two summary components: the physical component summary (PCS) and the mental component summary (MCS). All raw scales are linearly converted to a 0–100 scale, with higher scores indicating higher levels of health-related QoL. The SF-36 has been translated into many languages including Persian (Farsi), and its score has shown good psychometric properties in Iranians [19] and general population samples [20].

Quality of Life Enjoyment and Satisfaction Questionnaire Short Form (Q-LES-Q-SF)

The Q-LES-Q-SF is a 16-item self-report questionnaire to measure satisfaction and enjoyment in various domains of well-being [21]. Total score is computed by summing scores on the first 14 items, which are scored on a five-point Likert scale ranging from 1 (very poor) to 5 (very good) [21]. To facilitate interpretation, the total score is converted into a 0–100 scale with higher scores indicating greater enjoyment or satisfaction. Two general questions ask about satisfaction with medication use and overall life satisfaction and contentment. The Persian version of the Q-LES-Q-SF was developed, and the score was validated by Tagharrobi et al. [22].

Positive and Negative Affect Schedule (PANAS)

The PANAS is a self-report tool commonly used to assess positive affect (PA) and negative affect (NA) [23]. The PANAS has two 15-item scales: one for PA and one for NA. Responses are rated on a 5-point scale from very slightly (1) to very much (5) [23]. The PANAS has been translated into many languages including Persian; the Iranian version of the PANAS score has been found to be highly valid and reliable [24].

Young Mania Rating Scale (YMRS)

The severity of mania syndrome was determined by using the YMRS [25]. This scale contains 11 items, and each item measures specifically clinical condition related to mania syndrome over the previous 48 h [25].

Hamilton Rating Scale for Depression (HAM-D)

The HAM-D is a commonly used interviewer-rating scale to assess depressive symptoms [26]. The HAM-D has 17 items rating participants' experiences over the past week and behavior at interview. The Iranian version of the HAM-D score has demonstrated adequate reliability and validity for assessing depressive symptoms in Iranian participants with bipolar depression [27].

The Brief Quality of Life in Bipolar Disorder (Brief QoL.BD) questionnaire

The Brief QoL.BD is a 12-item disease-specific self-report measure designed to capture patients' subjective perceptions of QoL [12]. The 12 items include physical, sleep, mood, cognition, leisure, social, spirituality, finance, household, self-esteem, independence and identity (Table 2). Patients describe their experiences over the past 7 days on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). Total score is the sum of the 12 items, with higher scores indicating better QoL [12]. The original version of the Brief QoL.BD score has been found to be valid and reliable [12].

Translation and cultural adaptation

Permission was granted by the developers to translate the Brief QoL.BD into Persian (Farsi). The translation procedure was performed in accordance with Beaton's recommendations [28]. The first step was forward translation. Two bilingual translators whose mother tongue was Persian/Farsi (with divergent backgrounds in medicine and sociology) translated the original English version into Persian independently. Next, the translators and the project manager compared translations, resolved discrepancies and synthesized them into an interim version. Backward translation then involved independent translation of this interim version by two native English speakers who were not familiar with the original English version. Subsequently, the two translators and the project manager assessed agreement between the translations and also the original English version. Finally, to assess cross-cultural equivalence of the interim Persian version, an expert committee was formed. Members of the committee were a methodologist, public health professional, language professional, psychiatrist, mental health specialist, nurse and the translators. All translated versions were assessed and checked for discrepancies. A pre-final Persian version was developed and pilot-tested in 31 participants with BD and diverse educational backgrounds. Each participant was asked to complete the questionnaire and explain what they thought about questionnaire items. Recommended changes were made, and the final Persian version was administrated to 184 participants with BD.

Procedure

The Ethics Committee of the Qazvin University of Medical Sciences approved the present study. Potential study participants were progressively identified and invited to enroll in the study between February 2014 and March 2015. The study aims and procedure were explained to the participants, and informed consent was obtained. Study participants completed the baseline questionnaires. Current depressive symptoms were assessed by two board-certified and trained psychiatrists at each center using the HAM-D. The same questionnaires were readministered at 10 days, 3 and 6 months after baseline assessment.

Statistical analyses

Reliability

The reliability of the Persian version of the Brief QoL.BD scores was assessed by a series of analyses including internal consistency, reproducibility and agreement. To test item homogeneity, Spearman's rank–order correlation was calculated from inter-item and corrected item-to-total correlations. Internal consistency reliability was measured using Cronbach's alpha coefficient, with the criterion of $\geq$0.70 [29]. Brief QoL.BD score stability was assessed by administering the scale on two occasions with a 10-day interval between tests. The intraclass correlation coefficient (ICC) was calculated and evaluated against a minimum standard of ICC $\geq$ 0.70 [29].

We report minimal detectable change (MDC95 %), which estimates the smallest score change that likely ($p < 0.05$) corresponds to observable behavioral change and not simply measurement error [29].

Validity

Convergent validity was assessed using bivariate Pearson correlations between the Brief QoL.BD total score and the scores of the following scales: PCS and MCS subscales of the SF-36, PA and NA subscales of the PANAS, HAM-D total score, YMRS total score, Q-LES-Q-SF total score on 14 items and two general item scores of Q-LES-Q-SF

Table 2 Item descriptions and characteristics and test–retest reliability of the Brief QoL.BD item and total scores

Item: description	Test[a]	Retest[b]	Skewness	Kurtosis	ICC (95 %CI)	SEM	MDC95 %
Physical: Felt physically well	3.53 (0.15)	3.31 (0.64)	−0.70	0.88	0.78 (0.75–0.83)	0.37	1.01
Sleep: Woken up feeling refreshed	3.19 (0.47)	3.06 (0.76)	−0.97	0.70	0.88 (0.85–0.90)	0.30	0.84
Mood: Enjoyed things as much as I usually do	3.33 (0.45)	3.16 (1.15)	−0.74	0.61	0.84 (0.81–0.87)	0.34	0.93
Cognition: Had good concentration	3.22 (0.37)	2.99 (0.98)	−0.57	0.11	0.79 (0.74–0.83)	0.36	1.00
Leisure: Been interested in my leisure activities	3.36 (0.33)	3.41 (0.98)	−0.63	0.42	0.94 (0.92–0.95)	0.23	0.64
Social: Been interested in my social relationships	3.48 (0.51)	3.40 (0.91)	−0.91	0.57	0.88 (0.86–0.91)	0.30	0.84
Spirituality: Practiced my spirituality as I wished	3.23 (0.45)	3.20 (0.87)	−0.86	0.53	0.86 (0.83–0.89)	0.32	0.89
Finance: Had enough money for extras	3.20 (0.47)	3.02 (0.81)	−0.82	0.43	0.83 (0.79–0.86)	0.34	0.95
Household: Kept my home tidy	3.23 (0.56)	3.03 (0.74)	−0.74	0.50	0.79 (0.74–0.83)	0.36	1.00
Self-esteem: Felt accepted by others	3.25 (0.25)	3.18 (0.92)	−1.00	0.82	0.74 (0.69–0.79)	0.38	1.05
Independence: Travelled around freely (e.g., driving, using public transport)	2.59 (0.58)	2.41 (0.93)	−0.34	0.31	0.83 (0.79–0.86)	0.34	0.95
Identity: Had a clear idea of what I want and don't want	3.26 (0.46)	3.12 (0.99)	−0.80	0.23	0.76 (0.71–0.81)	0.37	1.03
Total score	39.51 (7.38)	38.16 (10.18)	−0.08	0.04	0.79 (0.74–0.83)	0.36	1.00

ICC intraclass correlation coefficient, *CI* confidence interval, *SEM* standard error of measurement, *MDC95 %* minimal detectable change based on 95 % CI

[a] Test at baseline

[b] Retest at 10th day

(satisfaction with medication use; overall life satisfaction and contentment). Based on Varni et al. [30] and Cheng et al. [31], we classified the correlation coefficients as small (0.1–0.29), medium (0.3–0.49) and large (≥0.5) [31].

In addition, the correlation between Brief QoL.BD total score and HAM-D total score was compared to that between Brief QoL.BD total score and YMRS total score using Fisher *r*-to-*z* transformation test [32]. Moreover, known-group validation was tested: independent *t* tests adjusting for multiple comparisons (Benjamini–Hochberg procedure [33]) as well as adjusting for age and gender were performed to test whether the Brief QoL.BD score could differentiate between subgroups of the patients. Based on the existing literature, it was hypothesized that patients with bipolar II disorder due to longer time spent depressed and higher ratio of depression to mania would report lower QoL scores than patients with bipolar I disorder [34, 35].

Construct validity of the Brief QoL.BD was further assessed via analyses of factor structure. The Brief QoL.BD is a short version of the full 12-scale, 56-item QoL.BD [12], and a single-factor model was described by the developers [12]. Therefore, we conducted a confirmatory factor analysis (CFA) to examine the data model fit of the structure. Considering the ordinal nature of the data, weighted least squares (WLS) was used to estimate the parameters of the CFA model. Moreover, both the polychoric correlations matrix and the asymptotic covariances matrix were used as input for the analyses. Several model fit indices were used: a nonsignificant χ^2 test, comparative fit index (CFI) > 0.90, root-mean-square error of approximation (RMSEA) ≤ 0.08, standardized root-mean-square residual (SRMR) ≤ 0.08, Bentler–Bonett normed fit index (NFI) > 0.90, non-normed fit index (NNFI) > 0.90, goodness-of-fit index (GFI) > 0.90 and adjusted goodness-of-fit index (AGFI) > 0.90 [36].

There is some evidence that QoL is poorer among women with BD relative to men [37, 38], raising the possibility of a gender difference in interpretation of QoL.BD items. Consequently, factorial invariance across genders was also assessed. Three hierarchical levels of factorial invariance were considered: configural invariance, metric invariance and scalar invariance. Configural invariance is achieved if a similar factor structure is found in both male and female samples, metric invariance is deduced if male and female samples demonstrate equal factor loadings, and scalar invariance additionally constrains equal item intercepts [39]. A nonsignificant χ^2 difference test suggests factorial invariance, while changes in the CFI, RMSEA and NNFI between the three levels <0.01 also indicate acceptable factorial invariance [40].

Responsiveness to change

Participants were receiving a range of psychosocial and pharmacological interventions, and it was hypothesized that they would report higher QoL scores at the 3- and 6-month time points relative to baseline [41]. Predicted

changes in the Brief QoL.BD scores were assessed using standardized response mean (SRM: mean change scores divided by pooled SD). Based on Cohen's guidelines, SRM < 0.2 is trivial, 0.2–0.5 is small, 0.5–0.8 is medium, and >0.8 is large [42]. All statistical analyses were performed using SAS 9.2 and LISREL 8.80.

Results

No difficulties were experienced during the translation process, and almost all (99 %) patients found the questionnaire items as well as the instructions easy to understand and acceptable. Because of the comprehensible items, the average rate of incomplete (missing and not applicable) data at the item level was 1.9 %, with a range of 0.1–4.1 %. Average time to complete the Brief QoL.BD was 1.8 ± 0.5 min.

Item characteristics and reliability

Scores on all twelve items of Brief QoL.BD correlated significantly with each other and the total score. The inter-item correlations varied ranged from 0.31 to 0.68 ($p < 0.05$). Correlations between each item score and corrected total scale score ranged from 0.47 to 0.79 ($p < 0.05$). Cronbach's alphas were 0.87 (12 items) and 0.89 for the first and second test administrations, respectively.

An average of 10.3 ± 2 days elapsed between administrations of the Brief QoL.BD. Table 2 shows the SEM and MDC95 % for each Brief QoL.BD item score as well as the total score. Neither individual item scores nor the total Brief QoL.BD score significantly differed between test and retest. The ICCs for the test–retest analysis were high, ranging from 0.74 (self-esteem) to 0.94 (leisure) (Table 2).

Validity

It was hypothesized that people diagnosed with BD II would have significantly lower scores on the Brief QoL.BD than people diagnosed with BD I. Using the method of validation comparing known groups, the data provide support for this hypothesis. Findings presented in Table 3 show that the average total Brief QoL.BD score of people with BD II was almost 7 points lower than that of people with BD I. In addition, people with BD II produced average scores on each item that were significantly lower than those of people with BD I.

Table 4 shows the Spearman's rank–order correlation coefficients between the Brief QoL.BD item scores and scores of external measures including the SF-36 subscales, PANAS subscales, HAM-D, YMRS and Q-LES-Q-SF. All coefficients were statistically significant at the 5 % level. Significant negative correlations were observed between the Brief QoL.BD item scores, HAM-D total score, YMRS total score and NA score in PANAS. Moreover, all Brief QoL.BD item scores correlated positively with the SF-36 summary scores, Q-LES-Q subscales scores as well as PA score in PANAS. Of the 117 correlations, there were 95 magnitudes above medium (≥0.3), and 22 were small (0.1–0.29). In addition, the absolute correlation between Brief QoL.BD total score and HAM-D total score ($r = -0.39$) was significantly greater than correlation between Brief QoL.BD total score and YMRS total score ($r = -0.23$; $p = 0.046$).

Goodness-of-fit measures in the CFA showed that the single-factor solution was adequate: $\chi^2 = 112.712, df = 54$, $p < 0.001$; RMSEA = 0.066 (95 % CI 0.049–0.082); CFI = 0.984; NFI = 0.973; NNFI = 0.987; SRMR = 0.042; GFI = 0.947; AGFI = 0.929. Factor loadings for the twelve items ranged from 0.42 to 0.77 and were all significantly different from zero (Table 5).

Invariance tests of the Brief QoL.BD across genders were conducted by a series of multigroup CFAs. Results indicated that configural, metric invariance and scalar invariance models had acceptable fit indices, except for the χ^2 test. Nevertheless, there was no significant difference between every two models in terms of χ^2 (configural vs. metric invariance: difference = 14.133, $p = 0.29$; metric invariance vs. scalar invariance: difference = 20.825, $p = 0.053$; configural vs. scalar invariance: difference = 34.958, $p = 0.07$); CFI, RMSEA and NNFI differences' values between every two models were <0.01. The overall change in CFI, RMSEA and NNFI measured from the least constrained model (configural model) to our most constrained model (equal factor loadings and item intercepts) was −0.001, 0.000 and 0.000 respectively.

Table 6 shows changes over time (T1–T3) in the Brief QoL.BD item scores. The results of the repeated measures ANOVA showed significantly longitudinal changes on all of the individual Brief QoL.BD item scores ($p < 0.01$), with the exception of finance. SRM values ranged from 0.02 to 0.58 in the Brief QoL.BD item scores as well as total score, suggesting that the improved QoL responds to the treatment over time (3–6 months). However, as we simultaneously taken the MDC95 % results on Table 2 into consideration, the responsiveness seemed trivial.

Discussion

Our primary aim was cross-cultural adaptation and assessment of validity and reliability of the Persian Brief QoL.BD score. Broadly, findings demonstrate that Persian

Table 3 Comparisons of the Brief QoL.BD items scores and total scores for bipolar disorder (BD) types I and II

Item or total scores	Mean (SD)		Cohen's *d* (effect size)
	BD type I ($n = 134$)	BD type II ($n = 50$)	
Physical[a]	3.75 (0.81)	3.22 (1.05)	0.56
Sleep[a]	3.53 (0.89)	2.85 (0.97)	0.73
Mood[a]	3.77 (0.91)	2.89 (0.84)	1.00
Cognition[a]	3.49 (0.83)	2.91 (0.86)	0.68
Leisure[a]	3.90 (0.87)	3.01 (0.93)	0.99
Social[a]	3.72 (0.73)	3.24 (0.81)	0.62
Spirituality[a]	3.57 (0.85)	3.09 (0.93)	0.54
Finance[a]	3.50 (0.88)	2.90 (0.94)	0.66
Household[a]	3.46 (0.88)	3.03 (0.64)	0.56
Self-esteem[a]	3.69 (0.79)	3.20 (0.89)	0.58
Independence[a]	2.80 (0.78)	2.67 (0.69)	0.18
Identity[a]	3.67 (0.77)	3.22 (0.89)	0.54
Total score[a]	42.93 (7.43)	36.06 (7.59)	0.91

[a] Statistically significant according to Benjamini–Hochberg procedure

Table 4 Inter-correlations between Brief QoL.BD and SF-36, Q-LES-Q, PANAS, YMRS and HAM-D scores

Brief QoL.BD	SF-36		Q-LES-Q			PANAS		YMRS	HAM-D
	PCS	MCS	SMU	Overall	OLSC	PA	NA		
Total score	0.39	0.61	0.24	0.41	0.24	0.27	−0.46	−0.23	−0.39
Physical	0.38	0.64	0.56	0.28	0.47	0.24	−0.53	−0.44	−0.25
Sleep	0.47	0.66	0.38	0.18	0.39	0.36	−0.46	−0.37	−0.38
Mood	0.35	0.56	0.44	0.25	0.42	0.40	−0.59	−0.31	−0.48
Cognition	0.31	0.69	0.31	0.21	0.37	0.53	−0.70	−0.21	−0.30
Leisure	0.37	0.57	0.30	0.38	0.38	0.37	−0.40	−0.10	−0.44
Social	0.42	0.51	0.41	0.26	0.52	0.29	−0.33	−0.43	−0.57
Spirituality	0.36	0.44	0.33	0.41	0.41	0.38	−0.43	−0.26	−0.51
Finance	0.41	0.70	0.29	0.38	0.59	0.40	−0.52	−0.29	−0.68
Household	0.33	0.51	0.43	0.68	0.50	0.28	−0.39	−0.30	−0.42
Self-esteem	0.31	0.61	0.49	0.36	0.48	0.58	−0.58	−0.42	−0.48
Independence	0.39	0.56	0.19	0.34	0.51	0.39	−0.18	−0.25	−0.65
Identity	0.46	0.43	0.32	0.31	0.40	0.52	−0.41	−0.23	−0.54

All p values <0.05

HAM-D Hamilton Depression Rating Scale, *MCS* mental component summary, *NA* negative affect, *PA* positive affect, *PCS* physical component summary, *PANAS* Positive and Negative Affect Scale, *Q-LES-Q-SF* Quality of Life Enjoyment and Satisfaction Questionnaire Short Form, *OLSC* overall life satisfaction and contentment, *SF-36* Short Form 36, *SMU* satisfaction with medication use, *YMRS* Young Mania Rating Scale

Brief QoL.BD score has acceptable psychometric properties. The new instrument shows its score had acceptable reliability including internal consistency and test–retest reliability, adequate convergent validity, known-group validity and responsiveness to intervention over time. With more than 110 million Persian-speaking people today, this instrument will be an important tool to permit rapid evaluation of well-being in Persian-speaking people with BD. For the first time in any cultural group, we also showed that meaning of the items in the Brief QoL.BD is perceived similarly by both genders. Therefore, differences in the scores between males and females might indeed reflect true gender differences in QoL in BD.

The internal consistency reliability of the Persian Brief QoL.BD is comparable to that of the English version [12]. Cronbach's alpha coefficients were 0.87 and 0.89 for Persian Brief QoL.BD scale measured 10 days apart. These coefficients exceed the criterion of 0.70 for acceptable consistency. Test–retest reliability established by intraclass coefficient was 0.79 for a 7- to 10-day interval, which exceeds the acceptable criterion value of 0.70.

Table 5 Goodness-of-fit indices and factorial invariance results of the confirmatory factor analysis (CFA)

Fit indices	Model			
	Baseline	Configural	Metric invariance	Scalar invariance
χ^2	112.712	186.582	200.715	221.540
df	54	108	120	132
Comparative fit index (CFI)	0.984	0.987	0.986	0.986
Root-mean-square error of approximation (RMSEA)	0.066	0.054	0.053	0.054
95 % CI for RMSEA	0.049–0.082	0.038–0.069	0.038–0.068	0.038–0.069
Standardized root-mean-square residual (SRMR)	0.042	0.029	0.013	0.024
Bentler–Bonett normed fit index (NFI)	0.973	0.964	0.963	0.963
Non-normed fit index (NNFI)	0.987	0.985	0.986	0.985
Goodness-of-fit index (GFI)	0.947	0.989	0.930	0.946
Adjusted goodness-of-fit index (AGFI)	0.929	0.932	0.911	0.919

Configural model freely estimates all parameters

Metric invariance model constrained all factor loadings as equal across gender

Scalar invariance model constrained all factor loadings and item intercepts as equal across gender

CI confidence interval

Table 6 Responsiveness of the Brief QoL.BD scores

	Baseline Mean (SD)	First follow-up[a] Mean (SD)	Second follow-up[b] Mean (SD)	SRM 0–3 months	SRM 0–6 months	*F*	*p* value
Physical	3.53 (0.15)	3.55 (0.21)	3.65 (0.27)	0.11	0.57	14.94	0.006
Sleep	3.19 (0.47)	3.24 (0.25)	3.33 (0.38)	0.14	0.33	17.00	0.003
Mood	3.33 (0.45)	3.36 (0.29)	3.46 (0.38)	0.08	0.31	16.09	0.001
Cognition	3.22 (0.37)	3.25 (0.39)	3.36 (0.42)	0.08	0.35	18.81	0.001
Leisure	3.36 (0.33)	3.48 (0.21)	3.58 (0.31)	0.07	0.38	16.87	0.001
Social	3.48 (0.51)	3.50 (0.39)	3.59 (0.42)	0.04	0.24	16.46	0.001
Spirituality	3.23 (0.45)	3.37 (0.28)	3.46 (0.34)	0.38	0.58	16.37	0.001
Finance	3.20 (0.47)	3.22 (0.55)	3.21 (0.59)	0.04	0.02	0.15	0.863
Household	3.23 (0.56)	3.27 (0.65)	3.39 (0.49)	0.07	0.30	22.36	0.001
Self-esteem	3.25 (0.25)	3.35 (0.44)	3.48 (0.55)	0.29	0.58	15.91	0.009
Independence	2.59 (0.58)	2.73 (0.47)	2.87 (0.65)	0.27	0.46	19.42	0.001
Identity	3.26 (0.46)	3.39 (0.51)	3.46 (0.57)	0.27	0.39	19.05	0.003
Total score	39.51 (7.38)	39.87 (7.92)	41.14 (8.60)	0.05	0.20	19.29	0.001

SRM standardized response mean: mean change scores divided by pooled SD

[a] First follow-up at 3rd month

[b] Second follow-up at 6th month

The Persian Brief QoL.BD scores showed acceptable known-group validity between patients with BD I and II disorder [35]. As expected, patients with BD II showed significantly poorer QoL than did patients with BD I. However, our results on known-group validity do not imply that clinicians should use the Brief QoL.BD as a diagnostic tool to distinguish patients with BD I from those with BD II. Our results add to the evidence that BD II may be associated with greater illness burden compared to BD I.

All correlations between the Persian Brief QoL.BD and other instruments were statistically significant, and most were greater than 0.30, indicating a medium effect size. The brief QoL.BD correlates with instruments that measure similar health-related constructs. However, the magnitude of the coefficients indicates that the Brief QoL.BD may also measure some facets of the quality of life construct that are unique to BD. Previous studies showed that among patients with BD, QoL has a stronger negative relation to depressive than manic symptoms [37, 38]. Our results support these findings by showing a larger correlation coefficient between QoL.BD total score and HAM-D score than that between QoL.BD total score and YMRS score.

Key advantages of our study include a thorough translation process, multicenter data collection, diverse socioeconomic and geographical background of the participants, repeated measures analyses and multifaceted evaluation of the instrument's validity. In particular, the diversity of the participants increases the external validity of the study by making it generalizable to a larger population. Extensive assessment of convergent validity was ensured using tools that measure various aspects of mood, psychopathology, QoL and functioning. Furthermore, in line with previous studies [12, 41], we found that the Brief QoL.BD scores were significantly improved over time (individual item scores and total score) among this sample of people attending university clinics for treatment. Specifically, the improved QoL was trivial to small (SRM = 0.04–0.38) for all item scores in the 3-month treatment, while small to medium (SRM = 0.20–0.58) in the 6-month treatment, except for the finance item score (SRM = 0.02). When we additionally consider the MDC95 % values, which suggest the smallest noticeable changes, all SRM values were smaller than their corresponding MDC95 % values. Hence, our results might indirectly indicate that the ordinal treatments for BD had weak effects on the QoL of our participants. The trend was also shown in the total score of QoL.BD, a much more reliable score than each individual item score of QoL.BD. Particularly, the SRM of the total score was nearly trivial (0.20), which suggested that although significant, the overall QoL was not improved noticeable.

There are some limitations in the study. First, our sample size on CFA was not large. Some researchers [43] suggest a minimum of 200 is essential, and we only had 184. However, 184 participants are unlikely to cause significant bias; in another study, Anderson and Gerging [44] found that a sample size of 100 would usually be sufficient for convergence. Second, it is unclear whether our results of responsiveness are attributable to the actual effects of intervention as the patients in our study did not receive systematic intervention. Moreover, our SRM results showed that the responsiveness of QoL.BD was trivial and hard to detect. In other words, the intervention in the study might not have promising effects on QoL for patients with BD. Hence, future studies may also want to examine whether symptom management alone is sufficient to improve the QoL for patients with BD, or whether a comprehensive treatment such as systematic intervention substantially improves their QoL. Future studies using systematic intervention are warranted to corroborate our findings. Third, our participants were all recruited from outpatient clinics; the results may not, therefore, be generalizable to inpatient settings. Therefore, future studies recruiting patients from diverse settings are warranted to corroborate our findings.

In summary, the Brief Persian QoL.BD is a psychometrically sound measure with acceptable validity and reliability in its score. Future studies might focus on actual performance of the Persian Brief QoL.BD in routine clinical practice, as well as its responsiveness to change in the context of clinical trials.

Funding This study was funded by Iran's National Elites Foundation (INEF) (Grant Number 15/72243).

Compliance with ethical standards

Conflict of interest The authors Linda Beckman, Amirhossein Modabbernia, Chung-Ying Lin, Mohammadhossein Yaghoubidoust, Bengt Fridlund, Erin E Michalak, Greg Murray and Amir H Pakpour declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Michalak, E. E., Yatham, L. N., Kolesar, S., & Lam, R. W. (2006). Bipolar disorder and quality of life: A patient-centered perspective. *Quality of Life Research, 5*, 25–37.
2. Michalak, E. E., Yatham, L. N., & Lam, R. W. (2005). Quality of life in bipolar disorder: A review of the literature. *Health and Quality of Life Outcomes, 3*, 72.
3. Murray, G., & Michalak, E. E. (2012). The quality of life construct in bipolar disorder research and practice: Past, present, and possible futures. *Bipolar Disorders, 14*, 793–796.
4. The World Health Organization Quality of Life Assessment. (1995). (WHOQOL): Position paper from the World Health Organization. *Social Science and Medicine, 41*, 1403–1409.
5. Food and Drug Administration (FDA). (2009). *Guidance for industry—Patient-reported outcome measures: Use in medical product development to support labeling claims*. Silver Spring, MD: FDA.
6. European Medicines Agency (EMA). (2005). *Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products.* London, UK: EMA. Committee for Medicinal Products for Human Use (CHMP).
7. Dean, B. B., Gerner, D., & Gerner, R. H. (2004). A systematic review evaluating health-related quality of life, work impairment, and healthcare costs and utilization in bipolar disorder. *Current Medical Research and Opinion, 20*, 139–154.
8. IsHak, W. W., Brown, K., Aye, S. S., Kahloon, M., Mobaraki, S., & Hanna, R. (2012). Health-related quality of life in bipolar disorder. *Bipolar Disorders, 14*, 6–18.
9. Namjoshi, M. A., & Buesching, D. P. (2001). A review of the health-related quality of life literature in bipolar disorder. *Quality of Life Research, 10*, 105–115.
10. Sajatovic, M., Jenkins, J. H., Cassidy, K. A., & Muzina, D. J. (2009). Medication treatment perceptions, concerns and expectations among depressed individuals with type I bipolar disorder. *Journal of Affective Disorders, 115*, 360–366.

11. Sachs, G. S., & Rush, A. J. (2003). Response, remission, and recovery in bipolar disorders: What are the realistic treatment goals? *Journal of Clinical Psychiatry, 64*(Suppl 6), 18–22.
12. Michalak, E. E., Murray, G., & Crest, B. D. (2010). Development of the QoL.BD: A disorder-specific scale to assess quality of life in bipolar disorder. *Bipolar Disorders, 12*, 727–740.
13. Wiebe, S., Guyatt, G., Weaver, B., Matijevic, S., & Sidwell, C. (2003). Comparative responsiveness of generic and specific quality-of-life instruments. *Journal of Clinical Epidemiology, 56*, 52–60.
14. Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology, 46*, 1417–1432.
15. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
16. First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition. (SCID-I/P)*. New York: Biometrics Research, New York State Psychiatric Institute.
17. Ware, J. E., Kosinski, M. A., & Keller, S. D. (1994). *SF-36 physical and mental health summary scales: A user's manual*. Boston, MA: The Health Institute, New England Medical Center.
18. Ware, J. E., & Gandek, B. (1998). Overview of the SF-36 health survey and the international quality of life assessment project. *Journal of Clinical Epidemiology, 51*, 903–912.
19. Pakpour, A. H., Nourozi, S., Molsted, S., Harrison, A. P., Nourozi, K., & Fridlund, B. (2011). Validity and reliability of short form-12 questionnaire in Iranian hemodialysis patients. *Iranian Journal of Kidney Diseases, 5*, 175–181.
20. Montazeri, A., Goshtasebi, A., Vahdaninia, M., & Gandek, B. (2005). The Short Form Health Survey (SF-36): Translation and validation study of the Iranian version. *Quality of Life Research, 14*, 875–882.
21. Stevanovic, D. (2011). Quality of Life Enjoyment and Satisfaction Questionnaire—Short form for quality of life assessments in clinical practice: A psychometric study. *Journal of Psychiatric and Mental Health Nursing, 18*, 744–750.
22. Tagharrobi, Z., Sharifi, K., Sooky, Z., & Tagharrobi, L. (2012). Psychometric evaluation of the Iranian version of Quality of Life Enjoyment and Satisfaction Questionnaire Short Form (Q-LES-QSF). *Payesh, 11*, 235–244.
23. Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.
24. Bakhshipoor, A., & DejhKam, M. (2010). Factorial analysis of Positive and Negative Affect Scale. *Journal of Psychology, 36*, 351–365.
25. Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A rating scale for mania: Reliability, validity and sensitivity. *British Journal of Psychiatry, 133*, 429–435.
26. Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23*, 56–62.
27. Shabani, A., Akbari, M., & Dadashi, M. (2010). Reliability and validity of the Bipolar Depression Rating Scale on an Iranian sample. *Archives of Iranian Medicine, 13*, 217–222.
28. Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*, 3186–3191.
29. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*, 34–42.
30. Varni, J. W., Seid, M., & Kurtin, P. S. (2001). PedsQL 4.0: Reliability and validity of the pediatric quality of life inventory version 4.0 generic core scales in healthy and patient populations. *Medical Care, 39*, 800–812.
31. Cheng, C.-P., Luh, W.-M., Yang, A.-L., Su, C.-T., & Lin, C.-Y. (2015). Agreement of children and parents scores on Chinese version of pediatric quality of life inventory version 4.0: Further psychometric development. *Applied Research in Quality of Life*. doi:10.1007/s11482-015-9405-z.
32. Lin, C.-Y., Luh, W.-M., Yang, A.-L., Su, C.-T., Wang, J.-D., & Ma, H.-I. (2012). Psychometric properties and gender invariance of the Chinese version of the self-report pediatric quality of life inventory version 4.0: Short form is acceptable. *Quality of Life Research, 21*, 177–182.
33. Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*, 491–507.
34. Maina, G., Albert, U., Bellodi, L., Colombo, C., Faravelli, C., Monteleone, P., et al. (2007). Health-related quality of life in euthymic bipolar disorder patients: Differences between bipolar I and II subtypes. *The Journal of Clinical Psychiatry, 68*, 207–212.
35. Albert, U., Rosso, G., Maina, G., & Bogetto, F. (2008). Impact of anxiety disorder comorbidity on quality of life in euthymic bipolar disorder patients: Differences between bipolar I and II subtypes. *Journal of Affective Disorders, 105*, 297–303.
36. Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.
37. de la Cruz, M. S., Lai, Z., Goodrich, D. E., & Kilbourne, A. M. (2013). Gender differences in health-related quality of life in patients with bipolar disorder. *Archives of Women's Mental Health, 16*, 317–323.
38. Robb, J. C., Young, L. T., Cooke, R. G., & Joffe, R. T. (1998). Gender differences in patients with bipolar disorder influence outcome in the medical outcomes survey (SF-20) subscale scores. *Journal of Affective Disorders, 49*, 189–193.
39. Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117–144.
40. Wu, T.-H., Chang, C.-C., Chen, C.-Y., Wang, J.-D., & Lin, C.-Y. (2015). Further psychometric evaluation of the Self-Stigma Scale-Short: Measurement invariance across mental illness and gender. *PLoS ONE, 10*, e0117592.
41. Murray, G., Leitan, N. D., Berk, M., Thomas, N., Michalak, E., Berk, L., et al. (2015). Online mindfulness-based intervention for late-stage bipolar disorder: Pilot evidence for feasibility and effectiveness. *Journal of Affective Disorders, 178*, 46–51.
42. Middel, B., & van Sonderen, E. (2002). Statistical significant change versus relevant or important change in (quasi) experimental design: Some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care, 2*, e15.
43. Su, C.-T., Ng, H.-S., Yang, A.-L., & Lin, C.-Y. (2014). Psychometric evaluation of the Short Form 36 Health Survey (SF-36) and the World Health Organization Quality of Life Scale Brief Version (WHOQOL-BREF) for patients with schizophrenia. *Psychological Assessment, 26*, 980–989.
44. Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*, 155–173.