

Development and Evaluation of a New Patient-Reported Instrument: the Bipolar Functional Status Questionnaire

J.F. Goldberg,¹ L.D. McLeod,² S.E. Fehnel,² L.R. Hamm,² K. Gilchrist³

¹ Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, US;

² RTI Health Solutions, RTI International, Research Triangle Park, North Carolina, US; ³ Health Economics & Outcomes Research, AstraZeneca, Wilmington, Delaware, US

ABSTRACT

Aims: The Bipolar Functional Status Questionnaire (BFSQ) is a novel patient-reported instrument designed to assess day-to-day functioning among patients with bipolar disorder. The development and psychometric evaluation of the BFSQ are described.

Methods: Constructs of the BFSQ were identified through a combination of literature review, expert consultation, and patient interviews. Iterative sets of interviews with additional patients were conducted to inform item reduction

and revisions. 596 patients completed the BFSQ, as well as a series of other patient- and clinician-reported health assessments during a multisite validation study. Participants included three patient groups: 148 patients who were hypomanic or recently recovered from a manic episode within the last month, 215 patients who were currently depressed or dysthymic within the last month, and 233 patients whose mood had been stable for a minimum of 2 months. Test-retest was assessed in 187 stable patients using data from a second administration. Additional analyses evaluated the factor structure, internal consistency, reliability, and validity.

Results: Exploratory and confirmatory factor analyses indicated that a one-factor structure best fit the data, providing support for a total score. Item-level descriptive statistics, Cronbach's alphas, and validity correlations all met standard criteria. Furthermore, the BFSQ demonstrated superior ability to discriminate among the three subgroups.

Conclusions: The BFSQ is a psychometrically sound measure of functional status among bipolar patients. Efforts are under way to gather evidence for its responsiveness. The ultimate goal is for the BFSQ to facilitate the identification of treatments, which maximize functional status and ultimately improve patient adherence to treatment.

BACKGROUND

- Individuals with bipolar disorder often have impaired psychosocial functioning, related potentially to subsyndromal mood symptoms, medication effects, or other factors.
- Few self-report measures exist that examine and account for multiple domains of functioning.
- The present study sought to develop and validate a new measure, the BFSQ, to address this need.

PHASE 1: INSTRUMENT DEVELOPMENT

Methods

Step 1. Construct Identification

- Literature review of functional status concepts (generic and disease-specific) and instrument review
- Interviews with well-managed bipolar manic and depressed adults (aged ≥ 18) in outpatient treatment settings; 20% nonwhite.
- Consultation with expert advisory board (measurement experts and psychiatrists).

Step 2. Item Development

- Multiple items were drafted for each construct; varying response scales and reference periods were considered. Draft items were reviewed by the advisory board and revised based on its feedback.

Step 3. Cognitive Pretest and Finalization

- Draft items were pretested during two iterative sets of face-to-face interviews with bipolar patients; cognitive pretesting entailed the process of asking participants to "think aloud" while responding to draft items and to answer probe questions. Based on the results of the interviews and input from the advisory board, final items were selected for the BFSQ.

Results

Step 1. Construct Identification

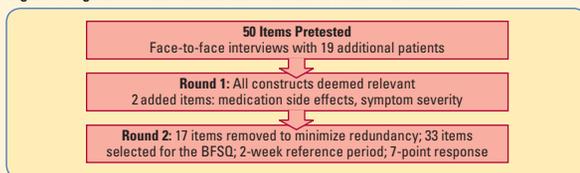
- Eight constructs were selected to guide the development of the BFSQ (Table 2).

Step 2. Item Development

- The advisory board reviewed and provided feedback on approximately 100 items; the draft questionnaire retained 50 items for cognitive pretesting.

Step 3. Cognitive Pretest and Finalization

Figure 1. Cognitive Pretest Results and Selection of Items for Evaluation



Sample Item:

During the past 2 weeks, how often did you have trouble concentrating on what you were doing?

0 1 2 3 4 5 6
 Never About half of the time Always

PHASE 2: PSYCHOMETRIC EVALUATION

Methods

Study Design

The BFSQ was administered in a multisite study, along with additional instruments selected for validation purposes. Investigators at each site had previous clinical trial experience specific to bipolar disorder and were tasked with determining patient eligibility with the same criteria used for the patient interviews.

At visit 1 (V1), subgroup classification was based on disease status (and assigned by the investigators) as follows:

- Stable/Euthymic:** no change in previous 2 months (clinician-reported); MADRS total score ≤ 9; and YMRS total score ≤ 9
- Depressed:** recently recovered from a major depressive episode (defined as discharged from a depression-related hospitalization < 1 month prior to study entrance) or currently in a depressed state (MADRS ≥ 10)
- Hypomanic:** recently recovered from a manic episode (defined as discharged from a mania-related hospitalization < 1 month prior to study entrance) or currently in a manic or hypomanic state (YMRS ≥ 10)

At visit 2 (V2), patients classified as stable at V1 who also completed the survey 7 to 14 days after V1 and were still stable were included in the test-retest subgroup.

CONCLUSIONS

The BFSQ is a psychometrically sound measure of functional status among bipolar patients. Efforts are under way to gather evidence for its responsiveness. The ultimate goal is for the BFSQ to facilitate the identification of treatments, which maximize functional status and ultimately improve patient adherence to treatment.

(PHASE 2: CONTINUED)

Step 1. Initial Evaluation and Item Reduction

- Item-level evaluations (using V1 data): Item responses were evaluated for differences among the three disease subgroups and for floor and ceiling effects; inter-item and item-to-total correlations were computed to identify items that were either redundant or not highly related to the underlying construct.

- Exploratory factor analyses (EFAs) were performed to investigate the BFSQ structure in an initial subsample.

- Elimination was considered for items demonstrating floor or ceiling effects or redundancy with other better-performing items.

Step 2. Psychometric Evaluation of the Final Item Set

- Confirmatory factor analyses (CFAs) performed in a remaining subsample

- Internal consistency (Cronbach's α)

- Test-retest reliability (intra-class correlation coefficients [ICCs])

- Construct validity: Correlations among BFSQ scores and other scales (clinical and patient-reported) at V1

- Discriminant validity

Results

Study Design

The majority of patients were white (79.2%) females (69.1%) aged 18 to 76 years (mean = 40.9 years; standard deviation [SD] = 11.7) (Table 1).

Table 1. YMRS and MADRS Scores by Disease Status Subgroups*

Instrument Subgroup	N	Visit 1			Visit 2		
		Mean (SD)	Median (Range)	N	Mean (SD)	Median (Range)	
YMRS	596	7.42 (6.7)	5 (0-46)	187	2.52 (2.4)	2 (0-11)	
Stable/Euthymic	233	2.92 (2.6)	2 (0-9)	187	2.52 (2.4)	2 (0-11)	
Hypomanic	148	16.15 (5.9)	15 (6-46)	—	—	—	
Depressed	215	6.30 (4.4)	5 (0-24)	—	—	—	
MADRS	596	12.37 (9.4)	9 (0-49)	187	4.50 (2.9)	5 (0-9)	
Stable/Euthymic	233	4.68 (2.9)	5 (0-9)	187	4.50 (2.9)	5 (0-9)	
Hypomanic	148	11.20 (6.4)	10 (0-37)	—	—	—	
Depressed	215	21.51 (7.7)	21 (2-49)	—	—	—	

* Disease status at the time of enrollment (V1) for patients from all 11 clinics in the United States.

Step 1. Initial Evaluation and Item Reduction

- Item-level evaluations: None of the items demonstrated floor or ceiling effects. Overall, the items were highly related.
 - The average inter-item correlation coefficient was 0.33, with 177 of the 528 item pairs (33×32/2) achieving a correlation of ≥ 0.4.
 - The highest inter-item correlation coefficient was 0.79. Nine item pairs were flagged for large inter-item correlations. Four item pairs were flagged for small item-to-total correlations.
 - Each item discriminated among the three disease status groups ($P < 0.0001$). Post hoc pairwise comparisons showed that some items discriminated between all possible pairs.

- EFA (1st subsample, n = 284): Based on results from the scree plot and proportion of variance explained, 1-, 2- and 3-factor models were evaluated. A 1-factor model fit best and was most easily interpretable.

- Item reduction: nine items were removed because they related to symptoms more than functional status; they did not perform as well (i.e., based on ability to discriminate among the disease status groups) as another item in an item pair that was highly correlated and designed to measure similar constructs; or they performed differently for either gender or ethnic subgroups.

Step 2. Psychometric Evaluation of the Final Item Set

- CFA (2nd subsample, n = 284): 1-factor model based on 24-items (of the original 33 items) was confirmed. Model fit indices provided further evidence of model fit (i.e., Goodness-of-Fit Index [GFI] = 0.96; Adjusted GFI = 0.95; Non-Normed-Fit Index [NNFI] = 0.97). Based on these results, BFSQ scores were derived so that higher scores indicated increased functional status. Nineteen items were reverse-scored prior to constructing the summed total BFSQ score.

Table 2. Constructs Identified for Item Development, Content of Final BFSQ, and CFA Factor Loadings

Construct	Subconstruct	Factor 1 Loading (SE)
Cognitive functioning	1. Concentration	0.77 (0.03)
	2. Memory	0.71 (0.03)
	3. Decision making	0.74 (0.03)
Sleep	1. Difficulty falling asleep	0.63 (0.04)
	2. Trouble staying asleep	0.56 (0.04)
	3. Nonrestorative sleep/not feeling rested in the morning	0.48 (0.05)
	4. Somnolence during the day	
Role functioning	1. Work and general productivity	0.75 (0.03)
	a. Missed time from work	
	b. Accomplished less/was less efficient	0.67 (0.04)
	c. Difficulty doing/performing work	0.80 (0.03)
	d. Limited in the kind of work (you can do)	
	2. Other	
	a. Family responsibilities	0.72 (0.04)
	b. Managing money	0.67 (0.04)
	b. Managing money	0.66 (0.04)
	c. Household responsibilities/chores (with similar themes as those for work)	
Emotional functioning	1. Feeling frustrated	0.77 (0.03)
	2. Feeling overwhelmed	
	3. Tolerance for others	
	4. Emotional stability/ability to control emotions	-0.38 (0.07)
Energy/vitality	1. Too much energy/too little energy (feeling tired)	-0.24 (0.07)
	2. Trouble getting out of bed in the morning	0.59 (0.05)
	3. Lacking motivation	0.87 (0.02)
Social Functioning	1. Social activities	-0.47 (0.05)
	2. Relationships	-0.59 (0.04)
	2. Relationships	0.61 (0.04)
	2. Relationships	0.62 (0.04)
	2. Relationships	-0.35 (0.06)
Personal management	1. Managing medications	
	2. Maintaining a healthy lifestyle	
	3. Taking care of appearance	0.46 (0.06)
	4. Dressing appropriately	
	5. Behaving appropriately in public	
Sexual functioning	1. Libido	

Note: Based on patient input, physical functioning was not addressed. Factor loadings are present for the final 24 BFSQ items.

Table 3. Reliability (Internal Consistency and Test-Retest) Results

Property	Results
Internal Consistency	Visit 1 (Overall) = 0.93
	Visit 1 (Stable/Euthymic) = 0.93
	Visit 1 (Hypomanic) = 0.92
	Visit 1 (Depressed) = 0.87
	Visit 2 (Stable) = 0.95
Test-Retest Reliability	ICC = 0.86 (0.82, 0.89; 95% CI)

- Construct validity: Stronger correlations were observed between the BFSQ and instruments measuring similar constructs compared to those measuring dissimilar constructs (Table 4).

Table 4. Pearson Correlations Between the 24-Item BFSQ and Additional Study Instruments

Additional Scale	Visit 1 (N = 596)	Visit 2 (N = 187)
YMRS	-0.26	-0.38
MADRS	-0.61	-0.50
SF-36 Bodily Pain	0.44	0.58
SF-36 Role Emotional	0.73	0.74
SF-36 General Health	0.50	0.53
SF-36 Mental Health	0.75	0.77
SF-36 Physical Function	0.34	0.49
SF-36 Role Physical	0.52	0.62
SF-36 Social Function	0.75	0.71
SF-36 Vitality	0.62	0.69
MOS Sleep Problems Index I	-0.72	-0.74
MOS Sleep Problems Index II	-0.75	-0.75
MOS Cognitive Functioning Scale	0.79	0.80
Sheehan Disability Scale	-0.71	-0.77

Note: All correlation coefficients were significant at $P < 0.0001$. V2 correlations were computed using only data from patients who remained stable at V2.

- Discriminant validity: On average, the stable/euthymic group had the highest and the depressed group had the lowest BFSQ scores. The group scores were statistically different, with subsequent pairwise tests indicating that BFSQ scores were statistically different between all possible pairs ($P < 0.0001$) (Table 5).

Table 5. Total BFSQ Score Descriptive Statistics with F-Statistic by Disease State (V1)

	Stable/Euthymic	Depressed	Hypomanic
Mean	92.11	61.12	73.47
SD	25.9	20.3	26.8
$F_{2,593}$	92.45; $P < 0.0001$		

CONTACT INFORMATION

Joseph Goldberg, MD
128 East Avenue
Norwalk, CT 06851

Phone: 203-854-9607
Fax: 203-854-9608
Email: jfgoldberg@yahoo.com

Presented at: 16th European Congress of Psychiatry
April 5-9, 2008, Nice, France

Research supported by AstraZeneca LP.