RELIABILITY AND VALIDITY OF THE BORDERLINE EVALUATION OF SEVERITY OVER TIME (BEST): A SELF-RATED SCALE TO MEASURE SEVERITY AND CHANGE IN PERSONS WITH BORDERLINE PERSONALITY DISORDER

Bruce Pfohl, MD, **Nancee Blum, MSW**, **Don St. John, PAC**, **Brett McCormick, MA**, **Jeff Allen, PhD**, and **Donald W. Black, MD**
From Department of Psychiatry, University of Iowa Roy J. and Lucille A. Carver College of Medicine, Iowa City

Abstract

A new self-rated scale to measure severity and change in persons with borderline personality disorder (BPD) is described. The Borderline Evaluation of Severity Over Time (BEST) was developed to rate the thoughts, emotions, and behaviors typical of BPD. Data were collected in the course of a randomized controlled trial (RCT) of Systems Training for Emotional Predictability and Problem Solving (STEPPS) for subjects with BPD. The instrument showed moderate test-retest reliability, high internal consistency, and high discriminant validity. Its 15 separate items showed a moderate or better correlation with the total score. The BEST was also sensitive to clinical change as early as week 4 of the RCT and correlated highly with other measures of illness severity. We conclude that the new scale is both reliable and valid in measuring severity and change in persons with BPD.

Several instruments have been developed in the past two decades to assess BPD, including clinician-administered diagnostic assessments such as the Structured Interview for DSM-IV Personality (SIDP-IV; Pfohl, Blum, & Zimmerman, 1994) or the Structured Clinical Interview for DSM-IV Disorders—II (SCID-II; First, Spitzer, Gibbon, & Williams, 1995). To our knowledge, in addition to our own scale, the Borderline evaluation of Severity Over Time (BEST), there are only three other instruments designed to measure acute severity and rate change during clinical trials. These include the Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD; Zanarini & Frankenburg, 2001), probably the most widely employed clinician-rated scale; the Borderline Personality Disorder Severity Index (Arntz et al., 2003), also clinician-rated; and the Borderline Symptom List (Bohus et al., 2007), a self-report questionnaire. All have preliminary evidence of reliability and validity. None were available at the time the BEST was created.

The BEST was developed by two of the authors (BP, NB) as a companion to the Systems Training for Emotional Predictability and Problem Solving (STEPPS) treatment program (Blum, Pfohl, St. John, Monahan, & Black, 2002; Black, Blum, Pfohl, & St. John, 2004). STEPPS was created in the mid-1990s to address the growing need for outpatient treatment programs for what was generally acknowledged to be a challenging disorder. The 20-week model combines cognitive-behavioral elements with skills training. Because self-appraisal is

Address correspondence to Dr. Donald W. Black, 2–126b MEB/Psychiatry Research, University of Iowa Carver College of Medicine, Iowa City, IA 52242; donald-black@uiowa.edu.

an important part of STEPPS, it was concluded that participants needed a quick, self-rated, symptom-based measure that could be filled out at the beginning of each session to yield a "snapshot" of the individual's current status: that is, since the last session has the patient been more emotionally stable, less impulsive, or less likely to have hurt him- or herself? Additionally, was the person more likely to have put into practice skills taught in the STEPPS program?

The scale includes 15 items and three subscales. All items are rated on a Likert-like scale. The first eight items comprise subscale A (Thoughts and Feelings), and involve assessments of mood reactivity, identity disturbance, unstable relationships, paranoia, emptiness, and suicidal thinking. The next four items comprise subscale B (Behaviors-Negative), which rates negative actions such as injuring oneself. Items on these subscales are rated from 1 (None/Slight) to 5 (Extreme). The final three items comprise subscale C (Behaviors-Positive), which rates actions such as following through on therapy plans. These items are rated from 5 (Almost Always) to 1 (Almost Never).

Subscales A and B are taken from the DSM-IV criteria. We created the two subscales to recognize that Thoughts and Feelings (A) are different from the Negative Behaviors (B) typical of the disorder. As the STEPPS program developed, it was hypothesized that negative behaviors (B) would improve more rapidly than thoughts and feelings (A), and that the use of the newly taught behavioral skills would be reinforced when subjects noted the improvement in their subscale B score. Section C was added to acknowledge the acquisition of positive behaviors, as well as to reinforce the use of new skills before seeing improvement in A and B, because scores in subscale C were expected to improve first. (Said one patient, "How do we get credit for doing positive things?") We felt that seeing this change (i.e., improvement in C) reinforces the continued use of new skills, which eventually would be evident from improvement in B, then A. We felt that patients would be discouraged if their expected use of new skills failed to produce *some* improvement. Thus, seeing improvement in C encourages them to keep using the skills despite noticeable lack of improvement in A and B.

To score the BEST, the total for each subscale is determined. The scores of subscales A and B are then added together and the total from subscale C is subtracted. A correction factor of 15 is added to yield the final score which can range from 12 (best) to 72 (worst). The BEST was designed to measure severity in an ill population, and was not designed as a diagnostic instrument. The scale is included in the Appendix.

We recently examined the reliability and validity of the BEST in a well-characterized sample of subjects with BPD enrolled in a randomized controlled trial (RCT) testing the efficacy of STEPPS (Blum et al., 2008). In this report, we examine the test-retest reliability of the instrument, both convergent and discriminant validity, and sensitivity to change.

METHODS

SUBJECTS AND STUDY DESIGN

Subjects age 18 years or older with DSM-IV (American Psychiatric Association, 1994) BPD ($n = 164$) were randomly assigned to receive STEPPS plus treatment as usual (TAU) or TAU alone. They were recruited through referral from our Iowa inpatient and outpatient psychiatric services, through our partial hospital program, through clinician referral, by word-of-mouth, and by advertising. The diagnosis was confirmed using the SIDP-IV (Pfohl et al., 1997). Subjects could not have a diagnosis of schizophrenia, schizo-affective disorder, psychotic mood disorder, or a primary neurological disorder; obvious cognitive impairment; or current (past month) substance abuse or dependence.

A comparison sample ($n = 28$) was recruited to approximate the age and gender distribution of subjects with BPD. The only requirement was that the individual not meet criteria for BPD as determined by administering the BPD module from the SIDP-IV. For reasons of convenience, subjects were employees recruited within the Department of Psychiatry and were asked to fill out the BEST based on their level of functioning within the last week. Because item #15 specifically refers to following therapy plans (as an indicator of "health"), that question was omitted for the comparison sample and the correction factor was changed to 10. All subjects included in this report gave written, informed consent according to procedures approved by the University of Iowa Institutional Review Board.

In addition to screening and baseline assessments, subjects enrolled in the STEPPS research study were reassessed at weeks 4, 8, 12, 16, and 20.

ASSESSMENTS

Subjects with BPD were also assessed with the ZAN-BPD (Zanarini & Frankenburg, 2001) to assess anger, mood reactivity, emptiness, identity disturbance, stress-related paranoid ideation/dissociation, efforts to avoid abandonment, suicidal/self-harm behavior, impulsivity, and unstable relationships; the Symptom Checklist-90-R (SCL-90-R; Derogatis, 1983) to assess a wide range of psychiatric symptoms; and the Social Adjustment Scale (SAS; Weissman & Bothwell, 1976) to rate social functioning (work, social and leisure activities, relationship with extended family, marital and parental role, and economic dependence). Other scales used to rate outcome included the Clinical Global Impression (CGI) severity scale and a patient-rated global improvement scale (Guy, 1976), the Global Assessment Scale (GAS; Endicott, Spitzer, Fleiss, & Cohen, 1976), and the Beck Depression Inventory (BDI; Beck, 1978).

DATA ANALYSIS

Reliability and validity are characteristics of a set of scores derived from a particular scale to measure a specifically-defined population for a given purpose (Thorndike, Cunningham, Thorndike, & Hagen, 1991). For this reason, we provide separate analyses for: (1) the patient sample; (2) the comparison sample; and (3) the combined patient and comparison samples. To provide evidence of reliability and validity for both untreated patients and those undergoing treatment and exhibiting a range of severity, we present data for weekly study visits as well as for screening and baseline visits. Weekly estimates of reliability and validity should provide a better clue to the real population parameters than possible with a baseline estimate only.

We chose not to combine the weekly ratings of each subject, thereby treating the rating—not the subject—as the unit of analysis. While combining the ratings would have increased our correlational power (because there is greater variation in scores within our treated subjects from baseline to week 20 than there is across subjects within any given visit), we believe this approach would have introduced positive bias, and therefore preferred to take a more conservative approach.

We used Cronbach's α coefficient to measure reliability. This statistic provides a measure of internal consistency, is an estimate of the lower bound of the reliability coefficient, and yields the average of all possible split-half procedures. We correlated each item from the BEST with the total score (minus that item). This indicates the degree to which each item behaves consistently with, and therefore measures the same thing as, the total score. Item-total score correlations provide evidence for both reliability and validity.

Test-retest reliability was examined by correlating screening with baseline scores. We felt this was justified because no subject had greater than 25% improvement in the BEST total score from screening to baseline.

The BEST exhibits face validity by assessing thoughts and behaviors typical of BPD. Evidence for content validity has been indirectly established because the items are derived from the DSM-IV.

Construct validity refers to the degree to which a set of scores measures what they purport to measure. We assessed construct validity by examining convergent and discriminant validity. BEST total scores were correlated with ZAN-BPD scores with the expectation that a strong correlation would emerge between the two scales, both of which measure BPD symptoms (convergent validity). We then correlated BEST total score with scales that measure constructs related to but different from BPD (discriminant validity). For the latter correlations, we expected BEST total scores to be strongly correlated with CGI and SAS scores, but even more strongly correlated with the ZAN-BPD scores. We expected BEST scores to be moderately to highly correlated with the BDI total scores, and moderately to weakly with the SAS total scores.

We also correlated weekly BEST scores with CGI severity, patient-rated global improvement, and GAS scores. The CGI severity ratings range from 1–7 and indicate the severity of mental illness of the patient compared with the particular population being studied. The severity of illness of patients with BPD should be strongly correlated with the BEST score, highly correlated with a general symptom scale such as the SCL-90-R, and moderately correlated with the SAS scores.

We measured the sensitivity of the BEST to clinical change by modeling mean assessment scores by weekly visit assuming a first-order autoregressive covariance structure. We contrasted each weekly visit's mean to the screening mean. In this way, we could assess when the time effect became significant. We also assessed the time effect for CGI severity, patient-rated global improvement, ZAN-BPD, and BDI scores.

All correlations were estimated with the Pearson correlation coefficient (R) from the SAS PROC CORR procedure (SAS Institute, Inc., 1999). Because item #15 was not applicable to the comparison sample, the reliability and validity statistics are based on the 14-item BEST score. For the analyses involving only those with BPD, the statistics are based on the 15-item score. For analyses that combine the comparison sample with the subjects with BPD, the statistics are based on the 14-item score.

RESULTS

BASELINE SEVERITY

Subjects with BPD and those in the comparison group did not differ significantly in gender or age. The BPD sample was 85% female with a median age of 29 ($SD = 9.5$) years. Corresponding figures for the comparison sample were 82% female with a median age of 32.5 ($SD = 13.3$) years. The age distributions were not significantly different (Mann-Whitney $\chi^2 = 2.8$, $df = 1$, $P = .10$).

Subjects with BPD had a mean (SD) of 7.6 (1.2) DSM-IV BPD criteria. Over three-quarters (76.2%) had prior suicide attempts, and slightly fewer (71.3%) prior self-harm acts such as cutting or burning. They had a mean (SD) of 4.9 (2.4) lifetime SCID disorders each, including 73% with lifetime major depression; 50.9% had current major depression.

The BEST total score significantly separated BPD and comparison subjects according to baseline severity (Mann-Whitney $\chi^2 = 62.3$, $df = 1$, $p < .001$). Each of the 14 items significantly differered between the groups as well. The items that best discriminated the groups were items #3 (extreme changes in how you see yourself), #4 (severe mood swings), and #12 (temper outbursts). For item #4, 30 of 133 (23%) subjects with BPD exhibited extreme difficulty with the item compared to 0 of 28 (0%) comparison subjects; 22 of 28 (79%) comparison subjects showed none/slight difficulty compared to 7 of 133 (5%) subjects with BPD.

INTERNAL CONSISTENCY AND ITEM ANALYSIS

Cronbach's α coefficients at baseline for subjects with BPD and comparison subjects were 0.86 and 0.90, respectively, indicating that test homogeneity was relatively high at baseline (Table 1). When subjects with BPD were combined with comparison subjects, the test homogeneity of the baseline scores remained high (α = 0.92). Cronbach's α coefficient for the borderline subjects was 0.89 after the first month of treatment, and remained high (0.90 to 0.92) during the 20-week treatment period. Item-total correlations and the corresponding overall measure of internal consistency, Cronbach's α coefficient, from these visits indicate that all items are measuring the same dimension.

TEST-RETEST RELIABILITY

Correlation between baseline and screening BEST total scores was moderate ($r = 0.62$, $n = 130$, $P < .001$). As mentioned earlier, we expected some test-retest instability due to real changes in borderline symptoms, but this was not seen. There was a mean (*SD*) of 53.1 (45.6) days between screening and baseline assessments.

CONVERGENT AND DISCRIMINANT VALIDITY

At the screening visit, the BEST correlated strongly with the ZAN-BPD score, SCL-90-R total score, the SAS total score, the CGI severity score, and both the GAS and BDI scores. To our surprise, the BEST correlated more strongly with the SCL-90-R total score than with any of the other scales at the screening visit, but this was only a matter of degree because all relationships were significant (Table 2). At each time point throughout the study (baseline through week 20), each instrument score remained significantly related to the BEST total score, yet the relationship between the BEST and the BDI scores produced the highest coefficients (0.67–0.80), while the relationship with the CGI severity score and the SAS total score produced the lowest (0.33–0.59, and 0.41–0.59, respectively).

SENSITIVITY TO CLINICAL CHANGE

The BEST total score was sensitive to clinical change that occurred among all subjects with BPD who participated in the STEPPS treatment study. In Table 3, we present observed and modeled means of BEST total score, CGI Severity, patient-rated global improvement, and BDI score. The modeled means, obtained using a repeated measures model with first-order autoregressive covariance structure, are estimates of the mean that would be observed if subjects were not lost to follow-up. The BEST total score decreased from a mean of 38.7 (SD = 11.3) at baseline to a mean of 32.9 (SD = 12.0) at week 20 of the study. The overall time effect for visits at baseline through week 20 was significant ($P < .001$). The CGI severity scale, the patient-rated global improvement scale, and the BDI were also sensitive to clinical change by week 20 ($P < .001$). However, follow up contrast tests which compared each weekly visit to the screening visit revealed that the CGI severity scores were not significantly different at week 4 from screening scores, and did not become significant until week 12 ($P = .002$). BEST scores on the other hand, showed an average decrease of 3.5 from screening to week 4 ($P = .001$), indicating that the BEST was sensitive to clinical change

that occurred early. The BDI also showed a significant difference by week 4, and changes in both the BDI and ZAN-BPD were significant by week 8.

DISCUSSION

Reliability and validity are critical properties for a clinical scale to demonstrate. Our recently completed RCT provided data to explore these properties for the BEST in a well-characterized sample of outpatients with BPD (Blum et al., 2008). Their symptom scores at the screening visit, comorbid depression, and history of self-harm acts and suicidal behavior suggest that the sample had moderate to severe BPD symptoms and depression. They compare favorably in severity to other groups of patients with BPD that have participated in clinical trials (Bateman & Fonagy, 1999; Verheul et al., 2003; Davidson et al., 2006; Linehan et al., 2006; Clarkin, Levy, Lenzenweger, & Kernberg, 2007).

The results indicate that the instrument has good test-retest reliability. Internal consistency is excellent, as shown by the moderate to high Cronbach's α coefficients across study visits, which argues strongly for construct validity. The scale also demonstrates excellent discriminant validity and is sensitive to clinical change occurring as early as week 4 of the study. The data show that BEST scores were the most robust indicators of illness severity for subjects with BPD. These results build on the preliminary data reported earlier in which we presented evidence that the BEST had good internal consistency and was sensitive to change (Blum et al., 2002).

There are several limitations to acknowledge. First, the data on subjects with BPD were collected in the course of a clinical trial, and this analysis was opportunistic. For this reason, the design used to examine test-retest reliability was not optimal despite the good results. It would have been preferable to give the BEST at a set interval following the first administration, rather than to collect it at screening and baseline, wherein the interval differed from subject to subject. Arntz et al. (2003) reported a similar experience in assessing their instrument, wherein three months had passed between the first and second administration. In each case, the correlation coefficients likely underestimate the true test-retest reliability of the respective instrument. Next, the comparison sample was gathered for convenience, and the only requirement was that the individual not meet criteria for BPD. Ideally, comparisons should involve other clinical populations screened to exclude persons with BPD, such as persons with major depression or panic disorder. The purpose would be to show that the BEST discriminates between syndromes and specifically taps items of interest to subjects with BPD. Additionally, several items (i.e., #13, #14) initially had low correlations with the total item score. This could be due to the fact that these items rate a person's use of skills being taught in the STEPPS program, which would initially be minimal. The fact that the correlations improved as the study progressed suggests this could be true. Alternatively, because these are scored in the opposite direction than the first 12 items, it may be that subjects not paying close attention could have miscoded the item.

Acknowledgments

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: American Psychiatric Association; 1994.

Arntz A, van den Hoorn M, Cornelis J, Verheul R, van den Bosch WM, de Bie AJ. Reliability and validity of the borderline personality disorder severity Index. Journal of Personality Disorders. 2003; 17:45–59. [PubMed: 12659546]

Bateman A, Fonagy P. Effectiveness of partial hospitalization in the treatment of borderline personality disorder—a randomized controlled trial. American Journal of Psychiatry. 1999; 156:1563–1569. [PubMed: 10518167]

Beck, AT. Depression inventory. Philadelphia, PA: Philadelphia Center for Cognitive Therapy; 1978.

Black DW, Blum N, Pfohl B, St John D. The STEPPS group treatment program for outpatients with borderline personality disorder. Journal of Contemporary Psychotherapy. 2004; 34:193–210.

Blum N, Pfohl B, St John D, Monahan P, Black DW. STEPPS: A cognitive behavioral systems based group treatment for outpatients with borderline personality disorder-A preliminary report. Comprehensive Psychiatry. 2002; 43:301–310. [PubMed: 12107867]

Blum N, Pfohl B, St John D, Stuart S, McCormick B, Allen J, Arndt S, Black DW. Systems training for emotional predictability and problem solving (STEPPS) for outpatients with borderline personality disorder: A randomized controlled trial and 1-year follow-up. American Journal of Psychiatry. 2008; 165:468–478. [PubMed: 18281407]

Bohus M, Limberger MF, Frank U, Chapman A, Kuhler T, Steiglitz RD. Psychometric properties of the borderline symptom list (BSL). Psychopathology. 2007; 40:126–132. [PubMed: 17215599]

Clarkin JF, Levy KN, Lenzenweger MF, Kernberg OF. Evalu- ating three treatments for borderline personality disorder: A multiwave study. American Journal of Psychiatry. 2007; 164:922–928. [PubMed: 17541052]

Davidson K, Tyrer P, Gumley A, Tata P, Norrie J, Palmer S, et al. A randomized controlled trial of cognitive behavior therapy for borderline personality disorder: Rationale for trial, method, and description of sam- ple. Journal of Personality Disorders. 2006; 20:431–449. [PubMed: 17032157]

Derogatis, LR. SCL-90-R administration, scoring, and procedures manual II. Towson, MD: Clinical Psychometric Research; 1983.

Endicott J, Spitzer RL, Fleiss JL, Cohen J. The global assess- ment scale: A procedure for measuring overall severity of psychiatric distur- bance. Archives of General Psychiatry. 1976; 33:766–771. [PubMed: 938196]

First MB, Spitzer RL, Gibbon M, Williams JBW. The struc- tured clinical interview for DSM-III-R personality disorders (SCID-II), I: De- scription. Journal of Personality Disorders. 1995; 9:83–91.

Guy, W. ECDEU assessment manual for psychopharmacology. Washington, DC: 1976. p. 217-222.US Department of HEW Publications, 76–338

Linehan MM, Comtois KA, Murray AM, Brown MZ, Gallop RJ, Heard HL, et al. Two year randomized controlled trial and follow up of dialectical behavior therapy vs. therapy by experts for suicidal behaviors and borderline personality disorder. Archives of General Psychiatry. 2006; 63:757–766. [PubMed: 16818865]

Pfohl, B.; Blum, N.; Zimmerman, M. Structured interview for DSM-IV personality disorders (SIDP-IV). Washington, DC: American Psychiatric Press; 1997.

SAS. SAS/STAT User's Guide: Version 8. Cary, NC: SAS Institute, Inc; 1999.

Thorndike, RM.; Cunningham, G.; Thorndike, RL.; Hagen, EP. Measurement and evaluation in psychology and education. 5. New York: Macmillan; 1991.

Verheul R, Van den Bosch LMC, Koeter MWJ, De Ridder MAJ, Stijnen T, Van den Brink W. Dialectical behavior therapy for women with borderline personality disorder: 12-month randomized clinical trial in The Netherlands. British Journal of Psychiatry. 2003; 182:135–140. [PubMed: 12562741]

Weissman MM, Bothwell S. Assessment of social adjustment by patient self-report. Archives of General Psychiatry. 1976; 33:1111–1115. [PubMed: 962494]

Zanarini, MC.; Frankenburg, FR. Zanarini rating scale for borderline personality disorder (ZAN-BPD). Belmont, MA: Harvard Medical School; 2001.

APPENDIX

QUESTIONS FROM THE BEST

Instructions: For the first 12 items, the highest rating (5) means that the item caused extreme distress, severe difficulties with relationships, and/or kept you from getting things done. The lowest rating (1) means it caused little or no problems. Rate items 13–15 (positive behaviors) according to frequency.

A. THOUGHTS AND FEELINGS

1   Worrying that someone important in your life is tired of you or is planning to leave you.

2   Major shifts in your opinions about others such as switching from believing someone is a loyal friend or partner to believing the person is untrustworthy and hurtful.

3   Extreme changes in how you see yourself. Shifting from feeling confident about who you are to feeling like you are evil or that you don't even exist.

4   Severe mood swings several times a day. Minor events cause major shifts in moods.

5   Feeling paranoid or like you are losing touch with reality.

6   Feeling angry.

7   Feelings of emptiness.

8   Feeling suicidal.

B. BEHAVIORS (NEGATIVE)

9   Going to extremes to try to keep someone from leaving you.

10   Purposefully doing something to injure yourself or making a suicide attempt.

11   Problems with impulsive behavior (not counting suicide attempts or injuring yourself on purpose). Examples include: over-spending, risky sexual behavior, substance abuse, reckless driving, binge eating, other _____ (*circle those that apply*).

12   Temper outbursts or problems with anger leading to relationship problems, physical fights, or destruction of property.

C. BEHAVIORS (POSITIVE)

13   Choosing to use a positive activity in circumstances where you felt tempted to do something destructive or self-defeating.

14   Noticing ahead of time that something could cause you emotional difficulties and taking reasonable steps to avoid/prevent the problem.

15   Following through with therapy plans to which you agreed (e.g., talk therapy, "homework assignments, coming to appointments, medications, etc.)

TABLE 1

Correlations Between Individual Items and Total Score on the BEST

				Week of study						
	Subjects with BPD							Comparison Subjects	Subjects with BPD + Comparison Subjects	
Item	SC (n = 150)	BL (n = 133)	4 (n = 120)	8 (n = 110)	12 (n = 104)	16 (n = 100)	20 (n = 97)	BL (n = 28)	BL (n = 161)	
Thoughts and Feelings										
1	0.51	0.54	0.64	0.63	0.60	0.64	0.67	0.73[a]	0.68[a]	
2	0.57	0.46	0.54	0.64	0.66	0.70	0.65	0.54[a]	0.59[a]	
3	0.58	0.52	0.65	0.64	0.77	0.73	0.61	0.54[a]	0.68[a]	
4	0.64	0.67	0.73	0.75	0.75	0.77	0.78	0.80[a]	0.80[a]	
5	0.44	0.58	0.66	0.68	0.75	0.65	0.65	0.79[a]	0.63[a]	
6	0.49	0.67	0.62	0.60	0.64	0.73	0.72	0.47[a]	0.78[a]	
7	0.51	0.56	0.70	0.67	0.76	0.76	0.80	0.71[a]	0.72[a]	
8	0.50	0.55	0.59	0.53	0.61	0.54	0.44	0.82[a]	0.61[a]	
Behaviors (Negative)										
9	0.48	0.61	0.60	0.57	0.71	0.58	0.62	0.81[a]	0.67[a]	
10	0.44	0.58	0.56	0.55	0.48	0.49	0.38	—	0.59[a]	
11	0.48	0.42	0.43	0.57	0.62	0.59	0.56	0.80[a]	0.57[a]	
12	0.49	0.49	0.55	0.68	0.64	0.61	0.56	0.61[a]	0.61[a]	
Behaviors (Positive)										
13	0.21	0.26	0.47	0.43	0.60	0.55	0.61	0.68[a]	0.40[a]	
14	0.28	0.36	0.53	0.44	0.61	0.62	0.49	0.57[a]	0.47[a]	
15	0.02	0.24	0.06	0.21	0.22	0.16	0.35			
Internal Consistency Cronbach's α	0.83	0.86	0.89	0.90	0.92	0.91	0.90	0.90[a]	0.92[a]	

Notes. —correlation undefined since there was no variability in item 15 in the comparison group.

[a] BEST score based on 14 items.

SC = screening; BL = baseline

TABLE 2

Convergent and Discriminant Validity: Correlations Between Total Score on the BEST and Other Scores that Assess Symptoms of BPD, General Psychiatric Symptomatology, Social Adjustment, Overall Severity of Psychopathology, and Depression

							Week of study							
	SC		BL		4		8		12		16		20	
Scale	r	P	r	P	r	P	r	P	r	P	r	P	r	P
ZAN-BPD	0.51	<.001					0.76	<.001					0.63	<.001
SCL-90 total	0.59	<.001					0.72	<.001					0.70	<.001
SAS total	0.42	<.001					0.41	<.001					0.59	<.001
CGI Severity	0.33	<.001	0.38	<.001	0.52	<.001	0.48	<.001	0.59	<.001	0.52	<.001	0.59	<.001
Patient-rated global Improvement					0.49	<.001	0.49	<.001	0.52	<.001	0.40	<.001	0.56	<.001
GAS	−0.23	.006					−0.51	<.001					−0.59	<.001
BDI	0.53	<.001	0.70	<.001	0.74	<.001	0.67	<.001	0.80	<.001	0.79	<.001	0.71	<.001
BDI (Comparison Subjects) ($n = 28$)			0.59[a]	.001										
BDI (Subjects with BPD + Comparison Subjects) ($n = 160$)			0.81[a]	<.001										

Notes.

[a] BEST score based on 14 items.

SC = screeing; BL = baseline; ZAN-BPD = Zanarini Rating Scale for Borderline Personality Disorder; SCL-90-R = Symptom Checklist-90-Revised; SAS = Social Adjustment Scale; CGI = Clinical Global Impression; GAS = Global Assessment Scale; BDI = Beck Depression Inventory

TABLE 3

Sensitivity to Clinical Change Among Subjects Diagnosed with BPD

							Week of study													
	SC		BL			4			8			12			16			20		
Scale	M/MM	SD/SE	M/MM	P	SD/SE	M/MM	P	SD/SE	M/MM	P	SD/SE	M/MM	P	SD/SE	M/MM	P	SD/SE	M/MM	P	SD/SE
BEST total (M)	41.5	10.7	39.1		11.2	37.5		12.2	35.9		12.0	35.1		13.1	33.5		12.9	32.1		12.0
BEST total (MM)	41.1	0.96	38.7	.005	0.99	37.5	.001	1.03	35.7	<.001	1.07	35.2	<.001	1.10	34.3	<.001	1.12	32.9	<.001	1.14
CGI Severity (M)	4.98	0.81	4.99		0.81	4.89		0.88	4.81		0.88	4.61		0.93	4.52		0.91	4.51		1.02
CGI Severity (MM)	4.95	0.07	4.99	.422	0.07	4.90	.451	0.07	4.84	.163	0.08	4.69	.002	0.08	4.61	<.001	0.08	4.59	<.001	0.08
Patient-rated Global Improvement (M)						3.63		1.29	3.40		1.384	3.10		1.36	3.12		1.33	2.81		1.35
Patient-rated Global Improvement (MM)						3.61		0.12	3.38	.093	0.13	3.12	.003	0.13	3.08	.003	0.13	2.87	<.001	0.13
ZAN-BPD (M)	18.2	7.2							13.7		6.6							12.3		6.90
ZAN-BPD (MM)	18.2	0.55							13.9	<.001	0.62							12.3	<.001	0.67
BDI (M)	31.4	12.8	29.3		13.5	27.0		14.3	26.5		12.8	24.9		15.3	24.8		14.4	23.7		15.0
BDI (MM)	31.0	1.13	29.3	.092	1.18	27.5	.008	1.22	26.2	<.001	1.27	25.1	<.001	1.30	25.7	.001	1.33	24.2	<.001	1.36

Notes. M = observed mean; SD = standard deviation; MM = modeled mean (repeated measures model with first-order autoregressive covariance structure); SE = standard error of modeled mean; P = *p*-value for comparing week's modeled mean to modeled mean at screening (for Patient-rated global improvement, compared to week 4 mean). SC = screening; BL = baseline; CGI = Clinical Global Impression; ZAN-BPD = Zanarini Rating Scale for Borderline Personality Disorder; BDI = Beck Depression Inventory