Published in final edited form as:
*Bipolar Disord*. 2009 December ; 11(8): 867–875. doi:10.1111/j.1399-5618.2009.00764.x.

# Stability of symptoms across major depressive episodes in bipolar disorder

**Roy H Perlis**[a], **Michael J Ostacher**[a], **Rudolf Uher**[b], **Andrew A Nierenberg**[a], **Francesco Casamassima**[a], **Christine Kansky**[a], **Joseph R Calabrese**[c], **Michael Thase**[d], and **Gary S Sachs**[a]

[a]Bipolar Clinical and Research Program, Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[b]Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK

[c]Department of Psychiatry, Case Western Reserve University, Cleveland, OH

[d]Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

## Abstract

**Objective—**Some studies suggest that depressive subtypes, defined by groups of symptoms, have predictive or diagnostic utility. These studies make the implicit assumption of stability of symptoms across episodes in mood disorders, which has rarely been investigated.

Corresponding author: Roy H. Perlis, MD MSc, Bipolar Clinical and Research Programs, Department of Psychiatry, Massachusetts General Hospital, 50 Staniford Street, 4th Floor, Boston, MA 02114, USA, Fax: 617-724-3028, rperlis@partners.org.

**Disclosures**

RHP has received research support from Eli Lilly & Co. and Elan/Eisai; advisory/consulting fees from AstraZeneca, Bristol-Myers Squibb, Eli Lilly & Co., Pfizer, and Proteus; speaking fees or honoraria from AstraZeneca, Bristol-Myers Squibb, Eli Lilly & Co., GlaxoSmithKline, and Pfizer; and has equity holdings and patents for Concordant Rater Systems, LLC. MJO reports research support from Pfizer; advisory/consulting fees from Pfizer and Concordant Rater Systems; and speaking fees from AstraZeneca, Bristol-Myers Squibb, Eli Lilly & Co., Forest, GlaxoSmithKline, Janssen, Pfizer, and the Massachusetts General Psychiatry Academy (MGHPA talks were supported through Independent Medical Education grants from the following pharmaceutical companies in 2008: AstraZeneca, Eli Lilly & Co., and Janssen). AAN has received grant support from Bristol-Myers Squibb, Cederroth, Cyberonics, Forest, GlaxoSmithKline, Janssen, Lichtwer Pharma, Eli Lilly & Co., NARSAD, NIMH, Pfizer, the Stanley Foundation, and Wyeth-Ayerst; has served as a consultant for Abbott Laboratories, BrainCells, Bristol-Myers Squibb, GHenaissance, GlaxoSmithKline, Innapharma, Janssen, Eli Lilly & Co., Novartis, Pfizer, Sepracor, Shire, and Somerset; and has received honoraria from Bristol-Myers Squibb, Cyberonics, Forest, GlaxoSmithKline, Eli Lilly & Co., and Wyeth-Ayerst; he has no relevant stock ownership to declare. JRC receives federal funding from the Department of Defense, Health Resources Services Administration, and NIMH; receives research funding or grants from the following private industries or nonprofit funds: Cleveland Foundation, NARSAD, Repligen, and Stanley Medical Research Institute; receives research grants from Abbott, AstraZeneca, Cephalon, GlaxoSmithKline, Janssen, and Eli Lilly & Co.; serves on the advisory boards of Abbott, AstraZeneca, Bristol-Myers Squibb, Dainippon Sumitomo Pharma, Forest, France Foundation, GlaxoSmithKline, Janssen, Organon, OrthoMcNeil, Repligen, Servier, Solvay/Wyeth, and Supernus Pharmaceuticals; and reports continuing medical education activities with AstraZeneca, Bristol-Myers Squibb, France Foundation, GlaxoSmithKline, Janssen, Johnson & Johnson, Organon, and Solvay/Wyeth. MT has served as an advisor or consultant to AstraZeneca, Bristol-Myers Squibb, Cephalon, Cyberonics, Eli Lilly & Co., GlaxoSmithKline, Janssen, MedAvante, Neuronetics, Novartis, Organon, Sepracor, Shire, Supernus, and Wyeth; has served on the speakers bureaus of AstraZeneca, Bristol-Myers Squibb, Cyberonics, Eli Lilly & Co., GlaxoSmithKline, Organon, Sanofi-aventis, and Wyeth; has provided expert testimony for Jones Day (Wyeth litigation) and Phillips Lytle (GlaxoSmithKline litigation); is a shareholder with MedAvante; receives income from royalties and/or patents with American Psychiatric Publishing, Guilford Publications, and Herald House; and his wife is the senior medical director for Advogent. GSS has received research support from Abbott, AstraZeneca, Bristol-Myers Squibb, Eli Lilly & Co., GlaxoSmithKline, Janssen, Memory Pharmaceuticals, NIMH, Novartis, Pfizer, Repligen, Shire, and Wyeth; has served on the speakers bureaus of Abbott, AstraZeneca, Bristol-Myers Squibb, Eli Lilly & Co., GlaxoSmithKline, Janssen, Memory Pharmaceuticals, Novartis, Pfizer, Sanofiaventis, and Wyeth; has served as an advisory board member or consultant for Abbott, AstraZeneca, Bristol-Myers Squibb, Cephalon, CNS Response, Elan Pharmaceuticals, Eli Lilly & Co., GlaxoSmithKline, Janssen, Memory Pharmaceuticals, Merck, Novartis, Organon, Pfizer, Repligen, Sanofi-aventis, Shire, Sigma-Tau, Solvay, and Wyeth; and his spouse is a shareholder with Concordant Rater Systems. RU, FC, and CK have no conflicts of interest to report.

**Methods—**We examined prospective data from a cohort of 3,750 individuals with bipolar I or II disorder participating in the Systematic Treatment Enhancement Program for Bipolar Disorder study, selecting a subset of individuals who experienced two depressive episodes during up to two years of follow-up. Across-episode association of individual depressive or hypomanic/mixed symptoms was examined using the weighted kappa measure of agreement as well as logistic regression.

**Results—**A total of 583 subjects experienced two prospectively observed depressive episodes, with 149 of those subjects experiencing a third. Greatest evidence of stability was observed for neurovegetative features, suicidality, and guilt/rumination. Loss of interest and fatigue were not consistent across episodes. Structural equation modeling suggested that the dimensional structure of symptoms was not invariant across episodes.

**Conclusion—**While the overall dimensional structure of depressive symptoms lacks temporal stability, individual symptoms including suicidality, mood, psychomotor, and neurovegetative symptoms are stable across major depressive episodes in bipolar disorder and should be considered in future investigations of course and pathophysiology in bipolar disorder.

**Keywords**

bipolar disorder; factor analysis; major depression; mixed state; psychosis; subtype; suicide; symptom stability

---

The DSM-IV includes subtypes, or illness specifiers, for major depressive episodes such as atypical and melancholic, which have been suggested to have predictive validity (1, 2). That is, depressive features may be informative about outcome or diagnosis (3). In particular, some investigators have reported that atypical depressive symptoms, most notably reversed neurovegetative symptoms, and melancholic features are more characteristic (and perhaps a hallmark) of bipolar disorder compared to major depressive disorder (MDD) (4-6).

This literature makes the implicit, but important, assumption that depressive features are *stable* across episodes, an assumption rarely examined prospectively in large cohorts. In the only study to examine more than one depressive recurrence, Coryell and colleagues (7) found some stability for psychotic, agitated versus retarded, and ‘endogenous’ depression; this cohort included ~120 subjects with bipolar disorder. Smaller studies in MDD identified modest correlation for neurovegetative symptoms (8), groups of ‘endogenous’ or anxious symptoms (9, 10), melancholia (11), and suicidality (10).

To our knowledge, no study has specifically examined stability of these symptoms in bipolar depression, and only one study considered more than one depressive episode. Beyond refining psychiatric nosology, understanding temporal stability may facilitate biological studies by clarifying ‘core’ symptoms of depression in mood disorders. It may also guide clinical practice if certain symptoms such as suicidality demonstrate stability from episode to episode. We therefore examined data from the multicenter Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) cohort study of bipolar disorder, utilizing the subset of individuals with up to three prospectively observed depressive episodes. We attempted to confirm the stability of neurovegetative symptoms and suicidality between episodes and explore the broader stability of mood symptoms.

# Methods

## Study overview

STEP-BD was a multicenter 'effectiveness' study conducted in the U.S. between 1999 and 2005 which evaluated prospective outcomes among individuals with bipolar disorder. Methods for the STEP-BD study as a whole are detailed elsewhere (12, 13).

## Participants

Study participation was offered to all bipolar patients seeking outpatient treatment at one of the participating study sites. Entry criteria included meeting DSM-IV criteria for bipolar disorder I, II, or not otherwise specified; cyclothymia; or schizoaffective disorder, bipolar type; and ability to provide informed consent. For individuals age 15-17, written assent was also required from parent or guardian. Hospitalized individuals were eligible to enter following discharge.

## Assessments

Bipolar diagnosis was determined using mood and psychosis modules from the Structured Clinical Interview for DSM-IV (SCID) as incorporated in the Affective Disorders Evaluation (ADE) and confirmed by a second clinical rater using the Mini International Neuropsychiatric Interview (MINI) (12). Comorbid Axis I diagnoses were also determined using the MINI. At each subsequent visit, clinicians assigned current mood status based upon the Clinical Monitoring Form (CMF) (14), which assesses DSM-IV criteria for depressive, manic, hypomanic, or mixed states in the prior 14 days. Each criterion is scored on a 0-2 scale, where 1 represents *threshold* or syndromal by DSM-IV mood episode criteria; fractional scores are used to indicate subthreshold symptoms. Greatest reliability has been observed for three anchor points: 0 (symptoms absent), 0.5 (symptoms present but below DSM-IV *threshold* because of severity or frequency), and 1 (symptoms present at threshold or syndromal levels). For this reason, and to maximize interpretability and consistency with DSM-IV, the present analysis focused on scores quantized to one of these three levels. The CMF also examines proportion of time with clinically significant anxiety and irritability in the prior 10 days; these were also examined as continuous measures. Finally, the CMF also includes the Clinical Global Impression scale for Current Severity (CGI-S).

## Intervention

Study clinicians in STEP-BD were trained to use model practice procedures, which included published pharmacotherapy guidelines (15), but they could prescribe any treatment which they felt to be indicated. Elsewhere, we have reported high concordance between treatment selection and guideline recommendations, indicating that patients received standard-of-care treatment when entering STEP-BD (16).

## Outcomes

Because STEP-BD was intended to mimic clinical practice, participants were seen as frequently as clinically indicated. Clinical status was assessed at each follow-up visit with the CMF and was used to define the mood states which represent the primary outcome measure. *Remission* (defined in other STEP-BD reports as recovery, or durable recovery) was defined as two or fewer syndromal features of mania, hypomania, or depression for at least eight weeks, consistent with standard DSM-IV criteria for partial or full remission and with criteria used in the prior NIMH Collaborative Study of Depression (17). *Recurrence* was defined as meeting full DSM-IV criteria for a manic, hypomanic, mixed, or depressive episode on any one follow-up visit.

### Statistical analysis

In total, 4,107 subjects entered STEP-BD, including 3,750 bipolar I or II subjects. From these, we identified those who experienced a prospective depressive episode, followed by a remission (i.e., eight or more weeks of 'recovered' status), followed by a second depressive episode. Where available, we then identified the next remission and third depressive episode.

Primary comparisons examined the first observed visit of the first and second depressive episodes. For comparability with previous reports, we calculated weighted kappa (for symptom measures) using quadratic weights, which is equivalent to intraclass correlation coefficient (18). Pearson's correlation was used for continuous measures. To provide a more clinically interpretable measure of stability across episodes, we used logistic regression to examine association between presence or absence of each symptom at threshold levels at the first and second episode, adjusted for overall depressive severity in terms of CGI-S (or linear regression for continuous measures, such as percent of days anxious). To account for varying episode and recovery durations, a second logistic regression model was fit which included terms for interepisode interval and episode duration, as well as one limited to subjects with an interval of at least 365 days between the first and second analyzed visit. In a secondary analysis, we also examined agreement between first and third, and second and third, episode symptoms for the subset of patients (n = 149) with three prospectively observed episodes. Finally, to examine potential moderating effects of comorbidity, post-hoc analyses were run with inclusion of terms for current comorbid Axis I anxiety disorder or substance use disorder and for rapid cycling in the year prior to study entry.

We also conducted exploratory factor analysis of symptom items from the initial visit and examined the fit of a confirmatory factor model derived from the first-episode data to the second-episode data. We used methods appropriate for categorical variables with models fitted to a matrix of polychoric correlations using the robust weighted least square estimator, which has been shown to provide robust unbiased estimates for binary data (19).

Analyses were conducted using Stata 10.0 (College Station, TX, USA); factor analyses were fitted in *M*plus, v 5.1 (20).

## Results

Of the 4,107 subjects entering STEP-BD, 3,750 met criteria for bipolar I or II disorder. In this cohort, median follow-up duration was 383.5 days [interquartile range (IQR) 121-777]; median number of follow-up visits was 10 (IQR 4-19). Of the 3,750 subjects in this subset, 2,344 experienced a major depressive episode during prospective follow-up, 1,301 subsequently achieved recovery (median time to recovery 161 days; IQR 105-278), and 583 experienced a second depressive episode (median time to recurrence 153 days; IQR 70-296). For these 583 subjects (66.2% bipolar I), mean age was 40.6 (SD 12.1) years, and mean onset age was 16.6 (SD 8.2) years. A total of 62.6% were female and 92.8% were Caucasian; 35.9% had a history of psychotic symptoms, 73.4% had a lifetime history of rapid cycling, 41.2% had a history of suicide attempt, and 149 experienced a third depressive episode (median time to recurrence following recovery 168 days; IQR 69-309).

Table 1 summarizes measures of agreement for individual mood symptoms between the first and second episode. Among depressive symptoms, the most prevalent in both episodes was loss of interest; the least prevalent was psychomotor agitation. Greatest stability was observed for neurovegetative symptoms, psychomotor symptoms, suicidal ideation, and depressed mood. Least stability was observed with loss of interest and fatigue. Percentage of days irritable and days anxious also demonstrated significant stability between episodes, as

did overall severity in terms of CGI and count of DSM-IV mood symptoms. We also examined hypomanic/mixed symptoms present during major depressive episodes. Distractibility and flight of ideas were the most prevalent of such symptoms; the paucity of such symptoms limited precision of estimates of agreement. Finally, although observed in only ~3% of episodes, psychosis was very stable between first and second episode.

To further clarify the role of recency effects (i.e., the stability of symptoms across episodes separated by another episode) for a subset of 149 subjects, we examined agreement between symptoms in episode 1 and 3, and episode 2 and 3 (Table 2). Results were generally consistent with those observed in the larger cohort. Incorporating current substance use disorders or Axis I anxiety disorders as covariates in the logistic regression models also did not meaningfully change results, nor did incorporating terms for number of intervening visits rather than intervening time between episodes (i.e., odds ratios for depressive symptoms changed by less than 10%) (Table 3).

Finally, we conducted exploratory factor analysis of symptom item data. An exploratory three-factor model (interest—sleep—guilt) provided an excellent fit for the initial-visit data (root mean square error of approximation < 0.001). However, a confirmatory three-factor model derived from the first-episode data did not fit the second-episode data. Likewise, constraining all factor loadings to be equal across the two episodes significantly deteriorated the fit of factor analytic models (whether these are derived from the first or from the second episode). This was due to substantial differences in symptom correlations in the first and second episodes. We concluded that symptom dimensions were not invariant, precluding a comparison of factor scores between episodes and mandating an analysis of one symptom at a time (as done above using kappas and logistic regression) to index stability and change between episodes.

## Discussion

We identified evidence of consistency of many symptoms between two episodes but with wide variation in the extent of correlation, and substantially less correlation when the episodes are separated by another depressive episode. The most robust effects were observed for neurovegetative symptoms, psychomotor symptoms, thoughts of suicide, and depressed mood. While weighted kappa values indicated only modest agreement (~0.15-0.28), odds ratios for these symptoms range from 1.5 to ~3.7, indicating that the odds of a symptom being present in a subsequent depressive episode are up to ~3.7-fold greater if a symptom is present in an initial episode. Conversely, fatigue and loss of interest in one episode showed little association with these symptoms in a subsequent episode.

Our results are generally consistent with the previous work of Oquendo and colleagues (10), who examined 78 inpatients with MDD across two episodes in the only prior study to investigate individual symptoms. Anxiety, guilt, worthlessness, and poor concentration were quite stable across episodes, while fatigue clearly was not. However, we found no evidence that anhedonia was correlated, contrary to their report. This discordance may be a result of difference in diagnosis (MDD versus bipolar disorder) or treatment setting (inpatient versus outpatient).

We also identified symptoms other than core DSM-IV depressive symptoms which may be stable across episodes. Anxiety and irritability are clearly correlated between episodes. Mixed/manic features, particularly distractibility, display some evidence of correlation between episodes, though considerably less than a previous small study of bipolar inpatients suggests (13); such estimates should be considered preliminary given the relative paucity of

such symptoms overall. Notably, total depression severity (in terms of CGI) displays modest, but significant, stability across episodes.

One clinically important finding from this analysis is the significant stability of suicidal ideation across adjacent depressive episodes. While prediction of suicide attempts is difficult (21), the substantial mortality associated with suicide in bipolar disorder (22) mandates a continued search for clinical predictors. Consistent with prior studies among inpatients (10) and recovered outpatients with MDD (23), our results suggest that the presence of suicidal ideation in one depressive episode should raise the clinician's concern for suicidal ideation in the subsequent episode.

We attempted to determine whether dimensions or groups of symptoms are correlated across the two episodes using exploratory and confirmatory factor analyses. As the relationships (correlations) between individual symptoms were not stable between episodes, however, it was not possible to examine the consistency of symptom dimensions. This inconsistency across episodes implies that using groups of symptoms to distinguish bipolar depression, or to parse it into subtypes, may be problematic because the relationships between symptoms change over time.

Taken together, this analysis does support the stability of many but not all depressive symptoms, as well as some symptoms, such as irritability and anxiety, not previously investigated for consistency in depressive episodes. Our findings are generally consistent with previous investigations (7, 8, 10), although we are unable to investigate the precise subtypes (e.g., endogenous and atypical depression) studied by Coryell and colleagues (7) because not all of these features were collected in STEP-BD. As a predominantly outpatient study, the prevalence of psychosis is low, though nonetheless it identifies stability across two (although not three) episodes, as others have reported in MDD (24) and bipolar disorder (25). We also find support for the concept of 'recency' proposed by Coryell and colleagues (7), which posits that symptom similarity diminishes with greater separation between episodes. Limiting our analysis to episodes separated by a year or more yielded less symptom stability, as did examining a third prospective episode.

Several caveats bear consideration in interpreting our results. First, as we elected to limit the analysis to syndromal (threshold) or subthreshold symptoms, we may underestimate the degree of correlation between episodes which would be detected using a broader ordinal scale. While STEP-BD did include the Montgomery-Åsberg Depression Rating Scale (26), these assessments were performed on a quarterly basis and so would not be able to capture recurrent episodes in most cases. We elected to focus on individual symptoms, rather than subtypes per se, because of abundant evidence of overlap even among recognized descriptors such as atypical and melancholic (27), a strategy further supported by the results of our SEM analysis. Second, the use of a single (cross-sectional) assessment in each episode fails to capture the fluctuation in symptoms which may be observed within a single episode. For example, some patients could initially experience hypersomnia, followed by insomnia, within a single episode. On the other hand, the approach employed here is more readily interpretable than one looking across all episodes [for example, using a time-series approach (28)]. Finally, we cannot entirely exclude medication effects, which are likely to be correlated across episodes—i.e., stability in patient treatment could account for some of these effects if symptoms are actually related to side effects. However, incorporating individual pharmacotherapies as covariates in regression models did not meaningfully change measures of association (results not shown).

In spite of these limitations, these results suggest the phenotypic complexity of bipolar depression. As Coryell and colleagues pointed out (7), they illustrate another means by

which psychiatric nosology may be validated even in the absence of biological gold standards. Moreover, further investigation of the neurobiology of the more stable symptoms, or groups of symptoms, may be particularly useful in elucidating the pathophysiology of bipolar disorder.

## Acknowledgments

We would like to thank the patients and families who have contributed their time to this study. This analysis was supported by National Institute of Mental Health (NIMH) R01 MH086026 (RHP) and a NARSAD Young Investigator Award (RHP). The STEP-BD project was funded in whole or in part with Federal funds from the NIMH under contract N01MH80001 to Gary S. Sachs (PI), Michael E. Thase (Co-PI), and Mark S. Bauer (Co-PI). Active STEP-BD sites and principal investigators included Baylor College of Medicine (Lauren B. Marangell); Case Western Reserve University (Joseph R. Calabrese); Massachusetts General Hospital and Harvard Medical School (Andrew A. Nierenberg); Portland VA Medical Center (Peter Hauser); Stanford University School of Medicine (Terence A. Ketter); University of Colorado Health Sciences Center (Marshall Thomas); University of Massachusetts Medical Center (Jayendra Patel); University of Oklahoma College of Medicine (Mark D. Fossey); University of Pennsylvania Medical Center (Laszlo Gyulai); University of Pittsburgh, Western Psychiatric Institute and Clinic (Michael E. Thase); University of Texas Health Science Center at San Antonio (Charles L. Bowden).

## References

1. Stewart JW, McGrath PJ, Quitkin FM, Klein DF. Atypical depression: current status and relevance to melancholia. Acta Psychiatr Scand Suppl. 2007:58–71. [PubMed: 17280572]
2. Stewart JW, Thase ME. Treating DSM-IV depression with atypical features. J Clin Psychiatry. 2007; 68:e10. [PubMed: 17474800]
3. Fava M, Uebelacker LA, Alpert JE, Nierenberg AA, Pava JA, Rosenbaum JF. Major depressive subtypes and treatment response. Biol Psychiatry. 1997; 42:568–576. [PubMed: 9376453]
4. Akiskal HS, Benazzi F. Atypical depression: a variant of bipolar II or a bridge between unipolar and bipolar II? J Affect Disord. 2005; 84:209–217. [PubMed: 15708418]
5. Perugi G, Akiskal HS, Lattanzi L, et al. The high prevalence of “soft” bipolar (II) features in atypical depression. Compr Psychiatry. 1998; 39:63–71. [PubMed: 9515190]
6. Mitchell PB, Wilhelm K, Parker G, Austin MP, Rutgers P, Malhi GS. The clinical features of bipolar depression: a comparison with matched major depressive disorder patients. J Clin Psychiatry. 2001; 62:212–216. [PubMed: 11305713]
7. Coryell W, Winokur G, Shea T, Maser JD, Endicott J, Akiskal HS. The long-term stability of depressive subtypes. Am J Psychiatry. 1994; 151:199–204. [PubMed: 8296889]
8. Nierenberg AA, Pava JA, Clancy K, Rosenbaum JF, Fava M. Are neurovegetative symptoms stable in relapsing or recurrent atypical depressive episodes? Biol Psychiatry. 1996; 40:691–696. [PubMed: 8894060]
9. Paykel ES, Prusoff BA, Tanner J. Temporal stability of symptom patterns in depression. Br J Psychiatry. 1976; 128:369–374. [PubMed: 1260234]
10. Oquendo MA, Barrera A, Ellis SP, et al. Instability of symptoms in recurrent major depression: a prospective study. Am J Psychiatry. 2004; 161:255–261. [PubMed: 14754774]
11. Melartin T, Leskela U, Rytsala H, Sokero P, Lestela-Mielonen P, Isometsa E. Co-morbidity and stability of melancholic features in DSM-IV major depressive disorder. Psychol Med. 2004; 34:1443–1452. [PubMed: 15724875]
12. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998; 59(Suppl 20):22–33. [PubMed: 9881538]
13. Sato T, Bottlender R, Sievers M, Schroter A, Kleindienst N, Moller HJ. Evaluating the inter-episode stability of depressive mixed states. J Affect Disord. 2004; 81:103–113. [PubMed: 15306135]
14. Sachs GS, Guille C, McMurrich SL. A clinical monitoring form for mood disorders. Bipolar Disord. 2002; 4:323–327. [PubMed: 12479665]

15. Sachs GS, Thase ME, Otto MW, et al. Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD). Biol Psychiatry. 2003; 53:1028–1042. [PubMed: 12788248]

16. Dennehy EB, Bauer MS, Perlis RH, Kogan JN, Sachs GS. Concordance with treatment guidelines for bipolar disorder: data from the systematic treatment enhancement program for bipolar disorder. Psychopharmacol Bull. 2007; 40:72–84. [PubMed: 18007570]

17. Keller MB, Lavori PW, Coryell W, Endicott J, Mueller TI. Bipolar I: a five-year prospective follow-up. J Nerv Ment Dis. 1993; 181:238–245. [PubMed: 8097229]

18. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement. 1973; 33:613–619.

19. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods. 2004; 9:466–491. [PubMed: 15598100]

20. Muthen, LK.; Muthen, B. Mplus. 5. Los Angeles, CA: Muthén & Muthén; 2008.

21. Oquendo MA, Galfalvy H, Russo S, et al. Prospective study of clinical predictors of suicidal acts after a major depressive episode in patients with major depressive disorder or bipolar disorder. Am J Psychiatry. 2004; 161:1433–1441. [PubMed: 15285970]

22. Osby U, Brandt L, Correia N, Ekbom A, Sparen P. Excess mortality in bipolar and unipolar disorder in Sweden. Arch Gen Psychiatry. 2001; 58:844–850. [PubMed: 11545667]

23. Antypa N, Van der Does AJ, Penninx BW. Cognitive reactivity: Investigation of a potentially treatable marker of suicide risk in depression. J Affect Disord. 200910.1016/j.jad.2009.06.013

24. Charney DS, Nelson JC. Delusional and nondelusional unipolar depression: further evidence for distinct subtypes. Am J Psychiatry. 1981; 138:328–333. [PubMed: 6110345]

25. Winokur G, Scharfetter C, Angst J. Stability of psychotic symptomatology (delusions, hallucinations), affective syndromes, and schizophrenic symptoms (thought disorder, incongruent affect) over episodes in remitting psychoses. Eur Arch Psychiatry Neurol Sci. 1985; 234:303–307. [PubMed: 3987739]

26. Montgomery SA, Ásberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979; 134:382–389. [PubMed: 444788]

27. Angst J, Gamma A, Benazzi F, Ajdacic V, Rossler W. Melancholia and atypical depression in the Zurich study: epidemiology, clinical characteristics, course, comorbidity and personality. Acta Psychiatr Scand Suppl. 2007:72–84. [PubMed: 17280573]

28. Zeger SL, Irizarry R, Peng RD. On time series analysis of public health and biomedical data. Annu Rev Public Health. 2006; 27:57–79. [PubMed: 16533109]

**Table 1**

Stability of symptoms between first and second major depressive episode in bipolar disorder

| Symptom | Prevalence episode 1 | Prevalence episode 2 | Agreement (observed) | Agreement (expected) | kappa | *Z*-score | p-value | Crude OR | 95% CI | Adjusted OR | 95% CI | OR (> 365 d) | 95% CI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Suicidal ideation | 18.4 | 20.4 | 78.6% | 69.9% | 0.288 | 6.963 | * | 3.59 | 2.26-5.69** | 3.64 | 2.29-5.80** | 3.58 | 1.88-6.84** |
| Hypersomnia | 36.2 | 37.9 | 65.2% | 54.9% | 0.230 | 5.555 | * | 2.25 | 1.59-3.20** | 2.29 | 1.61-3.26** | 2.15 | 1.33-3.48** |
| Psychomotor agitation | 18.2 | 17.2 | 75.8% | 69.9% | 0.196 | 4.730 | * | 1.91 | 1.15-3.17** | 1.94 | 1.16-3.25** | 1.64 | 0.84-3.23 |
| Depressed mood | 76.7 | 64.0 | 67.9% | 61.1% | 0.175 | 4.450 | * | 2.02 | 1.36-3.00** | 2.26 | 1.51-3.40** | 1.20 | 0.68-2.12 |
| Loss of appetite | 27.4 | 25.7 | 69.0% | 62.2% | 0.181 | 4.386 | * | 1.68 | 1.12-2.51** | 1.62 | 1.08-2.43** | 1.97 | 1.10-3.53** |
| Psychomotor retardation | 34.1 | 40.3 | 64.7% | 57.6% | 0.167 | 4.070 | * | 1.91 | 1.35-2.72** | 1.98 | 1.39-2.82** | 1.63 | 1.00-2.67 |
| Hyperphagia | 18.9 | 21.8 | 72.6% | 67.2% | 0.165 | 4.004 | * | 1.86 | 1.17-2.95** | 1.82 | 1.14-2.91** | 1.09 | 0.59-2.02 |
| Insomnia | 32.8 | 33.9 | 64.2% | 57.4% | 0.160 | 3.860 | * | 1.93 | 1.34-2.78** | 1.92 | 1.33-2.77** | 2.21 | 1.33-3.66** |
| Worthlessness | 72.9 | 71.0 | 74.2% | 70.2% | 0.135 | 3.261 | * | 1.67 | 1.13-2.48** | 1.67 | 1.13-2.48** | 1.10 | 0.63-1.92 |
| Poor concentration | 68.2 | 65.0 | 73.7% | 69.9% | 0.126 | 3.043 | * | 1.76 | 1.23-2.53** | 1.75 | 1.22-2.52** | 1.84 | 1.11-3.05** |
| Guilt/rumination | 40.8 | 41.5 | 62.1% | 57.6% | 0.107 | 2.590 | | 1.61 | 1.15-2.26** | 1.61 | 1.15-2.26** | 1.74 | 1.10-2.76** |
| Poor energy | 85.1 | 81.7 | 82.5% | 81.4% | 0.057 | 1.403 | | 1.30 | 0.74-2.31 | 1.27 | 0.71-2.26 | 0.85 | 0.37-1.92 |
| Loss of interest | 89.4 | 83.3 | 84.6% | 84.7% | -0.011 | -0.268 | | 0.96 | 0.45-2.06 | 0.95 | 0.44-2.04 | 0.50 | 0.17-1.50 |
| Distractibility | 32.8 | 40.8 | 67.9% | 58.1% | 0.235 | 5.735 | * | 2.50 | 1.75-3.57** | 2.52 | 1.76-3.61** | 2.27 | 1.39-3.69** |
| Flight of ideas | 9.4 | 13.0 | 80.9% | 77.0% | 0.172 | 4.251 | * | 3.58 | 1.89-6.76** | 3.58 | 1.89-6.80** | 1.86 | 0.70-4.94 |
| Talkativeness | 2.4 | 4.1 | 91.9% | 90.4% | 0.148 | 3.718 | * | 1.78 | 0.85-2.70** | 1.78 | 0.86-2.71** | 1.89 | 0.55-3.24** |
| Goal-directed activity | 1.0 | 4.3 | 93.1% | 92.8% | 0.031 | 0.882 | | 4.21 | 0.47-38.01 | 4.44 | 0.46-42.37 | nc | |
| High-risk behavior | 2.2 | 2.7 | 93.1% | 93.0% | 0.025 | 0.616 | | nc | | nc | | nc | |
| Decreased need for sleep | 3.3 | 2.4 | 93.2% | 93.4% | -0.025 | -0.609 | | nc | | nc | | nc | |
| Hallucinations | 1.5 | 2.6 | 96.6% | 96.0% | 0.150 | 3.752 | * | 11.36 | 2.01-64.20** | 11.99 | 2.08-69.08** | 55.64 | 3.02-1026.43** |
| Delusions | 1.4 | 1.5 | 98.1% | 97.1% | 0.343 | 8.306 | * | 56.01 | 10.83-289.69** | 83.09 | 13.86-497.99** | 25.24 | 2.08-305.96** |
| Psychosis | 2.6 | 3.6 | 95.2% | 94.0% | 0.198 | 4.857 | * | 11.36 | 3.25-39.65** | 13.21 | 3.67-47.59** | 7.25 | 0.74-71.32 |
| | **Mean (SD)** | **Mean (SD)** | | | | ***r*** | | **Beta** | **95% CI** | **Beta** | **95%CI** | **Beta** | **95% CI** |
| % Days anxious | 45.5 (40.9) | 49.7 (42.0) | – | – | – | 0.24 | * | 0.230 | 0.15-0.31** | 0.22 | 0.14-0.31** | 0.18 | 0.07-0.28** |
| Manic symptoms | 0.8 (1.0) | 0.7 (0.9) | – | – | – | 0.21 | * | 0.175 | 0.11-0.24** | n/a | n/a | 0.11 | 0.02-0.20** |

| Symptom | Prevalence episode 1 | Prevalence episode 2 | Agreement (observed) | Agreement (expected) | kappa | *Z*-score | p-value | Crude OR | 95% CI | Adjusted OR | 95% CI | OR (> 365 d) | 95% CI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| % Days irritable | 36.9 (38.9) | 34.0 (37.3) | – | – | – | 0.18 | * | 0.190 | 0.11-0.28** | 0.19 | 0.10-0.27** | 0.25 | 0.14-0.36** |
| CGI | 4.0 (0.9) | 4.0 (0.8) | – | – | – | 0.17 | * | 0.190 | 0.10-0.28** | n/a | n/a | 0.15 | 0.02-0.28** |
| Depressive symptoms | 5.8 (1.3) | 5.8 (1.2) | – | – | – | 0.08 | | 0.076 | 0.00-0.15** | n/a | n/a | 0.04 | -0.07-0.14 |

Columns 1 and 2 show the prevalence of each symptom in the first and second major depressive episode, respectively. Columns 3-7 show agreement between these two episodes, in terms of the kappa statistic. Remaining columns depict association between each symptom in the first and second episode, in terms of odds ratio (OR) and 95% confidence interval (CI). 'Adjusted' OR refers to adjustment for overall depression severity in terms of symptom burden (i.e., number of syndromal symptoms). 'OR > 365 days' refers to analyses of the subset of patients for whom the two visits are separated by at least 365 days, in order to examine the impact of 'recency' on symptom similarity.

nc = not calculable (i.e., cells containing zero); CGI = Clinical Global Impression Scale.

*Bonferonni-corrected $p < 0.05$ (i.e., $p < 0.002$).

**95% CI excludes 1 (i.e., $p < 0.05$).

**Table 2**

Agreement between symptoms in first and third, and second and third, major depressive episodes during prospective follow-up

| | Symptom prevalence | | | Agreement (Episode 1 versus 3) | | | | | Agreement (Episode 2 versus 3) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Symptom** | **Episode 1** | **Episode 2** | **Episode 3** | **Observed** | **Expected** | **kappa** | ***z*** | **p** | **Observed** | **Expected** | **kappa** | ***z*** | **p** |
| Hypersomnia | 30.9 | 36.9 | 39.6 | 66.9 | 56.2 | 0.246 | 3.033 | 0.0012 | 67.8 | 55.1 | 0.282 | 3.503 | 0.0002 |
| Poor concentration | 67.1 | 69.1 | 71.8 | 77.5 | 71.1 | 0.223 | 2.732 | 0.0032 | 75.7 | 71.4 | 0.150 | 1.841 | 0.0328 |
| Worthlessness | 73.8 | 75.0 | 76.5 | 79.9 | 75.4 | 0.184 | 2.241 | 0.0125 | 76.5 | 73.3 | 0.121 | 1.483 | 0.0691 |
| Insomnia | 34.7 | 34.5 | 32.4 | 64.6 | 56.8 | 0.179 | 2.168 | 0.0151 | 64.2 | 57.2 | 0.165 | 1.994 | 0.0231 |
| Psychomotor agitation | 16.8 | 20.1 | 17.4 | 74.3 | 69.3 | 0.163 | 1.992 | 0.0232 | 78 | 70.5 | 0.255 | 3.116 | 0.0009 |
| Guilt/rumination | 37.6 | 37.6 | 42.3 | 65.1 | 58.7 | 0.155 | 1.892 | 0.0293 | 66.9 | 57.6 | 0.221 | 2.707 | 0.0034 |
| Suicidal ideation | 21.5 | 26.2 | 19.5 | 71.1 | 66.3 | 0.142 | 1.743 | 0.0406 | 76.2 | 69.6 | 0.217 | 2.656 | 0.0040 |
| Poor energy | 87.9 | 81.9 | 81.2 | 85.6 | 83.3 | 0.134 | 1.715 | 0.0432 | 83.6 | 83.7 | -0.007 | -0.086 | 0.5343 |
| Hyperphagia | 19.5 | 20.8 | 19.5 | 70.3 | 67.5 | 0.086 | 1.051 | 0.1466 | 76.8 | 68.2 | 0.271 | 3.307 | 0.0005 |
| Psychomotor retardation | 37.6 | 38.9 | 37.6 | 61.4 | 58.0 | 0.081 | 0.986 | 0.1621 | 60.4 | 57.6 | 0.066 | 0.801 | 0.2115 |
| Depressed mood | 77.2 | 63.8 | 70.5 | 61.9 | 60.6 | 0.034 | 0.458 | 0.3235 | 75.8 | 66.5 | 0.279 | 3.497 | 0.0002 |
| Loss of appetite | 26.8 | 26.8 | 31.5 | 61.9 | 62.0 | -0.004 | -0.043 | 0.5174 | 70.1 | 60.4 | 0.246 | 3.012 | 0.0013 |
| Loss of interest | 90.6 | 85.8 | 89.3 | 85.5 | 86.0 | -0.039 | -0.496 | 0.6901 | 91.1 | 91.3 | -0.021 | -0.261 | 0.6030 |
| Distractibility | 29.5 | 42.3 | 34.2 | 65.8 | 57.5 | 0.195 | 2.480 | 0.0066 | 66.3 | 60.3 | 0.151 | 1.866 | 0.0310 |
| Flight of ideas | 8.1 | 14.8 | 10.1 | 78.7 | 76.4 | 0.098 | 1.251 | 0.1054 | 83.4 | 80 | 0.171 | 2.095 | 0.0181 |
| Goal-directed activity | 2.0 | 2.7 | 0.7 | 94.5 | 94.1 | 0.063 | 0.775 | 0.2192 | 95.1 | 95.4 | -0.054 | -0.665 | 0.7471 |
| Decreased need for sleep | 0.7 | 2.0 | 2.7 | 96.3 | 96.4 | -0.025 | -0.345 | 0.6349 | 95.5 | 95.6 | -0.027 | -0.394 | 0.6531 |
| Talkativeness | 3.4 | 2.0 | 1.3 | 91.1 | 91.4 | -0.032 | -0.390 | 0.6517 | 94.6 | 92.8 | 0.252 | 3.246 | 0.0006 |
| High-risk behavior | 1.3 | 4.0 | 2.0 | 93.6 | 93.9 | -0.039 | -0.522 | 0.6990 | 95.3 | 94.9 | 0.085 | 1.096 | 0.1365 |
| Hallucinations | 2.7 | 1.3 | 1.3 | 96.0 | 96.0 | -0.018 | -0.236 | 0.5935 | 96.0 | 96.0 | -0.018 | -0.236 | 0.5935 |
| Delusions | 2.0 | 1.3 | 2.0 | 96.6 | 96.7 | -0.016 | -0.204 | 0.5809 | 97.3 | 96.1 | 0.320 | 3.902 | 0.0001 |
| Psychosis | 4.0 | 2.7 | 3.4 | 93.3 | 93.5 | -0.033 | -0.415 | 0.6610 | 94.0 | 92.9 | 0.151 | 1.848 | 0.0323 |
| | **Mean (SD)** | **Mean (SD)** | **Mean (SD)** | | | | ***r*** | **p** | | | | ***r*** | **p** |
| % Days anxious | 45.4 (43.2) | 51.5 (43.7) | 47.4 (42.8) | | | | 0.31 | 0.0001 | | | | 0.29 | 0.0004 |
| % Days irritable | 34.3 (38.9) | 34.5 (38.9) | 31.5 (36.2) | | | | 0.20 | 0.0164 | | | | 0.34 | < 0.0001 |
| CGI | 4.05 (0.88) | 4.06 (0.73) | 3.97 (0.74) | | | | 0.11 | 0.1907 | | | | 0.14 | 0.0869 |
| Manic symptoms | 0.85 (0.99) | 0.70 (0.88) | 0.59 (0.81) | | | | 0.01 | 0.8920 | | | | 0.15 | 0.0700 |

| | Symptom prevalence | | | Agreement (Episode 1 versus 3) | | | | | Agreement (Episode 2 versus 3) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Symptom** | **Episode 1** | **Episode 2** | **Episode 3** | **Observed** | **Expected** | **kappa** | ***z*** | **p** | **Observed** | **Expected** | **kappa** | ***z*** | **p** |
| Depressive symptoms | 5.91 (1.24) | 5.93 (1.18) | 5.91 (1.04) | | | | -0.03 | 0.6905 | | | | 0.01 | 0.8920 |

Shaded rows indicate symptoms with nominally significant ($p < 0.05$) association between first and third episodes and second and third episodes.

**Table 3**

Stability of symptoms between first and second major depressive episode in bipolar disorder, adjusted for clinical covariates

| Symptom | Crude OR | 95%CI | Anxiety OR | 95% CI | Substance use OR | 95% CI | Rapid cycling OR | 95% CI |
|---|---|---|---|---|---|---|---|---|
| Depressed mood | 2.02 | 1.36-3.00* | 2.01 | 1.36-2.99* | 2.00 | 1.34-2.97* | 2.05 | 1.38-3.04* |
| Guilt/rumination | 1.61 | 1.15-2.26* | 1.61 | 1.14-2.25* | 1.59 | 1.13-2.23* | 1.62 | 1.16-2.28* |
| Suicidal ideation | 3.59 | 2.26-5.69* | 3.59 | 2.26-5.69* | 3.56 | 2.24-5.65* | 3.65 | 2.30-5.82* |
| Psychomotor agitation | 1.91 | 1.15-3.17* | 1.88 | 1.13-3.13* | 1.87 | 1.12-3.13* | 1.87 | 1.12-3.12* |
| Psychomotor retardation | 1.91 | 1.35-2.72* | 1.92 | 1.35-2.73* | 1.91 | 1.35-2.72* | 1.88 | 1.32-2.67* |
| Hypersomnia | 2.25 | 1.59-3.20* | 2.24 | 1.58-3.18* | 2.27 | 1.60-3.22* | 2.26 | 1.59-3.21* |
| Hyperphagia | 1.86 | 1.17-2.95* | 1.86 | 1.17-2.96* | 1.86 | 1.17-2.97* | 1.87 | 1.17-2.98* |
| Loss of interest | 0.96 | 0.45-2.06 | 0.99 | 0.46-2.13 | 0.96 | 0.45-2.06 | 0.98 | 0.46-2.09 |
| Worthlessness | 1.67 | 1.13-2.48* | 1.67 | 1.13-2.48* | 1.65 | 1.11-2.45* | 1.67 | 1.12-2.47* |
| Poor energy | 1.30 | 0.74-2.31 | 1.31 | 0.74-2.33 | 1.29 | 0.73-2.29 | 1.31 | 0.74-2.32 |
| Poor concentration | 1.76 | 1.23-2.53* | 1.76 | 1.22-2.53* | 1.82 | 1.26-2.62* | 1.76 | 1.22-2.53* |
| Insomnia | 1.93 | 1.34-2.78* | 1.89 | 1.31-2.73* | 1.93 | 1.34-2.78* | 1.95 | 1.35-2.81* |
| Loss of appetite | 1.68 | 1.12-2.51* | 1.71 | 1.14-2.57* | 1.67 | 1.12-2.50* | 1.68 | 1.12-2.51* |
| Talkativeness | 1.78 | 0.85-2.70* | 1.56 | 0.19-12.59 | 1.92 | 0.24-15.51 | 1.70 | 0.21-13.69 |
| Flight of ideas | 3.58 | 1.89-6.76* | 2.96 | 1.54-5.68* | 3.64 | 1.91-6.91* | 3.42 | 1.80-6.50* |
| Distractibility | 2.50 | 1.75-3.57* | 2.50 | 1.75-3.58* | 2.51 | 1.76-3.58* | 2.47 | 1.73-3.53* |
| Goal-directed activity | 4.21 | 0.47-38.01 | 3.36 | 0.36-31.15 | 4.28 | 0.47-39.04 | 3.86 | 0.42-35.34 |
| High-risk behavior | nc | | nc | | nc | | nc | |
| Decreased need for sleep | nc | | nc | | nc | | nc | |
| Hallucinations | 11.36 | 2.01-64.20* | 10.85 | 1.93-60.85* | 10.97 | 1.86-64.67* | 11.29 | 1.99-64.12* |
| Delusions | 56.01 | 10.83-289.69* | 60.06 | 11.26-320.42* | 47.33 | 9.15-244.78* | 55.40 | 10.18-301.36* |
| Psychosis | 11.36 | 3.25-39.65* | 11.56 | 3.28-40.72* | 10.44 | 2.94-37.05* | 10.64 | 3.03-37.45* |
| | **Beta** | **95%CI** | **Beta** | **95%CI** | **Beta** | **95%CI** | **Beta** | **95%CI** |
| CGI | 0.19 | 0.10-0.28* | 0.20 | 0.11-0.29* | 0.19 | 0.10-0.28* | 0.20 | 0.10-0.29* |
| Depressive symptoms | 0.08 | 0.00-0.15* | 0.09 | 0.00-0.18* | 0.09 | 0.00-0.17 | 0.09 | 0.00-0.18* |

| Symptom | Crude OR | 95%CI | Anxiety OR | 95% CI | Substance use OR | 95% CI | Rapid cycling OR | 95% CI |
|---|---|---|---|---|---|---|---|---|
| Manic symptoms | 0.18 | 0.11-0.24* | 0.23 | 0.14-0.33* | 0.25 | 0.16-0.35* | 0.24 | 0.14-0.33* |
| % Days anxious | 0.23 | 0.15-0.31* | 0.21 | 0.13-0.28* | 0.23 | 0.15-0.30* | 0.23 | 0.15-0.31* |
| % Days irritable | 0.19 | 0.11-0.28* | 0.18 | 0.10-0.27* | 0.19 | 0.11-0.27* | 0.19 | 0.11-0.27* |

OR = odds ratio; CI = confidence interval; nc = not calculable (i.e., cells containing zero); CGI = Clinical Global Impression Scale.

*95% CI excludes 1 (i.e., p<0.05).



*Bipolar Disord*. Author manuscript; available in PMC 2013 February 07.