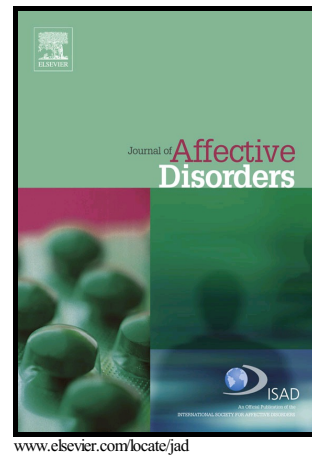# Author's Accepted Manuscript

An Approach to Revealing Clinically Relevant Subgroups Across the Mood Spectrum

Meredith L. Wallace, Burcin Simsek, David J. Kupfer, Holly A. Swartz, Andrea Fagiolini, Ellen Frank

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

An Approach to Revealing Clinically Relevant Subgroups Across the Mood Spectrum

Meredith L. Wallace[1,2], Burcin Simsek[2], David J. Kupfer[1], Holly A. Swartz[1], Andrea Fagiolini[3],

Ellen Frank[1]

1. Department of Psychiatry, University of Pittsburgh, Pittsburgh PA

2. Department of Statistics, University of Pittsburgh, Pittsburgh PA

3. Department of Molecular Medicine, University of Siena, Italy

Correspondence concerning this article should be addressed to Meredith L. Wallace, Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213, Phone: (412) 383-5125, Fax: (412) 383-5112, email: lotzmj@upmc.edu

Background: Individuals diagnosed with bipolar 1 disorder (BP1), bipolar 2 disorder (BP2), or major depressive disorder (MDD) experience varying levels of depressive and (hypo)manic symptoms. Clarifying symptom heterogeneity is meaningful, as even subthreshold symptoms may impact quality of life and treatment outcome. The MOODS Lifetime self-report instrument was designed to capture the full range of depressive and (hypo)manic characteristics.

Methods: This study applied clustering methods to 347 currently depressed adults with MDD, BP2, or BP1 to reveal naturally occurring MOODS subgroups. Subgroups were then compared on baseline clinical and demographic characteristics and as well as depressive and (hypo)manic symptoms over twenty weeks of treatment.

Results: Four subgroups were identified: (1) high depressive and (hypo)manic symptoms (N=77, 22%), (2) moderate depressive and (hypo)manic symptoms (N=115, 33%), (3) low depressive and moderate (hypo)manic symptoms (N=82, 24%), and (4) low depressive and (hypo)manic symptoms (N=73, 21%). Individuals in the low lifetime depressive / moderate lifetime (hypo)manic subgroup had poorer quality of life and greater depressive symptoms over the course of treatment. Individuals in the high and moderate severity subgroups had greater substance use, longer duration of illness, and greater (hypo)manic symptoms throughout treatment. Treatment outcomes were primarily driven by individuals diagnosed with MDD.

Limitations: The sample was drawn from three randomized clinical trials. Validation is required for this exploratory study.

Conclusions: After validation, these subgroups may inform classification and personalized treatment beyond categorical diagnosis.

Key words: *subtype; cluster; major depressive disorder; bipolar disorder; dimensional construct; mood severity*

Major depressive disorder (MDD), bipolar 2 (BP2), and bipolar 1 (BP1) disorder diagnoses are currently based on categorical conceptualizations of the number and duration of symptoms and their severity. That is, individuals are required to meet clinically-defined threshold levels of (hypo)manic and/or depressive symptoms in order to receive one of these diagnoses. However, evidence of heterogeneity in symptom severity within unipolar and bipolar disorder diagnoses has been accumulating (e.g., see Cassano et al., 2004 and Fagiolini et al., 2007), particularly regarding sub-clinical levels of symptoms. Cassano et al. (1999) emphasize the importance of considering the "full range of characteristics of subthreshold mania", as even incomplete manifestations of (hypo)mania can have clinical relevance. For example, individuals with MDD may experience (hypo)manic symptoms that do not present in such a way as to meet the clinical threshold for bipolar 1 or 2 disorder (Cassano et al., 2004), and individuals who do meet the criteria for BP1 and BP2 may have varying levels of depression and (hypo)mania (Fagiolini et al., 2007). To treat MDD, BP1 and BP2 most effectively, it is important to consider the full continuum of depressive and (hypo)manic symptoms, rather than rely only on whether these symptoms meet a syndromal threshold.

With the recognition of heterogeneity in symptom severity *within* MDD, BP2 and BP1 diagnoses, the potential for some individuals to have similar symptom profiles *across* these diagnoses should also be considered. To this end, nosologists have posed the question of whether MDD, BP2, and BP1 diagnoses are separated by natural boundaries, or whether psychopathology may be better characterized by a continuum of depressive, hypomanic, and manic symptoms. For example, Angst and Cassano (2005) suggest a schema for describing the "mood spectrum", that is, a gradient of depressive, hypomanic, and manic symptoms across a horizontal plane and symptom severity (normal, sub-threshold, threshold non-psychotic, and threshold psychotic) on a

vertical plane. Beyond theoretical discussion, ample empirical evidence is available to challenge the classic unipolar-bipolar distinction, including the frequent misdiagnosis of bipolar disorder as unipolar disorder (Altamura et al. 2015; Ghaemi et al., 1999, 2001) and genetic similarities across MDD, BP2, and BP1 (Dell'Osso et al., 2014; Duffy et al., 2000; Lin et al., 2011; McGruffin and Katz, 1989).

This study searches for natural subgroups across the mood spectrum (MDD, BP2, BP1) based on the full continuum of lifetime mood severity symptoms, including those that are subthreshold. To accomplish this, we considered 347 depressed adults diagnosed with MDD, BP2, or BP1, and used clustering methods to reveal subgroups of individuals with similar symptom profiles based on the MOODS Lifetime instrument (Cassano et al., 2002; Fagiolini et al., 1999). The MOODS Lifetime instrument was created specifically to capture a continuum of self-reported depressive and (hypo)manic symptoms that may be reported across the mood spectrum over an individual's lifetime, with a particular emphasis on the subthreshold manifestations that are likely to contribute to much of the heterogeneity observed within our existing diagnoses. If novel subgroups could be identified based on the MOODS instrument, we aimed to determine whether they were related to demographic and clinical information, comorbidities, quality of life, and treatment outcome. After validation, these findings could suggest ways to improve current classification, inform existing treatment approaches, and develop new personalized treatments for each subgroup (Jablensky, 2016; Philips, 2016).

METHODS

Participants

The sample used in the present study includes 347 depressed adults aged 18-65 [mean (SD) = 38.5 (12.1)] diagnosed with unipolar depression (N=190, 54.8%), bipolar 1 disorder (N=71, 20.5%), or bipolar 2 disorder (N=86, 24.8%), and consists of 63% female and 82% Caucasian participants. All participants were in a DSM-defined major depressive episode at the time of the MOODS and at screening for study entry, and were selected from one of three parent randomized clinical trials: 1) Bipolar Disorder Center for Pennsylvanians Study (Fagiolini et al., 2009; Kupfer et al., 2009), 2) Bipolar 2 Study (MH84831, PI: Swartz), and 3) Depression Phenotypes Study (Frank et al., 2011). All three of these studies were based out of the same clinic with largely overlapping staff. Sample characteristics across and within studies are provided in Table 1. The institutional review board at the University of Pittsburgh reviewed and approved all study procedures and all participants gave informed consent.

Bipolar Disorder Center for Pennsylvanians (BDCP) study

The BDCP study (Fagiolini et al., 2009; Kupfer et al., 2009) followed 463 individuals aged ≥ 12 with bipolar disorder (BP1, BP2, bipolar disorder not otherwise specified, or schizoaffective disorder, bipolar type) longitudinally for between 1 – 3 years. The study was conducted at sites in Pittsburgh, PA and DuBois, PA. Diagnoses were made based on the Structured Clinical Interview for DSM-IV (SCID; First et al., 2001; American Psychological Association, 2000) for adults or the Schedule for Affective Disorders and Schizophrenia for School Aged Children, Present and Lifetime version for adolescents aged 12-18 (Kaufman et al., 1997). Relevant exclusion criteria included: schizophrenia, current schizoaffective disorder, and a current substance use disorder related to their mood disorder. Although acute inpatient hospitalization was not an exclusion criterion, participation was ceased if the research staff was unable to manage a subject's care while hospitalized. Participants were seen at varying

frequencies, ranging from weekly to every 2 months, depending on their level of symptoms and stability. When participants met DSM-IV criteria for a mood episode or experienced a clinical worsening they were randomized to either specialized care for bipolar disorder (SCBD) or SCBD plus an enhanced clinical intervention (SCBD+ECI), and remained in this treatment arm for the remainder of their time in the study.   Full study details are provided elsewhere (Fagiolini et al., 2009; Kupfer et al., 2009).

From this parent study, we included in the present study 91 adults aged ≥ 18 diagnosed with either BP1 (N=71) or BP2 (N=20), who met DSM-IV criteria for a major depressive episode at study entry (when the Lifetime MOODS was administered), and who had at least 90% items observed on each of the seven MOODS Lifetime subscales. Of these 91, 43% (N=39) were randomized to SCBD+ECI, 56% (N=51) were randomized to SCBD, and one was not randomized.

Bipolar 2 (BP2) Study

The BP2 study (MH84831, PI: Swartz) randomized depressed adults aged 18 – 65 with BP2 to either a bipolar-specific psychotherapy (interpersonal and social rhythm therapy, IPSRT) plus quetiapine or IPSRT plus placebo. At screening, all participants met DSM-IV criteria for BP2, were in a DSM-IV-defined current major depressive episode, and had a rating of ≥15 on the 17-item Hamilton rating scale for depression (HRSD; Hamilton, 1950).   Relevant exclusion criteria included: need for acute inpatient hospitalization, schizophrenia, a current schizoaffective disorder, antisocial or borderline personality disorder, a current substance use disorder related to their mood disorder, and a primary eating or obsessive compulsive disorder diagnosis (i.e., the

eating or obsessive compulsive disorder was associated with greater impairment than the mood disorder). Participants were followed approximately every week for up to 20 weeks.

From this parent study, we included 66 participants with at least 90% items on each of the seven MOODS Lifetime subscales observed prior to randomization. Of these, 44% (N=29) were randomized to IPSRT plus placebo and 56% (N=37) were randomized to IPSRT plus quetiapine.

Depression Phenotypes (DP) study

The DP study (Frank et al., 2011) randomized 318 currently depressed adults aged 18-65 diagnosed with a history of unipolar depression to either interpersonal psychotherapy (IPT) or selective serotonin reuptake inhibitor (SSRI) pharmacotherapy at two sites (Pittsburgh, PA and Pisa, Italy). All participants met DSM-IV criteria for major depression confirmed by the SCID and had a rating of ≥15 on the 17-item HRSD at screening. Relevant exclusion criteria included: need for acute inpatient hospitalization, schizophrenia, current schizoaffective disorder, a current substance use disorder related to their mood disorder, antisocial personality disorder, and a primary eating disorder diagnosis (i.e., the eating disorder was associated with the greater impairment). The study began with a variable length acute treatment phase defined as at least 12 weeks of treatment and 3 weeks of stable remission, after which participants entered a 6-month continuation treatment phase. If the initial treatment did not bring about response by week 6 or remission by week 12, the other treatment was added. Full study details are provided elsewhere (Frank et al., 2011).

From this parent study, the present study included 190 participants from the Pittsburgh site with at least 90% items observed on each of the seven Lifetime MOODS subscales prior to

randomization. 51% (N=99) were randomized to receive IPT as their initial treatment and the

remaining 49% (N=96) received SSRI as their initial treatment.

MOODS Lifetime instrument

The lifetime version of the MOODS instrument (Fagiolini et al., 1992) was used to reveal

new, naturally occurring subgroups in our sample. This measure allows for a continuous and

unitary approach to measuring depressive and (hypo)manic symptoms that may have occurred

over the course of an individual's life. The MOODS Lifetime instrument consists of 7

subdomains: mood manic (possible range: 0 – 27; observed range: 1 - 27), mood depressive

(possible/observed range: 0 – 28), energy manic (possible/observed range 0 – 9), energy

depressive (possible/observed range: 0 – 12), cognition manic (possible range: 0 – 27; observed

range: 0 - 26), cognition depressive (possible/observed range: 0 – 22), and rhythmicity (possible

range: 0 – 29; observed range = 2 - 28). Higher scores on each subdomain indicate greater

symptom severity. Ten individuals were missing up to 10% of items on at least one subscale

(individuals with >10% missing items on any one subscale were excluded from the sample). In

these instances, we imputed each missing item in the subscale with the mean of the observed

items in the subscale to allow the total subscale score to be calculated.

Baseline clinical and demographic characteristics

We compared subgroups across a variety of clinical and demographic characteristics that

were observed in all three randomized clinical trials. Demographic information included age,

gender, and race. Clinical characteristics related to illness history included age at first depressive,

manic, or hypomanic episode; years since first depressive, manic, or hypomanic episode;

whether the first episode occurred prior to age 15; and family history of mental illness (MDD,

bipolar disorder, anxiety, or schizophrenia). Additional clinical characteristics included quality

of life and life satisfaction based on the Quality of Life Enjoyment and Satisfaction

Questionnaire (QLESQ; Endicott et al., 1993), and the total Panic-Agoraphobic Spectrum (PAS;

Cassano, 1999) Lifetime score. Diagnostic characteristics included lifetime and past month

DSM-IV diagnoses of anxiety disorders, obsessive compulsive disorder, substance use disorders,

eating disorders, and post-traumatic stress disorder.

Treatment outcomes

The BP2 and DP studies used the 25-item HRSD to capture current depressive symptoms

and the Young Mania Rating Scale (YMRS; Young et al., 1978) to capture current (hypo)manic

symptoms. The BDCP study used the Bipolar Disorder Visit Form (BDVF; Kupfer et al., 2009)

to capture current (hypo)manic and depressive symptoms. Although these depressive and

(hypo)manic symptom data collection instruments are different, they capture many highly

overlapping symptoms (see Supplement). Thus, we created combined outcome measures of

current depressive and (hypo)manic symptoms that were consistent across all studies. The

combined depression score contained 11 symptoms: depressed mood, anhedonia, insomnia,

hypersomnia, appetite increase, appetite decrease, worthlessness or guilt, psychomotor agitation,

psychomotor retardation, fatigue, suicide attempt or ideation. The combined (hypo)mania score

contained 7 symptoms: elevated mood, irritability motor hyperactivity, decreased sleep, rapid

speech, inflated self-esteem, distractibility. On both depression and (hypo)mania scales, each

symptom was rated 0 (not present), 1 (subthreshold), or 2 (present). The total combined

depression scale score ranged from 0 to 22 and the total combined (hypo)mania scale score

ranged from 0 to 14. Because the (hypo)mania scale was skewed, we used a square root

transformation in analyses. Full details regarding the correspondence between specific BDVF,

HRDS, and YMRS items and the combined depressive and (hypo)manic symptom outcomes are provided in the supplement.

After examining the longitudinal data, we chose to consider current depression and (hypo)mania scores at baseline, 4, 8, 12, 16, and 20 weeks post-randomization (allowing scores within an interval of +/- 2 weeks at each time point) as the longitudinal outcome. These intervals were selected because they provided the greatest coverage and consistency across the three studies. Participants selected from the DP and BDCP studies were observed up to a median of 20 weeks [(Q1, Q3) = (20, 20)] in the present study. Participants selected from the BP2 study were observed up to a median of 20 weeks [(Q1, Q3) = (12, 20)] weeks in the present study.

Baseline levels of current depressive and (hypo)manic symptoms in the full sample and within each study are provided in Table 1. At baseline, participants in the BDVF study had lower current depressive symptoms compared to those in either DP or BP2 studies. This reflects the fact that the DP and BP2 studies required a minimum score on the HRSD to enter the study, while the BDCP study did not have this requirement. Participants in the BP2 study had higher current depressive and (hypo)manic symptoms relative to participants in the BDCP and DP studies.

Data Analysis

We applied a clustering method called "mixture modeling" (Fraley & Raftery, 2002) to identify subgroups with similar symptoms based on the seven continuously measured MOODS subdomains. Generally speaking, clustering is a technique that is used to divide a sample into more meaningful (homogenous) subgroups based on a set of selected characteristics, when

information such as the number of subgroups is not known *a priori*. These subgroups can be thought of as being separated by natural boundaries. The specific clustering method of mixture modeling is particularly useful for revealing these meaningful subgroups because it is a probability-based model, and thus, inherently allows for a more informed selection of the number of subgroups.

We considered models with one through six subgroups and any of 14 possible covariance structures. Models were compared using both Bayesian Information Criterion (BIC; Kass and Raftery, 1995) and the Bootstrap Likelihood Ratio Test (BLRT; McLachlan, 1987). The BLRT compares sequential numbers of clusters (e.g., 1 versus 2, 2 versus 3, etc.) through a likelihood ratio test statistic (LRTS). A significant LRTS p-value indicates the model with the additional cluster is a better fit for the data. These models were fit using the `mclust` package (Fraley & Raftery, 2002; Fraley et al., 2012) in `R` version 3.2.2 (R Core Team, 2015).

We performed an omnibus ANOVA, Chi-square, or Fisher's Exact test of subgroup differences for each comparison variable and then used the Benjamini-Hochberg (BH) approach (Benjamini & Hochberg, 1995; Glickman et al., 2014) to adjust for multiple comparisons within each set of analyses. If a test remained statistically significant after BH adjustment, we used Cohen's d effect sizes to evaluate whether pairwise differences between groups were clinically meaningful ($d > |.2|$). Finally, to evaluate the effects of subgroup beyond diagnosis and study, we regressed each comparison variable on relevant indicators for subgroup, study, and diagnosis, and noted any differences in inference after adjusting for these factors. Logistic and linear regression models were used for binary and continuous variables, respectively. `R` version 3.2.2 was used for these analyses.

We used linear mixed effects models to assess whether subgroup predicted (i.e., indicated a different outcome, regardless of treatment) or moderated (i.e., indicated different treatment effects for different subgroups) the longitudinal outcomes within each randomized clinical trial. At a minimum, these models included days since randomization, subgroup, and treatment assignment. Because the BDCP study included both BP1 and BP2 participants, the model for this study also included diagnosis. Focusing on the effect of subgroup membership within each study ensures that our findings are not conflated by differences across studies. However, to determine whether our findings might be robust across studies, we also fit models using the aggregated sample. At a minimum, models for the aggregated sample included days since randomization, subgroup, relevant indicators for diagnosis, and treatment assignment (keeping all 6 treatments across the three studies separate). Because treatments were unique within each study, including treatment also adjusts for variability resulting from study differences. For all models, we considered two- and three-way interactions (e.g., interactions with time to allow for different trajectories based on diagnosis, treatment, or subgroup), and a random intercept and slope were used to account for within-subject correlation resulting from repeated measures. Final models were selected based on the Akaike Information Criteria (Akaike, 1974). After fitting each final model, we used Cohen's d effect sizes estimated from post-hoc subgroup comparisons to reveal which subgroup differences were clinically meaningful ($|d| \geq .2$). SAS version 9.4 Proc Mixed (SAS Institute; Cary, NC) was used for these analyses.

RESULTS

Both the BIC and BLRT indicated that a four subgroup solution was the best fit for the data. BLRT tests confirmed that the four-cluster solution was the best fit for the data. Specifically, they indicated that the two-cluster model was superior to the one-cluster model

(LRTS = 120.7, p=.001), the three-cluster model was superior to the two-cluster model (LRTS = 42.8, p=.001), and the four-cluster model was superior to the three-cluster model (LRTS= 95.9, p=.001). However, the five-cluster model was not superior to the four-cluster model (LRTS = 25.1, p =.07), thereby indicating the four-cluster solution was the best fit for the data. The median (25%, 75%) uncertainty of the four-cluster model was .09 (.02, .15), indicating that the vast majority of observations were well classified.

Figure 1 illustrates the standardized means and 95% confidence intervals of each of the four subgroups on the seven MOODS subdomains. Individuals in subgroup 1 (N=77, 22%) had the highest severity based on all subdomains; we refer to this subgroup as "high severity" ("H"; 40% BP1; 46% BP2, 14% MDD). Individuals in subgroup 2 (N=115, 33%) had moderate severity based on all subdomains; we refer to this subgroup as "moderate severity" ("M"; 27% BP1; 33% BP2; 40% MDD). Individuals in subgroup 4 (N=73, 21%) had the lowest severity based all subdomains; we refer to this subgroup as "low severity" ("L"; 7% BP1; 11% BP2; 82% MDD). Individuals in subgroup 3 (N=82, 24%) had moderate (hypo)manic symptoms (energy-manic scores similar to those in the moderate severity group) but lower depressive symptoms (energy-depressive, mood-depressive, and cognition-depressive scores similar to those in the low severity subgroup). Thus, we refer to this subgroup as "low depressive / moderate (hypo)manic severity" ("LM"; 5% BP1; 6% BP2; 89% MDD).

Of the participants with BP1 (N=71), most were assigned to either the high (44%, N=31) or moderate (44%, N=31) severity subgroups. Similarly, of the participants with BP2 (N=86), most were assigned to either the high (41%, N=35) or moderate (44%, N=38) severity subgroups. The participants with MDD (N=190) were primarily distributed across the moderate (24%, N=46), low depression / moderate (hypo)mania (38%, N=73), and low (32%, N=60)

severity subgroups. However, it is notable that some (6%, N=11) MDD participants were assigned to the high severity subgroup.

Comparison on other clinical and demographic characteristics (Tables 2 and 3)

The subgroups revealed through clustering differed with respect to clinical but not demographic characteristics. Individuals in the high severity subgroup were primarily diagnosed as having BP1 or BP2 (86%, N=66), although it is notable that individuals diagnosed with MDD (14%, N=11) are also included in this group. They had the youngest age of their first depressive and (hypo)manic episodes, the longest duration since their first depressive episode, and were most likely to have had their first episode prior to age 15. They reported the highest severity of panic-agoraphobic symptoms based on the PAS and had the highest likelihood of a lifetime comorbid substance use disorder. In contrast, individuals in the low severity subgroup were primarily diagnosed as having MDD (82%, N=60). These individuals had the oldest age at first depressive or (hypo)manic episode, the fewest years since their first depressive episode, and were least likely to have had their first episode prior to age 15. They reported the lowest severity of panic-agoraphobia symptoms based on the PAS, the highest quality of life and life satisfaction, and had the lowest likelihood of a lifetime comorbid substance use diagnosis. Individuals in the moderate and low depressive / moderate (hypo)manic subgroups fell in the middle with respect to most characteristics, but with some important exceptions. Notably, individuals in the low depressive / moderate (hypo)manic severity subgroup reported the lowest quality of life and life satisfaction relative to all other subgroups. Importantly, these inferences did not change after controlling for indicators of diagnosis and study, suggesting the influence of subgroup on these characteristics above and beyond differences related clinical diagnosis and study.

Because of the clinically relevant finding that 14% (N=11) of individuals in the high severity subgroup were diagnosed with MDD, we aimed to determine whether these "high severity" individuals with MDD did indeed differ from than those with MDD in other subgroups in the way that would be expected based on their subgroup. As shown in Table 3, very similar patterns across subgroups were observed based on only the MDD sample as were seen in the full sample. Those with MDD in the high severity subgroup had the highest levels of lifetime mania and depression based on the MOODS, the highest lifetime levels of panic-agoraphobic symptoms based on the PAS, the earliest age at of first depressive episode, longest duration of illness, and the highest likelihood of a lifetime substance use disorder. Those with MDD in the low depressive / moderate (hypo)manic subgroup had similar lifetime (hypo)manic symptoms (energy-mania, cognition-mania) to those with MDD in the moderate severity subgroup but similar life depressive symptoms (mood-depressive, energy-depressive, cognition-depressive) to those with MDD in the low severity subgroup. These individuals also reported the worst quality of life and the lowest life satisfaction based on the QLESQ. An interesting difference not observed in the full sample was that those in the low depressive / moderate (hypo)mania group were also most likely to be female.

Comparison on treatment outcome (Figures 2 and 3)

In the DP study (MDD participants only), subgroup was a significant predictor of current depression symptoms (p<.001) over 20 week of treatment after controlling for time and treatment. However, subgroup was not associated with a different trajectory of depressive symptoms over time nor did it moderate the treatment effect. Post-hoc pairwise effect size

comparisons indicated that individuals in the low depressive / moderate (hypo)manic subgroup had meaningfully higher current depression scores throughout treatment relative to those in the high ($|d| = 0.26$), moderate ($|d| = 0.32$), and low ($|d| = 0.80$) severity subgroups.   Subgroup was also a predictor of (hypo)manic symptoms in this sample of MDD participants ($p=.05$) after controlling for treatment and time. Individuals in the high and moderate severity groups had greater (hypo)manic symptoms over time than those in the other two subgroups ($|d| = 0.31$ for H-LM; .29 for H-L, .32 for M-LM, .29 for M-L). In the BP2 study and BDCP studies, subgroup was not a predictor of depressive ($p=.222$ for BP2; $p=0.795$ for BDCP) or (hypo)manic ($p=.063$ for BP2, $p=0.110$ for BDCP) symptoms.

In the aggregated sample, subgroup was a predictor of both depressive ($p<.001$) and (hypo)manic ($p=.016$) symptoms over the 20 weeks of treatment. Both models included treatment, time, diagnosis, and the treatment-by-time interaction. Because each study used a different set of treatments, the inclusion of treatment in the model also adjusts for study differences. Pairwise comparisons for the depressive model indicated that the low depressive / moderate (hypo)manic subgroup had more depressive symptoms over the course of treatment relative to those in the moderate and low severity subgroups ($|d|$ for LM-M = .29; LM-L = .57) and similar symptoms as those in the high severity subgroup ($|d| = .12$). Those in the high and moderate severity subgroups also had greater depressive symptoms than those in the low severity subgroup ($|d| = .25$ for M-L, .33 for H-L). Pairwise comparisons for the (hypo)manic model indicated that those in the high and moderate severity subgroups had more (hypo)manic symptoms over the course of treatment than those in the other subgroups ($|d| = .27$ for H-LM, .25 for H-L, .28 for M-LM, .26 for M-L). Additional details regarding model results and pairwise comparisons for all models are provided in the supplement.

DISCUSSION

This study attempted to determine whether clustering methods could reveal clinically relevant subgroups across the mood spectrum [Major Depressive Disorder (MDD), Bipolar 2 Disorder (BP2), and Bipolar 1 Disorder (BP1)] based on lifetime mood severity as captured by the seven subdomains on the MOODS Lifetime self-report instrument. We found four subgroups: (1) high depressive and (hypo)manic symptom severity, (2) moderate depressive and (hypo)manic symptom severity, (3) low depressive and moderate (hypo)manic symptom severity, and (4) low (hypo)manic and depressive symptom severity. Individuals in the low depressive / moderate (hypo)manic severity subgroup reported the worst quality of life and the poorest life satisfaction. Individuals in the high and moderate severity subgroups were more likely to have substance use disorders. Among individuals diagnosed with MDD, we observed that: (1) those in the low depressive / moderate (hypo)manic severity subgroup experienced more current depressive symptoms throughout 20 weeks of treatment relative to individuals in the other subgroups, and (2) those in the high and moderate severity subgroups reported more current (hypo)manic symptoms throughout treatment.

Although using subgroups based on the Lifetime MOODS to predict current mood symptoms may at first seem tautological, we emphasize that the Lifetime MOODS instrument measures a different construct than current mood symptoms, which are captured by instruments such as the Hamilton Rating Scale for Depression, the Young Mania Rating Scale, or the DSM-IV. The most important differences are that the Lifetime MOODS: (1) emphasizes subthreshold symptoms; and (2) considers the presence of each symptom on its own over one's lifetime and

does not require symptoms to occur within a pre-specified duration of time (e.g., all in the previous week).   Given these differences, it would not be particularly surprising to find that the MOODS did not correspond to current treatment outcome. This makes our observance that MOODS subgroup membership is related to treatment outcome particularly novel and interesting.

Our findings have potential implications for classification of mood disorders.   A number of individuals with MDD were clustered into subgroups that were primarily populated by individuals with bipolar disorder diagnoses (i.e., the high and moderate severity subgroups), reporting lifetime (hypo)manic symptoms at higher levels than those with MDD in other subgroups. This finding is consonant with recent changes to the DSM-5 (American Psychiatric Association, 2013) that allow for subsyndromal mixed presentations, both on the manic and depressive sides. This finding is also consistent with previous work by Cassano et al. (2004), who found that individuals with recurrent MDD reported significant levels of (hypo)manic symptoms over their lifetimes (based on the MOODS) without reaching the categorical threshold for an actual bipolar diagnosis.

The findings presented herein also suggest important treatment considerations, especially for those diagnosed with MDD. First, because individuals with MDD in the high and moderate severity subgroups had the worst treatment outcomes with respect to (hypo)manic symptoms, we hypothesize that they may benefit from some of the treatment approaches, both psychopharmacologic and psychotherapeutic, more typically reserved, especially in the US, for those with a bipolar diagnosis. Second, somewhat surprisingly, those in the low depressive / moderate (hypo)manic severity subgroup were characterized as having the worst quality of life and the lowest life satisfaction, even though some other subgroups were associated with higher

average levels of lifetime mood symptoms.   This finding may be explained by the observance

that, although individuals with BP1 often come to treatment when manic, it is typically the

depressive mood states that drive treatment engagement for individuals with MDD and BP2.

Thus, the individuals with low lifetime depression severity paired with moderate (potentially

subthreshold) lifetime (hypo)manic symptoms are likely to be those with illness that is

unrecognized and undertreated over the course of their lives. Although these individuals

presented to the current trials with a major depressive episode (as was required for the research

protocols), their low level of *lifetime* depressive symptoms may have prevented receipt of

treatment for subthreshold yet clinically meaningful (hypo)manic symptoms, thereby

diminishing their quality of life. If clinicians were to emphasize the full continuum of both

depressive and (hypo)manic symptoms through use of the MOODS (for instance, if a patient

were to complete the form during a visit to a primary care office), these subthreshold

(hypo)manic symptoms may be brought to light earlier in the course of illness and treated

appropriately. Third, individuals with MDD who had low lifetime depressive yet moderate

(hypo)manic symptoms had the highest current depressive symptoms over 20 weeks of

treatment. Because the lifetime MOODS does not measure the duration of (potentially

subthreshold) symptoms over the course of the individual's life, this group that endorses

moderate levels of lifetime (hypo)mania and relatively lower levels (but not necessarily duration)

of depressive symptoms may represent individuals whose depressive symptoms, while not

severe, are nonetheless more chronic in nature and less likely to respond fully to conventional

treatment within a 20-week time frame. Individuals with this subacute high chronicity profile

may actually require longer treatment to achieve full remission of their depressive symptoms

than those with a "spiking" or episodic presentation.

Our findings should be considered in light of the fact that the sample came from three different randomized clinical trials. Other than mood diagnosis, the most notable difference across studies is that the BP2 and DP studies required a score of $\geq$ 15 on the HRSD to enter, while the BDCP participants did not. Also, the DP and BP2 studies had somewhat more stringent exclusion criteria than the BDCP study, especially with respect to current comorbid diagnoses. However, we do not expect these differences to impact the *subgroups* we revealed because: 1) the MOODS instrument captures lifetime (not current) symptoms; 2) all individuals were adults in a DSM-defined major depressive episode; 3) differences in exclusion criteria impacted only a small number of individuals (e.g., see Table 1); and 4) the majority of the individuals across the three studies came from the same research clinic in Pittsburgh and had largely overlapping research teams. Although we cannot rule out the possibility that more subtle and unmeasurable study differences may have induced artificial homogeneity within each study, our finding that subgroup characterizations did not change after controlling for study (or when considering only individuals with MDD) is empirical evidence that study differences had minimal impact on the subgroups we revealed.   However, because we do recognize the potential impact that study design differences (e.g., treatments, outcome measures, sampling scheme) may have on the longitudinal outcomes, we primarily focus on relating subgroup to treatment outcome within each study, rather than across studies.

Oher limitations to note in the present study are the small sample sizes within some of the subgroups tested in the longitudinal models (thereby decreasing power to detect effects), a larger proportion of individuals diagnosed with MDD than either BP1 or BP2 (although the proportion of MDD versus BP was roughly similar), and the exploratory nature of the study.   Regarding the latter limitation, it would be ideal to divide the sample into model development and validation

sets, fit a clustering model within each, and confirm that the two samples revealed similar results. Although our sample size of 347 is certainly large enough for the models presented herein, we do not expect that it is not large enough to divide the sample in half and fit reliable clustering models within each subsample, especially given the smaller number of individuals with BP1 and BP2 in the sample.   As such, our findings – while an important first step towards hypothesis generation regarding classification and treatment of mood disorders – do require additional validation.

Beyond the subgroups we revealed in the present sample, our work is potentially even more important insomuch as it demonstrates methods that can provide important information regarding classification and treatment personalization in other samples. Along these lines, it will be important to use methods such as those demonstrated herein to determine whether subgroups could be detected based on instruments more commonly used in large trials (e.g., HRSD and YMRS) but which are less sensitive to subthreshold symptoms. Subsequently, it would be possible evaluate how one or more of these subgroups differentially responds to a specific medication, such as lithium or citalopram. This would allow a clinician to know, for example, which specific treatment to provide for an individual diagnosed with MDD but with higher (hypo)manic symptoms. In the spirit of the National Institute of Mental Health's "Research Domain Criteria" (RDoC) initiative (Insel, 2014), one could also apply the approach presented herein to a range of measures captured not only through self-report but also genetic, neuroimaging, physiological, and behavioral measures, for example. This would further enhance our ability to improve classification and personalize treatment by clarifying disease mechanisms underlying the observed subgroups, and leading to at least one reliable 'antibiogram' that can inform our treatment choice.

References

1. Cassano, G.B., Rucci, P., Frank, E., Fagiolini, A., Dell'Osso, L., Shear, M.K., Kupfer, D.J., 2004. The mood spectrum in unipolar and bipolar disorder: arguments for a unitary approach. Am. J. Psychiatry. 161(7), 1264-1269.

2. Fagiolini, A., Frank, E., Rucci, P., Cassano, G.B., Turkin, S., Kupfer, D.J., 2007. Mood and anxiety spectrum as a means to identify clinically relevant subtypes of bipolar 1 disorder. Bipolar Disord. 9(5), 462-467.

3. Cassano, G.B., Dell'Osso, L., Frank, E., Miniati, M., Fagiolini, A., Shear, K., Pini, S., Maser, J., 1999. The bipolar spectrum: a clinical reality in search of a diagnostic criteria and an assessment methodology. J. Affect. Disord. 54(3), 319-328.

4. Angst, J., Cassano, G., 2005. The mood spectrum: Improving the diagnosis of bipolar disorder. Bipolar Disord. 7(Suppl 4), 4-12.

5. Altamura, A.C., Muoli, M., Caldiroli, A., Caron, L., Cumerlato, M.C., Dobrea, C., Cigliobianco, M., Zanelli Quarantini, F., 2015. Misdiagnosis, duration of untreated illness (DUI) and outcome in bipolar patients with psychotic symptoms: a naturalistic study. J. Affect. Disord. 182, 70-75.

6. Ghaemi, S.N., Sachs, G.S., Chiou, A.M., Pandurangi, A.K., Goodwin, K., 1999. Is bipolar disorder still underdiagnosed? Are antidepressants or overutilized? J. Affect. Disord. 52(1-3), 135-144.

7. Ghaemi, S.N., Ko, J.Y., Goodwin, F.K., 2001. The bipolar spectrum and the antidepressant view of the world. J. Psychiatr. Pract. 7(5), 287-297.

8. Liu, Y., Blackwood, D.H., Caesar, S., de Geus, E.J., Farmer, A., Ferreira, M.A., Ferrier, I.N., Fraser, C., Gordon-Smith, K., Green, E.K., Grozeva, D., Gurling, H.M., Hamshere, M.L., Heutink, P., Holmans, P.A., Hoogendijk, W.J., Hottenga, J.J., Jones, L., Jones, I.R., Kirov, G., Lin, D., McGuffin, P., Moskvina, V., Nolen, W,A., Perlis, R.H., Posthuma, D., Scolnick, E.M., Smit, A.B., Smit, J.H., Smoller, J.W., St Clair, D., van Dyck, R., Verhage, M., Willemsen, G., Young, A.H., Zandbelt, T., Boomsma, D.I., Craddock, N., O'Donovan, M.C., Owen, M.J., Penninx, B.W., Purcell, S., Sklar, P., Sullivan, P.F., Wellcome Trust Case-Control Consortium, 2011. Meta-analysis of genome-wide association data of bipolar disorder and major depressive disorder. Mol. Psychiatry 16, 2-4.

9. Dell'Osso, B., D'Addario, C., Carlotta Palazzo, M., Benatti, B., Camuri, G., Galimberti, D., Fenoglio, C., Scarpini, E., Di Francesco, A., Maccarrone, M., Altamura, A.C., 2014. Epigenetic modulation of BDNF gene: differences in DNA methylation between unipolar and bipolar patients. J. Affect. Disord. 166, 330-333.

10. Duffy, A., Grof, P., Robertson, C., Alda, M., 2000. The implications of genetic studies of major mood disorders for clinical practice. J. Clin. Psychiatry 61(9), 630-7.

11. McGuffin, P., Katz, R., 1989. Genetics of depression and manic-depressive disorder. Br. J. Psychiatry 15, 294-304.

12. Fagiolini, A., Dell'Osso, L., Pini, S., Armani, A., Bouanani, S., Rucci, P., Cassano, G.B., Endicott, J., Maser, J.D., Shear, M.K., Grochocinski, V.J., Frank, E., 1999. Validity and reliability of a new instrument for assessing mood symptomatology: the Structured

Clinical Interview for Mood Spectrum (SCI-MOODS). Int. J. Methods. Psych. Res. 8(2), 71-82.

13. Cassano, G.B., Frank, E., Miniati, M., Rucci, P., Fagiolini, A., Pini, S., Shear, M.K., Maser, J.D., 2002. Conceptual underpinnings and empirical support for the mood spectrum. Psychiatr. Clin. North Am. 25(4), 699-712.

14. Jablensky, A., 2016. Psychiatric classifications: validity and reliability. Word Psychiatry 15, 26-31.

15. Philips, M.R., 2016. Would the use of dimensional measures improve the utility of psychiatric diagnoses? World Psychiatry 15(1), 383-389.

16. Kupfer, D.J., Axelson, D.A., Birmaher, B., Brown, C., Curet, D.E., Fagiolini, A., Frank, E., Friedman, E.S., Grochocinski, V.J., Houck, P.R., Kilbourne, A.M., Mulsant, B.H., Pollock, B.G., Reynolds, C.F. 3rd, Stofko, M.G., Swartz, H.A., Thase, M.E., Turkin, S.R., Whyte, E.M., 2009. Bipolar disorder center for Pennsylvanians: implementing an effectiveness trial to improve treatment for at-risk patients. Psychiatr. Serv. 60(7), 888-897.

17. Fagiolini, A., Frank, E., Axelson, D.A., Birmaher, B., Cheng, Y., Curet, D.E., Friedman, E. S., Gildengers, A.G., Goldstein, T., Grochocinski, V.J., Houck, P.R., Stofko, M.G., Thase, M.E., Thompson, W.K., Turkin, S.R., Kupfer, D.J., 2009. Enhancing outcomes for patients with bipolar disorders results from the Bipolar Disorder Center for Pennsylvanians Study. Bipolar Disord. 11(4), 382-390.

18. Frank, E., Cassano, G.B., Rucci, P., Thompson, W.K., Kraemer, H,C., Fagiolini, A., Maggi, L., Kupfer, D.J., Shear, M.K., Houck, P.R., Calugi, S., Grochocinski, V.J., Scocco, P., Buttenfield, J., Forgione, R.N., 2011.

Predictors and moderators of time to remission of major depression with interpersonal

psychotherapy and SSRI pharmacotherapy. Psychol. Med. 41(1), 151-62.

19. First, M.B., Spitzer, R.L., Gibbon, M. Williams, J.B.W., 2001. User's Guide for the

Structured Clinical Interview for DSM-IV Axis I Disorders - Research

Version. Biometrics Research, New York State Psychiatric Institute, New York.

20. American Pschiatric Association, 2000. Task force for the handbook of psychiatric

measures: handbook of psychiatric measures. American Psychiatric Association,

Washington DC.

21. Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., Ryan, N.,

1997. Schedule for affective disorder and schizophrenia for school-age children, present and

lifetime version. (KSADS-PL). Initial reliability and validity data. J. Am. Acad. Child Adoslc.

Psychiatry, 36, 980-988.

22. Hamilton, M., 1950. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 25, 56-62.

23. Endicott, J., Nee, J., Harrison, W., Blumenthal., R., 1993. Quality of life enjoyment and

satisfaction questionnaire: a new measure. Psychopharmacol. Bull. 29(2), 321-326.

24. Cassano, G.B., Banti, S., Muari, M., Dell'Osso, L., Miniati, M., Maser, J.D., Shear,

M.K., Frank. E., 1999. Internal consistency and discriminant validity of the Structured

Clinical Interview for Panic Agoraphobic Spectrum (SCI-PAS). Int. J. Methods Psych.

Res. 8(3), 138-145.

25. Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania:

Reliability, validity and sensitivity. Br. J. Psychiatry 133, 429-435.

26. Fraley, C., Raftery, A.E. 2002. Model-based clustering, discriminant analysis and density

estimation. J. Am. Stat. Assoc. 97, 611-631.

27. Fraley, C., Raftery, A.E., Murphy, T.B., Lucca, S., 2012. mclust Version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. Technical Report No. 597, Department of Statistics, University of Washington.

28. R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

29. Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90(430), 773-795.

30. McLachlan, G.J., 1987. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. J. R. Stat. Soc. Ser. C Appl. Stat. 36(3), 318-324.

31. Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289-300.

32. Glickman, M.E., Rao, S.R., Schultz, M.R., 2014. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J. Clin. Epidemio. 67(8), 850-857.

33. Akaike, H., 1974. A new look at the statistical model identification. IEEE T. Automat. Contr. 19(6), 717-723.

34. American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders: DSM-5. American Psychiatric Association, Washington, D.C.

35. Insel, T.R., 2014. The NIMH Resarch Domain Criteria (RDoC) project: precision medicine for psychiatry. Am. J. Psychiatry 171(4), 395-397.

Table 1. Study comparison based on clinical and demographic characteristics.   Abbreviations: BDCP = Bipolar Disorder Center for Pennsylvanians; BP2 = Bipolar 2 Study, DP =Depression Phenotypes Study; QLESQ = Quality of Life Enjoyment and Satisfaction Questionnaire; PAS=Panic-Agoraphobic Spectrum; SCID = Structured Clinical Interview for DSM-IV. FE in place of statistic indicates that the Fisher's Exact test was used because of small sample sizes within each cell.

	Full sample (N=347)	BDCP (N=91; 78% BP1, 22% BP2)	BP2 (N=66; 100% BP2)	DP (N=190; 100% MDD)	F or Chi-square Statistic (p-value)	Pairwise Comparisons ($|d| > .2$)
Lifetime MOODS, *mean (SD)*						
Mood-Manic	19.19 (4.76)	20.66 (4.37)	20.5 (4.01)	18.04 (4.88)	13.31 (<0.001)	(1, 2)>3
Mood-Depressive	13.54 (6.83)	17.88 (5.39)	18.38 (5.21)	9.78 (5.54)	100.65(<0.001)	(1, 2)>3
Energy-Manic	6.54 (2.48)	7.35 (1.92)	7.48 (2.02)	5.83 (2.64)	19.39 (<0.001)	(1, 2)>3
Energy-Depressive	6.16 (3.66)	8.88 (2.52)	8.61 (2.68)	4.02 (2.96)	124.19(<0.001)	(1, 2)>3
Cognition-Manic	15.17 (5.07)	18.05 (3.86)	16.76 (5.19)	13.24 (4.69)	38.62(<0.001)	1>2>3
Cognition-Depressive	9.02 (5.24)	12.33 (4.48)	12 (4.36)	6.39 (4.34)	75.37(<0.001)	(1, 2)>3
Rhythm	16.91 (5.01)	19.42 (4.43)	19.13 (4.83)	14.94 (4.47)	39.93(<0.001)	(1, 2)>3
Clinical Characteristics, *mean(SD) or %(N)*						
Age of First Depressive Episode (N=333)	21.17 (10.91)	18.39 (7.8)	16.22 (5.68)	24.13 (12.44)	17.87 (<0.001)	3>1>2
Age of First Manic or Hypomanic Episode (N=144)	20.94 (7.83)	21.83 (7.88)	19.81 (7.68)	NA	2.37 (0.126)	
First Episode ≤ Age 15 (N=338)	34.91 (118)	44.83 (39)	50.77 (33)	24.73 (46)	19.44 (<0.001)	(1, 2)>3
Years since First Depressive Episode (N=333)	17.37 (12.74)	22.13 (11.98)	17.12 (11.08)	15.36 (13.11)	8.41 (<0.001)	1>2>3
Years Since First (hypo)manic Episode (N=144)	16.7 (10.95)	19.01 (10.98)	13.73 (10.25)	NA	8.7 (0.004)	1>2
Family History of MDD, BP1, BP2, Anxiety, or Schizophrenia (N=248)	96.37 (239)	94.32 (83)	100 (57)	96.12 (99)	FE (0.208)	

Quality of Life (QLESQ; N=324)	38.95 (8.11)	40.43 (9.84)	37.81 (7.91)	38.71 (7.33)	1.94 (0.145)	
Life Satisfaction (QLESQ; N=338)	2.55 (0.84)	2.68 (1.06)	2.56 (0.82)	2.49 (0.72)	1.59 (0.206)	
Panic-Agoraphobic Symptoms (PAS; N=345)	33.29 (20.11)	41.11 (20.63)	41.63 (19.78)	26.73 (17.51)	25.71 (<0.001)	(1, 2)>3
SCID Diagnoses (Lifetime)						
Any Anxiety Disorder	62.25 (216)	54.95 (50)	69.7 (46)	63.16 (120)	3.69 (0.158)	
Obsessive Compulsive Disorder	6.05 (21)	5.49 (5)	4.55 (3)	6.84 (13)	FE (0.867)	
Substance Use Disorder	35.16 (122)	58.24 (53)	34.85 (23)	24.21 (46)	31.26 (<0.001)	1> 2> 3
Eating Disorder	12.1 (42)	17.58 (16)	15.15 (10)	8.42 (16)	5.57 (0.062)	
Posttraumatic Stress Disorder	12.39 (43)	15.38 (14)	15.15 (10)	10 (19)	2.21 (0.33)	
SCID Diagnoses (Past Month)						
Any Anxiety Disorder	50.43 (175)	37.36 (34)	63.64 (42)	51.11 (99)	11.03 (0.004)	2>1
Obsessive Compulsive Disorder	2.59 (9)	1.1 (1)	3.03 (2)	3.16 (6)	FE (0.652)	
Substance Use Disorder	2.31 (8)	4.4 (4)	3.03 (2)	1.05 (2)	FE (0.151)	
Eating Disorder	2.59 (9)	4.4 (4)	4.55 (3)	1.05 (2)	FE (0.073)	
Posttraumatic Stress Disorder	4.03 (14)	4.4 (4)	6.06 (4)	3.16 (6)	FE (0.503)	
Demographics						
Female (vs. Male)	63.11 (219)	68.13 (62)	65.15 (43)	60 (114)	1.89 (0.388)	
White (vs. Non-white)	82.42 (286)	82.42 (75)	74.24 (49)	85.26 (162)	4.11 (0.128)	
Age	38.37 (12.17)	40.87 (11.6)	32.96 (10.79)	39.05 (12.38)	9.14 (<0.001)	(1, 3)>2
Current Mood Symptom Scores						
Depressive Symptoms	12.76 (3.48)	6.20 (4.06)	13.79 (2.64))	13.43 (2.39)	94.89 (<0.001)	2>3>1
Square Root of (Hypo)anic Symptoms	1.02 (0.86)	0.80 (1.08)	1.79 (0.73)	0.72 (0.64)	50.06 (<0.001)	2>3>1

Table 2. Subgroup comparisons based on clinical and demographic characteristics.
Abbreviations: BP1 = Bipolar 1 Disorder; BP2 = Bipolar 2 Disorder, MDD = Major Depressive
Disorder; QLESQ = Quality of Life Enjoyment and Satisfaction Questionnaire; PAS=Panic-
Agoraphobic Spectrum; SCID = Structured Clinical Interview for DSM-IV. FE in place of
statistic indicates that the Fisher's Exact test was used because of small sample sizes within cells.

	Full sample (N=347)	High Severity (N=77)	Moderate Severity (N=115)	Low Depressive / Moderate (Hypo)manic Severity (N=82)	Low Severity (N=73)	F or Chi-square Statistic (p-value)	Pairwise Comparisons (|d| > .2)
Lifetime MOODS, *mean (SD)*							
Mood-Manic	19.19 (4.76)	22.65 (2.24)	20.08 (3.57)	18.99 (3.3)	14.38 (5.78)	60.26 (<0.001)	1>2>3>4
Mood-Depressive	13.54 (6.83)	21.05 (3.41)	15.71 (4.83)	7.89 (4.12)	8.54 (5.12)	156.57 (<0.001)	1> 2 > (3,4)
Energy-Manic	6.54 (2.48)	8.27 (0.81)	7.37 (1.48)	7.29 (1.21)	2.56 (1.66)	281.21 (<.0001)	1> (2, 3) > 4
Energy-Depressive	6.16 (3.66)	10.31 (1.13)	7.92 (1.9)	2.48 (1.7)	3.16 (2.46)	335.24 (<0.001)	1>2 > (3,4)
Cognition-Manic	15.17 (5.07)	19.93 (2.7)	15.89 (3.52)	14.74 (3.98)	9.48 (4.49)	101.72 (<0.001)	1>2>3>4
Cognition-Depressive	9.02 (5.24)	14.77 (3.04)	10.13 (4.01)	5.24 (3.4)	5.45 (4.16)	115.83 (<0.001)	1>2 > (3,4)
Rhythm	16.91 (5.01)	21.93 (2.13)	17.4 (3.98)	15.61 (3.89)	12.33 (4.82)	82.32 (<0.001)	1>2>3>4
Clinical Characteristics, *mean(SD) or %(N)*							
Age of First Depressive Episode (N=333)	21.17 (10.91)	16.92 (7.18)	19.99 (9.15)	21.01 (11.6)	27.61 (12.91)	13.91 (<0.001)	4>(2,3)>1
Age of First Manic or Hypomanic Episode (N=144)	20.94 (7.83)	19.15 (6.81)	21.15 (7.64)	25 (11.83)	25.77 (7.76)	3.76 (0.012)	(3,4)>(1,2)
First Episode ≤ Age 15 (N=338)	34.91 (118)	53.33 (40)	36.11 (39)	32.93 (27)	16.44 (12)	22.37 (<0.001)	1>(3,4); 2>4
Years since First Depressive Episode (N=333)	17.37 (12.74)	20.87 (12.13)	18.08 (12.21)	18.93 (14.04)	10.87 (10.29)	9.2 (<0.001)	(1,2,3)>4; 1 > 2
Years Since First (hypo)manic Episode (N=144)	16.7 (10.95)	17.89 (11.92)	16.03 (10.06)	22.22 (10.86)	10.31 (7.26)	2.65 (0.051)	
Family History of MDD, BP1, BP2, Anxiety, or	96.37 (239)	97.22 (70)	97.62 (82)	96.23 (51)	92.31 (36)	FE (0.512)	

Schizophrenia (N=248)							
Quality of Life (QLESQ; N=324)	38.95 (8.11)	37.76 (8.56)	39.33 (8.77)	35.91 (7.07)	42.94 (5.86)	10.97 (<0.001)	4>(1,2)>3
Life Satisfaction (QLESQ; N=338)	2.55 (0.84)	2.45 (0.84)	2.68 (0.93)	2.21 (0.73)	2.83 (0.65)	9.16 (<0.001)	(2, 4)>1 >3
Panic-Agoraphobic Symptoms (PAS; N=345)	33.29 (20.11)	49.46 (18.93)	33.94 (18.48)	29.35 (17.74)	19.79 (13.78)	37.44 (<0.001)	1>2>3>4
SCID Mood Diagnoses							
MDD	54.76 (190)	14.29 (11)	40 (46)	89.02 (73)	82.19 (60)	122.06 (<0.001)	(3,4) > 2 > 1
BP2	24.78 (86)	45.45 (35)	33.04 (38)	6.1 (5)	10.96 (8)	44.7 (<0.001)	(1,2)>(3, 4)
BP1	20.46 (71)	40.26 (31)	26.96 (31)	4.88 (4)	6.85 (5)	42.07 (<0.001)	(1,2)>(3, 4)
BP1 or BP2	45.2 (157)	85.71 (66)	60 (69)	11 (9)	17.8 (13)	122.06 (<0.001)	1>2> (3,4)
Other SCID Diagnoses (Lifetime)							
Any Anxiety Disorder	62.25 (216)	63.64 (49)	62.61 (72)	57.32 (47)	65.75 (48)	1.3 (0.729)	
Obsessive Compulsive Disorder	6.05 (21)	9.09 (7)	3.48 (4)	8.54 (7)	4.11 (3)	FE (0.264)	
Substance Use Disorder	35.16 (122)	49.35 (38)	46.96 (54)	19.51 (16)	19.18 (14)	30.81 (<.001)	(1,2)>(3, 4)
Eating Disorder	12.1 (42)	16.88 (13)	12.17 (14)	9.76 (8)	9.59 (7)	2.51 (0.473)	
Posttraumatic Stress Disorder	12.39 (43)	19.48 (15)	9.57 (11)	15.85 (13)	5.48 (4)	8.53 (0.036[a])	
Other SCID Diagnoses (Past Month)							
Any Anxiety Disorder	50.43 (175)	51.95 (40)	47.83 (55)	50 (41)	53.42 (39)	0.65 (0.885)	
Obsessive Compulsive Disorder	2.59 (9)	3.9 (3)	0.87 (1)	4.88 (4)	1.37 (1)	FE (0.262)	
Substance Use Disorder	2.31 (8)	3.9 (3)	2.61 (3)	1.22 (1)	1.37 (1)	FE (0.686)	
Eating Disorder	2.59 (9)	5.19 (4)	3.48 (4)	1.22 (1)	0 (0)	FE (0.168)	
Posttraumatic Stress Disorder	4.03 (14)	6.49 (5)	2.61 (3)	4.88 (4)	2.74 (2)	FE (0.543)	
Demographics							
Female (vs. Male)	63.11 (219)	63.64 (49)	58.26 (67)	75.61 (62)	56.16 (41)	8.19 (0.042[a])	
White (vs. Non-white)	82.42 (286)	79.22 (61)	81.74 (94)	89.02 (73)	79.45 (58)	3.49 (0.322)	

Age	38.37 (12.17)	37.33 (11.82)	38.22 (11.97)	39.72 (12.67)	38.16 (12.38)	0.54 (0.657)

a. Not significantly different after Benjamini Hochberg correction for multiple comparisons

Table 3. Subgroup comparisons based on clinical and demographic characteristics among MDD participants only. Abbreviations: BP1 = Bipolar 1 Disorder; BP2 = Bipolar 2 Disorder, MDD = Major Depressive Disorder; QLESQ = Quality of Life Enjoyment and Satisfaction Questionnaire; PAS=Panic-Agoraphobic Spectrum; SCID = Structured Clinical Interview for DSM-IV. FE in place of statistic indicates Fisher's Exact test was used because of small sample sizes within cells.

	Full sample (N=190)	High Severity (N=11)	Moderate Severity (N=46)	Low Depressive / Moderate (Hypo)manic Severity (N=73)	Low Severity (N=60)	F or Chi-square Statistic (p-value)	Pairwise Comparisons (|d| > .2)
Lifetime MOODS, *mean (SD)*							
Mood-Manic	18.04 (4.88)	21.73 (1.62)	19.85 (3.77)	19.11 (3.29)	14.67 (5.81)	19.24 (<0.001)	1>2>3>4
Mood-Depressive	9.78 (5.54)	18.55 (3.5)	13.75 (4.58)	7.49 (3.89)	7.9 (5)	36.93 (<0.001)	1> 2 > (3,4)
Energy-Manic	5.83 (2.64)	8.55 (0.69)	7.04 (1.59)	7.29 (1.25)	2.62 (1.66)	145.4 (<0.001)	1> (2, 3) > 4
Energy-Depressive	4.02 (2.96)	9.91 (1.22)	6.96 (1.4)	2.32 (1.58)	2.75 (2.19)	119.12 (<0.001)	1>2 > (3,4)
Cognition-Manic	13.24 (4.69)	18 (2.1)	15.15 (3.07)	14.62 (3.9)	9.22 (4.26)	35.95 (<0.001)	1>(2, 3) > 4
Cognition-Depressive	6.39 (4.34)	13.36 (2.2)	8.46 (4.09)	5.01 (3.3)	5.22 (4.17)	22.59 (<0.001)	1>2 > (3,4)
Rhythm	14.94 (4.47)	21.18 (2.09)	16.43 (2.86)	15.55 (3.99)	11.92 (4.33)	25.76 (<0.001)	1>2>3>4
Clinical Characteristics, mean(SD) or %(N)							
Age of First Depressive Episode (N=186)	24.13 (12.44)	22.9 (11.26)	22.65 (11.49)	20.97 (11.8)	29.23 (12.71)	5.55 (<0.001)	4>(1,2,3)
First Episode ≤ Age 15 (N=186)	24.51 (25)	25 (1)	28.57 (8)	16.22 (6)	30.3 (10)	FE (0.017)	(2,3)>4
Years since First Depressive Episode (N=190)	15.36 (13.11)	17.2 (11.86)	17.37 (13.39)	18.38 (14.02)	9.94 (10.24)	5.49 (< 0.001)	(1,2,3)>4
Family History of MDD, BP1, BP2, Anxiety, or Schizophrenia (N=103)	96.12 (99)	100 (8)	100 (19)	95.74 (45)	93.1 (27)	FE (0.758)	

Quality of Life (QLESQ; N=189)	38.71 (7.33)	37.73 (10.45)	38.72 (6.32)	35.79 (7.2)	42.51 (5.86)	10.61(<0.001)	4>(1,2)>3
Life Satisfaction (QLESQ; N=189)	2.49 (0.72)	2.55 (0.82)	2.5 (0.69)	2.22 (0.73)	2.8 (0.58)	7.82 (<0.001)	4>(1,2)>3
Panic-Agoraphobic Symptoms (PAS; N=190)	26.73 (17.52)	42.18 (21.21)	32.26 (17.08)	28.22 (18.06)	17.85 (11.62)	11.28 (<0.001)	1>2>3>4
DSM-IV Diagnoses (Lifetime)							
Any Anxiety Disorder	63.16 (120)	54.55 (6)	60.87 (28)	58.9 (43)	71.67 (43)	FE (0.409)	
Obsessive Compulsive Disorder	6.84 (13)	9.09 (1)	4.35 (2)	9.59 (7)	5 (3)	FE (0.589)	
Substance Use Disorder	24.21 (46)	36.36 (4)	34.78 (16)	20.55 (15)	18.33 (11)	FE (0.148)	
Eating Disorder	8.42 (16)	0 (0)	6.52 (3)	9.59 (7)	10 (6)	FE (0.671)	
Posttraumatic Stress Disorder	10 (19)	9.09 (1)	6.52 (3)	16.44 (12)	5 (3)	FE (0.13)	
DSM-IV Diagnoses (Past Month)							
Any Anxiety Disorder	52.11 (99)	36.36 (4)	50 (23)	52.05 (38)	56.67 (34)	FE (0.647)	
Obsessive Compulsive Disorder	3.16 (6)	9.09 (1)	0 (0)	5.48 (4)	1.67 (1)	FE (0.18)	
Substance Use Disorder	1.05 (2)	0 (0)	0 (0)	1.37 (1)	1.67 (1)	FE (.999)	
Eating Disorder	5.66 (6)	0 (0)	10.34 (3)	5.13 (2)	2.94 (1)	FE (0.17)	
Posttraumatic Stress Disorder	3.16 (6)	0 (0)	2.17 (1)	5.48 (4)	1.67 (1)	FE (0.704)	
Demographics							
Female (vs. Male)	60 (114)	54.55 (6)	45.65 (21)	73.97 (54)	55 (33)	FE (0.014)	3>4
White (vs. Non-white)	85.26 (162)	81.82 (9)	86.96 (40)	87.67 (64)	81.67 (49)	FE (0.762)	
Age	39.05 (12.38)	39.35 (13.52)	39.7 (12.67)	38.85 (12.13)	38.72 (12.56)	0.06 (0.979)	

Figure Legends

Figure 1. Standardized mean MOODS Lifetime scores with 95% confidence intervals. H=High symptom severity; M = Moderate symptom severity; LM = Low depressive / moderate (hypo)manic symptom severity; L = Low symptom severity.
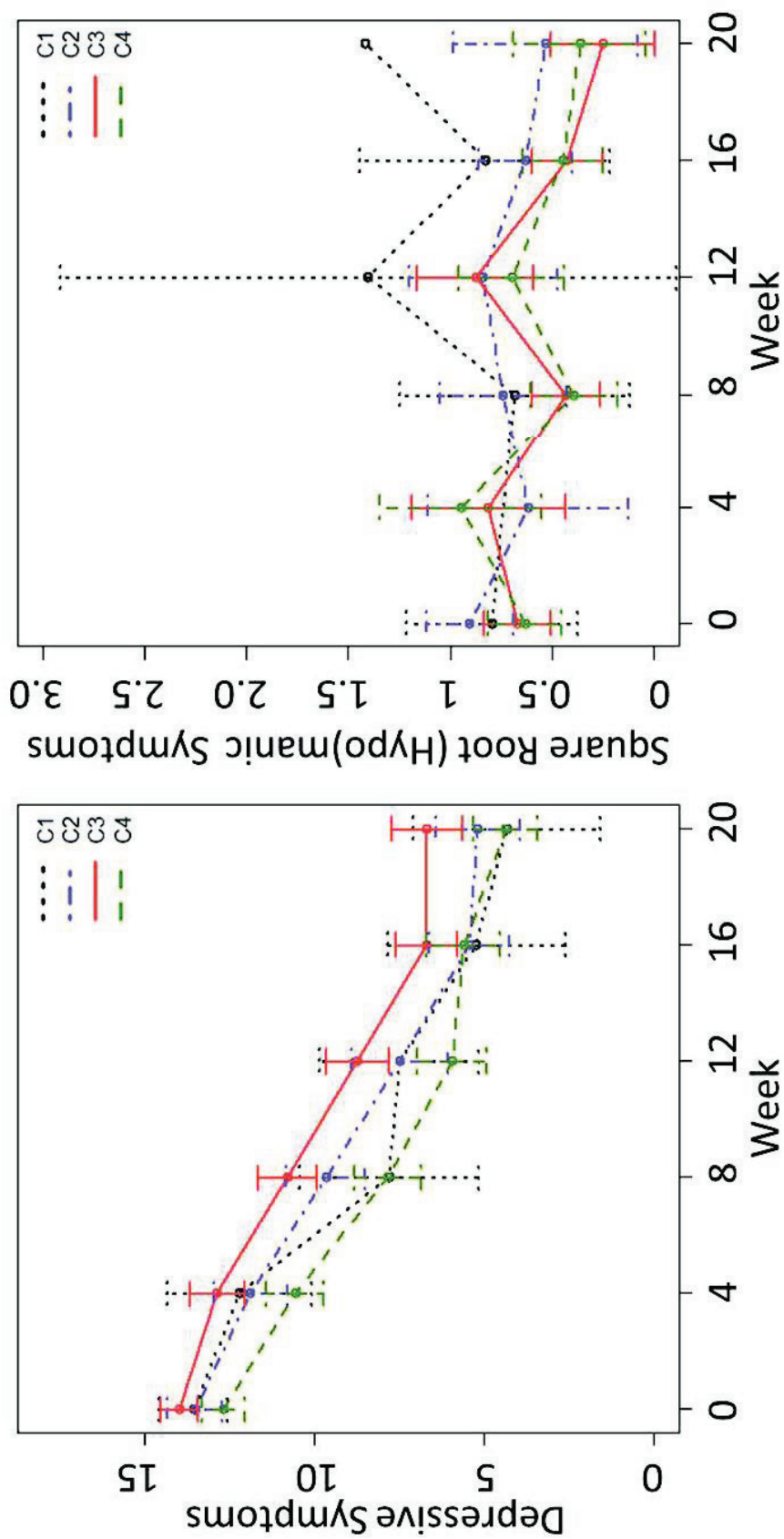
Figure 2. Observed mean (95% CI) depressive and (hypo)manic symptoms over 20 weeks of treatment in the MDD sample only.

Figure 2. Observed mean (95% CI) depressive and (hypo)manic symptoms over 20 weeks of treatment in the aggregated sample.

35

Figure 2. Observed mean (95% CI) depressive and (hypo)manic symptoms over 20 weeks of treatment in the MDD sample only.
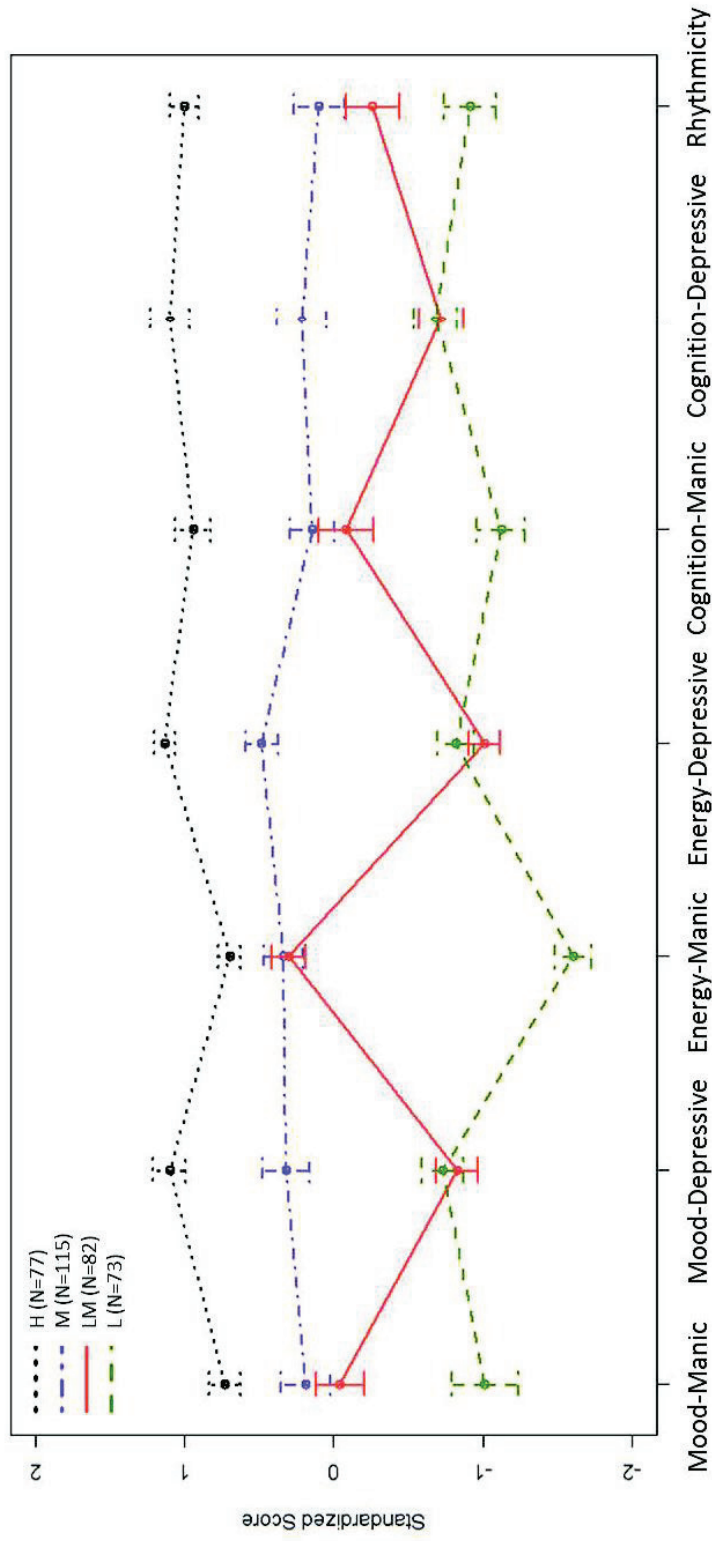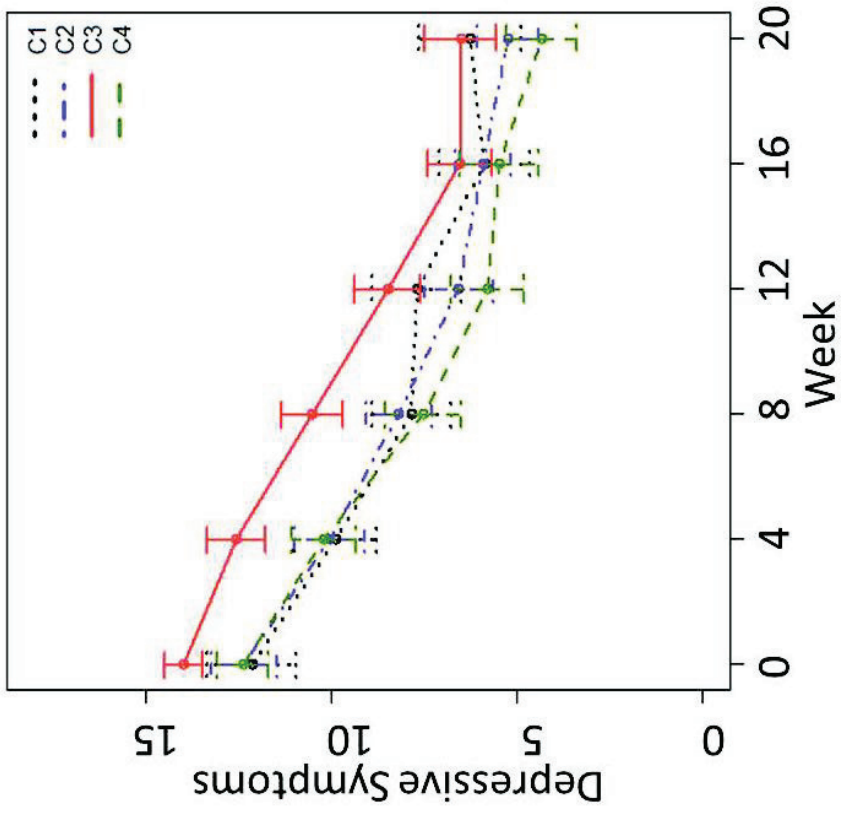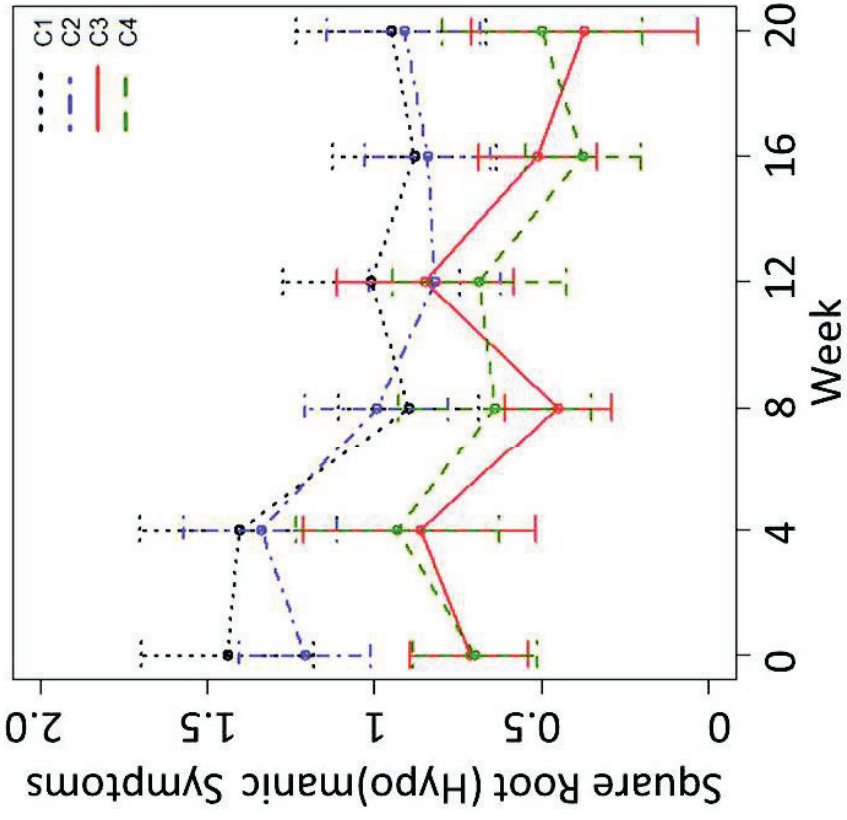
Figure 3. Observed mean (95% CI) depressive and (hypo)manic symptoms over 20 weeks of treatment in the aggregated sample.

37

Highlights:

- Rural case study on decentralized generation and storage technology (DGST) benefits
- Cost optimization model and scenarios developed to assess DGSTs until 2050
- Small hydro and photovoltaics (PV) increase self-sufficiency of community
- Storage enables full hydro potential usage and increased PV penetration
- Carbon price policies effective in mitigating local fossil fuel emissions