Smartphone Monitoring of Mood Instability in Young Depressed Patients: A Latent-class Analyses

Vibha Anand, PhD[1], Bo Hu, PhD[2], Amit Anand, MD[3]

[1]Center for Computational Health, IBM T.J. Watson Research Center, Cambridge, MA, USA
[2] Quantitative Health Sciences, Cleveland Clinic, Ohio, USA
[3] Center for Behavioral Health, Cleveland Clinic Foundation, Ohio, USA

Abstract

This study captured daily and weekly mood ratings using a smartphone from bipolar disorder (BD) and unipolar major depression disorder (MDD) subjects at high (HRMDD) and low risk (LRMDD) for developing Bipolar Disorder (BD) and healthy controls (HC). Method: 40 subjects (18 – 30 yr) (6 BD, 13 HRMDD, 16 LRMDD and 5 HC) were studied and a total of 2401 daily and 744 weekly ratings were collected. HRMDD and LRMDD subjects were naturalistically treated with antidepressants. We investigate if latent-class analyses of ratings can detect mood instability among MDD and BD groups. Results: Our analyses revealed four underlying mood states correlating with clinical mood states. There was a trend for greater number of state changes in BD and HRMDD subjects compared to LRMDD and HC groups. Conclusion: Smartphone ratings may adequately capture mood instability in BD subjects and at risk HRMDD subjects and offers a prudent way for monitoring development of serious manic symptoms.

Introduction

Depression can manifest itself in two forms – unipolar or major depression (MDD) in which subjects only suffer from episodes of depression (low mood and lack of pleasure, decreased energy, sleep/appetite problems, decreased concentration, hopelessness and suicidal ideation) and bipolar disorder (BD in which patients suffer from periods of depression but also suffer from periods of mania (elation or irritability, excessive energy, lack of need for sleep, rapid speech, grandiosity, poor judgment and impulsivity). [1] BD is a more serious illness as it is frequently associated with more severe depressive symptoms and with mania that can lead to serious social and legal consequences. Furthermore, the mood instability from periods of depressive symptoms to periods mania can lead to an unstable and chaotic lifestyle.

One critical problem in the treatment of MDD, particularly in young patients, is that some MDD patients may have a hidden risk for developing a (hypo) manic episode and conversion of diagnosis to BD. Many of these vulnerable subjects have a family history of BD or have sub-threshold BD symptoms. [2,3] It is critical to closely follow-up these patients for changes in mood symptoms particularly during antidepressant treatment as the response to treatment can be unpredictable in these patients. Some studies suggest a propensity for precipitation of mania or rapid antidepressant response [4] while others suggest that antidepressants can be safely given to MDD groups at high risk of developing BD. [5] Clinical appointments spaced out over several months may not be able to capture fluctuations in mood or occurrence of (hypo)manic symptoms. Patient reports at these appointments with clinicians are at best subjective retrospective recollections compounded by lack of recognition of (hypo)mania symptoms by the patients in themselves.

Therefore, it is critically important to develop new methods to measure any increase in mood fluctuations in MDD patients that have been recently started on antidepressants. These new methods need to be more sensitive to changes occurring over time as well as be more objective in terms of how they measure changes in mood instability between periods of depressive, (hypo)manic and euthymic states. In this regard, the recent availability of smartphone technology and the emerging field of behavioral informatics has provided a new way of acquiring and measuring mood data. Smartphone applications are being increasingly used to monitor mood symptoms and also deliver

interventions in MDD and BD. [6–10] However, few studies have investigated at-risk mood disorder populations to monitor effect of antidepressant treatment using smartphone technology.

In this study, we used a smartphone application to monitor the effect of antidepressant on mood instability in terms of fluctuations between mania and depression, in young MDD subjects who are at a high risk of developing BD (HRMDD) and compared them to young patients with depression at low risk of developing bipolar disorder (LRMDD) as well as with subjects already diagnosed with BD and Healthy Controls (HC). We perform latent class analyses of smartphone survey ratings. The ratings represent time-series of survey scores collected from subjects while under treatment or from HC. Using a data-driven approach, we aimed to find latent states that may correlate with clinical mood states in these groups. Furthermore, we aimed to find group-level differences in their (latent) state transitions. Our hypothesis was that HRMDD group will be similar to the BD group in that they will show an increase in mood instability over time and will have higher spikes of (hypo) manic symptoms compared to LRMDD and HC groups.

Methods

Medication-free subjects were recruited from an ongoing study for treatment of MDD in young adults in which after a baseline assessment, open-label treatment with antidepressants was given and patients were followed up with periodic assessments and ratings of depression and mania scores. In addition to these procedures, the subjects were consented separately to take part in a smartphone mood monitoring study in which they could record their mood on a daily as well as weekly basis using a proprietary smartphone app developed by the company - Ginger.io. (San Francisco, California)

Inclusion criteria for MDD subjects: Ages 15-30 years and able to give voluntary informed consent; 2) Satisfy criteria for DSM-IV-TR Major Depressive Episode using a Structured Interview; 3) Never met criteria for mania or hypomania; 4) 17-item Hamilton Depression Rating Scale score (HDRS) [11] > 15 and < 25; 5) Young Mania Rating Scale (YMRS) [12] score < 10; 6) Able to be managed as outpatients during the study as ascertained by the following – i. Clinical Global Severity Scale [13] < 5 i.e. moderately ill, ii. No significant suicidal or homicidal ideation or grossly disabled; 7) Have a smartphone on which the ginger.io mood application can be used; 8) willing to record their mood and activity using the Ginger.io application. **Inclusion criteria for BD subjects: 1)** Satisfy criteria for DSM-IV-TR Bipolar Disorder Depression using a Structured Interview; 2) Rest of the criteria same as that for MDD subjects; **Inclusion criteria for Healthy Control (HC) subjects:** 1) no personal or family history in first degree relative of psychiatric illness; 2) no significant neurological disorder; 3) no history of alcohol of substance dependence in the past 12 months; 3) on no psychotropic medication. **Ascertainment of HRMDD vs LRMDD** is described in a previous publication. [14]

Exclusion criteria for all: 1) meeting DSM-IV criteria for schizophrenia, schizoaffective disorder, or an anxiety disorder as a primary diagnosis; 2) use of psychotropics in the past 2 weeks; use of fluoxetine in the past 5 weeks; 3) acutely suicidal or homicidal or requiring inpatient treatment; 4) meeting DSM-IV criteria for substance dependence within the past year, except caffeine or nicotine; 5) positive urinary toxicology screening at baseline; 6) use of alcohol in the past 1 week; serious medical or neurological illness; 7) current pregnancy or breast feeding.

Smartphone application: The Ginger.io Behavior Platform as implemented at the time of the study was used by researcher groups, clinicians, patients and providers alike to collect data. In this paper, we are reporting results from active data collection using surveys and other self-reported data gathered through user input. This was achieved in two ways: (i) the user receives a notification that a survey is available and upon clicking the notification, they are transferred to the survey page, (ii) the user can then manually launch the application and they are shown all the surveys available at that time. The data described above is encrypted and transmitted over a secure 128-bit SSL 3.0 connection using the HTTPS protocol for transmissions between mobile application, web application and secure servers.

Smartphone surveys: *Daily ratings* were for the following items on a 1 – 5 visual analog scale with 1 signifying "Not at all" and 5 "Signifying Most of the day" on the other. The daily ratings for (hypomania) ratings were questions for the last 24 hours for the following items - *1) feeling much more excited or full of energy than usual; 2) feeling so good or so hyper that other people thought you were not your normal self or you were so hyper that you got into trouble; 3) feeling so irritable or grouchy that you either started arguments, shouted at people or hit people; 4) sleeping far less than usual and still not feel tired or sleepy. Weekly ratings* were also done to assess bipolar symptoms for the past 1 week on the following items and a visual analogue scale from 1 – 7 with 1 signifying "Not at all" on one end and 7 signifying "all the time" on the other. These surveys were done on the following items – *1) did you feel happier or more cheerful than usual; 2) did you feel more self-confident than usual; 3) did you sleep far less than usual and still not feel tired or happy; 4) did you talk more than usual; 5) were you more active either socially, sexually, at work,*

home or school than usual. The mania items were summed up to get one score for daily and weekly ratings each respectively.
Treatment: after the baseline ratings MDD patients were started on antidepressant treatment usually with selective serotonin reuptake inhibitor antidepressant, most frequently fluoxetine, unless patient wanted another antidepressant. Patients were given "real world" open label treatment and dosage was adjusted depending on tolerance or response. Augmentation with another antidepressant such as bupropion or change to another antidepressant such as a dual uptake inhibitor e.g. duloxetine was also done depending on tolerance and treatment response. Bipolar subjects were followed up in the regular outpatient clinic and treated with antidepressants and mood stabilizers as indicated.

Statistical Methods and Analyses

Two types of analyses were done for measurement of mood instability: 1) *A priori* defined criteria; and 2) *Latent-class* analyses of weekly rating score time series using a data-driven approach.

***A priori* criteria for mood instability:** The aim for these analyses was to measure daily and weekly sub-threshold (hypo)mania symptoms scores in a clinical setting – 1) a priori defined criteria for spikes in scores (>2 threshold and >25% increase from baseline) and fluctuations (change in two consecutive ratings of more than 25%). We also measured differences in subject groups using statistical methods (ANOVA).

Latent-class Analyses: using Hidden Markov Model (HMM) [15], mainly to test the hypotheses that the number of latent states in the *weekly mood ratings* correspond to the number of mood states inferred (using overall clinical ratings and assessment) by the clinician. We also hypothesized that the transitions of a subject in underlying latent states (as measured by percent times a subject switches states over the course of the study) will vary in expectation with clinical classification of mood states. Accordingly, HRMDD subjects would have the highest switch rate (as the BD subjects in the study were likely being treated with mood stabilizers).

Modeling: The aims of modeling are to a) discover the number of latent states that may represent clinical mood status in weekly mood ratings dataset; and b) gain insights as to the characteristics of these states in relationship to the smartphone surveys and subjects' sub-groups or class label of HRMDD, LRMDD, BD, HC. HC were studied to study the differences from normal for each of the patient group. In this regard, please note that the modeling did not target any particular variable from the smartphone survey or subject-level data.

For the first aim, we constructed several HMMs using the python library (https://pypi.org/project/hmms) and the weekly mood ratings data. The 5 distinct weekly rating questions (measured on a scale of 1 to 7) were summed to get a total score of 1 to 35 per subject for each weekly assessment. Then for learning a model and evaluating its performance, the dataset was randomly split into training and test (held-out) sets of weekly mood rating sequences (70-30% split). To address the need for learning a model from a larger sample or all available data, we use 10-fold cross validation approach, i.e. we used 10 random splits (samples of training and test sets) and performed independent experimental setups (as detailed below). This is in accordance with accepted machine learning practice of evaluating a model to avoid overfitting. [16]

Our primary interest was to find an explainable set of latent states that may correspond with clinical mood status (i.e. euthymia, depression, and range of mania symptoms). Thus, we limit our exploration to the 5 latent states (varied from 2 to 6). It's possible that a greater number of latent states may increase the model fit but may also reduce explain-ability. Thus, we sought to compare 5 separate latent state models in the 10 experimental setups. We selected the number of latent states based on a model that maximizes the likelihood of the data (i.e. held-out set). This model can then be explored further for latent state characteristics and decoding of weekly mood ratings. Below we give details of the experimental setup, performance metric and model selection and model-based decoding of the weekly mood ratings data.

Experimental setup: In each experiment, we parameterize a model to learn a fixed (N) number of latent states (using the training set), among other model parameters (i.e. initial state, emission probabilities and transition matrix) for an HMM using the Expectation-Maximization (Baum-Welch) algorithm [17] implemented in the python library. To measure model performance, we compute the log-likelihood (probability) of data (i.e. total score of weekly mood rating sequence) for every subject in the held-out set under the model (parameters). We also start each experiment from random initialization (seed) to mitigate the issue of local maxima and repeat the above procedures for every latent state we wished to explore, i.e. we varied N from 2 to 6 latent states in all the 10 experimental setups. Overall,

we performed 50 experiments and computed log-likelihood of held-out set in each.

Performance metric and Model selection: Individual held-out subject probabilities were computed in each experiment, and log summed up for obtaining the log-likelihood (LL) of the (held-out) test set, i.e. *predictive LL of the model*. This score was used for model comparison of all the latent-states explored. In general, the higher the *predictive LL* score, the better the model fit to the number of latent states and the observational data (i.e. weekly mood rating sequences). In our setup, we selected a model based on the *median value* of the *predictive LL* for all the latent state explored in the 10 random experimental setups. The model that best fit the observational data (D) with the number of latent states (N) is the one with maximum median LL, i.e. $argmax(P(D, N \mid \Theta))$, where Θ are the learned model parameters. Here we refer to this model as the selected model.

Model based decoding: The selected model was then used to decode an individual subject's weekly mood rating sequence, i.e. the selected model assigned a latent state to every subject's weekly mood rating. Here we use the state assignment to draw observational insights and its association to other subject-level characteristics, such as their (known) class label or sub-group. Please note that these analyses are post-hoc and sub-group label was not used for modeling in anyway. In particular, we examine percent (latent) state change by these sub-groups, i.e. number of times the latent state change occurs over the observational period which is defined as the number of subject-level measures. We also examine clinician interpretation of the modeling states and any age or gender differences in these states.

All data were collected in accordance with IRB approval. Cleveland Clinic Foundation IRB approved this study.

Results

A total of 2401 records from 40 subjects (6 BD, 13 HR or HRMDD, 16 LR or LRMDD and 5 HC) were available for daily mood ratings. A total of 744 records were available from 38 patients (6 BD, 13 HRMDD, 14 LRMDD and 5 HC) for the weekly ratings for 2 to 44 weeks. Please see Table 1 for subject characteristics.

Table 1: Study Subject Characteristics

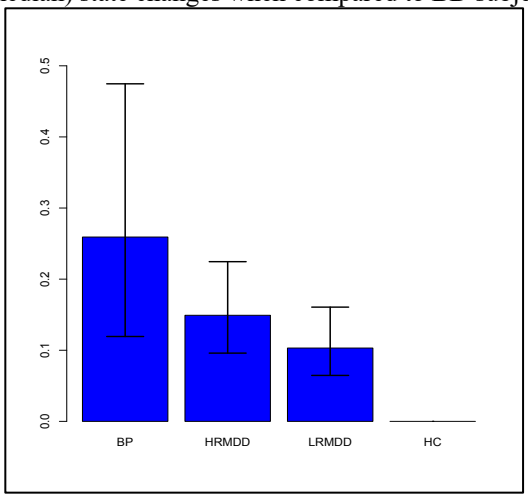
Characteristics	BD (N=6)	HR (N=13)	LR (N=16)	HC (N=5)	p-value
Age (mean (sd))	24.3(4.1)	22.9(3.6)	25.6(3.3)	23.6 (4.6)	ns
Gender (Female (n) %)	4(67)	12(92)	9(56)	1(20)	0.026
Race (Caucasian (n) %)	4(67)	7(54)	13(81)	3(60)	0.005
Years of illness (mean (sd))	11(5.4)	9.2(6)	10.9(5.2)		ns
Age at first episode (mean (sd))	14(3.6)	13.7(4.5)	14.7(4.5)		ns
Medication free period in weeks (mean (sd))	76.3(84.5)	66.7(73.4)	88.9(137)		ns
Number of Depressive Episodes (mean (sd))	14.2(16.4)	11.7(12.3)	41.1(45.6)		0.05
Duration of medication in days (mean sd))		475.2(270)	520.8(230)		ns

A priori defined percent mood fluctuations: For percent mood fluctuations between two consecutive ratings, daily ratings showed differences between groups (BD: 13%, HRMDD, 5%, LRMDD: 6%, HC: 0%) and this difference was significantly different only between BD and HRMDD ($p = 0.02$). For percent mood fluctuations in weekly ratings, (Figure 1), differences were seen between groups (BD: 25%, HRMDD: 15%, LRMDD: 10% and HC: 0%) and for patient groups this difference was significant between BD and LRMDD ($p = 0.05$).

A priori defined percent mood spikes: For percent of manic spikes, significant differences were seen between groups for daily ratings (BD: 35%, HRMDD: 6%, LRMDD: 11%, HC: 0%). This difference was significant between BD and HRMDD ($p=0.001$) and BD and LRMDD ($p=0.05$) but not different between HRMDD and LRMDD. For weekly ratings, differences were also seen between different patient groups (BD: 49%, HRMDD: 19%, LRMDD: 27%, HC: 2%). For daily ratings this difference was significant only between BD and HRMDD ($p = 0.001$)

Latent-class analyses of weekly mood ratings: The *predictive LL* was largest for the model with 4 latent states, followed by 5, 2, 3 and 6 latent state models (median values: -3.77, -4.28, -4.46, -4.49, -4.80 respectively). Thus, using a data-driven approach, a 4-state model was selected for further interrogation. A 4-state model also lends itself well to the clinician's interpretation of mood status as euthymic, depressed, and range of mania symptoms. Therefore, further results here are derived from the best model among the 4-state models (in terms of predictive LL). Below we describe latent state changes as observed in the data.

Percent latent states change: Please refer to Figure 2 for the following text. As expected, healthy controls showed no changes (as reflected by latent states in our model), and LRMDD and HRMDD subjects show lower number of (median) state changes when compared to BD subjects.



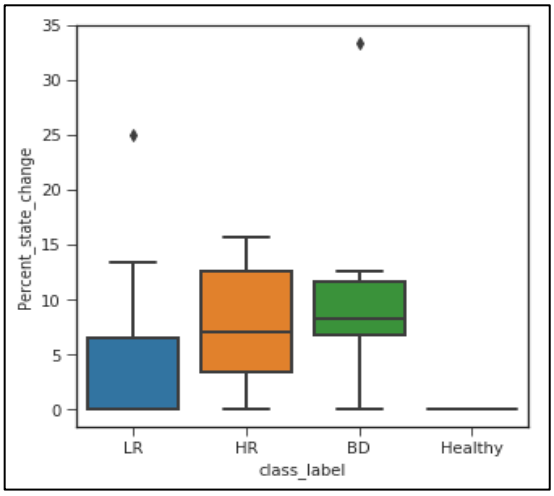


Figure 1: A priori criteria percent mood fluctuations **Figure 2: Percent latent states change**

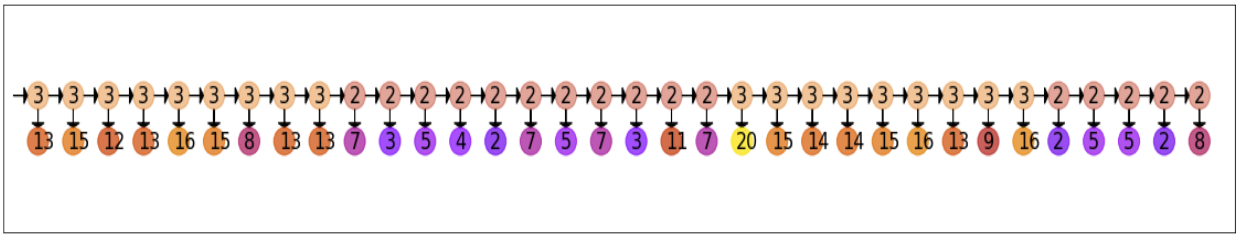
Note: BP refers to BD sub-group (Left), LR is LRMDD sub-group, HR is HRMDD sub-group (Right)

We further observed the following latent state characteristics as in Table 2. It is likely State 0 corresponds to euthymic mood status as in a HC (i.e. negligible mania symptoms), State 1 corresponds to minimal mania symptoms status, State 2 is few mania symptoms mood status and State 3 is mild mania symptoms mood status. This classification can be made on the basis of higher mania symptom scores as seen in states 2 and 3 (Table 2).

Table 2: Latent State characteristics

Variable (mean, sd)	State 0	State 1	State 2	State 3
Total weekly score (1-35)	1.20 (±4.70)	2.10 (±2.11)	5.31 (±2.84)	13.33 (±3.92)
Feel happier	0.29 (±1.14)	0.47 (±0.75)	1.40 (±0.99)	2.75 (±1.27)
Feel self-confident	0.28 (±1.11)	0.41 (±0.70)	1.27 (±1.05)	2.84 (±1.24)
Sleep less than usual	0.16 (±0.76)	0.40 (±0.79)	0.67 (±1.18)	2.13 (±1.46)
Talk more than usual	0.21 (±0.98)	0.31 (±0.63)	0.77 (±0.96)	2.74 (±1.26)
Active more than usual	0.26 (±1.08)	0.4 (±0.69)	1.22 (±1.15)	2.88 (±1.35)

Furthermore, its clinically known that subjects in the BD sub-group continue to have mild spikes even when on medications. [18] This is also depicted as higher latent state changes in the BD sub-group in Figure 2. We depict an example decoding of one BD subject's weekly mood rating sequence (into latent states) in Figure 3 and table below.



Week #	Total score	Latent State
1	13	3
2	15	3
3	12	3
4	13	3
5	16	3
6	15	3
7	8	3
8	13	3
9	13	3

Week #	Total score	Latent State
10	7	2 (*)
11	3	2
12	5	2
13	4	2
14	2	2
15	7	2
16	5	2
17	7	2
18	3	2

Week #	Total score	Latent State
19	11	2
20	7	2
21	20	3
22	15	3 (*)
23	14	3
24	14	3
25	15	3
26	16	3
27	13	3

Week #	Total score	Latent State
28	9	3
29	16	3
30	2	2 (*)
31	5	2
32	5	2
33	2	2
34	8	2

Figure 3: Latent states (top of figure) and total (emission) score* (bottom of figure) for a period of 34 weeks. *Total Scores are sum of 7 items (1-5 rating scale) on weekly mood rating scale. Table below figure shows total scores with latent states alternating between mild mania symptoms at week #1 to few mania symptoms at week #10 to mild mania symptoms again at week #22 and back to few mania symptoms at week #30 according to the model.

A typical BD sub-group subject switches states over the course of time (3 times or 9% for the subject in Figure 3 – marked with * in Table below). We found that overall the mood status (as measured by latent states change) fluctuated up to a third (i.e. 33% of times) in all subjects studied. There was no state fluctutaion in HC subjects, and on average 5% in LRMDD, 7% in HRMDD and 11% in BD subject sub-groups. This trend in latent state fluctuations was similar to that calculated directly using our 'a priori criteria' for mood score fluctuations, as mentioned above.

Discussion

The results of this study indicate the feasibility of assessing daily and weekly mood instability in subjects, particularly undergoing antidepressant treatment, using a smartphone. A few other studies have also reported good feasibility of capturing mania symptoms using daily ratings. [8] However, in this study, we also conducted weekly mood ratings. In general, daily mood and weekly mood ratings changes as measured with *a priori* criteria were in similar direction indicating that weekly mood ratings may suffice. Weekly ratings may decrease the burden on the subject in terms of answering daily surveys. [6]

In terms of differences among patient groups based on daily and weekly changes in mania symptoms, the BD group showed the greatest number of changes, the LRMDD group the least and the HRMDD group showed intermediate fluctuations (Figure 1) based on *a priori* criteria for spikes and fluctuations. Adding to face validity of these measures – the healthy controls showed little or no mania spikes or fluctuations. However, the difference between HRMDD and LRMDD for mood spikes and fluctuations, as defined, was not significant. This could indicate no difference between the two groups in terms of mood instability or that the *a priori* measures as defined were not constructed properly to show the difference. In terms of continuous measures, it is difficult to define mood spikes or fluctuations beyond consecutive ratings as many such measures could be constructed using 3, 4 or more number of ratings grouped at a time.

In this regard, conceptualizing mood and mood disorders as complex and non-linear systems can add to our current understanding of mood regulation [18] and a data-driven latent class model which explores hidden states can be much more objective and informative. The classification of mood ratings by this model also has face-validity as assessed by the clinician author (AA). Using a 4-state model, we found that both HRMDD and BD subjects have a greater number of state changes than LRMDD subjects. The direction of differences between groups (BD>HRMDD>LRMDD>HC) was the same using the 'a priori criteria' of fluctuations in mood symptoms and that for the latent mood states. These results suggest that latent class analyses of time series of mood states available from smartphone ratings provides a better analytical framework to identify mood instability.

Limitations of the study were the small number of subjects that were studied though a large number of data points were obtained. Future studies are needed with larger number of subjects particularly within each diagnostic subgroup. Comparison groups also need to be closely matched in terms of age, gender and ethnicity. Another limitation was that subjects started their smartphone ratings at different durations of treatment therefore future prospective studies are needed with all subjects at the same point of their treatment preferably at the start of the study. This study involved open label treatment with a variety of medications along with a naturalistic follow-up. Randomized controlled trials, possibly with placebo will be needed to further confirm the bipolar like mood instability in HRMDD patients while on antidepressant treatment. Smartphone ratings also need to be correlated with clinical ratings that are the gold standard in terms of mood ratings. Correlational analysis between smartphone ratings and clinical ratings along with clinical assessment of severity of illness need to be done to further validate the use of mobile ratings.

Conclusion

In summary, the results of the study indicate feasibility of doing smartphone ratings at a daily and weekly basis in mood disorder subjects. Weekly ratings may be adequate to decrease the rating burden on the subjects. In terms of clinical findings, bipolar disorder patients seem to have the most mood instability compared to MDD subjects. Within the MDD group, while on antidepressant treatment, subjects thought to be at high risk of developing BD (HRMDD) exhibited higher mood instability than those at low-risk (LRMDD). As the HRMDD subjects are at a risk of developing full-blown manic symptoms, weekly smartphone ratings may be used to monitor their mood to predict such events.

Acknowledgements

Part of this study was supported by NIMH grant to AA (R01MH093420).

Address for correspondence

Amit Anand, MD
Center for Behavioral Health, Cleveland Clinic
9500 Euclid Avenue P-57 Mailbox
Cleveland, OH 44122
Email: ananda@ccf.org
Fax: 216-636-2995
Phone: 216-636-2840

References

[1] American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders., 5th edition, Arlington, VA, 2013.

[2] K.R. Merikangas, J. He, M. Burstein, S.A. Swanson, S. Avenevoli, L. Cui, C. Benjet, K. Georgiades, and J. Swendsen, Lifetime Prevalence of Mental Disorders in US Adolescents: Results from the National Comorbidity Study-Adolescent Supplement (NCS-A), *J. Am. Acad. Child Adolesc. Psychiatry*. **49** (2010) 980–989. doi:10.1016/j.jaac.2010.05.017.

[3] J.I. Nurnberger, M. McInnis, W. Reich, E. Kastelic, H.C. Wilcox, A. Glowinski, P. Mitchell, C. Fisher, M. Erpe, E.S. Gershon, W. Berrettini, G. Laite, R. Schweitzer, K. Rhoadarmer, V.V. Coleman, X. Cai, F. Azzouz, H. Liu, M. Kamali, C. Brucksch, and P.O. Monahan, A high-risk study of bipolar disorder. Childhood clinical phenotypes as precursors of major mood disorders, *Arch. Gen. Psychiatry*. **68** (2011) 1012–1020. doi:10.1001/archgenpsychiatry.2011.126.

[4] R. Nusslock, and E. Frank, Subthreshold bipolarity: diagnostic issues and challenges, *Bipolar Disord.* **13** (2011) 587–603. doi:10.1111/j.1399-5618.2011.00957.x.

[5] M.M. Sidor, and G.M. Macqueen, Antidepressants for the acute treatment of bipolar depression: a systematic review and meta-analysis, *J. Clin. Psychiatry*. **72** (2011) 156–167. doi:10.4088/JCP.09r05385gre.

[6] E. Dogan, C. Sander, X. Wagner, U. Hegerl, and E. Kohls, Smartphone-Based Monitoring of Objective and Subjective Data in Affective Disorders: Where Are We and Where Are We Going? Systematic Review, *J. Med. Internet Res.* **19** (2017) e262. doi:10.2196/jmir.7006.

[7] M. Faurholt-Jepsen, M. Frost, C. Ritz, E.M. Christensen, A.S. Jacoby, R.L. Mikkelsen, U. Knorr, J.E. Bardram, M. Vinberg, and L.V. Kessing, Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial, *Psychol. Med.* **45** (2015) 2691–2704. doi:10.1017/S0033291715000410.

[8] M. Faurholt-Jepsen, C. Ritz, M. Frost, R.L. Mikkelsen, E. Margrethe Christensen, J. Bardram, M. Vinberg, and L.V. Kessing, Mood instability in bipolar disorder type I versus type II-continuous daily electronic self-monitoring of illness activity using smartphones, *J. Affect. Disord.* **186** (2015) 342–349. doi:10.1016/j.jad.2015.06.026.

[9] H. Javelot, A. Spadazzi, L. Weiner, S. Garcia, C. Gentili, M. Kosel, and G. Bertschy, Telemonitoring with respect to mood disorders and information and communication technologies: overview and presentation of the PSYCHE project, *BioMed Res. Int.* **2014** (2014) 104658. doi:10.1155/2014/104658.

[10] J. Torous, and A.C. Powell, Current research and trends in the use of smartphone applications for mood disorders, *Internet Interv.* **2** (2015) 169–173. doi:10.1016/j.invent.2015.03.002.

[11] W. Guy, National Institute of Mental Health (U.S.), Psychopharmacology Research Branch, and Early Clinical Drug Evaluation Program, ECDEU assessment manual for psychopharmacology, U.S. Dept. of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs, Rockville, Md., 1976.

[12] R.C. Young, J.T. Biggs, V.E. Ziegler, and D.A. Meyer, A rating scale for mania: reliability, validity and sensitivity, *Br. J. Psychiatry J. Ment. Sci.* **133** (1978) 429–435.

[13] M.K. Spearing, R.M. Post, G.S. Leverich, D. Brandt, and W. Nolen, Modification of the Clinical Global Impressions (CGI) Scale for use in bipolar illness (BP): the CGI-BP, *Psychiatry Res.* **73** (1997) 159–171.

[14] P. Koirala, B. Hu, M. Altinay, M. Li, A.L. DiVita, K.A. Bryant, H.S. Karne, J.G. Fiedorowicz, and A. Anand, Sub-threshold bipolar disorder in medication-free young subjects with major depression: Clinical characteristics and antidepressant treatment response, *J. Psychiatr. Res.* **110** (2019) 1–8. doi:10.1016/j.jpsychires.2018.12.006.

[15] L. Rabiner, and B. Juang, An introduction to hidden Markov models, *IEEE ASSP Mag.* **3** (1986) 4–16. doi:10.1109/MASSP.1986.1165342.

[16] G.C. Cawley, and N.L.C. Talbot, On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation, *J Mach Learn Res*. **11** (2010) 2079–2107.

[17] I. Baum, Y. Petrie, G. Soules, and N. Weiss, A maximisation technique occurring in the statistical analysis of probabilistic functions of Markov chains., *Ann Math Stat.* (1970) 164–171.

[18] A. Ortiz, and M. Alda, The perils of being too stable: mood regulation in bipolar disorder, *J. Psychiatry Neurosci. JPN*. **43** (2018) 363–365. doi:10.1503/jpn.180183.