Published in final edited form as:

J Dual Diagn. 2012 January 1; 8(4): 326–332. doi:10.1080/15504263.2012.723318.

A Pilot Study of Mood Ratings Captured by Mobile Phone Versus Paper-and-Pencil Mood Charts in Bipolar Disorder

Colin A. Depp, PhD^{1,2}, Daniel H. Kim, BA¹, Laura Vergel de Dios, BS¹, Vicki Wang, BS¹, and Jennifer Ceglowski, MS¹

¹Department of Psychiatry, University of California, San Diego, San Diego, California, USA

²San Diego Veterans Healthcare Administration, San Diego, California, USA

Abstract

Objective—Patient reported mood charts are frequently used in management of bipolar disorder. Although mood charts have recently been programmed in electronic devices such as mobile phones, little is known about the impact of the method of data capture on the psychometric properties and validity of these data.

Methods—In an ongoing pilot study, a sample of outpatients with bipolar disorder were randomized to either complete mood charts on a mobile phone or a standard paper-and-pencil mood chart as part of a 12 week-intervention (primary outcomes for the trial await study completion). We compared these conditions across single item rating of mood state, and we hypothesized that mobile phone based data capture would produce greater compliance to mood ratings, variability between and within participants, and concurrent validity with blinded clinician-rated affective symptom severity.

Results—A total of 56 participants were randomized and 40 participants were included in the analyses. There were no significant differences between conditions on demographic or clinical variables. The rate of compliance was significantly higher in paper-and-pencil versus mobile phone ratings. Ratings demonstrated significantly more variability within individuals in the mobile phone condition. Mobile phone mood ratings were significantly correlated with clinician-rated depressive symptom severity across the study and with manic symptom severity at the Week 6 assessment, whereas paper-and-pencil ratings were not significantly associated with clinician-rated depression or mania.

Conclusions—Although preliminary, our results suggest a lower rate of compliance with mobile phones compared to paper-and-pencil daily mood rating in bipolar disorder, yet a greater ability to capture variability and concurrent validity in quantifying affective symptoms. This clinical trial is registered at <http://www.clinicaltrials.gov> as NCT01670123.

Keywords

ecological momentary assessment; experience sampling; bipolar disorder depression; mania; health technology; psychometrics

There is a great deal of interest in the ways in which mobile devices such as cellular phones can be used in the context of mental health assessment and intervention (Burns et al., 2011; Ehrenreich, Righter, Rocke, Dixon, & Himelhoch, 2011; Heron & Smyth, 2009). For

Corresponding Author: Colin A. Depp, PhD, Department of Psychiatry, UCSD School of Medicine, 9500 Gilman Drive (0664), La Jolla, CA 92093-0664, USA. cdepp@ucsd.edu.

Disclosures: The authors report no financial relationships with commercial interests.

illnesses such as bipolar disorder that are heterogeneous and have a variable course over time, frequent data collection by mobile devices represents a potentially powerful tool for self-monitoring outside of the clinic setting, so as to facilitate measurement-based personalized care. Many such mobile health approaches are based on the framework of ecological momentary assessment (EMA; Shiffman, Stone, & Hufford, 2008), in which patients are asked to complete momentary ratings of their immediate experience repeatedly over time and within the day.

Little is known in bipolar disorder, however, about how data obtained from mobile telephones compare to that gathered by traditional retrospective paper-pencil self-monitoring tools (e.g., “mood charts”). Given that mood charts are frequently used in clinical management of bipolar disorder, it would be useful to understand how the data obtained through electronically captured momentary ratings of mood differ, particularly in regard to patient compliance with these procedures as well as psychometric properties and concurrent validity with clinician ratings. A number of studies outside of bipolar disorder have found generally equivalent compliance between paper-and-pencil and electronic data capture methods (Shiffman, 2006). However, electronic data capture methods gathering momentary data may increase participant burden, particularly among patients with serious mental illness who may be more unfamiliar with technology than general population comparison samples. In addition to rates of compliance, the manner in which participants respond to devices versus paper-and-pencil diaries may produce differences in the validity of the data obtained. In an elegant study, Stone et al. used hidden photoelectric sensors to show that paper-and-pencil diaries are commonly associated with “backfilling” - respondents complete batches of daily ratings at a single time, and thus introduce potential for additional retrospective biases (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2003).

Moreover, understanding the psychometric properties of electronic momentary self-ratings of mood is of great importance, particularly if these ratings are to be leveraged for momentary interventions. In bipolar disorder, self-ratings obtained from electronic data generally capture mood state (e.g., depressed, manic) and ancillary symptoms such as sleep quality/quantity (Bauer et al., 2004). Although there have been examples of terminal based computerized mood monitoring in bipolar disorder (Bauer, et al., 2004), only a handful of studies have employed ecological momentary assessment by mobile devices in bipolar disorder (Bopp et al., 2010) and have reported a comparison with self-rated retrospective or clinician reports. In a small two-week study, the correlation between depressive symptoms obtained in clinical ratings and electronic momentary -based depression ratings on a three-times daily single-item rating of depressed mood was high (Depp et al., 2010). In schizophrenia, Ben Zeev et al. found that momentary ratings of psychotic symptoms were largely similar to that in a weekly retrospective summary, yet with some evidence that retrospective reports were associated with systematic overestimation of the intensity of symptoms (Ben-Zeev, McHugo, Xie, Dobbins, & Young, 2012). Differences among momentary and clinician-based reports may reflect the influence of cognitive biases that arise in retrospective summarization of past events. A sizable literature suggests that retrospective self-ratings are heavily influenced by “peak” moments, such as more intense feeling states, or “recency” effects, with greater salience of moments that occur closest in time to the assessment (Schooler & Hertwig, 2005).

To better understand the impact of electronic data capture in self-monitoring of bipolar disorder, we used preliminary data from a randomized controlled trial in which adults with bipolar disorder were assigned to either complete mood and sleep ratings on a mobile phone or with a paper-and-pencil mood chart for up to 12 weeks. We contrasted the mean rates of compliance between phone and paper-and-pencil data capture. We also compared the mean values, within-subject variability over time, and correlation with clinician-rated measures of

depression and mania between the mobile phone and paper-and-pencil data. Based on prior literature, we did not predict a significant difference in rates of compliance between conditions (although our study was not powered to detect non-inferiority). We hypothesized that mood ratings obtained on mobile phone would be associated with greater within- and between- person variability and would be more highly correlated with clinician ratings of depression and mania than paper-and-pencil data.

Methods

Sample Characteristics

Study data were derived from a sample of 40 outpatients with bipolar disorder who were participants in an ongoing randomized controlled trial of mobile phone-enhanced psychoeducation for bipolar disorder. Participants were outpatients with either bipolar disorder I or II recruited from various sources including flyers and advertisements placed in clinics and online, referrals from other studies enrolling people with bipolar disorder, and community presentations at mental health agencies such as the Depression and Bipolar Support Alliance. To be eligible, participants needed to: 1) be age 18 and older, 2) meet diagnostic criteria for bipolar disorder as established by the MINI International Neuropsychiatric Interview (Sheehan, Lecrubier, Sheehan, & Amorim, 1998), 3) be receiving stable outpatient psychiatric treatment, and 4) have visual acuity and manual dexterity sufficient to operate a touch screen device. Participants were excluded if they: 1) met criteria for any substance use disorder in the prior 3 months, 2) were psychiatrically hospitalized in the prior month, or 3) scored in the severe range for either depressive symptoms (Montgomery Asberg Depression Rating Scale score over 32) or manic symptoms (Young Mania Rating Scale score over 20). There was a complete discussion of the study with potential participants and all participants provided written informed consent. This study was approved and monitored by the UCSD Human Subjects Protections Program.

Study Protocol

At baseline, participants were randomized with equal probability to a mobile phone arm ($n=18$) or a paper-and-pencil chart condition ($n=22$). Each participant was trained in the electronic/paper mood charts in their first session (an in-person meeting with project staff). Participants were informed that their compliance was essential to the success of this study and that the investigators would be able to monitor compliance remotely (phone condition) or would collect the completed paper-and-mood charts at the end of the study. Therefore, compliance was encouraged; however lapses in compliance did not result in withdrawal from the study.

In the mobile phone condition, participants were provided with an internet-enabled “smartphone” (Samsung Fascinate), which was programmed to send twice-daily requests to complete a mobile web-enabled survey of current momentary mood and related experiences. Persons randomized to the phone condition received invitations to complete assessments at a random time within two 3-4-hour blocks in the mornings and evenings. At the outset of the study, participants could select the earliest and latest time they would wish to be alerted, so as not to interfere with their typical sleep/wake cycle. Once prompted to respond, participants had 15 minutes to complete the survey, after which they received a reminder prompt if no response was provided. The survey “expired” and could not be completed after two hours. Partial responses were logged in that participants did not need to complete all of the questions for the data to be captured. At the outset of the study, participants were told to fill out the assessments as soon as they were received.

On the other hand, in the paper-and-pencil condition, participants were provided a binder with all of the mood charts for the subsequent twelve weeks. Participants were told to complete the paper-and-pencil mood charts every day, although they were not told to complete it at any particular time of day. In addition, participants were advised that these charts would be collected at the end of the study.

The paper-and-pencil mood charts and mobile assessments contained several identical questions, including the scales, about overall mood (see below) and eight different affect ratings (e.g., sad, energetic). Since these affect items were added to the paper-and-pencil mood charts after the study start, comparative analyses could not be performed, and the focus was on the single overall mood question. The phone condition additionally asked three questions about location, primary activity, and social context. The layout of the paper-and-pencil mood chart was adapted from the NIMH mood chart (Denicoff et al., 2000), and was completed once per day. These paper-and-pencil mood charts were collected at the end of 12 weeks (or earlier if participants opted to withdraw from the study). Thus, for the present study the participants did not complete both paper-and-pencil and electronic mood charts.

Both conditions were asked to complete these ratings for 12 weeks in conjunction with participation in a four-session psychoeducational intervention adapted from the Life Goals program (Bauer, McBride, Chase, Sachs, & Shea, 1998), which focused on education about the illness, identification of early warning signs, and active coping strategies for manic and depressive symptoms. The therapist and content of these sessions were identical between the two conditions. Participants in both conditions were asked to engage in self-monitoring of mood symptoms on a daily basis for a total of 12 weeks. Both groups of participants were assessed with standard clinical ratings (described below) at baseline, 6 weeks, and 12 weeks (a 24 week rating was also obtained but is not reported here as no data were recorded between 12 and 24 weeks). For the present study, we included all participants who completed at least one post-baseline assessment. A total of four participants in both the paper-and-pencil and phone conditions did not complete the 12 week assessment, but all participants completed the baseline and 6-week assessments.

Measures

Demographics and diagnosis—All participants were assessed at baseline with the Mini International Neuropsychiatric Interview (MINI)(Sheehan, et al., 1998) to establish diagnosis of bipolar disorder. Final diagnosis was made by combining information from the MINI and chart reviews from treating provider records, and were confirmed in consensus meetings. At baseline, participants provided information on basic demographics as well as diagnosis and treatment history, and current participation in treatment including medications.

Standard clinical ratings—Participants were assessed at baseline, 6 weeks, and 12 weeks post-baseline using the Montgomery Asberg Depression Rating Scale (MADRS) (Montgomery & Asberg, 1979) and the Young Manic Rating Scale (YMRS)(Young, Biggs, Ziegler, & Meyer, 1978). However, the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)(Randolph,1998) was only completed at baseline. The MADRS is a 10-item clinician-rated scale for depression that is widely used in assessing the severity of bipolar depression. The YMRS is an 11-item clinician-rated scale that is the most commonly used scale for quantifying the severity of mania. The RBANS is a popularly used brief neurocognitive battery with four alternate forms to detect and characterize cognitive decline. Clinical ratings were conducted by one psychometrist, who was blinded to group assignment. A separate project coordinator collected all of the mood charts and managed data obtained from the phones to ensure blinding. The MADRS and YMRS were

interviewer-administered, and as part of our research protocols, raters were trained to reliability to gold standard on these instruments by more senior raters. Consistent with the instructions for the MADRS and YMRS, these instruments cover the preceding week.

Mood rating—Participants rated their current mood state on a 9-point bipolar anchored scale. A value of one represented “most ever depressed”, a 2: “severely depressed”, a 3: “moderately depressed”, and 4: “mildly depressed”. A value of 5 represented “euthymic or even mood”. Values for mania were transposed from depression with 9 being “most ever manic”. The scaling was identical for phone and paper-and-pencil conditions. Written descriptors of each of the anchors were provided, adapted from the NIMH Life Chart Method (LCM) mood chart (Denicoff, et al., 2000). Note that the NIMH LCM is a 7-point scale with three levels of severity for mania and depression, centered around euthymia. We added “Most Ever” depressed or manic ratings so that participants could signify a crisis on the device that branched to a link, which enabled connection to the San Diego County Crisis Line.

Statistical Analyses

All values were assessed for normality for parametric statistical analyses. Demographic and clinical characteristics were compared with Analyses of Variance (ANOVA) for continuous variables and Pearson Chi-square for categorical variables. Characteristics on which the two groups differed were employed as covariates in subsequent analyses. To assess the rate of compliance across conditions, we calculated for each participant the proportion of surveys answered versus the total number to be completed over the period in which the participant was asked to complete forms (generally 12 weeks, in case of attrition at 6 weeks for eight participants). We also examined the effect of time in study on compliance, using the generalized estimating equations procedure, with a Poisson Link Function for binary data. In this analysis, the predictor variable was days on study and the outcome was with-subject adherence (yes or no) across the study period. In regard to psychometric properties, we contrasted within-person variability in two ways, owing to the fact that variability can be observed both in the amount of fluctuation over time in symptoms as well as the range of symptom severity expressed. To address fluctuation, we assessed within-patients variability in mood ratings by calculating the within-person standard deviation for each participant. We then used the generalized linear models procedure to compare within-person standard deviation between conditions, while controlling for mean severity level (mood rating). In addition, we calculated the Inter-Quartile range for the mood rating. Finally, to assess the external validity of mood ratings obtained by mobile phone or paper-and-pencil charts, we correlated mean mood ratings for each participant with mean total scores on the MADRS and YMRS across the study period, and repeated this analysis restricting to the 6 week assessment and mean data across weeks 0 to 6 on the mood charts. All analyses were set to alpha level of 0.05.

Results

Comparison of Sample Characteristics Across Conditions

Sample characteristics are displayed in Table 1. There were no significant differences in the composition of the mobile phone and paper-and-pencil conditions in regard to demographics, clinical history, or clinical ratings. On average, the sample was middle-aged, college educated, and had experienced symptoms of bipolar disorder for approximately 20 years. In terms of clinical ratings, the average level of severity of depression was in the mild range based on a cut-off of 10 on the MADRS (Tohen et al., 2009), manic symptoms were sub-syndromal to mild based on a cut-off of 7 on the YMRS (Tohen, et al., 2009), and participants were free of cognitive impairment (less than one standard deviation on the

RBANS). At baseline, all participants were taking a mood stabilizer (e.g., lithium, divalproex), anti-psychotic, or anti-depressant. A total of 70% ($n=28$) reported taking mood stabilizers, 48% of subjects ($n=19$) were prescribed atypical antipsychotics (only one subject was taking a typical antipsychotic), and 58% ($n=23$) were taking antidepressants.

Comparison of Compliance to Mood Ratings across Conditions

A total of 3 participants in the paper-and-pencil condition did not return forms, leaving 19 participants to be analyzed in that condition along with 18 participants in the phone condition. The mean number of observations per person was 51.2 ($SD=27.1$) in the paper-and-pencil condition and 72.3 ($SD=61.5$) in the mobile phone condition. The rate of compliance was significantly and substantially higher in the paper-and-pencil condition than in the phone condition; $t(35)=5.8$, $p<0.001$. The mean rate of compliance (number of surveys completed/number possible within the time frame) in the mobile phone condition was 42.1% ($SD=26.6\%$; range 4.8% to 93.0%). In contrast, compliance in the paper-and-pencil condition averaged 82.9% ($SD=14.1\%$; range 48%-100%). Factoring in the three participants who did not return paper-and-pencil mood charts, the rate of compliance would be 71.6%, still substantially greater than in the mobile phone condition. We examined time trends in compliance using generalized estimating equations with a Poisson link function. We found a significant effect of time on study in the phone condition, such that additional days on study were associated with lower adherence (estimate=-0.002, $SE=0.0002$, $p<0.001$). There was no significant effect of days on study with adherence in the paper-and-pencil condition (estimate=0.00007, $SE=0.0006$, $p=0.296$).

Comparison of Within-Person and Between-Person Variability Across Conditions

To estimate within-person variability in mood ratings, we calculated the Within-person Standard Deviation for each participant and compared these values across condition, controlling for level using the GLM procedure. There was a significant group effect, $F(1,32)=4.8$ $p=0.036$, indicating larger Within-person Standard Deviation values in the phone condition compared to the paper-and-pencil condition. Indicators of between-person variability were also greater in the phone condition, for the Inter-Quartile range on the mood rating scale covered two scale anchors on the 1 to 9 scale in the phone condition as opposed to the width of one scale anchor in the paper-pencil condition, thus indicating greater variation between participants in assigning ratings to the scale. Additionally, the standard deviation of mood rating Within-person Standard Deviation values of the phone condition (0.41) was nearly double that in the paper-and-pencil condition (0.22), signifying greater variability between participants.

Comparison of the Correlation of Mood Rating with Standard Clinical Ratings

To assess the concurrent validity of phone and paper-and-pencil obtained ratings, we conducted Pearson correlations between averaged mood ratings and mean within-person values of MADRS and YMRS scores averaged across assessments. In the phone condition, mood ratings correlated significantly with MADRS scores ($r=-0.567$, $p=0.014$), but not with manic symptoms ($r=0.294$, $p=0.236$). In contrast, neither MADRS nor YMRS scores correlated with paper-and-pencil ratings of mood ($r=-.243$, $p=0.346$ and $r=0.452$, $p=0.069$, respectively). We conducted additional correlational analyses with mood ratings and clinician ratings on the MADRS and YMRS values obtained at the Week 6 assessment only with corresponding data from the six weeks prior to assessment in each condition. In the phone condition, mood ratings were significantly correlated with both MADRS total score ($r=-0.542$, $p=0.028$) and YMRS total score ($r=0.520$, $p=0.032$). In contrast, the data obtained from the paper-and-pencil condition were not significantly correlated with either MADRS ($r=-0.094$, $p=0.701$) or YMRS ($r=0.396$, $p=0.093$).

Discussion

Our randomized study on the rates of compliance, psychometric properties, and concurrent validity of mobile phone versus paper-and-pencil mood ratings among outpatients with bipolar disorder suggested several strengths and weaknesses of mood monitoring via mobile technology. The rate of compliance to mobile phones was approximately half of that seen among participants who completed paper-and-pencil ratings. Rates of compliance appeared to decline over the course of the study in the phone condition, but not in the paper-and-pencil condition. On the other hand, as hypothesized, mood ratings captured by mobile phone evidenced more variability within- and between- patients. Moreover, mood ratings on the mobile phone were significantly associated with clinician rated depression averaged over the study and manic symptom severity at six weeks, whereas paper-and-pencil ratings were not (although these associations were not statistically different). Thus, our small study suggests a tradeoff for researchers selecting between mobile phone-based and paper-and-pencil based ratings of mood in bipolar disorder, with perhaps greater variability and validity obtained via mood ratings captured through mobile devices, yet diminished compliance.

There are a number of limitations to this study. Although the randomized study groups did not differ along any of the demographic or clinical characteristics studied, an ideal design for comparison of data collection methods would include a cross-over in which groups complete both paper-and-pencil and mobile phone based mood ratings with the same frequency of ratings per day. In addition, we cannot rule out the influence of the brief psychoeducational intervention (which was identical between conditions) on the expression of symptoms over the course of the study. Moreover, the data is derived from a small sample, limiting power to detect differences among conditions or to examine potential moderators (e.g., diagnostic subtype). Our sample was comprised of outpatients who were active participants in psychiatric care and who were experiencing, on average, a low level of severity of depression and mania. Thus, these results should be interpreted as preliminary, awaiting further study in a broader sample of those with bipolar disorder, and a study specifically designed for the purpose of disentangling the effect on compliance and validity of data capture method (i.e., phone/paper) versus sampling frame and frequency (i.e., momentary/daily global rating).

Despite these limitations, several potentially important questions arise from the differences observed between these data capture approaches. For one, why was the rate of compliance higher in the paper-and-pencil condition? Our finding of lower rate of compliance in the mobile phone condition compared to previous electronic momentary assessment studies in bipolar disorder that occurred over one to two weeks provides for some caution that longer-term electronic momentary assessment (e.g. 12 weeks) may need additional participant support and motivation. Indeed, there was some evidence of “fading” in adherence in the phone condition that was not present in the paper-and-pencil condition. One explanation for the lower rate of compliance is that because the mobile phone involves momentary prompts to provide mood ratings that expire after a brief period, and subsequently the opportunity to miss prompts to respond is much greater than associated with paper-and-pencil mood charts that can be completed *ad libitum*.

However, if the “timing out” of responding to surveys was the sole reason for differences in compliance, one would expect the psychometric properties and degree of within-variability and concurrent validity to be comparable between conditions. We found significantly greater variability in the mobile phone condition as well as significant concurrent validity with clinician ratings that were not seen in the paper-and-pencil condition. One explanation is that participants in the paper-and-pencil condition likely engaged in “back filling”, i.e., filled

in mood ratings for multiple days retrospectively and rated themselves for past dates based on their present frame of reference. Thus, paper-and-pencil mood charts that allow for backfilling may be more akin to retrospective reports than momentary ratings obtained via mobile phone. It is notable that mood ratings on the mobile device indicated significantly more severe depressive symptoms than seen in the paper-and-pencil condition despite a lack of differences in clinician rated depression. This differs from prior studies that observed greater severity in retrospective reports in patients who had unipolar depression (Ben-Zeev, Young, & Madsen, 2009), although recalled levels of fatigue were higher than momentary ratings in a sample of patients who were chronically fatigued (Friedberg & Sohl, 2008).

Although preliminary, our findings suggest that future development of mobile phone applications involving longer term monitoring of mood and related symptoms in bipolar disorder should identify and adapt to patient- and device-level barriers to compliance. In addition, it would be useful to examine whether these results extend to other symptom clusters, such as energy/activation, impulsivity, or lack of need for sleep in regard to concurrent validity with clinician-rated instruments (Bauer et al., 1991). Our results question the external validity of paper-and-pencil daily mood charts, which are frequently utilized on clinical management of bipolar disorder. Momentary ratings may be less subject to retrospective biases, and at least in our study, may better correspond to clinician rated symptoms. Future longitudinal research should examine variability between and within people across states in regard to accuracy in reporting symptoms in various states of illness. In order for the potential of mobile health to enhance the quality of care of people with bipolar disorder, it is essential to ensure that patients perceive utility in self-monitoring, and to maximize reliability and validity of the self-reported data upon which interventions are delivered.

Acknowledgments

This study was supported by National Institute of Mental Health Grants MH091260, MH077225, and MH080002, a UCSD Von Leibig Center Southern California Healthcare Technology Acceleration Award, and by the Department of Veterans Affairs.

References

- Bauer M, Crits-Christoph P, Ball W, Dewees E, McAllister T, Alahi P, Whybrow P. Independent assessment of manic and depressive symptoms by self-rating Scale characteristics and implications for the study of mania. *Archives of General Psychiatry*. 1991; 48(9):807–812.10.1001/archpsyc.1991.01810330031005 [PubMed: 1929771]
- Bauer M, Grof P, Gyulai L, Rasgon N, Glenn T, Whybrow PC. Using technology to improve longitudinal studies: Self-reporting with ChronoRecord in bipolar disorder. *Bipolar Disorder*. 2004; 6(1):67–74.10.1046/j.1399-5618.2003.00085.x
- Bauer MS, McBride L, Chase C, Sachs G, Shea N. Manual-based group psychotherapy for bipolar disorder: A feasibility study. *Journal of Clinical Psychiatry*. 1998; 59(9):449–455. [PubMed: 9771814]
- Ben-Zeev D, McHugo GJ, Xie H, Dobbins K, Young MA. Comparing retrospective reports to real-time/real-place mobile assessments in individuals with schizophrenia and a nonclinical comparison group. *Schizophrenia Bulletin*. 2012; 38(3):396–404.10.1093/schbul/sbr171 [PubMed: 22302902]
- Ben-Zeev D, Young MA, Madsen JW. Retrospective recall of affect in clinically depressed individuals and controls. *Cognition & Emotion*. 2009; 23(5):1021–1040.10.1080/02699930802607937
- Bopp JM, Miklowitz DJ, Goodwin GM, Stevens W, Rendell JM, Geddes JR. The longitudinal course of bipolar disorder as revealed through weekly text messaging: A feasibility study. *Bipolar Disorders*. 2010; 12(3):327–334.10.1111/j.1399-5618.2010.00807.x [PubMed: 20565440]

- Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, Mohr DC. Harnessing context sensing to develop a mobile intervention for depression. *Journal of Medical Internet Research*. 2011; 13(3):e55.10.2196/jmir.1838 [PubMed: 21840837]
- Denicoff KD, Leverich GS, Nolen WA, Rush AJ, McElroy SL, Keck PE, Post RM. Validation of the prospective NIMH-Life-Chart Method (NIMH-LCM-p) for longitudinal assessment of bipolar illness. *Psychological Medicine*. 2000; 30(6):1391–1397.10.1017/S0033291799002810 [PubMed: 11097079]
- Depp C, Mausbach B, Granholm E, Cardenas V, Ben-Zeev D, Patterson TL, Jeste D. Mobile interventions for severe mental illnesses: Design and preliminary data from three approaches. *Journal of Nervous and Mental Disease*. 2010; 198(10):715–721.10.1097/NMD.0b013e3181f49ea3 [PubMed: 20921861]
- Ehrenreich B, Richter B, Rocke DA, Dixon L, Himelhoch S. Are mobile phones and handheld computers being used to enhance delivery of psychiatric treatment? A systematic review. *Journal of Nervous and Mental Disease*. 2011; 199(11):886–891.10.1097/NMD.0b013e3182349e90 [PubMed: 22048142]
- Friedberg F, Sohl SJ. Memory for fatigue in chronic fatigue syndrome: The relation between weekly recall and momentary ratings. *International Journal of Behavioral Medicine*. 2008; 15(1):29–33.10.1007/BF03003071 [PubMed: 18444018]
- Heron KE, Smyth JM. Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*. 2009; 15(1):1–39.10.1348/135910709X466063 [PubMed: 19646331]
- Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*. 1979; 134:382–389.10.1192/bjp.134.4.382 [PubMed: 444788]
- Randolph, C. Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). San Antonio, Harcourt, TX: The Psychological Corporation; 1998.
- Schooler LJ, Hertwig R. How forgetting aids heuristic inference. *Psychological Review*. 2005; 112(3):610–628.10.1037/0033-295X.112.3.610 [PubMed: 16060753]
- Sheehan D, Lecrubier Y, Sheehan P, Amorim J. The Mini-International Neuropsychiatric Interview: The development and validation of a structured diagnostic psychiatric interview for the DSM-IV and ICD-10. *Journal of Clinical Psychiatry*. 1998; 59(20):22–33. [PubMed: 9881538]
- Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annual Review of Clinical Psychology*. 2008; 4:1–32.10.1146/annurev.clinpsy.3.022806.091415
- Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. *Controlled Clinical Trials*. 2003; 24(2):182–199.10.1016/j.bbr.2011.03.031 [PubMed: 12689739]
- Tohen M, Frank E, Bowden CL, Colom F, Ghaemi SN, Yatham LN, Berk M. The International Society for Bipolar Disorders (ISBD) Task Force report on the nomenclature of course and outcome in bipolar disorders. *Bipolar Disorders*. 2009; 11(5):453–473.10.1111/j.1399-5618.2009.00726.x [PubMed: 19624385]
- Young RC, Biggs J, Ziegler V, Meyer D. A rating scale for mania: Reliability, validity, and sensitivity. *British Journal of Psychiatry*. 1978; 133:429–435.10.1192/bjp.133.5.429 [PubMed: 728692]

Table 1
Sample Characteristics

Variable	Phone Condition (n=18) M(SD) or %	Paper-and-pencil Condition (n=22) M(SD) or %	Test Statistic	p-Value
Age	44.0 (14.0)	46.1 (13.5)	$F(1,38)=0.2$	.641
Gender (% Female)	55.6%	63.6%	$\chi^2=0.3$	.604
Ethnicity				
Caucasian	66.7%	77.3%	$\chi^2=8.3$	.080
African American	5.6%	18.2%		
Hispanic	22.2%	0.0%		
Asian	0.0%	4.5%		
Bi/Multi Racial	5.6%	0.0%		
Educational Attainment (Years)	14.3 (2.3)	15.0 (2.1)	$F(1,38)=0.9$	.337
Diagnosed with Bipolar I (%)	90.9%	89.9%		
Age of Onset of Depression (years)	20.8 (11.3)	19.2 (8.7)	$F(1,38)=0.3$	.615
Age of Onset of Mania/Hypomania (years)	19.2 (11.8)	23.2 (10.6)	$F(1,38)=1.3$	.263
RBANS Score	85.0 (16.9)	92.9 (11.7)	$F(1,37)=3.0$	.092
Mean MADRS Score ^I	11.3 (7.2)	9.3 (4.9)	$F(1,36)=0.9$	.333
Mean YMRS Score ^I	8.5 (6.8)	6.1 (3.6)	$F(1,38)=1.6$	.211

Note. RBANS: Repeatable Battery for the Assessment of Neuropsychological Status; MADRS:

Montgomery Asberg Depression Rating Scale; YMRS: Young Mania Rating Scale

^IThese values are averaged across the three assessments at 0, 6, and 12 weeks.