Original Article

Smartphone data as an electronic biomarker of illness activity in bipolar disorder

Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as an electronic biomarker of illness activity in bipolar disorder. Bipolar Disord 2015:00:000–000. © 2015 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd.

Objectives: Objective methods are lacking for continuous monitoring of illness activity in bipolar disorder. Smartphones offer unique opportunities for continuous monitoring and automatic collection of real-time data. The objectives of the paper were to test the hypotheses that (i) daily electronic self-monitored data and (ii) automatically generated objective data collected using smartphones correlate with clinical ratings of depressive and manic symptoms in patients with bipolar disorder.

Methods: Software for smartphones (the MONARCA I system) that collects automatically generated objective data and self-monitored data on illness activity in patients with bipolar disorder was developed by the authors. A total of 61 patients aged 18–60 years and with a diagnosis of bipolar disorder according to ICD-10 used the MONARCA I system for six months. Depressive and manic symptoms were assessed monthly using the Hamilton Depression Rating Scale 17-item (HDRS-17) and the Young Mania Rating Scale (YMRS), respectively. Data are representative of over 400 clinical ratings. Analyses were computed using linear mixed-effect regression models allowing for both between individual variation and within individual variation over time.

Results: Analyses showed significant positive correlations between the duration of incoming and outgoing calls/day and scores on the HDRS-17, and significant positive correlations between the number and duration of incoming calls/day and scores on the YMRS; the number of and duration of outgoing calls/day and scores on the YMRS; and the number of outgoing text messages/day and scores on the YMRS. Analyses showed significant negative correlations between self-monitored data (i.e., mood and activity) and scores on the HDRS-17, and significant positive correlations between self-monitored data (i.e., mood and activity) and scores on the YMRS. Finally, the automatically generated objective data were able to discriminate between affective states.

Conclusions: Automatically generated objective data and self-monitored data collected using smartphones correlate with clinically rated depressive and manic symptoms and differ between affective states in patients with bipolar disorder. Smartphone apps represent an easy and objective way to monitor illness activity with real-time data in bipolar disorder and may serve as an electronic biomarker of illness activity.

Over the last decade in bipolar disorder research there has been an increasing shift in illness paradigm from a focus on affective episodes to an increasing focus on inter-episodic mood instability (1, 2). Many patients with bipolar disorder continue to experience subsyndromal mood swings on
a daily basis, with euthymic patients suffering more from mood instability than healthy subjects do (2–4). Mood instability at a subclinical level is associated with impaired global functioning and a high risk of relapse (1, 5). Continuous real-time monitoring and assessment of mood instability, as well as other items that may reflect or correlate with illness activity, would be useful because it could allow for and facilitate early intervention on subsyndromal symptoms (6, 7).

Regular cell phones using text messages (7), computers (8–10), personal digital assistants (PDAs) (11, 12), and smartphones (13) have been suggested as easy and inexpensive tools to continuously and electronically self-monitor subjective items related to illness activity in long term by patients with bipolar disorder.

Regarding objective measures of illness activity, studies have suggested that decreased activity in speech may be a sensitive and valid measure of the prodromal symptoms of depression and effect of treatment (14–16); and conversely, studies have suggested that increased activity in speech predicts switch to hypomania (17). Similarly, changes in the level of engagement in social activities (18) and the amount of physical activity/the level of psychomotor retardation (19–21) represent central aspects of illness activity in bipolar disorder that could be evaluated objectively. Longitudinal assessment of these behavioral changes could provide objective markers of changes in illness activity in bipolar disorder. These objective measures of illness activity could be collected and quantified automatically and longitudinally in the long term with a low level of intrusiveness using smartphones, since patients already carry them for most of the day and use them for communication and other traceable activities. In this way, data from smartphones can reflect and provide information on the patients’ habitual behavior.

Smartphones offer unique opportunities for continuous monitoring of depressive and manic symptoms through the collection of both real-time self-monitored data on illness activity and automatically generated objective data on behavior such as speech activities, social activities and physical activities that can be collected automatically by smartphones (22, 23). In 2010, we started to develop a smartphone-based self-monitoring system [the MONitoring, treAtment and pRediCtion of bipolAr disorder episodes system (the MONARCA I system)] designed for patients with bipolar disorder. We hypothesized that self-monitoring of items reflecting illness activity in real time would allow for a closer and continuous monitoring of illness activity in the long term with the possibility of early intervention. Further, we hypothesized that collection of automatically generated objective data from the smartphone would reflect the level of and changes in speech activities, social activities, and physical activities across affective states and levels of illness activity.

In addition to being a smartphone-based self-monitoring tool, this system included a two-level feedback loop between patient and mental health care providers allowing for early intervention on prodromal symptoms based on the information from the collected smartphone data (details on the MONARCA I system are described elsewhere) (22, 23). A first pilot study on the MONARCA I system, by our group, showed that patients found the system acceptable to use, with a higher adherence to continuous daily self-monitoring than for self-monitoring with a paper-based version (22, 23). In a second pilot study on a new group of patients using the MONARCA system, a significant correlation between the level of self-monitored depressive symptoms and scores on the Hamilton Depression Rating Scale–17 item (HDRS-17) (24) was found (25). However, no correlation between the level of self-monitored manic symptoms and scores on the Young Mania Rating Scale (YMRS) (26) was found (25). Possible reasons for the latter negative finding could be that the sample size in the pilot study was too small or that the patients in the pilot study presented with low levels of manic symptoms. Alternatively, patients with (hypo-) manic symptoms may simply not be able to validly monitor symptoms due to decreased illness insight (27). To investigate whether smartphone data could serve as a biomarker of illness activity in patients with bipolar disorder, the above-mentioned subsequent pilot study also examined the correlation between automatically generated objective data and the level of clinically rated depressive and manic symptoms. Data showed a significant correlation between mobility (the number of changes in cell tower ID/day) and the HDRS-17, suggesting that the patients stayed at home more when they had higher HDRS-17 scores. Thus, the pilot studies gave an indication of a potential usefulness of the smartphone software as a usable self-monitoring tool and an automatic electronic biomarker of illness activity in patients with bipolar disorder that should be investigated further.

The effect of the MONARCA I system as an intervention was investigated in a randomized controlled trial (RCT), and data in the present paper were collected as part of the RCT investigating the effect of daily electronic self-monitoring using smartphones on depressive and manic symptoms in patients with bipolar disorder (28, 29).
results concerning the effects of the MONARCA I system investigated in an RCT are reported elsewhere (29), and the current study is a secondary analysis of data from the MONARCA trial.

The objectives of the present study were based on findings from the literature suggesting that measures of speech activities, social activities and physical activities may reflect the level of illness activity in patients with bipolar disorder (14–21), the results from analyses in the MONARCA pilot studies (25), and the technical possibilities available at the time of the design of the MONARCA trial. Thus, we aimed to investigate the hypotheses that (i) daily electronic self-monitored data collected using smartphones correlate with the level of clinically rated depressive and manic symptoms in patients with bipolar disorder, and (ii) automatically generated objective data on speech activities, social activities and physical activities collected using smartphones correlate with the level of clinically rated depressive and manic symptoms in patients with bipolar disorder and thus represent a type of electronic biomarker.

Methods

Study design, participants, and settings

Patients with bipolar disorder were recruited from a specialized mood disorder clinic, The Copenhagen Clinic for Affective Disorder, Psychiatric Center Copenhagen, Copenhagen, Denmark, from September 2011 to March 2013 as part of an RCT investigating the effect of daily electronic self-monitoring using smartphones on depressive and manic symptoms (28, 29). Treatment at the clinic includes two years of combined evidence-based psychopharmacological treatment and supporting therapy. Further details about the treatment program at the clinic and the effect of this are described elsewhere (30).

The patients were invited to participate in the study following referral to the clinic. The inclusion criteria were a bipolar disorder diagnosis according to the International Classification of Diseases 10th version (ICD-10) using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (31), age between 18 and 60 years, a HDRS-17 score ≤ 17 (24) and a YMRS score ≤ 17 (26) at the time of inclusion. The exclusion criteria were pregnancy, a lack of Danish language skills, an inability to learn the technicalities necessary for using a smartphone, unwillingness to use an Android smartphone as their primary cell phone, severe physical illness and schizophrenia, or any other ICD-10 F2 diagnosis according to the SCAN interview.

The reason for having a cut-off score on the HDRS-17 and the YMRS of ≤ 17 at the time of inclusion was that we aimed to include patients who were able to manage the technical aspects of learning how to use a smartphone and hypothesized that including patients with severe depressive and manic symptoms at inclusion would complicate this process.

The exclusion criterion ‘an inability to learn the technicalities necessary for using a smartphone’ was chosen since, at the time of the conceptualization of the study design in 2010, there had not been published much scientific literature on technical IT barriers and skills in patients with bipolar disorder, and thus perhaps patients who had never owned a regular cell phone or never used a computer would have to spend too much time simply learning how to interact with the smartphone for normal communicative purposes (e.g., make phone calls and sending text messages), and would not be able to use the smartphone for self-monitoring. The technical knowledge of the patients was assessed by the study nurse and the IT technician.

The smartphone used in this study was HTC Desire S (HTC Corporation, New Taipei City, Taiwan), which runs on the Android operating system, and all included patients were offered basic demonstrations on how to interact with the smartphone according to their needs. The patients used their own SIM card during the study. Economic costs due to data traffic from using the MONARCA I system were refunded to the patients by the study. Patients did not receive other economic compensation for participating in the study.

Collection of smartphone data

The intervention group. Because data in this study were collected as part of an RCT, all of the included patients were randomized 1:1 to either an intervention group or a control group. Patients in the intervention group were provided with a smartphone with the MONARCA I system installed and were instructed by the study nurse orally, were given a hands-on demonstration and received a detailed written manual, with pictures, on how to use the system for daily self-monitoring during a six-month study period. The patients in the intervention group were on a daily basis prompted to provide the self-monitored data upon hearing an alarm at a self-chosen time during the day. If the patients forgot to provide the self-monitoring data, it was possible to do so retrospectively for up to two days. The following self-monitored items were evaluated on a daily basis: mood (scored from depressive to manic on a scale from
The intervention group and the control group. In addition to the daily electronic self-monitored data collected in the intervention group, the following automatically generated objective data were collected from all the patients included in the study regardless of randomization group (meaning both the patients in the intervention group and the patients in the control group) throughout the day and night on a daily basis during the six-month study period, and did not require the patients to actively interact with the smartphone in any way: the number of incoming calls/day; the duration of incoming calls/day (sec/day); the number of incoming text messages/day; the number of outgoing calls/day; the duration of outgoing calls/day (sec/day); and the number of outgoing text messages/day. Thus, patients in the control group were also loaned a smartphone to use with their own SIM card for the entire study period and instructed to use the smartphone for their usual communicative purposes. The smartphones loaned to the patients in the control group did not have access to the MONARCA self-monitoring system or the feedback loop. The collected automatically generated objective data were chosen because of predefined hypotheses (stated in the Introduction section) that these data would reflect the level of and changes in psychopathology in bipolar disorder, e.g., changes in speech activity due to changes in depressive and manic symptoms would be reflected by the number and duration of incoming and outgoing phone calls; changes in psychomotor activity due to changes in depressive and manic symptoms would be reflected by the duration of incoming and outgoing phone calls; and changes in social activity due to changes in depressive and manic symptoms would be reflected by changes in the number of text messages.

Furthermore, at the beginning of the study, it was not technically possible to collect further objective measures using the available version of the MONARCA I system. Thus, the automatically generated objective data were the data that could be collected, and this process was approved by the ethical committee at the time of the study.

Outcome assessments

The bipolar disorder diagnosis according to ICD-10 was confirmed before inclusion using SCAN (31) and clinical and sociodemographic inclusion data were collected. Patients were invited to visit the researchers once a month for the six-month study period for clinical assessments of depressive and manic symptoms using both the HDRS-17 and the YMRS, respectively. We chose to assess the level of depressive and manic symptoms using the clinical rating scales HDRS-17 and YMRS since they address the severity of depressive and manic symptoms for the past four days and are two of the most widely used clinical rating scales to assess symptom severity in efficacy studies. All outcome assessments were carried out by two physicians (MF-J or Anne Sophie Jacoby) who were blinded to the patients’ randomization status (unaware of whether the patients were randomized to the intervention group or the control group), did not have access to any of the collected smartphone data (blinded to the smartphone data) and were not involved in the treatment of the patients. Thus, the researchers were blinded to the collected smartphone data and randomization group.

Statistical methods

The results concerning daily electronic self-monitored data are based on 30 patients with bipolar disorder who were randomized to the intervention group of the MONARCA I trial (29), with outcome assessments provided once a month for six months; these results thus represent over 200 clinical ratings. The results concerning automatically generated objective data are based on 61 patients with bipolar disorder who were from both the intervention group and the control group of the MONARCA I trial (23), with outcome assessments provided once a month for six months; these results thus represent over 400 clinical ratings. Because the data were collected as part of an RCT, the sample size and power calculation calculations were based on the estimation of a clinically relevant difference in the predefined primary outcomes (i.e., differences in depressive and manic symptoms between the intervention group and the control group during the six-month study period). Outcomes, sample size and power calculation were predefined in a study protocol that was published before data collection was completed and thus defined the size of the study (28).

The statistical analyses were defined a priori and computed using linear mixed-effect regression models. These regression models allow for both
within individual variation and between individual variations over time of the different dependent variables considered in the regression analyses. For each regression analysis, we employed a two-level linear mixed-effects regression model where the first level of the model represented the repeated measurements of the smartphone data within individuals over time and the second level of the model represented the between individuals variation of the smartphone data over time. All employed models included a random intercept and a patient-specific random effect to account for patient-specific correlations between the dependent variables (the smartphone data) and the independent variables (the level of depressive and manic symptoms) over time. All other covariates were specified as fixed effects in the employed models. Model assumptions were checked visually by means of residuals and quantile-quantile (Q-Q) plots. For all analyses defined a priori, we firstly considered an unadjusted analysis, and secondly an analysis adjusted for age and sex as possible confounding covariates. Averages of the self-monitored data and automatically generated objective data were analyzed for the day of the clinical assessment and the three previous days, because the HDRS-17 and the YMRS ratings address symptoms over the three previous days. In further analyses, affective states were categorized into sub-categories: asymptomatic (HDRS-17 and YMRS ≤7); mild depression/hypomania (HDRS-17 between 7 and 14; YMRS between 7 and 14); and moderate to severe depression/mania (HDRS-17 and YMRS ≥14). In addition, we analyzed associations between the automatically generated objective data and sub-items of the HDRS-17: mood (sub-item 1), psychomotor retardation (sub-item 8), and psychomotor agitation (sub-item 9), and sub-items of the YMRS: mood (sub-item 1), activity (sub-item 2), and speech (sub-item 6).

Further adjustment for the randomization group as a possible confounding variable was investigated in all of the employed models. Data were entered using the data entry program Epidata® (The EpiData Association, Odense, Denmark) and the statistical software program STATA version 12.1 (StataCorp LP, College Station, TX, USA) was used for the statistical analyses. The significance level of the p-values in the statistical analyses was set to 0.05 (two-sided).

Ethical considerations

The study was approved by the Regional Ethics Committee in the Capital Region of Denmark (H-2-2011-056) and the Danish Data Protection Agency (2013-41-1710). Electronic data generated from the smartphones were stored on a secure server at the Capital Region, Denmark (I-suite number RHP-2011-03). After oral and written presentations of the study had been given to all potential participants, written informed consent was obtained. The participants were informed that they could withdraw from the study at any time without this affecting their course of treatment at the clinic. The smartphones were provided to the participants free of charge and any economic costs incurred due to data traffic from the MONARCA I system were refunded by the study.

**Results**

**Background characteristics**

During the period from September 2011 to March 2013, 123 patients with a diagnosis of bipolar disorder who were not pregnant and were aged between 18 and 60 years were assessed for eligibility; of these, 78 (63.4%) patients were included in the study (see Fig. 1). Of the eligible patients, three were excluded because they did not meet the inclusion criteria (persistent HDRS-17 score ≥17), six were excluded because some exclusion criteria were met (two had a lack of Danish language skills, and four were unwilling to use an Android smartphone as their primary cell phone during the study), 32 patients were excluded because they did not want to participate for various reasons (e.g., expected participation to be too time consuming, or did not want to participate in a research project) and four patients were inaccessible by repeated phone calls and letters. We did not exclude any patient from the study due to an inability to learn the technicalities necessary for using a smartphone. Two patients dropped out after three months of follow-up, and 3.7% of the patient visits to the researcher for outcome assessment were missing.

Information on the clinical and sociodemographic characteristics of patients at inclusion is presented in Table 1. Figures 2 and 3 present the levels of depressive and manic symptoms during the six-month study period and are represented by scores on the HDRS-17 and the YMRS, respectively.

**Self-monitored data collected using smartphones and clinically rated depressive and manic symptoms**

Table 2 presents the results of linear mixed effect regression models for the self-monitored data collected using smartphones and the level of clinically
rated depressive and manic symptoms measured using the HDRS-17 and the YMRS, respectively.

The results from the unadjusted models and the models adjusted for age and sex only differed marginally, i.e., the coefficients, 95% confidence intervals (CIs), and p-values were very similar in all tests. Overall, the self-monitored mood and the self-monitored activity level correlated significantly with scores on both the HDRS-17 and the YMRS, whereas the self-monitored sleep length correlated significantly with scores on the YMRS, and the self-monitored stress level correlated significantly with scores on the HDRS-17.

There was a significant negative correlation between the self-monitored mood and scores on the HDRS-17 in both the unadjusted model and the model adjusted for age and sex (unadjusted model $B = -0.055$, 95% CI: $-0.067$ to $-0.042$, $p < 0.001$; adjusted model $B = -0.058$, 95% CI: $-0.071$ to $-0.045$, $p < 0.001$), indicating that for every score that increased 10 points on the HDRS-17 in the adjusted model, the self-monitored mood was 0.58 lower on a mood scale from 0 to 3. The same significant negative correlation between the self-monitored mood and scores on the HDRS-17 was observed when restricting the model to include only scores on the HDRS-17 mood item (HDRS-17 sub-item 1) (adjusted model $B = -0.38$, 95% CI: $-0.46$ to $-0.31$, $p < 0.001$).

There was a significant positive correlation between the self-monitored mood and scores on the YMRS in both the unadjusted model and the model adjusted for age and sex (unadjusted model $B = 0.39$, 95% CI: 0.016–0.062, $p < 0.001$; adjusted model $B = 0.039$, 95% CI: 0.017–0.062, $p < 0.001$). The same significant positive correla-
tion between the self-monitored mood and scores on the YMRS was observed when restricting the model to include only scores on the YMRS mood item (YMRS sub-item 1) (adjusted model $B = 0.38$, 95% CI: 0.24–0.53, $p < 0.001$).

Models regarding self-monitored sleep length, activity level, and stress level. Table 2 presents the results from models regarding self-monitored sleep length, activity level and stress level in the same way as results from models of self-monitored mood. In relation to depressive symptoms, the self-monitored activity level correlated negatively with scores on the HDRS-17, and the self-monitored stress level correlated positively with scores on the HDRS-17, whereas there was no significant correlation in relation to the self-monitored sleep length. In relation to manic symptoms, the self-monitored sleep length correlated negatively with scores on the YMRS and the self-monitored activity level correlated positively with scores on the YMRS, whereas there was no significant correlation in relation to the self-monitored stress level.

Automatically generated objective data collected using smartphones and clinically rated depressive and manic symptoms

Table 3 presents the results from models regarding the automatically generated objective data that were collected using smartphones and the level of clinically rated depressive and manic symptoms using the HDRS-17 and the YMRS, respectively. The results from the unadjusted models and the models adjusted for age and sex only differed marginally, i.e., the coefficients, 95% CIs, and p-values were very similar in all tests. Overall, the duration of incoming calls/day correlated positively and significantly with scores on both the HDRS-17 and
the YMRS, whereas the duration of outgoing calls/day correlated positively and significantly with scores on the YMRS and borderline significantly with scores on the HDRS-17.

There was a significant positive correlation between the duration of incoming calls/day and scores on the HDRS-17 in both the unadjusted model and the model adjusted for age and sex (unadjusted model $B = 19.96$, 95% CI: $4.12–35.80$, $p = 0.014$; adjusted model $B = 17.15$, 95% CI: $1.00–33.30$, $p = 0.037$), indicating that for every score that increased 10 points on the HDRS-17 in the adjusted models there was an increase in the duration of incoming calls/day of 171.5 (10.0; 333.0) sec. Further, there was a significant positive correlation between the duration of incoming calls/day and scores on the YMRS in both the unadjusted model and the model adjusted for age and sex (unadjusted model $B = 28.54$, 95% CI: $5.17–51.90$, $p = 0.017$; adjusted model $B = 30.38$, 95% CI: $7.04–53.71$, $p = 0.011$), indicating that for every score that increased 10 points on the YMRS in the adjusted models there was an increase in the duration of incoming calls/day of 303.8 (70.4; 537.1) sec.

Table 4 presents the results from models regarding automatically generated objective data and sub-components of the level of clinically rated depressive and manic symptoms, as represented by scores on sub-items on the HDRS-17 and the YMRS, respectively. For the HDRS-17, items concerning mood (sub-item 1), psychomotor retardation (sub-item 8) and psychomotor agitation (sub-item 9) were selected, and for the YMRS, items concerning mood (sub-item 1), activity (sub-item 2) and speech (sub-item 6) were selected. These items on the clinical rating scales were selected because they represent central and objectively measurable parts of depression and mania. Scores on the activity item on the YMRS (sub-item 2) correlated positively and significantly with the automatically generated objective data in relation to the number of incoming and outgoing calls/day and the number of outgoing text messages/day. Scores on the psychomotor retardation item on the HDRS-17 (sub-item 8) correlated positively and significantly with the duration of outgoing calls/day in the unadjusted model and borderline significantly in the adjusted model.

Table 5 presents the results from models regarding the automatically generated objective data and affective states by the HDRS-17 and the YMRS categorized into the subcategories of asymptomatic (HDRS-17 and YMRS ≤3), mild depression/hypomania (HDRS-17 and YMRS 7–14) and moderate to severe depression/mania (HDRS-17 and YMRS ≥14). For the HDRS-17, patients with moderate to severe depression showed a significantly higher duration of outgoing calls/day than did asymptomatic patients in both the unadjusted and the adjusted models (unadjusted model $B = 452.17$, 95% CI: $149.56–754.78$, $p = 0.003$; adjusted model $B = 421.57$, 95% CI: $111.55–731.60$, $p = 0.008$).
### Table 3. Correlations between automatically generated objective data collected using smartphones and depressive and manic symptoms measured using the HDRS-17 and YMRS, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted</th>
<th></th>
<th></th>
<th>Adjusted</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>95% CI</td>
<td>p-value</td>
<td>Coefficient</td>
<td>95% CI</td>
<td>p-value</td>
</tr>
<tr>
<td><strong>Incoming calls (no./day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDRS-17</td>
<td>0.022</td>
<td>−0.010 to 0.054</td>
<td>0.18</td>
<td>0.024</td>
<td>−0.009 to 0.057</td>
<td>0.15</td>
</tr>
<tr>
<td>YMRS</td>
<td>0.060</td>
<td>0.016–0.100</td>
<td>0.007</td>
<td>0.062</td>
<td>0.018–0.110</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>Duration of incoming calls (sec/day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDRS-17</td>
<td>19.96</td>
<td>4.12–35.80</td>
<td>0.014</td>
<td>17.15</td>
<td>1.00–33.30</td>
<td>0.037</td>
</tr>
<tr>
<td>YMRS</td>
<td>28.54</td>
<td>5.17–51.90</td>
<td>0.017</td>
<td>30.38</td>
<td>7.04–53.71</td>
<td>0.011</td>
</tr>
<tr>
<td><strong>Incoming text messages (no./day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDRS-17</td>
<td>−0.037</td>
<td>−0.18 to 0.14</td>
<td>0.61</td>
<td>−0.029</td>
<td>−0.17 to 0.11</td>
<td>0.69</td>
</tr>
<tr>
<td>YMRS</td>
<td>0.087</td>
<td>−0.10 to 0.28</td>
<td>0.37</td>
<td>0.10</td>
<td>−0.088 to 0.290</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>Outgoing calls (no./day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDRS-17</td>
<td>0.031</td>
<td>−0.047 to 0.110</td>
<td>0.44</td>
<td>0.030</td>
<td>−0.050 to 0.110</td>
<td>0.46</td>
</tr>
<tr>
<td>YMRS</td>
<td>0.15</td>
<td>0.045–0.250</td>
<td>0.005</td>
<td>0.15</td>
<td>0.043–0.250</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>Duration of outgoing calls (sec/day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDRS-17</td>
<td>28.27</td>
<td>10.15–46.40</td>
<td>0.002</td>
<td>26.33</td>
<td>7.68–44.98</td>
<td>0.006</td>
</tr>
<tr>
<td>YMRS</td>
<td>23.87</td>
<td>−3.08 to 50.83</td>
<td>0.083</td>
<td>24.97</td>
<td>−2.11 to 52.04</td>
<td>0.071</td>
</tr>
<tr>
<td><strong>Outgoing text messages (no./day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDRS-17</td>
<td>0.014</td>
<td>−0.16 to 0.19</td>
<td>0.88</td>
<td>0.022</td>
<td>−0.15 to 0.19</td>
<td>0.80</td>
</tr>
<tr>
<td>YMRS</td>
<td>0.22</td>
<td>−0.006 to 0.450</td>
<td>0.057</td>
<td>0.24</td>
<td>0.019–0.470</td>
<td>0.034</td>
</tr>
</tbody>
</table>

CI = confidence interval; HDRS-17 = Hamilton Depression Rating Scale–17 item; YMRS = Young Mania Rating Scale.

*a* Averages of the smartphone data were analyzed for the current day and three days before ratings with the HDRS-17 and YMRS, as these rating scales address symptoms over the last four days.

*b* Analyses on all study participants; total N = 61.

*c* Adjusted for age and sex.

### Table 4. Correlations between automatically generated objective data collected using smartphones and sub-items on the HDRS-17 and YMRS, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted</th>
<th></th>
<th></th>
<th>Adjusted</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>95% CI</td>
<td>p-value</td>
<td>Coefficient</td>
<td>95% CI</td>
<td>p-value</td>
</tr>
<tr>
<td><strong>Incoming calls (no./day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YMRS sub-item 1 (mood)</td>
<td>0.27</td>
<td>−0.30 to 0.56</td>
<td>0.075</td>
<td>0.27</td>
<td>−0.022 to 0.57</td>
<td>0.070</td>
</tr>
<tr>
<td>YMRS sub-item 2 (activity)</td>
<td>0.30</td>
<td>0.086–0.520</td>
<td>0.006</td>
<td>0.30</td>
<td>0.085–0.52</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>Duration incoming calls (sec/day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDRS-17 sub-item 1 (mood)</td>
<td>32.34</td>
<td>−62.51 to 127.19</td>
<td>0.50</td>
<td>16.64</td>
<td>−79.05 to 112.34</td>
<td>0.73</td>
</tr>
<tr>
<td>HDRS-17 sub-item 8 (psychomotor retardation)</td>
<td>162.69</td>
<td>−23.88 to 349.26</td>
<td>0.087</td>
<td>147.30</td>
<td>−40.39 to 334.98</td>
<td>0.12</td>
</tr>
<tr>
<td>HDRS-17 sub-item 9 (psychomotor agitation)</td>
<td>−14.25</td>
<td>−218.80 to 190.30</td>
<td>0.89</td>
<td>−19.66</td>
<td>−223.52 to 184.19</td>
<td>0.85</td>
</tr>
<tr>
<td>YMRS sub-item 1 (mood)</td>
<td>134.52</td>
<td>−22.90 to 291.94</td>
<td>0.094</td>
<td>145.76</td>
<td>−11.31 to 302.82</td>
<td>0.669</td>
</tr>
<tr>
<td>YMRS sub-item 6 (speech)</td>
<td>85.74</td>
<td>−10.81 to 182.30</td>
<td>0.082</td>
<td>93.41</td>
<td>−2.98 to 189.79</td>
<td>0.058</td>
</tr>
<tr>
<td><strong>Duration outgoing calls (sec/day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDRS-17 sub-item 1 (mood)</td>
<td>60.95</td>
<td>−49.61 to 171.52</td>
<td>0.28</td>
<td>46.97</td>
<td>−65.53 to 159.48</td>
<td>0.41</td>
</tr>
<tr>
<td>HDRS-17 sub-item 8 (psychomotor retardation)</td>
<td>221.50</td>
<td>5.97–437.02</td>
<td>0.044</td>
<td>201.03</td>
<td>−17.68 to 419.73</td>
<td>0.072</td>
</tr>
<tr>
<td>HDRS-17 sub-item 9 (psychomotor agitation)</td>
<td>171.10</td>
<td>−61.01 to 403.21</td>
<td>0.15</td>
<td>170.31</td>
<td>−62.11 to 402.73</td>
<td>0.15</td>
</tr>
<tr>
<td>YMRS sub-item 1 (mood)</td>
<td>112.01</td>
<td>−70.08 to 294.10</td>
<td>0.23</td>
<td>121.14</td>
<td>−61.67 to 303.95</td>
<td>0.19</td>
</tr>
<tr>
<td>YMRS sub-item 6 (speech)</td>
<td>58.81</td>
<td>−51.52 to 169.13</td>
<td>0.30</td>
<td>64.54</td>
<td>−46.40 to 175.49</td>
<td>0.25</td>
</tr>
<tr>
<td><strong>Outgoing calls (no./day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YMRS sub-item 1 (mood)</td>
<td>0.54</td>
<td>−0.17 to 1.24</td>
<td>0.13</td>
<td>0.53</td>
<td>−0.18 to 1.23</td>
<td>0.14</td>
</tr>
<tr>
<td>YMRS sub-item 2 (activity)</td>
<td>0.64</td>
<td>0.13–1.15</td>
<td>0.014</td>
<td>0.63</td>
<td>0.12–1.14</td>
<td>0.016</td>
</tr>
<tr>
<td><strong>Outgoing text messages (no./day)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YMRS sub-item 1 (mood)</td>
<td>0.88</td>
<td>−0.60 to 2.36</td>
<td>0.24</td>
<td>0.92</td>
<td>−0.55 to 2.40</td>
<td>0.22</td>
</tr>
<tr>
<td>YMRS sub-item 2 (activity)</td>
<td>1.49</td>
<td>0.38–2.59</td>
<td>0.008</td>
<td>1.52</td>
<td>0.42–2.62</td>
<td>0.007</td>
</tr>
</tbody>
</table>

CI = confidence interval; HDRS-17 = Hamilton Depression Rating Scale–17 item; YMRS = Young Mania Rating Scale.

*a* Averages of the smartphone data were analyzed for the current day and three days before ratings with the HDRS-17 and YMRS, as these rating scales address symptoms over the last four days.

*b* Analyses on all study participants; total N = 61.

*c* Adjusted for age and sex.
For the YMRS, patients with mania showed a significantly higher number of incoming calls/day than did asymptomatic patients in both the unadjusted and the adjusted models (unadjusted model $B = 0.95$, 95% CI: 0.076–1.82, $p = 0.033$; adjusted model $B = 0.97$, 95% CI: 0.10–1.84, $p = 0.029$).

Additionally, patients with mania showed a borderline significantly higher number of outgoing text messages/day than did asymptomatic patients in both the unadjusted and the adjusted models.

Overall, further adjustment for the randomization group (intervention group or control group) in each of the models presented in Tables 2–5 did not change the estimates.

Discussion
In the present longitudinal study, we investigated correlations between smartphone data and the level of depressive and manic symptoms, respectively, using repeated measurements in outpatients with bipolar disorder. In accordance with our a priori hypotheses, we found that automatically generated objective data on changes in speech, social and physical activities and self-monitored data collected using smartphones correlated with the level of depressive and manic symptoms assessed with the HDRS-17 and the YMRS, respectively.

This is the first study investigating both automatically generated objective data and self-monitored data collected using smartphones in relation to the level of clinically rated depressive and manic symptoms in a larger study of patients with bipolar disorder.

The most intriguing and novel results from the present study were that (i) several of the automatically generated objective data correlated with scores on the HDRS-17 and the YMRS, and (ii) the levels of automatically generated objective data were able to discriminate between affective states (asymptomatic versus moderate to severe depression or mania), suggesting that such automatically generated objective data may be used as electronic biomarkers for the longitudinal evaluation and monitoring of illness activity in patients with bipolar disorder.

Interestingly, in addition to the results for the automatically generated objective data, the present study showed that patients with bipolar disorder were able to validly evaluate a number of symptoms of illness activity on a daily basis using the MONARCA I system, e.g., their daily electronic self-monitored data correlated with their scores on the HDRS-17 and the YMRS.

Notably, the length of self-monitored sleep did not correlate with the level of depressive symptoms. Bipolar depression often presents with more atypical symptoms than does unipolar depression (32). Thus, patients with bipolar depression may suffer from both increased and decreased sleep length. This could explain why there was neither a positive nor a negative correlation between self-monitored sleep length and depressive symptoms. As expected, in the case of mania, patients with manic symptoms reported lower sleep length.
Correlations between self-monitored data on depressive and manic symptoms that were collected electronically and the level of clinically rated depressive and manic symptoms that were measured using the HDRS-17 and the YMRS, respectively, have been reported previously by the authors and others (25, 33, 34). However, data on electronic self-monitoring from other studies were collected using a computer-based web interface and thus required the patients to log on to a separate ‘monitoring device’, and more importantly allows for the collection of automatically generated objective data that may serve as an electronic biomarker, as presented in the present paper. The automatically generated objective data were sampled on a daily basis outside laboratory settings, thus reflecting the patients’ habitual use of the smartphone and level of communication. Furthermore, the patients did not need to actively interact with the smartphone for this automatic and objective part of the data collection, this reducing the risk of missing data; reducing the risk of bias of the automatically generated objective data (e.g., only collecting automatically generated objective data when in a certain affective state); reducing the level of intrusiveness and the number of demands on the patients (e.g., causing disruptions to everyday life and requesting the patients to record self-monitored items); and minimizing the risk of fatigue in adherence to the daily monitoring procedure in the long term. Furthermore, using smartphones, and not separate ‘monitoring devices’, for continuous monitoring lowers the risk of stigmatization due to the risk of simply seeing the monitoring devices in the patient’s hand (11). Also, a system capable of collecting automatically generated objective data reflecting the level of illness activity could potentially mean that patients could choose not to evaluate the self-monitored items, since information on the level of depressive and manic symptoms would still be collected and thus allow for and facilitate early intervention when signs of deterioration emerge.

Accordingly, this type of automatic data collection using smartphones has clear advantages for everyday deployment and use in the treatment and monitoring of patients with bipolar disorder, a chronic and serious disorder with the need for continuous treatment often for many years (35, 36).

The present study investigated correlations firstly between self-monitored data and the level of depressive and manic symptoms and secondly between automatically generated objective smartphone data and the level of depressive and manic symptoms. The study used a design with repeated measurement in each patient during a six-month study period. In contrast to cross-sectional designs, studies with a longitudinal design with repeated measurements allow for estimation of the influence of within individual alterations in the level of smartphone data according to the level of depressive and manic symptoms. Furthermore, the present study is the first to investigate whether automatically generated objective data collected using smartphones correlate with illness activity in patients with bipolar disorder.

Interestingly, the data from this study showed that the duration of phone calls correlated significantly with the level of both depressive and manic symptoms, respectively, and these correlations were found even when adjusting for the randomization group, where part of the intervention in the MONARCA I trial consisted of a feedback loop between patients and clinicians using phone calls and text messages to establish contact with each other. Alterations in the level of psychomotor activity in bipolar disorder have been reported (21, 37–39), and a possible explanation for the correlations between the duration of incoming and outgoing phone calls and the level of depressive and manic symptoms in the present study may be that patients with manic symptoms speak on the phone for a longer time due to pressure of speech, and patients with depressive symptoms may speak for a longer time due to psychomotor retardation, thus taking longer to express themselves. Analyses on the sub-item level of the clinical rating scales confirmed that the activity sub-item on the YMRS correlated significantly with automatically generated objective data. Interestingly, the number of incoming calls was increased when patients had manic symptoms, suggesting that perhaps the social network of the patient may be activated by the patient’s condition and is engaged in higher amounts of communicative activities.

Limitations

A number of limitations to the present study should be mentioned. The present study included a rather small sample of patients defined according to the study protocol of the MONARCA I trial. However, the design of the study as longitudinal
with repeated measurements of both the smartphone data and the level of depressive and manic symptoms over a six-month study period increases the statistical power, and allows estimation of within individual variation in smartphone data over time according to the level of depressive and manic symptoms. Also, despite the small sample size, an important strength of the present study is that this is a unique patient population with a serious psychiatric disorder.

Furthermore, because of the small sample in this study, our non-significant finding could be due to type II errors, and future studies should include more patients.

Although the present study supports the use of automatically generated objective data in addition to self-monitored data to monitor the level of illness activity in bipolar disorder with real-time data collected in naturalistic settings in the long term, future studies should identify which of the automatically generated objective data or combinations of these most validly identify emerging depressive and (hypo)manic episodes. This could provide a smartphone-based electronic ‘biomarker’ that could function as a diagnostic test with a high sensitivity and specificity that is able to discriminate more accurately between affective states in patients with bipolar disorder.

At the time of the present study, the MONARCA I system was only available for smartphones running the Android operating system, thus excluding patients who were not willing or able to use this type of smartphone. Extending the development of the software to include iPhones and Windows phones would be of great interest. The patients included in this study were rather young and could therefore represent a more motivated and technically oriented subpopulation who are able to interact with any smartphone or system offered. However, in recent years, many people have purchased and learned how to interact with smartphones.

Data collection and continuous monitoring of this type require high data security and a high degree of trust between patients and mental health care providers so that the patients do not have the feeling of being ‘watched’. However, none of the patients participating in this study complained that they felt watched in their everyday life.

During the study period, the patients were not directly asked if they did not use the smartphone themselves or if they shared it with others. Patients were instructed to use the smartphone for communication as usual and signed a contract stating that they were lent a smartphone to use with their own SIM card during the study period. Thus, the possibility cannot be excluded that some patients may have lent the smartphone to friends or family members at some point during the study, but this is not believed to be the case to any large extent.

Finally, during the six-month study, patients were invited to a face-to-face meeting for outcome assessments to complete the HDRS-17 and the YMRS once a month, a rather high frequency that could be a potential source of bias influencing the patients’ level of depressive and manic symptoms.

Perspectives

Frequently, patients with bipolar disorder present too late with moderate or full-blown episodes of depression or mania. Patients, relatives and clinicians may, in the future, be able to detect emerging depressive and manic episodes at earlier stages using daily electronic monitoring on smartphones. However, the effect of early intervention on depressive and manic symptoms using smartphone-based electronic monitoring must be further investigated in RCTs.

Investigating seasonal patterns in the level of self-monitored depressive and manic symptoms, other self-monitored items and automatically generated objective data using longer follow-up periods for each patient and computing time-series analysis would be interesting and could possibly provide new insight.

Conclusions

This study indicates that both automatically generated objective data and self-monitored data collected on a daily basis using smartphones correlate with the level of clinically rated depressive and manic symptoms and differ between affective states in patients with bipolar disorder. Thus, measuring the level of activity on a smartphone may be used as an objective way to measure alterations in affective states. Smartphones provide an easy and objective way to monitor illness activity and could serve as an electronic biomarker for depressive and manic symptoms in patients with bipolar disorder.

Acknowledgements

The authors would like to thank the patients for participating in this study; Rie Lambæk Mikkelsen, M.D. and Ulla Knorr, M.D., Ph.D. from the Copenhagen Clinic for Affective Disorder, Psychiatric Center Copenhagen, Rigshospitalet, Denmark for helping with the recruitment of patients; and Anne Sophie Jacoby, M.D. from The Copenhagen Clinic for Affective Disorder, Psychiatric Center Copenhagen, Rigshospitalet, Denmark for contributing to parts of the data collection.
Author contributions

MF-J, MV, and LVK conceived the study and authored the protocol. MF and JEB together with LVK, MV, EMC, and MF-J and a number of patients with bipolar disorder designed the MONARCA I system. MF and JEB handled all technical matters. MF-J and Anne Sophie Jacoby undertook the clinical assessments of participants. MF-J and LVK performed the statistical analyses and MF-J wrote the first draft of the manuscript. All authors contributed to and have approved the final version of the manuscript.

Disclosures

The MONARCA project was funded by the EU 7th Frame Program, the Mental Health Services in Copenhagen, Tryg-fonden Denmark, the Gert Einar Joergersen foundation, and the A.P. Moeller and Hustru Chastine Mc-Kinney Moeller foundation for general purposes. The funders had no role in the study design, data collection, or writing of the manuscript. MF-J has been a consultant for Eli Lilly & Co. MV has been a consultant for Eli Lilly & Co., Lundbeck, AstraZeneca, and Servier. EMC has been a consultant for Eli Lilly & Co., AstraZeneca, Servier, Bristol-Myers Squibb, Lundbeck, and Medilink. LVK has, within the past three years, been a consultant for Lundbeck and AstraZeneca. MF and JEB report no financial or other relationships relevant to the subject of this article.

References

Faurholt-Jepsen et al.


31. Wing JK, Babor T, Brugha T et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 1990; 47: 589–593.


