ORIGINAL RESEARCH

# Daily self-reported and automatically generated smartphone-based sleep measurements in patients with newly diagnosed bipolar disorder, unaffected first-degree relatives and healthy control individuals

Sharleny Stanislaus [1,2] Maj Vinberg,[1] Sigurd Melbye,[1,2] Mads Frost,[3] Jonas Busk,[4] Jakob Eyvind Bardram,[4] Maria Faurholt-Jepsen,[1] Lars Vedel Kessing[1,2]

[1]Copenhagen Affective Disorder Research Center (CADIC), Psychiatric Center Copenhagen, Rigshospitalet, Copenhagen, Denmark
[2]Faculty of Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[3]Monsenso Aps, Copenhagen, Denmark
[4]Copenhagen Center for Health Technology (CACHET), Department of Health Technology, Technical University of Denmark, Lyngby, Denmark

**Correspondence to**
Dr Sharleny Stanislaus, Rigshospitalet, Psychiatric Center Copenhagen, Copenhagen 2100, Denmark; sharleny.stanislaus.01@regionh.dk

## ABSTRACT

**Objectives** (1) To investigate daily smartphone-based self-reported and automatically generated sleep measurements, respectively, against validated rating scales; (2) to investigate if daily smartphone-based self-reported sleep measurements reflected automatically generated sleep measurements and (3) to investigate the differences in smartphone-based sleep measurements between patients with bipolar disorder (BD), unaffected first-degree relatives (UR) and healthy control individuals (HC).

**Methods** We included 203 patients with BD, 54 UR and 109 HC in this study. To investigate whether smartphone-based sleep calculated from self-reported bedtime, wake-up time and screen on/off time reflected validated rating scales, we used the Pittsburgh Sleep Quality Index (PSQI) and sleep items on the Hamilton Depression Rating Scale 17-item (HAMD-17) and the Young Mania Rating Scale (YMRS).

**Findings** (1) Self-reported smartphone-based sleep was associated with the PSQI and sleep items of the HAMD and the YMRS. (2) Automatically generated smartphone-based sleep measurements were associated with daily self-reports of hours slept between 12:00 midnight and 06:00. (3) According to smartphone-based sleep, patients with BD slept less between 12:00 midnight and 06:00, with more interruption and daily variability compared with HC. However, differences in automatically generated smartphone-based sleep were not statistically significant.

**Conclusion** Smartphone-based data may represent measurements of sleep patterns that discriminate between patients with BD and HC and potentially between UR and HC.

**Clinical implication** Detecting sleep disturbances and daily variability in sleep duration using smartphones may be helpful for both patients and clinicians for monitoring illness activity.

**Trial registration number** clinicaltrials.gov (NCT02888262).

## INTRODUCTION

In addition to alterations in mood and activity, changes in the sleep–wake patterns are a central symptom of bipolar disorder (BD).[1] Sustaining a stable sleep–wake balance is essential for obtaining and maintaining mood stability in patients with BD.[2] Apart from alterations in sleep during depression, (hypo)mania and mixed episodes, emerging evidence suggests that sleep disturbances are also present in between mood episodes[3–5] and associated with impaired quality of life, global functioning[6] and increased risk of relapse.[7 8] In the prodromal phase of BD, sleep disturbances seem to be prominent symptoms[9] present to a higher degree in first-degree relatives compared with control individuals[10] and suggested to be a risk factor for BD.[11 12] Overall, these observations underline the critical role of sleep in both diagnoses, monitoring and treatment of patients with BD.

Self-reports of sleep are primarily based on clinical interviews, questionnaires and sleep diaries.[13] However, these methods rely on retrospective information with the risk of recall bias and backfilling.[14] The gold standard for objective sleep assessment is polysomnography,[15] a method not suitable for daily monitoring. A wrist-worn actigraphy is another widely available device correlated with polysomnography[16] which is suitable for daily sleep–wake monitoring in naturalistic settings. However, actigraphy assesses sleep–wake rhythm solely based on one parameter (acceleration) and as with any wearable long-term monitoring suffers from adherence concerns.[17] Also, due to the Hawthorne effect, awareness of monitoring might cause alterations in sleep–wake rhythm.[18 19] Therefore, an unobtrusive and easily accessible platform is necessary to monitor sleep–wake patterns longitudinally, especially in patients with BD presenting with complex behavioural alterations in episodes of different polarities and duration.

Smartphones offer a unique platform to collect self-reported sleep information repeatedly during naturalistic settings. Recently, smartphone-based apps have been developed, aiming to estimate sleep–wake patterns based on both self-reported

information and automatically generated data.[20–22] The use of information concerning when the screen is turned on/off has been found to be a reliable and low-cost method to infer sleep–wake patterns.[23 24] Additionally, the duration the screen is turned on per day has been associated with reduced duration and quality of sleep in healthy individuals.[25] Nevertheless, smartphone-based self-reported and automatically generated sleep measurements have not been systematically compared with validated rating scales addressing sleep disturbance. Furthermore, it has not been investigated whether daily smartphone-based self-reported and automatically generated sleep measurements differ between patients with BD, unaffected first-degree relatives (UR) and healthy control individuals (HC).

### Objectives

The present study had three overall aims: First, to investigate if smartphone-based self-reported and automatically generated sleep measurements reflected validated measures of sleep including (a) a validated sleep questionnaire and (b) sleep disturbance assessed by trained clinicians according to sleep items of rating scales addressing the severity of depressive and manic symptoms, respectively. Second, to investigate associations between automatically generated smartphone-based sleep measurements and self-reported smartphone-based sleep. Third, to investigate whether smartphone-based self-reported and automatically generated sleep measurements would be able to discriminate between patients with *newly* diagnosed BD, UR and HC.

We hypothesised that (1) smartphone-based self-reported and automatically generated sleep (measured as mean hours slept between 12:00 midnight and 06:00, and daily variability in hours slept between 12:00 midnight and 06:00) would be associated with global measures of sleep; (2) daily smartphone-based self-reported and automatically generated sleep measures would be associated and (3) patients with BD would have higher day-to-day variability in sleep timing and more sleep disturbances compared with HC and UR with UR representing an intermediary group.

### METHODS

#### Study design

The present study is a part of the Bipolar Illness Onset (BIO) study.[26] Three groups of participants were included in the study: patients with BD, UR and HC. The data included in the present paper were collected from participants enrolled in the BIO study from September 2016 to February 2019. All participants were invited for assessments at baseline and annually for up to 3 years. Additionally, patients with BD were invited for assessments after a shift in a mood episode.

#### Study participants

The schedules of Clinical Assessment in Neuropsychiatry (SCAN) interview[27] was used to confirm the diagnosis of BD (or the lack of) according to the International Classification of Diseases 10th version (ICD-10).[28]

All newly diagnosed patients with BD living in the Capital Region of Denmark, who were offered a 2-year programme at the Copenhagen Affective Disorder Clinic Copenhagen, Denmark, were invited to participate in the study.[29] The criteria for inclusion were newly diagnosed BD/single manic episode according to the ICD-10 and the age of 15–70 years.

Unaffected relatives, including siblings or children to the patients with BD, were invited to participate in the study.

Exclusion criteria were any previous or current psychiatric diagnosis lower than F34.0 according to ICD-10 (ie, organic mental disorders, mental and behavioural disorders due to psychoactive substance use, schizophrenia or other psychotic disorders, affective disorders).

HC, without a treatment requiring psychiatric disorder in the individual or among the individuals' first-degree family members, were recruited among blood donors, age 15–70 years, from the Blood Bank at Rigshospitalet, Copenhagen, Denmark.

Observer-based rating scales and self-reported questionnaires were administered at each visit with the researchers. Due to the longitudinal design of the study with different number of visits with the researchers for each participant, some participants provided repeated self-reported and clinician-based rating scales.

#### Observer-based rating scales

Sleep items from two observer-based clinical rating scales, namely the Hamilton Depression Rating Scale 17-item (HAMD-17)[30] and the Young Mania Rating Scale (YMRS),[31] were used to assess the degree of sleep disturbances for the past 3 days. A summary measure of subitem 4–6 of HAMD-17 was used to evaluate the degree of insomnia and item 4 of the YMRS was used as an estimate of reduced sleep duration.

#### Patient-reported questionnaires

The Pittsburgh Sleep Quality Index (PSQI), a self-reported retrospective questionnaire, was used to evaluate overall sleep quality for the past month.[32] It consists of 19 items comprising 7 sleep factors, including estimates of sleep duration, subjective sleep quality, sleep disturbances, sleep efficiency, sleep latency, daytime dysfunction and use of sleeping medication. Higher scores indicate reduced sleep quality.[32] The PSQI has been found to have a good test–retest reliability for the global score.[33]

#### Smartphone-based monitoring

All included participants installed an app on their iPhone or Android smartphones (the Monsenso app). Automatically generated smartphone-based data, including information regarding screen on/off, was obtainable only from Android smartphones. All participants daily entered bedtime (defined as the time the light was turned off and the participant tried to sleep) and wake-up time (final awakening time) (figure 1). Participants were asked to enter smartphone-based data in real time, that is, daily. If the participant forgot to enter data for any specific day, it was possible to enter data retrospectively for up to 2 days. Self-reports could be filled out the whole day until 12:00 midnight. Using the reminder functionality of the smartphone, the participants were reminded daily to fill in the survey. The study was observational, and participants did not receive feedback via the app. Patients with BD were asked to fill in self-assessment as long as possible and preferably for at least 3 months, whereas the UR and HC were asked to fill in daily assessment for at least 1 month. Based on the smartphone-based self-reported bedtime and wake-up time, we calculated sleep duration (bedtime – wake-up time), mid-sleep (bedtime +sleep duration/2) and the number of hours slept between 12:00 midnight and 06:00. Daily variation in sleep duration and the number of hours slept between 12:00 and 06:00 were calculated by applying the root mean successive difference (RMSSD) method, taking the square root of the sum of the squared differences between daily and previous day's sleep evaluation.[34] The automatically generated screen features (the number of times the screen was turned on and off per day and the intervals between on and off times) were
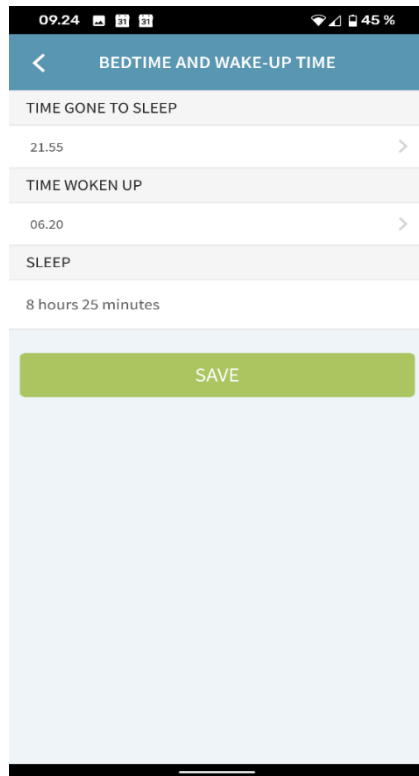
**Figure 1** Sleep registration on smartphone-based application, Monsenso.

used to estimate the time the screen was off between 12:00 and 06:00. We used the screen on/off features based on the assumption that most people lock their phones during bedtime and unlock their phones after awakening. Therefore, the screen on/off time may be a good estimate for sleep.[25]

**Statistical methods**

All hypotheses and statistical analyses were planned a priori. Differences in demographic variables and clinical characteristics were analysed using mixed models for continuous variables and $\chi^2$ for categorical variables.

For each measure of interest, a two-level linear mixed-effect model, which accommodates both the variation of the variables of interest within patients (intraindividual variation) and between individuals (interindividual variation), was employed. Familial relationships and repeated measures for each participant were accounted for by adding family number and identification number as random factors. In analyses comparing mean differences between the three groups (patients with BD, UR and HC), groups were added as a fixed factor. For each measure of interest, an unadjusted model and a model adjusted for age and sex were conducted. The linear mixed-effect model also allows us to include all observations and not just complete data sets. Thus, all available daily smartphone-based assessments of sleep and automatically collected smartphone-based data were included in the statistical analyses.

For the objectives of the study, first we investigated whether the number of hours and daily variability in hours slept between 12:00 and 06:00, calculated based on smartphone-based self-reported and automatically generated sleep measurements, were associated with sleep quality measured with PSQI and clinically rated sleep items of the HAMD-17 and the YMRS. Summary measures of smartphone-based sleep indices were calculated for the same time period as addressed by the self-reported and clinical rating scales. Second, we investigated whether the estimation of sleep duration, calculated from the automatically collected smartphone data during the hours between 12:00 and 06:00, was associated with same-day self-reported sleep between 12:00 and 06:00. Third, we investigated whether clinician-rated and smartphone-based measurements of sleep differed between patients with BD, UR and HC.

Since the field has been investigated in a few studies only, and due to the explorative nature of the present study, adjustment for multiple testing was not conducted. Thus, p values below 0.05 were considered statistically significant. Model assumptions were checked graphically by using residuals and predicted values. All statistical analyses were conducted using the Statistical Package of the Social Sciences (SPSS) V.22.

**FINDINGS**

**Socio-demographic and clinical characteristics**

A total of 240 patients with BD, 66 UR and 117 HC were included in the BIO-study from September 2016 to February 2019. Of these, a total of 203 patients with BD, 54 UR and 109 HC individuals were included in the present study and provided self-reported smartphone-based data concerning sleep. The main reasons for not willing to participate being that the study was too time-consuming, surveillance concerns and not owning a smartphone and did not want to borrow a smartphone. Automatically smartphone-generated data were only obtainable from Android smartphones and measurements of the screen on/off were provided for 71 patients with BD, 15 UR and 32 HC. Socio-demographic and clinical characteristics for the participants are presented in table 1. In total, the participants from the three groups provided smartphone-based self-reported data for 49 153 days (patients with BD 31 687; UR 5873; HC 11 593) and automatically generated smartphone-based data were collected for a total of 29 001 days (patients with BD 17 822 days; UR 3426 days; HC 7753 days).

**Daily smartphone-based sleep measurements**

As presented in table 2, we found that shorter smartphone-based self-reported sleep duration between 12:00 and 06:00 was associated with a worsening of sleep quality measured using the total PSQI score (B = −0.07, 95% CI: −0.10 to −0.05, p<0.001). Also, there was a statistically significant negative association between self-reported smartphone-based hours slept between 12:00 and 06:00 and sleep items on both the HAMD-17 (B=−0.10, 95% CI: −0.15 to −0.05, p<0.001) and the YMRS (B=−0.30, 95% CI: −0.43 to −0.17, p<0.001). Moreover, smartphone-based self-reported sleep duration was associated with sleep duration evaluated using the PSQI (adjusted for age and sex, B=0.73, 95% CI: 0.64 to 0.82, p<0.001).

There were no associations between automatically generated smartphone-based estimates of hours slept between 12:00 and 06:00 and any self-rated or clinician-rated sleep measurements.

**Daily automatically generated versus self-reported smartphone-based sleep measurements**

As can be seen in table 3, the automatic sampling of sleep (between 12:00 and 06:00) using the screens on/off events is a proxy for sleep. There was a statistically significant positive association with smartphone-based self-reported sleep time (between 12:00 and 06:00) and hence might represent a potential marker for sleep duration (B=0.28, 95% CI: 0.26 to 0.30, p<0.001).

**Table 1** Socio-demographic and clinical characteristics of patients with BD, UR and HC at baseline, n=366

| | BD | UR | HC | BD vs HC, p value | BD vs UR, p value | UR vs HC, p value |
|---|---|---|---|---|---|---|
| Participants, n | 203 | 54 | 109 | | | |
| Age, years | 28 (24–35) | 26 (22–31) | 26 (24–36) | 0.39 | 0.028 | 0.014 |
| Female sex, % (n) | 69.0 (140) | 56.6 (30) | 63.3 (69) | 0.31 | 0.064 | 0.34 |
| Education, years | 15 (13–17) | 15 (13–17) | 16 (15–17) | <0.001 | 0.27 | 0.09 |
| Full-time employment, % (n) | 32.0 (64) | 48.1 (25) | 50.0 (52) | 0.002 | 0.031 | 0.82 |
| Student, % (n) | 40.5 (81) | 46.2 (24) | 38.5 (40) | 0.73 | 0.46 | 0.36 |
| HAMD-17 | 9 (5–15) | 2 (0–3) | 0 (0–2) | <0.001 | <0.001 | 0.60 |
| YMRS | 2 (0–7) | 0 (0–2) | 0 (0–1) | <0.001 | <0.001 | 0.60 |
| FAST, total score | 22 (13–30) | 1 (0–3) | 0 (0–2) | <0.001 | <0.001 | 0.20 |
| BD type II, % (n) | 66.5 (135) | – | – | – | – | – |
| Age of onset, years | 17 (14–21) | – | – | – | – | – |
| Illness duration, years* | 10 (5–14) | – | – | – | – | – |
| Untreated BD, years† | 4 (1–10) | – | – | – | – | – |
| No. of prior depressive episodes | 5 (3–13) | – | – | – | – | – |
| No. of prior hypomanic episodes | 5 (2–15) | – | – | – | – | – |
| No. of prior manic episodes | 1 (1–2) | – | – | – | – | – |
| No. of prior mixed episodes | 1 (1–2) | – | – | – | – | – |
| No. of total episodes | 12 (6–27) | – | – | – | – | – |
| Episode at baseline‡ | | | | | | |
| Full or partial remission, % (n) | 60.6 (123) | – | – | – | – | – |
| Hypomania/mania, % (n) | 5.9 (12) | – | – | – | – | – |
| Depression, % (n) | 33.0 (67) | – | – | – | – | – |
| Mixed, % (n) | 0.5 (1) | – | – | – | – | – |
| Prescribed medication at baseline | | | | | | |
| Lithium, % (n) | 33.0 (67) | – | – | – | – | – |
| Antiepileptic treatment, % (n) | 49.8 (101) | – | – | – | – | – |
| Antipsychotic treatment, % (n) | 38.9 (79) | – | – | – | – | – |
| Antidepressant treatment, % (n) | 27.1 (55) | – | – | – | – | – |
| Other§, % (n) | 7.9 (16) | – | – | – | – | – |
| No psychotropic medicine, % (n) | 17.7 (36) | – | – | – | – | – |

Continuous variables are presented as mean (SD) or median (IQR) and p values are calculated based on differences in mean between the three groups using mixed models. Categorical data are presented as % (n) and p values are calculated by using the $\chi^2$ test.

*Illness duration was defined as the time from the first episode to the time of inclusion.

†Untreated BD was defined as time from the first mania, hypomania or mixed episode to time of diagnosis.

‡Episode at baseline according to HAMD-17 and YMRS, full or partial remission was defined as HAMD-17 and YMRS <14, scores on HAMD-17 or YMRS ≥14 were defined as depression or hypomania/mania and as mixed if both HAMD-17 and YMRS ≥14.

§Other includes benzodiazepines and melatonin.

BD, bipolar disorder; FAST, Functional Assessment Short Test; HAMD-17, Hamilton Depression Rating Scale 17-item; HC, healthy control individuals; UR, unaffected first-degree relatives; YMRS, Young Mania Rating Scale.

**Differences in daily smartphone-based sleep measurements between patients with newly diagnosed BD, UR, and HC**

As can be seen from the upper part of table 4, the day-to-day variability in total sleep duration was higher in patients with BD compared with UR and HC, respectively (model adjusted for age and sex: all p values <0.001). Furthermore, in patients with BD, the total sleep time between 12:00 and 06:00 was shorter compared with UR and HC. Interestingly, the total sleep duration did not differ between the three groups. However, this may be explained by patients with BD having more variability in sleep duration (figure 2). The total PSQI scores and subitems addressing sleep on the HAMD-17 were statistically significantly higher in UR compared with HC individuals. Other measures were not statistically significant between UR and HC individuals. In exploratory subanalysis, including only visits when participants were in full or partial remission (HAMD-17 and YMRS <14), the differences between groups in the PSQI, the HAMD-17 and the YMRS

scores were comparable with the results presented in table 4 (results not presented).

The lower part of table 4 shows that between 12:00 and 06:00 patients with BD had decreased number of hours with the screen turned off, increased day-to-day variability in hours with the screen turned off and increased number of times the screen was turned on compared with HC. Although, none of these differences reached statistical significance.

**DISCUSSION**

This study showed three things. First, smartphone-based self-reported hours slept between 12:00 midnight and 06:00 reflected measures of sleep according to the PSQI, and sleep items on the HAMD and the YMRS, whereas automatically smartphone-generated sleep measures calculated based on the screen on/off time did not. Second, automatically smartphone-generated calculation of hours slept between 12:00 and 06:00 was associated with daily smartphone-based self-reported

**Table 2** Associations between smartphone-based sleep indices* and a self-reported sleep questionnaire†, and clinically rated depressive‡ and manic§ symptoms, respectively

| | Model 1¶ | | | Model 2¶ | | |
|---|---|---|---|---|---|---|
| | B | 95% CI | P value | B | 95% CI | P value |
| Smartphone-based self-reported hours slept between 12:00 and 06:00** | | | | | | |
| HAMD sleep item, n=471 | −0.10 | −0.15 to −0.05 | **<0.001** | −0.10 | −0.15 to −0.05 | **<0.001** |
| YMRS sleep item, n=471 | −0.30 | −0.43 to −0.17 | **<0.001** | −0.30 | −0.43 to −0.17 | **<0.001** |
| PSQI, n=383 | −0.08 | −0.11 to −0.05 | **<0.001** | −0.07 | −0.10 to −0.05 | **<0.001** |
| Smartphone-based self-reported daily variability in hours slept between 12:00 and 06:00†† | | | | | | |
| PSQI, n=248 | 0.05 | 0.02 to 0.08 | **0.001** | 0.05 | 0.02 to 0.08 | **0.001** |
| Automatically smartphone-generated hours slept between 12:00 and 06:00 | | | | | | |
| HAMD sleep item, n=136 | −0.03 | −0.10 to 0.045 | 0.46 | −0.02 | −0.094 to 0.05 | 0.57 |
| YMRS sleep item, n=136 | −0.08 | −0.25 to 0.07 | 0.29 | −0.08 | −0.25 to 0.08 | 0.31 |
| PSQI, n=115 | −0.002 | −0.037 to 0.33 | 0.92 | −0.003 | −0.038 to 0.033 | 0.89 |
| Automatically smartphone-generated daily variability in hours slept between 12:00 and 06:00 | | | | | | |
| PSQI, n=96 | −0.027 | −0.078 to 0.024 | 0.30 | −0.027 | −0.079 to 0.024 | 0.30 |

All participants are included in the analyses.

*Summary measures of smartphone-based sleep indices were calculated for the same time period as addressed by the self-reported and clinical rating scales: for the PSQI the past 28 days, and for the HAMD and YMRS the past 3 days.

†PSQI: total score.

‡HAMD: sum of items 4–6 addressing sleep.

§YMRS: subitem 4 addressing sleep.

¶Model 1: Unadjusted. Model 2: Adjusted for age and sex.

**Mean hours slept between 12:00 and 06:00.

††Daily variability in hours slept between 12:00 and 06:00 calculated for the past 14 days and 28 days, respectively, by using the RMSSD score.

Automatically smartphone-generated sleep between 12:00 and 06:00 calculated by using the screen on/off feature.

Daily variability in hours slept between 12:00 and 06:00 calculated based on screen on/off features and by using the RMSSD score.

HAMD, Hamilton Depression Rating Scale; PSQI, The Pittsburgh Sleep Quality Index; RMSSD, root mean successive difference; YMRS, The Young Mania Rating Scale.

hours slept between 12:00 and 06:00. Third, according to smartphone-based self-reported and automatically generated measurements, patients with BD slept less between 12:00 and 06:00 and with more interruption and variability in daily sleep patterns compared with HC, although differences in automatically smartphone-generated measures did not reach statistical significance. UR had numerically intermediary levels between patients and HC on most smartphone-based sleep measurements but did not differ statistically significant from HC.

### Estimation of sleep from smartphone-based data

This study overall indicates that smartphones could be a potential platform for long-term sleep monitoring. Also this finding is supported by other studies supporting the use of smartphones for sleep monitoring in patients with schizophrenia and healthy individuals with sleep problems.[21 35] The lack of statistical significant difference in automatically

estimated sleep between the three groups could be explained by decreased amount of automatically smartphone-generated data with decreased statistical power, which increases the risk of type II error and the fact that we exclusively used the screen on/off time to estimate sleep. Participants are not necessarily at sleep when the screen is off, which could result in a higher estimation of sleep time using smartphone-based sensor data compared with self-reported. Especially bedtime might be estimated to be earlier than self-reported bedtime.[21] Therefore, screen off time between 12:00 and 06:00 could be estimated using more advanced machine learning methods[24] or possibly by combining screen features with other smartphone-based features, suitable for sleep monitoring such as ambient light, Global Positioning System locations, Fitbit watch, etc. to improve sleep characterisation.[21 35–37] However, these methods can be heavily battery consuming, which limits realistic use by patients during naturalistic settings.

**Table 3** Associations between self-reported smartphone-based sleep and automatically smartphone-generated sleep between 12:00 and 06:00

| | Model 1* | | | Model 2* | | |
|---|---|---|---|---|---|---|
| | B | 95% CI | P value | B | 95% CI | P value |
| Self-reported hours slept between 12:00 midnight and 06:00† | | | | | | |
| Hours slept between 12:00 midnight and 06:00 (screen on/off)‡ | 0.28 | 0.26 to 0.30 | <0.001 | 0.28 | 0.26 to 0.30 | <0.001 |
| Self-reported smartphone-based variability§ | | | | | | |
| Variability in hours slept between 12:00 midnight and 06:00 (screen on/off)¶ | 0.14 | −0.03 to 0.32 | 0.10 | 0.05 | −0.12 to 0.23 | 0.53 |

All participants regardless of study group are included in the analyses (n=113).

*Model 1: Unadjusted. Model 2: Adjusted for age and sex.

†Self-reported smartphone-based measure of hours slept between 12:00 midnight and 06:00, calculated based on bedtime and wake-up time.

‡Automatically smartphone-generated measure of hours slept between 12:00 midnight and 06:00 calculated as total time screen is off between 12:00 and 06:00.

§Daily variability in self-reported hours slept between 12:00 midnight and 06:00, calculated by using the RMSSD score.

¶Daily variability in automatically smartphone-generated measures of hours slept between 12:00 and 06:00, calculated based on time screen is off between 12:00 and 06:00 and by using the RMSSD score.

RMSSD, root mean successive difference.

**Table 4** Estimated differences in sleep indices between patients with BD, UR and HC (n=366)

| | Model 1* | | | | | | | | | Model 2* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | BD | | UR | | HC | | BD/HC | BD/UR | UR/HC | BD | | UR | | HC | | BD/HC | BD/UR | UR/HC |
| | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI | P value | P value | P value | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI | P value | P value | P value |
| **Self-reported smartphone-based sleep** BD=203, UR=54, HR=109 | | | | | | | | | | | | | | | | | | |
| Bedtime (hours:min) | 00:25 | 00:15 to 00:34 | 23:59 | 23:41 to 00:16 | 23:52 | 23:39 to 00:05 | <0.001 | 0.003 | 0.56 | 00:23 | 00:13 to 00:32 | 23:49 | 23:32 to 00:05 | 23:51 | 23:38 to 00:04 | <0.001 | <0.001 | 0.80 |
| Wake-up time (hours:min) | 08:21 | 08:11 to 08:30 | 7.53 | 7.35 to 8.11 | 7.40 | 7.26 to 7.54 | <0.001 | 0.003 | 0.25 | 8.18 | 8.08 to 8.27 | 7.40 | 7.23 to 7.57 | 7.39 | 7.26 to 7.51 | <0.001 | <0.001 | 0.89 |
| Mid-sleep (hours:min) | 04:23 | 04:14 to 04:32 | 03:56 | 03:39 to 04:18 | 03:46 | 03:34 to 03:58 | <0.001 | 0.001 | 0.36 | 04:21 | 04:12 to 04:29 | 03:45 | 03:29 to 04:00 | 3.45 | 3.33 to 3.56 | <0.001 | <0.001 | 0.99 |
| Total sleep time (hours:min) | 07:56 | 07:49 to 08:03 | 07:53 | 07:39 to 08:06 | 07:47 | 07:38 to 07:57 | 0.15 | 0.74 | 0.46 | 07:56 | 07:48 to 08:03 | 07:51 | 07:37 to 08:05 | 07:47 | 07:38 to 07:57 | 0.17 | 0.55 | 0.65 |
| Total sleep time variability† (hours) | 2.09 | 1.99 to 2.19 | 1.86 | 1.68 to 2.04 | 1.73 | 1.59 to 1.87 | <0.001 | 0.014 | 0.26 | 2.10 | 2.00 to 2.20 | 1.80 | 1.63 to 1.98 | 1.77 | 1.63 to 1.90 | <0.001 | 0.001 | 0.73 |
| Wake-up after 08:00‡ | 55.9 | 52.1 to 59.8 | 43.3 | 35.8 to 50.8 | 42.7 | 37.5 to 48.0 | <0.001 | 0.002 | 0.90 | 57.4 | 53.7 to 61.2 | 40.1 | 33.1 to 47.2 | 45.0 | 40.1 to 50.0 | <0.001 | <0.001 | 0.26 |
| Hours slept between 12:00 and 06:00 (hours:min)§ | 04:59 | 04:53 to 05:05 | 05:15 | 05:05 to 05:25 | 05:24 | 05:16 to 05:32 | <0.001 | 0.003 | 0.15 | 05:01 | 04:55 to 05:07 | 05:21 | 05:10 to 05:31 | 05:25 | 05:17 to 05:33 | <0.001 | <0.001 | 0.50 |
| Daily variability in hours slept between 12:00 and 06:00¶ (hours) | 1.47 | 1.38 to 1.56 | 1.22 | 1.05 to 1.40 | 1.30 | 1.18 to 1.43 | 0.032 | 0.011 | 0.46 | 1.50 | 1.41 to 1.59 | 1.16 | 1.00 to 1.33 | 1.36 | 1.24 to 1.47 | 0.042 | 0.001 | 0.063 |
| **Pittsburgh Sleep Quality Index** BD=241, UR=81, HC=163 | | | | | | | | | | | | | | | | | | |
| Total PSQI score** | 8.66 | 8.23 to 9.09 | 5.13 | 4.36 to 5.90 | 3.81 | 3.26 to 4.36 | <0.001 | <0.001 | 0.006 | 8.56 | 8.12 to 8.99 | 5.01 | 4.24 to 5.78 | 3.77 | 3.22 to 4.32 | <0.001 | <0.001 | 0.011 |
| **Automatically smartphone-generated measures of sleep††** BD=71 UR=15, HC=32 | | | | | | | | | | | | | | | | | | |
| Hours with screen turned off between 12:00 and 06:00 am | 05:28 | 05:21 to 05:35 | 05:40 | 05:25 to 05:54 | 05:39 | 05:28 to 05:49 | 0.082 | 0.13 | 0.94 | 05:29 | 05:22 to 05:36 | 5.41 | 05:26 to 05:56 | 05:39 | 05:29 to 05:49 | 0.10 | 0.12 | 0.82 |
| Variability in hours slept between 12:00 and 06:00 | 1.17 | 1.01 to 1.34 | 1.18 | 0.84 to 1.53 | 1.07 | 0.81 to 1.32 | 0.49 | 0.96 | 0.59 | 1.16 | 0.99 to 1.33 | 1.14 | 0.79 to 1.49 | 1.09 | 0.84 to 1.34 | 0.62 | 0.90 | 0.81 |
| Screen turned on (number/daily between 12:00 and 06:00) | 5.71 | 4.43 to 7.0 | 5.28 | 2.48 to 8.07 | 4.40 | 2.49 to 6.30 | 0.26 | 0.78 | 0.61 | 5.50 | 4.21 to 6.80 | 4.79 | 2.02 to 7.55 | 4.39 | 2.53 to 6.25 | 0.34 | 0.64 | 0.81 |
| **Sleep measures on HAMD and YMRS** BD=337, UR=91, HC=183 | | | | | | | | | | | | | | | | | | |
| HAMD subitems 4–6 | 2.33 | 2.12 to 2.52 | 1.24 | 0.85 to 1.63 | 0.67 | 0.40 to 0.95 | <0.001 | <0.001 | 0.019 | 2.26 | 2.05 to 2.46 | 1.17 | 0.79 to 1.56 | 0.65 | 0.38 to 0.92 | <0.001 | <0.001 | 0.029 |
| YMRS subitem 4 | 0.39 | 0.31 to 0.46 | 0.14 | −0.08 to 0.29 | 0.07 | −0.03 to 0.17 | <0.001 | 0.002 | 0.44 | 0.38 | 0.31 to 0.46 | 0.13 | −0.02 to 0.27 | 0.08 | −0.02 to 0.18 | <0.001 | 0.002 | 0.61 |

*Model 1: Unadjusted, Model 2: Adjusted for age and sex.
†Day-to-day variability in sleep duration calculated by using the RMSSD score.
‡Proportion of days participants woke up after 08:00.
§Hours where the participants had self-reported sleep between12:00 and 06:00.
¶Day-to-day variability in hours slept between 12:00 and 06:00, calculated by using the RMSSD score.
**PSQI: the total score, all visits are used in the calculations.
††Automatically smartphone-generated sleep measures.
HAMD: sum of items 4–6 addressing sleep.
YMRS: subitem 4 addressing sleep.
BD, bipolar disorder; HAMD, Hamilton Depression Rating Scale; HC, healthy control individuals; PSQI, Pittsburgh Sleep Quality Index; UC, unaffected first-degree relatives; YMRS, Young Mania Rating Scale.
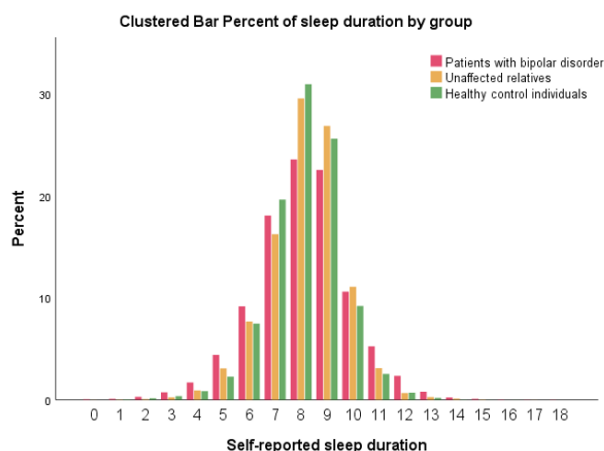
**Figure 2** Clustered bar chart illustrating percentage of time with self-reported sleep duration by group.

## Differences in sleep measurements between study groups

Interestingly, we found that patients with *newly* diagnosed BD have more irregular sleep patterns and reduced sleep quality compared with UR and HC, respectively. This could indicate that sleep disturbances are present at early illness stages. Specifically, patients with BD had a delayed sleep pattern (mean 26 min later bedtime and 41 min delayed wake-up time) and slept less between 12:00 and 06:00 than HC. Also, patients with BD had more variability in daily sleep patterns. The latter is in accordance with findings in a meta-analysis.[5] Interestingly, the mean PSQI total score for UR was 5.01, just above the cut-off for poor sleepers.[32] The UR group scores for three of the subitems on the PSQI sleep latency, sleep disturbance and sleep quality were significantly different from those of the HC group. This is in accordance with several other studies that have reported sleep disturbances in individuals at risk of developing BD.[10 11 38] Studies point towards that especially sleep onset latency, awakening during night time as well as overall sleep quality could be trait and state marker for BD.[3 11 39 40] The present study did not include smartphone-based self-assessment of sleep satisfaction, but used questionnaire-based measures of sleep quality. Combining both smartphone-based subjective sleep reports and automatically generated sleep estimations could potentially provide a more sensitive and specific estimate of sleep variability, sleep latency and sleep disturbance and may be clinically useful in detecting emerging patterns of sleep disturbance and represent a potential diagnostic, risk and treatment marker for BD.

## Limitations

First, we investigated associations between smartphone-based sleep measures against PSQI and sleep items on HAMD-17 and YMRS. Additionally, future studies comparing automatically generated smartphone-based sleep measurements against actigraphy-based sleep measurements or polysomnography would be highly relevant for validating objective sleep-monitoring methods. Second, participants were included regardless of mood state. Additionally, the patients with BD included in the present study presented rather low levels of depressive and manic symptoms during the study period. Thus, the associations and differences in sleep parameters when comparing the patients with BD, UR and HC will not reflect associations and differences during more severe affective states. However, the patients with BD still had sleep disturbances even though most were in remission.

Third, linear mixed-effect models were used for the statistical analyses using all available data points. However, missing data were not accounted for in the statistical analyses, and thus were assumed to be missing at random. It should be considered that the patients during depressive and manic periods had alterations in adherence to self-monitoring, and thus the missing data may contain clinically useful information.[41] Fourth, we investigated estimated sleep time between 12:00 and 06:00. A larger time window would have been interesting to investigate, especially for those with a delayed sleep phase. However, we were interested in how much time the participants slept during nighttime. It is also more likely that participants are awake after 06:00, without using the phone, because they are at work, school, etc.; therefore, a single attribute, as screen off time, may not be an adequate measure of sleep in the morning. There exist several sensors that could be suitable for monitoring sleep and integrating data from more than one of these sensors or applying more advanced machine learning algorithms may possibly represent a more accurate sleep measurement that can capture total sleep duration instead of only an estimate of sleep during nighttime. Also, phone usage may vary vastly between individuals. Therefore, it may be that the best estimate of sleep measures is a combination of self-reported and automatically generated sleep measurements. Fifth, we collected automatically generated smartphone data across multiple platforms, and there may be possible variations in automatically generated smartphone data between platforms, which we did no account for in the present study. Sixth, due to the sample size of the present study, we did not investigate potential differences between participants using Android smartphones and iPhone smartphones in smartphone-based self-assessed sleep measures. There may be potential demographic differences between these groups given the comparative costs of the devices. Seventh, despite that the three groups were matched on sex and age, work status differed, which probably influence sleep–wake patterns and may explain parts of the variability between patients with BD and UR/HC. Eighth, patients with BD were taking psychotropic medication, and this was not accounted for in the analyses. Finally, all patients were followed in the Copenhagen Affective Disorder Clinic, where stabilisation of the sleep–wake pattern is an essential part of the treatment. Therefore, sleep disturbances in this group might not be representable of all patients with BD.

## Strengths

This is the first large observational study with 366 systematically recruited participants, including patients with *newly* diagnosed BD and UR. Diagnosis (and lack of diagnosis) was verified by using a SCAN interview. On a daily basis, participants reported sleep in real time, and because the smartphone system was used, recall bias was minimised. In addition, the participants were clinically evaluated and completed questionnaires during the study period. The clinical assessments used included the HAMD-17 and the YMRS, which, in contrast to the questionnaires, are clinical rating scales administered by an experienced clinical researcher.

## CONCLUSIONS

Daily self-reported smartphone-based data may represent measurements of sleep patterns that differ between patients with BD and HC and potentially between UR and HC. Automatically smartphone-generated sleep measurements may hold promises but should be investigated in more detail and potentially include

additional smartphone-based sensor data, more advanced feature extractions and algorithms to estimate circadian rhythm 24/7.

**ORCID iD**
Sharleny Stanislaus http://orcid.org/0000-0003-1073-6436

## REFERENCES

1 Harvey AG. Sleep and circadian rhythms in bipolar disorder: seeking synchrony, harmony, and regulation. *Am J Psychiatry* 2008;165:820–9.
2 Murray G, Harvey A. Circadian rhythms and sleep in bipolar disorder. *Bipolar Disord* 2010;12:459–72.
3 Geoffroy PA, Scott J, Boudebesse C, *et al*. Sleep in patients with remitted bipolar disorders: a meta-analysis of actigraphy studies. *Acta Psychiatr Scand* 2015;131:89–99.
4 Harvey AG, Schmidt DA, Scarnà A, *et al*. Sleep-related functioning in euthymic patients with bipolar disorder, patients with insomnia, and subjects without sleep problems. *Am J Psychiatry* 2005;162:50–7.
5 Ng TH, Chung K-F, Ho FY-Y, *et al*. Sleep-wake disturbance in interepisode bipolar disorder and high-risk individuals: a systematic review and meta-analysis. *Sleep Med Rev* 2015;20:46–58.
6 Gruber J, Harvey AG, Wang PW, *et al*. Sleep functioning in relation to mood, function, and quality of life at entry to the systematic treatment enhancement program for bipolar disorder (STEP-BD). *J Affect Disord* 2009;114:41–9.
7 Cretu JB, Culver JL, Goffin KC, *et al*. Sleep, residual mood symptoms, and time to relapse in recovered patients with bipolar disorder. *J Affect Disord* 2016;190:162–6.
8 Jackson A, Cavanagh J, Scott J. A systematic review of manic and depressive prodromes. *J Affect Disord* 2003;74:209–17.
9 Van Meter AR, Burke C, Youngstrom EA, *et al*. The bipolar prodrome: meta-analysis of symptom prevalence prior to initial or recurrent mood episodes. *J Am Acad Child Adolesc Psychiatry* 2016;55:543–55.
10 Duffy A, Alda M, Crawford L, *et al*. The early manifestations of bipolar disorder: a longitudinal prospective study of the offspring of bipolar parents. *Bipolar Disord* 2007;9:828–38.
11 Levenson JC, Axelson DA, Merranko J, *et al*. Differences in sleep disturbances among offspring of parents with and without bipolar disorder: association with conversion to bipolar disorder. *Bipolar Disord* 2015;17:836–48.
12 Ritter PS, Marx C, Bauer M, *et al*. The role of disturbed sleep in the early recognition of bipolar disorder: a systematic review. *Bipolar Disord* 2011;13:227–37.
13 Buysse DJ, Ancoli-Israel S, Edinger JD, *et al*. Recommendations for a standard research assessment of insomnia. *Sleep* 2006;29:1155–73.
14 Stone AA, Shiffman S, Schwartz JE, *et al*. Patient compliance with paper and electronic diaries. *Control Clin Trials* 2003;24:182–99.
15 Baglioni C, Nanovska S, Regen W, *et al*. Sleep and mental disorders: a meta-analysis of polysomnographic research. *Psychol Bull* 2016;142:969–90.
16 Kaplan KA, Talbot LS, Gruber J, *et al*. Evaluating sleep in bipolar disorder: comparison between actigraphy, polysomnography, and sleep diary. *Bipolar Disord* 2012;14:870–9.
17 Finkelstein EA, Haaland BA, Bilger M, *et al*. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. *Lancet Diabetes Endocrinol* 2016;4:983–95.
18 Saunders KEA, Bilderbeck AC, Panchal P, *et al*. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. *Eur Psychiatry* 2017;41:115–21.
19 Wickström G, Bendix T. The "Hawthorne effect"--what did the original Hawthorne studies actually show? *Scand J Work Environ Health* 2000;26:363–7.
20 Fino E, Mazzetti M. Monitoring healthy and disturbed sleep through smartphone applications: a review of experimental evidence. *Sleep Breath* 2019;23:13–24.
21 Min J-K, Doryab A, Wiese J. Toss'n'turn: smartphone as sleep and sleep quality detector. Proceedings of the SIGCHI conference on human factors in computing systems, ACM, 2014:477–86.
22 Wang R, Chen F, Chen Z. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing, ACM, 2014:3–14.
23 Abdullah S, Matthews M, Murnane EL. Towards circadian computing: early to bed and early to rise makes some of US unhealthy and sleep deprived. Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing, ACM, 2014:673–84.
24 Cuttone A, Bækgaard P, Sekara V, *et al*. Sensiblesleep: a Bayesian model for learning sleep patterns from smartphone events. *PLoS One* 2017;12:e0169901.
25 Christensen MA, Bettencourt L, Kaye L, *et al*. Direct measurements of smartphone screen-time: relationships with demographics and sleep. *PLoS One* 2016;11:e0165331.
26 Kessing LV, Munkholm K, Faurholt-Jepsen M, *et al*. The bipolar illness onset study: research protocol for the bio cohort study. *BMJ Open* 2017;7:e015462.
27 Wing JK, Babor T, Brugha T, *et al*. Scan. schedules for clinical assessment in neuropsychiatry. *Arch Gen Psychiatry* 1990;47:589–93.
28 World Health Organisation. *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Genova: World Health Organisation, 1992.
29 Kessing LV, Hansen HV, Hvenegaard A, *et al*. Treatment in a specialised out-patient mood disorder clinic V. standard out-patient treatment in the early course of bipolar disorder: randomised clinical trial. *Br J Psychiatry* 2013;202:212–9.
30 Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;6:278–96.
31 Young RC, Biggs JT, Ziegler VE, *et al*. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 1978;133:429–35.
32 Buysse DJ, Reynolds CF, Monk TH, *et al*. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989;28:193–213.
33 Backhaus J, Junghanns K, Broocks A, *et al*. Test-retest reliability and validity of the Pittsburgh sleep quality index in primary insomnia. *J Psychosom Res* 2002;53:737–40.
34 Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. *Psychol Assess* 2009;21:463–75.
35 Staples P, Torous J, Barnett I, *et al*. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. *NPJ Schizophr* 2017;3:37.
36 Abdullah S, Matthews M, Frank E, *et al*. Automatic detection of social rhythms in bipolar disorder. *J Am Med Inform Assoc* 2016;23:538–43.
37 Chen Z, Lin M, Chen F. Unobtrusive sleep monitoring using smartphones. Proceedings of the 7th International Conference on pervasive computing technologies for healthcare, Institute for computer sciences, Social-Informatics and …, 2013:145–52.
38 Scott J, Naismith S, Grierson A, *et al*. Sleep–wake cycle phenotypes in young people with familial and non-familial mood disorders. *Bipolar Disord* 2016;18:642–9.
39 Tazawa Y, Wada M, Mitsukura Y, *et al*. Actigraphy for evaluation of mood disorders: a systematic review and meta-analysis. *J Affect Disord* 2019;253:257–69.
40 la Cour Karottki NF, Coello K, Stanislaus S, *et al*. Sleep and physical activity in patients with newly diagnosed bipolar disorder in remission, their first-degree unaffected relatives and healthy controls. *Int J Bipolar Disord*. In Press 2020;8:16.
41 Gershon A, Kaufmann CN, Torous J, *et al*. Electronic ecological Momentary assessment (EMA) in youth with bipolar disorder: demographic and clinical predictors of electronic EMA adherence. *J Psychiatr Res* 2019;116:14–18.