Published in final edited form as:

Behav Sleep Med. 2013 ; 11(4): . doi:10.1080/15402002.2012.685997.

Comparison of five actigraphy scoring methods with bipolar disorder

Carole Boudebesse, MD^{1,2,3}, Marion Leboyer, MD, PhD^{2,3}, Amy Begley, MA¹, Annette Wood, BA¹, Jean Miewald, BA¹, Martica Hall, PhD¹, Ellen Frank, PhD¹, David Kupfer, MD¹, and Anne Germain, PhD¹

¹Department of Psychiatry, University of Pittsburgh Medical Center, United States

²INSERM 955 unit, Groupe Henry Mondor-Albert Chenevier, France

³Fondation FondaMental, French National Science Foundation

Abstract

The goal of this study was to compare five actigraphy scoring methods in a sample of 18 remitted patients with bipolar disorder. Actigraphy records were processed using five different scoring methods relying on the sleep diary; the event-marker; the software-provided automatic algorithm; the automatic algorithm supplemented by the event-marker; visual inspection (VI) only. The Algorithm and the VI methods differed from the other methods for many actigraphy parameters of interest. Particularly, the Algorithm method yielded longer sleep duration, and the VI method yielded shorter sleep latency compared to the others methods. The present findings provide guidance for the selection of signal processing method based on sleep parameters of interest, time-cue sources and availability, and related scoring time costs for the study.

Introduction

Sleep disturbances and circadian abnormalities are frequent in bipolar disorder (Harvey et al. 2008), and can be characterized with actigraphy (Morgenthaler et al., 2007; Ancoli-Israel et al., 2003). A handful of studies reported greater variability across days in activity pattern (Jones et al., 2005), sleep patterns (Axelson et al., 2003), sleep duration and nocturnal wake time (Millar et al., 2004), as well as longer sleep time and lower daytime activity among euthymic patients with bipolar disorder compared to insomniac and good sleepers (Harvey et al., 2005; Millar et al., 2004; Salvatore et al., 2008). However, these studies have used different scoring methods to identify relevant sleep parameters. The increasing use of actigraphy in psychiatric samples (e.g., Harvey et al., 2005; Millar et al., 2004) raises the question of standardization.

The goal of the present study was to compare five different approaches for processing actigraphy data on sleep-wake parameter estimates in a sample of remitted patients with type I bipolar disorder. The scoring methods relied on a single or different combinations of approaches to identify bedtime and rise time. Sleep diary data based on participants' subjective sleep perception (Diary); 2. Event marker (Marker, based on participants' subjective perception) 3. Software-provided automatic algorithm (Algorithm; a computer-defined approach); 4. Automatic algorithm supplemented by the event marker (Algorithm +

Corresponding Author: Anne Germain PhD, Associate Professor of Psychiatry and Psychology, University of Pittsburgh School of Medicine. 3811 O'Hara Street, Room E-1118, Pittsburgh PA 15213. germax@upmc.edu. Telephone: 412-246-6413. FAX: 412-246-5300.

The authors have no conflict of interest to report

Marker, which combined the objective and subjective approaches); and 5. Visual Inspection (VI) by a trained scorer, which comprises the researcher's signal perception.

Methods

Participants were adult outpatients suffering from Bipolar Disorder 1 who were recruited from outpatient clinics at the University of Pittsburgh. All provided written, informed consent to participate in an ongoing federally funded study looking at the effects of an augmented treatment of bipolar disorder on medical risk outcomes (MH081003; PI: Kupfer). Data collected from all available participants (n= 18) enrolled in the parent study between December 2008 and March 2009 were used in the present study.

After obtaining informed written consent, diagnostic and mood assessments were conducted by a trained psychiatric nurse. Primary diagnosis of the participants was type I bipolar disorder, determined by the Structured Interview for DSMIV Axis I Disorders (SCID; First et al., 1995). Remission was defined by a score < 6 on the Montgomery-Asberg Depression Rating Scale (MADRS; Samuels et al., 1996) and a score < 7 on the Young Mania Rating Scale (YMRS; Young et al., 1978). Participants were excluded from the parent study if they met any of the following exclusion criteria: Ultra-rapid cycling bipolar I disorder, pervasive developmental disorder, antisocial personality disorder, current substance dependence or abuse, organic mental disorder.

All participants wore the actigraph device (AW-64, Phillips Respironics) continuously on non-dominant wrist, for a mean of 7.83 days (SD=0.92). Participants were instructed to press the event-marker when they went to bed to sleep and when they got out of bed to start the day. The actigraph accelerometer sensitivity was set at 0.05 g for 3 to 11 Hz. The analog signal was digitized by digital integration method. We used a wake threshold of 40 counts per minute. Data was sampled in one-minute epochs and analyzed with the sleep detection algorithm provided by actiware 5.0 software (Phillips Respironics). Sleep/wake parameters can be determined by the algorithm only or by the combination algorithm with event markers or sleep diaries.

In addition, all participants completed the Pittsburgh Sleep Diary (Monk et al., 1994) concurrently with the actigraphy recording. The Pittsburgh Sleep Diary is a diary of sleep-wake behavior and asks questions regarding the time at which the participant went to bed, attempted to fall asleep, the number, cause, and duration of nocturnal awakenings, final time out of bed, and estimated total sleep time.

In order to process the actigraphy signal by Actiware 5.0 Software (Phillips Respironics), start time and end time of the rest intervals need to be defined for each day, and for each record. Once the rest interval is identified, the active interval and the sleep interval are computed. Rest interval, sleep interval and active interval are terms that specifically apply to the device and software used in the study. The rest interval reflects time spent in bed, or the total time of inactivity in bed or with the intention to sleep (excluding theoretically time spent in bed to read or watch TV for example). The sleep interval and active interval refer to the total sleep time and the detected period of daytime wakefulness, respectively.

The onset of the rest interval (bedtime) was defined by the time when a participant went to bed with the intention to sleep (and when s/he was instructed to push the event marker, and log the clock time on the sleep diary the following morning). The end of the rest interval (rise time) was defined by the time when the participant got out of bed to start his/her day (and when s/he was instructed to push the event marker again, and log this time in the sleep diary). Total sleep time for each night was automatically computed when rest intervals were defined.

A newly trained scorer first completed a training course with an expert actigraphy data manager (A. Wood). This included didactic readings, demonstrations of actigraphy data collection, data uploads, and data processing according to various scoring methods used in the current study derived from the published approaches as well as from developed or adapted scoring methods used over the course of previous studies. The newly trained rater was then provided with a training set of actigraphy records that were reviewed and discussed with the expert scorer when discrepancies or questions were identified. The scorer was certified by the expert manager after additional records were scored with a 90% efficiency or greater.

All actigraphy records were first inspected visually to identify and reject signal artifacts due to periods of missing data (i.e., when the actigraph was not worn but stored, before the device is given to participants, or immediately after receiving an actigraph from a participant), that often found at the beginning or/and at the end of actigraphy records. Each artifact-free record was then processed with each of the five scoring methods.

First, the sleep diary based method (Diary) was used on all records. The rest interval times provided by participants in the sleep diary were first manually entered by a trained technician to define the rest intervals of each record. When participants forgot to fill in the diary, rest interval times provided by the event marker were used. For the 127 nights collected from the 18 participants, substitution with the event marker to replace missing diary data occurred three times for bed times (0.02%), and five times (0.04%) for rise times.

Second, the event marker method (Marker) was applied to all 18 original records. With this method, the rest intervals were first entered manually using accurate times at which the event marker was pressed at the onset and offset of each rest interval. When participants forgot to press the event marker, or when the times provided by the event marker did not match the signal at visual inspection, start and end times provided by participants in their sleep diary were used to determine the start and end of the rest intervals. Event markers were available for 103 of the 127 nights for bed times, and for 109 of the 127 nights for rise times.

Third, we used the Algorithm scoring method, using options available in Actiware 5.0 Software applied to the records to automatically set major rest interval, and detect only one rest interval per day. The rest interval of each day was based upon the activity values, without using sleep diaries, event-markers, or corrections by the scorers.

Fourth, we used the Algorithm + Marker scoring method, using options available in Actiware 5.0 Software to automatically set major rest interval, to detect only one rest interval per day, and to use event marker to adjust rest interval start and end time. With this method, the longest rest interval of each day was first identified by the automatic algorithm based upon the activity values, supplemented by event markers. Specifically, if an event marker was identified within 30 minutes of the algorithm-identified rest interval onset or offset, the start or end times were adjusted to match the event marker time. If no event-marker was found within 30 minutes of the algorithm-defined start or end times of the rest interval, the algorithm defaulted to using the automatic algorithm only.

Finally, the visual inspection method (VI) was applied to all records. Human visual inspection of the actigraphy signal relied exclusively of the trained record reader to identify the start and end of all rest intervals. Only one technician quoted all the actigraphy records used in the study.

Each actigraphy record from 18 participants was processed five times, using each of the five scoring methods described above. Means for individual participants were calculated for each rest, sleep, and daily activity interval. Actigraphy parameters of interest were the rest

interval start time, rest interval duration (or total time spent in bed), total sleep time, sleep onset, , sleep latency (SL) , wake after sleep onset (WASO), fragmentation index, sleep efficiency (SE), and average activity count/epoch (or average activity levels per one-minute epoch of the active interval). Data were examined for normality and transformations were applied when necessary. Analyses of variance (ANOVAs) with repeated measures were used to compare the five methods on each of the parameters of interest. Data analysis was conducted using SAS v9.2 PROC MIXED.

Results

The mean age of the participants was 44.90 year old (SD = 7.73), and 61.11% (n=11) were women. The mean body mass index was 42.48 (SD = 8.16). The mean age of bipolar disorder onset was 21.39 (SD= 6.63). The mean Hamilton Depression Rating Scale total score was 4.55 (SD= 2.83), and the mean Young Mania Rating scale total score was 2.33 (SD= 2.80). As expected, all of the 18 participants but one were taking psychotropic medications at the beginning of the study.

Results of the comparison of the five actigraphy scoring methods are presented in Table 1. The Algorithm and the VI methods differed from the other methods for many actigraphy parameters of interest. The Algorithm method provided longer estimates for the rest interval duration and WASO compared to the Algorithm +Marker and the VI methods. The Algorithm method provided significantly earlier sleep onset times compared to the Algorithm + Marker method. The Algorithm method provided also longer estimates of total sleep time compared to the other four scoring methods. Finally, the Algorithm scoring method provided higher estimates of sleep fragmentation and of average activity count/epoch compared to the Algorithm + Marker scoring method. In addition, the VI scoring method provided significantly shorter estimates for sleep latency compared to all other methods. Estimates of sleep efficiency showed variability across the five scoring methods. The Diary and the Marker methods provided significantly lower parameter estimates than the Algorithm and the Algorithm + Marker scoring methods, which provided SE estimates significantly lower than the VI scoring method. After Bonferroni correction for multiple comparisons, all significant statistical differences remain significant except for the difference of sleep onset between the Algorithm method and the Algorithm + Marker method.

Discussion

The present findings highlight the notion that careful consideration should be given to the selection of actigraphy scoring methods. Differences observed with the VI method could be explained by the use of decreases in activity observed on the actigraphy signal to identify bed times, which can mask difficulties falling asleep by shortening sleep latency estimates. In turn, the latter may artificially impact sleep efficiency estimates. Differences observed with the Algorithm method could be explained as well by less accurate bed time detection in a population with low activity level, as it has been described in bipolar disorder. Consequently rest interval duration and total sleep time would be increased. However, the Algorithm method avoids memory bias, which can be encountered with the Diary method and data loss, which can happen with the Diary method and the Marker method when the participant forgets to fill in the diary or to push the marker. Even though the diary can help to define more accurate bed time in populations with low activity level or spending extended time in bed without sleeping, its use may be problematic in unmotivated, impaired, or illiterate groups. Reliance on information derived from sleep diaries (that may be missing or completed retrospectively), thus, can compromise the value of actigraphy as an objective sleep measurement method. Furthermore, the Marker, the Diary and Visual Inspection

methods reduce reliance of on actigraphy only as an independent source of sleep data. Finally the Diary method and the Marker method partially overlap in their time source. These two mixed methods are not fully independent, but they illustrate ecological way of actigraphy processing in a manner that limits data loss.

In large studies, the Algorithm method and its objective and automated scoring approach offers a complete absence of human interaction for scoring the actigraphy signal, and limits difficulties inherent to inter-scorer reliability, and personnel and time costs. Nevertheless, even with these automated scoring methods, some level of visual inspection remains necessary to capture and correct inaccurate interval detection or missing data. Visual inspection and correction, then, diminishes the objective quality of the Automatic method.

Of note, the time and effort necessary to each of the five methods also differ, with the Algorithm and Algorithm + Marker methods being empirically more rapid (around 15 minutes / record) and consuming less personnel and time resources. The Marker, Diary, and VI methods take more time to process, with at least 25 minutes necessary for each record.

Limitations to this preliminary study include the lack of comparison with PSG, the gold standard for sleep parameter measurement. The absence of concurrent PSG studies limits interpretation and prevents conclusions regarding the accuracy of the five scoring methods. Without PSG validation, this study cannot point out the most robust actigraphy method. Another limitation of the present study is the use of a single actigraphy device and Software (Actiwatch AW-64 and Actiware5 software; Phillips Respironics). These same five scoring methods may yield different observations with other devices. Moreover, our test sample included 18 remitted bipolar patients with high BMI and taking psychotropic medication. The results are limited to this population and may not apply to bipolar patients in manic or depression episodes, or in populations without psychiatric or sleep disorder. Replications are needed in larger and different samples to draw more definitive conclusions on the generalizability of the present findings to other samples, other devices, and other software.

In summary, the present findings showed differences in sleep estimates among the five actigraphy scoring methods. Without PSG validation, this study can only suggest that the selection of the actigraphy data processing would benefit from careful consideration of the research and/or clinical goals, primary actigraphy parameters of interests, preference regarding the source of time cues (from subjectively reported diary-based data to fully objective, automated algorithm-derived time estimates), and processing time and personnel costs.

Acknowledgments

The work was supported by the National Institute of Health MH080696, the Department of Defense Congressionally Directed Medical and Research Program (W81XWH-06-1-0257), INSERM (Institut National de la Santé et de la Recherche Médicale) and the Fondation FondaMental, French National Science Foundation. The authors are especially grateful to Dr. Jan Scott, MD, from the University of Newcastle, UK, for her insightful scientific and editorial comments on the manuscript.

References

- Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*. 2003; 26:342–392. [PubMed: 12749557]
- Axelsson DA, Bertocci MA, Lewin DS, Trubnick LS, Birmaher B, Williamson DE, Ryan ND, Dahl RE. Measuring mood and complex behavior in natural environments: use of ecological momentary assessment in pediatric affective disorders. *Journal of Children Adolescent Psychopharmacology*. 2003; 13:253–266.

- First, M.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV Axis I Disorders-Patient Edition (SCID-I/P). New York State Psychiatric Institute; New York: 1995. (Version 2.0 edition).
- Germain A, Kupfer DJ. Circadian rhythm disturbances in depression. *Human Psychopharmacology*. 2008; 23:571–585. [PubMed: 18680211]
- Harvey AG. Sleep and circadian rhythms in bipolar disorder: seeking synchrony, harmony, and regulation. *American Journal of Psychiatry*. 2008; 165:820–829. [PubMed: 18519522]
- Harvey AG, Schmidt DA, Scarna A, Semler CN, Goodwin GM. Sleep-related functioning in euthymic patients with bipolar disorder, patients with insomnia, and subjects without sleep problems. *American Journal of Psychiatry*. 2005; 162:50–57. [PubMed: 15625201]
- Jones SH, Hare DJ, Evershed K. Actigraphic assessment of circadian activity and sleep patterns in bipolar disorder. *Bipolar Disorders*. 2005; 7:176–186. [PubMed: 15762859]
- Millar A, Espie CA, Scott J. The sleep of remitted bipolar outpatients: a controlled naturalistic study using actigraphy. *Journal of Affective Disorders*. 2004; 80:145–153. [PubMed: 15207927]
- Monk TH, Reynolds CF, Kupfer DJ, Buysse DJ, Coble PA, Hayes AJ, Machen MA, Petrie SR, Ritenour AM. The Pittsburgh Sleep Diary. *Journal of Sleep Research*. 1994; 3:111–120.
- Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*. 1979; 134:382–389. [PubMed: 444788]
- Morgenthaler T, Alessi C, Friedman L, Owens J, Kapur V, Boehlecke B, Brown T, Chesson A Jr, Coleman J, Lee-Chiong T, Pancer J, Swick TJ. Standards of Practice Committee; American Academy of Sleep Medicine. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep*. 2007; 4:519–529. [PubMed: 17520797]
- Salvatore P, Ghidini S, Zita G, De PC, Lambertino S, Maggini C, Baldessarini RJ. Circadian activity rhythm abnormalities in ill and recovered bipolar I disorder patients. *Bipolar Disorders*. 2008; 10:256–265. [PubMed: 18271904]
- Samuels S, Katz IR, Parmelee PA, Boyce A, DiFilippo S. Use of the Hamilton and Montgomery-Asberg Depression Scales in institutionalized elderly patients. *American Journal of Geriatric Psychiatry*. 1996; 4:237–246.
- Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *British Journal of Psychiatry*. 1978; 133:429–435. [PubMed: 728692]

Table 1

Comparisons of five actigraphy scoring methods for 18 patients with bipolar disorder.

	Diary Mean (SD)	Marker Mean (SD)	Algorithm Mean (SD)	Algorithm + Marker Mean (SD)	Visual Inspection Mean (SD)	F statistics (df=4,17)	Corrected [‡] p value
Rest interval onset (clock time)	23:53 (1:38)	23:52 (1:41)	23:46 (2:37)	00:15 (2:23)	00:04 (1:57)	F=2.70, p=0.066	0.087
Rest interval duration * (minutes)	488.83 (83.92)	490.82 (78.79)	540.54 (105.32)	493.17 (95.93)	469.09 (73.22)	F=6.78, p=0.002 A+M, VI < A	0.011
Sleep onset (clock time)	00:11 (1:41)	00:05 (1:43)	23:57 (2:36)	00:26 (2:23)	00:06 (1:57)	F=3.10, p=0.044 A+M > A	0.087
Sleep latency [§] (minutes)	17.84 (12.63)	13.56 (9.56)	11.40 (5.43)	10.97 (6.80)	2.45 (3.25)	F=27.62, p<0.0001 A+M, A, D; M> VI	0.0009
Total sleep time * (minutes)	458.27 (80.42)	461.20 (78.64)	521.75 (106.05)	471.69 (95.51)	463.82 (73.60)	F=4.53, p=0.011 A > A+M, D, M, VI	0.034
Wake after sleep onset [‡] (minutes)	52.16 (25.84)	52.36 (26.04)	56.47 (24.14)	47.98 (26.71)	46.61 (23.03)	F=6.14, p=0.003 A+M, VI < A	0.015
Sleep efficiency (%)	83.52 (5.48)	83.46 (5.36)	86.05 (4.07)	86.09 (4.71)	89.03 (4.38)	F=25.27, p<0.0001 D, M < A+M, A < VI	0.0009
Sleep fragmentation *	30.98 (12.61)	31.03 (12.38)	31.43 (10.75)	29.16 (11.43)	29.55 (12.02)	F=5.21, p=0.006 A+M < A	0.025
Average activity count / epoch	239.72 (89.62)	243.77 (92.38)	248.05 (76.11)	241.95 (81.26)	242.76 (88.96)	F=7.72, p=0.001 AM<A; D < M	0.007

Means and standard deviations reported in original units. A = Algorithm; A+M = Algorithm + Marker; M+D = Marker + Diary; D+M = Diary + Marker; VI = Visual Inspection

[‡] Stepwise Bonferroni corrected p-values

[‡] Transformation used for data used for analysis: square root (x)

* Transformation used for data used for analysis: ln(x)

[§] Transformation used for data used for analysis: ln (x+1)