Journal of Affective Disorders 217 (2017) 299–305

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy

Jesse D. Cook*, Michael L. Prairie, David T. Plante

Department of Psychiatry, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA

ARTICLE INFO

Keywords:
Fitbit
Depression
Sleep
Actigraphy
Polysomnography

ABSTRACT

Background: Sleep disturbance is a common and important component of affective illness. Fitness activity trackers are emerging as alternative means to estimate sleep in psychiatric patients; however, their ability to quantify sleep in mood disorders has not been empirically evaluated. Thus, this study sought to evaluate the utility of the Fitbit Flex (FBF) to estimate sleep in patients with major depressive disorder (MDD) relative to gold standard polysomnography (PSG) and a widely-used actigraph (Actiwatch-2; AW-2).

Methods: Twenty-one patients with unipolar MDD wore the FBF and AW-2 during in-laboratory PSG. Bland-Altman analysis compared sleep variables among devices. Epoch-by-epoch analysis further evaluated sensitivity, specificity, and accuracy for the FBF and AW-2 relative to PSG.

Results: The FBF demonstrated significant limitations in quantifying sleep and wake, relative to PSG. In the normal setting, the FBF significantly overestimated sleep time and efficiency, and displayed poor ability to correctly identify wake epochs (i.e. low specificity). In the sensitive setting, the FBF significantly underestimated sleep time and efficiency relative to PSG. Performance characteristics of the FBF were more similar to the AW-2 in the normal compared to sensitive setting.

Limitations: Participants were young to middle aged and predominantly female, which may limit generalizability of findings. Study design also precluded ability to assess longitudinal performance of FBF.

Conclusions: The FBF is not an adequate substitute for PSG when quantifying sleep in MDD, and the settings of the device sizably impact its performance relative to PSG and other standard actigraphs. The limitations and capabilities of the FBF should be carefully considered prior to clinical and research implementation.

1. Introduction

Sleep disturbance is very common in patients with Major Depressive Disorder (MDD). It has been estimated that up to 90% of individuals with MDD experience reduced sleep quality during a depressive episode (Tsuno et al., 2005). Depression can be accompanied by a diverse range of sleep disturbances including insomnia (difficulty falling asleep, maintaining sleep, or waking up too early) and/or hypersomnolence (excessive daytime sleepiness and/or sleep duration (Soehner et al., 2014). Objective changes in sleep continuity and duration in MDD, as measured by polysomnography, are robust physiological indicators of sleep disturbance in the disorder (Benca et al., 1992; Steiger and Kimura, 2010; Pillai et al., 2011; Plante et al., 2017). Sleep disturbance is also associated with treatment resistance, symptomatic relapse, suicidality, and impaired daytime function, underscoring its importance in the course of affective illness (Baglioni et al., 2011; McCall et al., 2010; Nadorff et al., 2013; Perlis et al., 1997; Riemann and Voderholzer, 2003; Szklo-Coxe et al., 2010). Thus, the ability to quantify sleep duration and continuity in patients with MDD is of potentially high value in the assessment and treatment of patients with mood disorders.

Polysomnography (PSG) is considered the gold standard for objective sleep measurement, however its widespread applicability is limited by its time-intensiveness, high cost, and intrusiveness (Meltzer et al., 2015; Montgomery-Downs et al., 2012). Furthermore, PSG is typically unable to provide information on longitudinal sleep-wake patterns over a multiple night assessment period (Meltzer et al., 2015). The use of validated actigraphs that utilize wrist-worn accelerometry to quantify movement as a surrogate measure for sleep and wake can circumvent some of these shortcomings of PSG due to their relatively low-cost, non-intrusiveness, and ambulatory capabilities (de Souza et al., 2003; Montgomery-Downs et al., 2012). Although actigraphy has been validated in its ability to identify sleep/wake times and patterns in adult populations (Morgenthaler et al., 2007) actigraphic devices may tend to overestimate total sleep time (de Souza et al., 2003; Montgomery-Downs et al., 2012). This deficiency stems largely from

* Correspondence to: Wisconsin Psychiatric Institute and Clinics, 6001 Research Park Blvd., Madison, WI 53719, USA.
E-mail address: jdcook4@wisc.edu (J.D. Cook).

http://dx.doi.org/10.1016/j.jad.2017.04.030
Received 29 September 2016; Received in revised form 22 February 2017; Accepted 19 April 2017
Available online 19 April 2017

an inability to correctly identify wake epochs (Marino et al., 2013) as these devices rely upon an accelerometer (movement detector) as the sole measurement for sleep/wake designation, which inherently leaves them vulnerable to classifying periods of inactivity as sleep regardless of vigilance state. In addition, most validated actigraphs used in clinical and research settings generally require patients to return the device periodically (typically 2–4 weeks) in order for data to be retrieved for evaluation, making their use over more extended periods cumbersome.

The rise of commercially-available fitness activity trackers has provided another low-cost, field-based, and user-friendly alternative that may prove useful in evaluating sleep for both clinical and research purposes (de Zambotti et al., 2015; Meltzer et al., 2015; Montgomery-Downs et al., 2012). These mass-marketed devices are gaining a broader acceptance in both general and patient populations, and practitioners have begun to integrate their use in the assessment and treatment of affective disorders, despite limited research evaluating their use in patients with psychiatric disorders (Vahia and Sewell, 2016). Besides their low cost, these devices typically leverage direct-to-consumer cloud-based platforms and/or mobile technologies to allow for continuous data collection and retrieval over time. Considering the widespread availability of these devices and their potential impact on the management of psychiatric illness, comparison of their performance in estimating sleep against gold standard PSG and other validated actigraphs is a vital area of inquiry.

In a previous investigation conducted in healthy adults, a fitness activity tracker (the inaugural version of the Fitbit) overestimated total sleep time and congruently demonstrated an inability to correctly identify wake epochs when compared against a commonly used brand of actigraphy (Actiwatch 64) and PSG (Montgomery-Downs et al., 2012). Although not many validation studies have been performed on fitness activity trackers in the domain of sleep, this demonstration of an overestimation for total sleep time and inability to accurately identify wake epochs has been corroborated by studies on women with insomnia (de Zambotti et al., 2015) and adolescents referred for clinical PSG (Meltzer et al., 2015). Contrary to the results of these investigations, one study in healthy, young adults demonstrated comparable results in estimation of total sleep time for multiple fitness activity trackers, relative to PSG (Mantua et al., 2016). However, because epoch-by-epoch comparisons were not performed, the full performance characteristics of the fitness trackers utilized in this study, relative to PSG, could not be determined. The inconsistent results of existing validation studies - particularly in regards to the estimation of total sleep time and identification of wake periods - suggests a need to further investigate the true capabilities of these devices, with an emphasis on elucidating their utility within specific disorders. To our knowledge, no prior research has evaluated the validity of a commercially-available fitness activity tracker in persons with affective illness.

Thus, to further extend this line of inquiry into patients with affective illness, the primary aim of this investigation was to examine the utility of a commercially-available fitness activity tracking device, the Fitbit Flex (FBF, Fitbit Inc.; San Francisco, CA), against both PSG and validated actigraphy, the Actiwatch 2 (AW-2; Phillips Respironics), in a well-characterized cohort of adult patients with MDD.

2. Methods

2.1. Participants, inclusion/exclusion criteria, and study design

A convenience sample of twenty-one, right-handed unmedicated patients with unipolar Major Depressive Disorder (MDD) was recruited as part of a larger study investigating electroencephalographic biomarkers of sleep disturbance in neuropsychiatric disorders. After an initial phone screening, participants completed an in-person medical, sleep, and psychiatric evaluation that included the Structured Clinical Interview for DSM-IV (SCID) (First et al., 2002), semi-structured sleep disorders evaluation, and physical exam, performed by a physician board certified in psychiatry and sleep medicine (DTP). Exclusion criteria included the following: smoking of greater than 15 cigarettes per day; > 3 caffeinated beverages per day; significant sleep, neurologic, or medical disorder; history of significant head trauma or loss of consciousness > 30 min; and imminent risk of self-harm or suicide. Women who were pregnant, breastfeeding, < 6 months post-partum, or planning to become pregnant during the study were also excluded. Participants were also excluded if they met DSM-IV criteria for alcohol or substance abuse/dependence within the preceding 6 months. Additionally, if patients met criteria for other Axis I psychiatric disorders, MDD had to be considered the primary disorder for study inclusion. Participants completed additional self-report instruments including the Beck Depression Inventory (BDI-II) (Beck et al., 1996), Pittsburgh Sleep Quality Index (PSQI) (Buysse et al., 1989), and Insomnia Severity Index (ISI) (Bastien et al., 2001). Eligible participants were then scheduled for an in-laboratory PSG at least one week but no more than one month after their in-person screening visit. All participants provided informed consent and were instructed to maintain their usual sleep-wake schedules for the duration of their time in the study. This study was approved by the Institutional Review Board of the University of Wisconsin-Madison.

2.2. In laboratory overnight visit procedures

Participants arrived at approximately 18:00 on the night of their PSG for set-up, at which point, a wrist-worn Actiwatch 2 (AW-2) and Fitbit Flex (FBF) were both placed adjacently on the participant's non-dominant (left) wrist. Polysomnographic data were collected using an integrated recording system that utilized a 256-channel EEG net (Electrical Geodesics, Eugene, OR) along with other standard recording sensors including electrooculogram (EOG), sub-mental electromyogram (EMG), electrocardiogram (ECG, bilateral tibial EMG, respiratory inductance plethysmography, pulse oximetry, and a positition sensor (Alice® Sleepware; Phillips Respironics, Murrysville, PA). A registered sleep technologist, blind to the FBF and AW-2 staging output, staged all sleep recordings using 30-second epochs according to standard criteria based on 6 EEG channels at approximate 10–20 locations (F3, F4, C3, C4, O1, and O2) referenced to the mastoids, electrooculogram, and sub-mental electromyogram according to American Academy of Sleep Medicine criteria (Berry et al., 2014). Bedtimes were tailored to each participant's habitual sleep pattern, with lights-off (participant actively trying to fall asleep) occurring between approximately 22:00 and 23:00. Participants were allowed to sleep ad-libitum, remaining undisturbed throughout the night and not awoken at a prescribed time the following morning. Lights on was determined based on the participant's stated desire to terminate the nocturnal sleep period upon awakening. Polysomnography and accelerometer data were collected within a local network of computers time synchronized to an external atomic clock through frequent restart.

2.3. Data analysis

PSG was considered the gold standard measure of sleep duration and continuity. PSG lights-off and lights-on times were used as the start and end points for the AW-2 and FBF rest periods to maintain consistency (Meltzer et al., 2015). The following sleep variables were calculated for PSG, FBF, and AW-2: total sleep time (TST; total duration of sleep during period of time in bed), sleep onset latency (SOL; time from lights-off to the first epoch of sleep), wake after sleep onset (WASO; total duration of wake time after sleep onset), and sleep efficiency (SE; equal to TST divided by total time in bed). AW-2 data were analyzed utilizing the medium threshold (value = 40) with five minute immobility time for sleep onset/offset since this setting has been shown to produce the most accurate output, relative to PSG (Chae et al., 2009). FBF data were analyzed using both the normal and sensitive settings, since prior studies in pediatric sleep apnea have suggested the