Journal of Psychosomatic Research 97 (2017) 38–44

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

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

Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers

Seung-Gul Kang[a,*], Jae Myeong Kang[a], Kwang-Pil Ko[b], Seon-Cheol Park[c], Sara Mariani[d], Jia Weng[d]

[a] Department of Psychiatry, Gil Medical Center, Gachon University, School of Medicine, Incheon, Republic of Korea
[b] Department of Preventive Medicine, Gachon University School of Medicine, Incheon, Republic of Korea
[c] Department of Psychiatry, Inje University College of Medicine and Haeundae Paik Hospital, Busan, Republic of Korea
[d] Division of Sleep & Circadian Disorders, Brigham & Women's Hospital, Harvard Medical School, USA

ARTICLE INFO

Keywords:
Fitbit Flex
Polysomnography
Insomnia
Sleep tracker
Validation

ABSTRACT

Objectives: To compare the accuracy of the commercial Fitbit Flex device (FF) with polysomnography (PSG; the gold-standard method) in insomnia disorder patients and good sleepers.
Methods: Participants wore an FF and actigraph while undergoing overnight PSG. Primary outcomes were intraclass correlation coefficients (ICCs) of the total sleep time (TST) and sleep efficiency (SE), and the frequency of clinically acceptable agreement between the FF in normal mode (FFN) and PSG. The sensitivity, specificity, and accuracy of detecting sleep epochs were compared among FFN, actigraphy, and PSG.
Results: The ICCs of the TST between FFN and PSG in the insomnia (ICC = 0.886) and good-sleepers (ICC = 0.974) groups were excellent, but the ICC of SE was only fair in both groups. The TST and SE were overestimated for FFN by 6.5 min and 1.75%, respectively, in good sleepers, and by 32.9 min and 7.9% in the insomnia group with respect to PSG. The frequency of acceptable agreement of FFN and PSG was significantly lower ($p = 0.006$) for the insomnia group (39.4%) than for the good-sleepers group (82.4%). The sensitivity and accuracy of FFN in an epoch-by-epoch comparison with PSG was good and comparable to those of actigraphy, but the specificity was poor in both groups.
Conclusions: The ICC of TST in the FFN–PSG comparison was excellent in both groups, and the frequency of agreement was high in good sleepers but significantly lower in insomnia patients. These limitations need to be considered when applying commercial sleep trackers for clinical and research purposes in insomnia.

1. Introduction

Insomnia is one of the most common distressing and clinically important sleep disturbances. The Majority of people with insomnia disorder are interested in their sleep duration and are required to keep a daily sleep diary and assess their subjective total sleep time (TST) and sleep efficiency (SE) if they undergo cognitive behavior therapy for insomnia (CBT-I). However, it has been shown that these patients tend to underestimate their sleep duration [1].

The above-described situation has prompted great interest among both clinicians and insomnia patients in how to obtain accurate sleep measurements. The recognized gold standard for this purpose is polysomnography (PSG). However, PSG is time-consuming and expensive for diagnostic and treatment purposes on a routine basis. Since PSG is typically only applied for one or two nights, it may not accurately reflect the typical sleep states and patterns. To overcome these limitations, actigraphy is frequently used [2]. However, actigraphs are difficult for general consumers to use because they are not only costly but also require the use of specialized software, the ability to export sleep data, and sufficient experience to interpret the sleep information.

Commercial sleep trackers have become increasingly popular among general users since they are extremely user-friendly, provide immediate information about sleep and activity via wireless synchronization to smartphones, and come in attractive designs at affordable prices. These merits of commercial sleep trackers have attracted proactive attention from sleep-disorder patients as well as healthy individuals.

Despite such rapidly growing interest, the accuracy of using the Fitbit Flex device (FF; Fitbit Inc., San Francisco, CA, USA)—which is

* Corresponding author at: Department of Psychiatry, Gil Medical Center, Gachon University, School of Medicine, 21, Namdong-daero 774 beon-gil, Namdong-gu, Incheon 21565, Republic of Korea.
E-mail address: kangsg@gachon.ac.kr (S.-G. Kang).

http://dx.doi.org/10.1016/j.jpsychores.2017.03.009
Received 1 December 2016; Received in revised form 21 March 2017; Accepted 21 March 2017

one of the most popular commercial sleep trackers—and similar wearable devices for sleep measurements in patients with sleep disorders has not been verified. Previous validation studies of commercial sleep trackers compared the first Fitbit device (the Fitbit Classic) [3], the FF [4], the Fitbit Ultra [5], and Jawbone UP [6,7] with the PSG. The results obtained in these studies varied with the trackers used, the age and clinical state of the subjects, the research method, and the type of statistical analysis. It has been argued that validation studies conducted on clinical populations with sleep disorders are needed [3]. To the best of our knowledge, most previous studies have involved healthy subjects, with none of them focusing on insomnia disorder patients.

The aims of this study were (i) to determine the validity of the FF compared to PSG in insomnia disorder patients and good sleepers using intraclass correlation coefficients (ICCs) calculation, and measure the frequency of clinically acceptable agreement in the two groups, defined as differences of ≤ 30 min and ≤ 5% between the methods for TST and SE, respectively; (ii) to compare sleep measurements (i.e., TST, SE, sleep onset latency [SOL], and wake after sleep onset [WASO]) among the FF, actigraphy, and PSG; and (iii) to perform an epoch-by-epoch validation of the FF and actigraphy—in terms of sensitivity, specificity, and accuracy—using PSG-based sleep scoring as the gold standard.

2. Methods

2.1. Participant enrollment

All the insomnia and healthy subjects were recruited from the Gil Medical Center in Incheon, South Korea by posting a notice on the bulletin board at Gil Medical Center, an online advertisement on the Internet, and by referral or advice to participate in the study from physicians.

We applied the following criteria in recruiting insomnia disorder participants: (i) aged 18–60 years, (ii) history of illness lasting at least 3 months and meeting the diagnostic criteria for insomnia disorder in the Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders [8], (iii) a Pittsburgh Sleep Quality Index (PSQI) total score of ≥ 8, (iv) an apnea-hypopnea index (AHI) of < 15 in PSG, and (v) not having taken any hypnotics or psychotropic medication or been treated with CBT-I during the previous 2 weeks.

The following inclusion criteria were applied in recruiting good sleepers: (i) aged 18–60 years, (ii) no symptoms or history of sleep disorders including insomnia, (iii) PSQI total score of ≤ 4 at screening, (iv) no evidence of any sleep disorder in PSG, and (v) having never taken any hypnotics or psychotropic medication during their lifetime.

The following common exclusion criteria were applied for both the insomnia and good-sleepers groups: (i) suspected of having a major sleep disorder other than insomnia disorder based on the medical history or the score on a screening scale; (ii) shift workers or travelers experiencing frequent jet lag; (iii) high risk of sleep apnea according to the Berlin Sleep Questionnaire or moderate-to-severe diurnal sleepiness as indicated by a score of ≥ 13 on the Epworth Sleepiness Scale (ESS) [9,10]; (iv) ≥ 15 periodic limb movements index during sleep in PSG; (v) diagnosed with other major psychiatric disorders in a clinical interview; (vi) serious medical or neurological conditions that could have affected participation in this study; or (vii) inability to wear an FF or an actigraph on the nondominant wrist due to physical handicap or other reason.

The subjects were screened through telephone interviews and an Internet website featuring screening scales that were constructed for this study. After the initial screening, board-certified psychiatrists specializing in sleep medicine evaluated the eligibility of each participant in a face-to-face interview. This interview examined potential symptoms and signs of insomnia disorder according to diagnostic criteria and the medical history, and included a semistructured interview for assessing sleep and psychiatric disorders. We obtained written informed consents from all of the included patients, and the institutional review board of Gil Medical Center approved this study.

2.2. Clinical questionnaire at screening and the comparison test

All participants completed the questionnaire regarding their demographic information, medical and psychiatric illnesses, sleep information including the frequency and consequences of insomnia, sleep duration and sleep–wake schedule, consumption of caffeine and alcohol, and smoking. The patients completed the Korean version of PSQI [11], ESS [12], and Berlin Questionnaire [13] at screening, and then completed the PSQI on the evening of the test day.

2.3. PSG, the FF, and actigraphy

Overnight PSG was performed using a portable PSG device (Embletta X100, Embla, Broomfield, CO, USA). This device was used for eight-channel in-home unattended PSG measuring the electroencephalogram (EEG), electrooculogram, electromyogram, nasal pressure, thoracic and abdominal effort, oximetry, and body position. A certified PSG technician attached the device and sensors to the participant, and the overnight recordings were performed unattended. All participants also wore an FF and an actigraph on their nondominant wrist. The placement of the actigraph and FF in relation to the wrist was randomly assigned. The unblinded simple randomization was performed using a Microsoft Excel spreadsheet. Both devices were placed on and removed from the wrist of each participant by the sleep technician. Prior to the test, all of the devices and computers used were time-synchronized. The data collected from the three devices between the lights-off and lights-on times of PSG were compared and analyzed. The outcome measures of interest in this study were TST, SE (TST/minutes between lights off and lights on), SOL (length of time [minutes] from lights off to sleep), and WASO (quantified in minutes being awake after sleep onset) as measured by PSG, an FF, and actigraphy.

A certified PSG technologist analyzed the PSG data. Manual scoring was performed using a scoring platform (RemLogic, version 3.4.0, Embla Systems, Kanata, ON, Canada) in accordance with the AASM manual (version 2.0.2) [14], and hypopnea during sleep was scored using the acceptable rules in that manual. Sleep stages and events were scored visually by a certified PSG technologist based on the AASM criteria, and all PSG data was confirmed by a sleep-specialist medical doctor. The scorers completed the interscorer reliability program of the AASM (http://www.aasmnet.org/isr/) before starting the study. Their mean reliability score was 94.5% for the diagnostic test.

The FF is a commercially available activity and sleep-tracking wristband that tracks the intensity and frequency of movements utilizing a triaxial accelerometer. Detailed information about the FF can be found on the manufacturer's website (help.fitbit.com). The FF functions include monitoring steps, distance, active minutes, and sleep measurements including TST, sleep latency, SE, and the number of awakenings. There are two kinds of sleep-detecting modes in the FF, named ‘normal’ (FFN) and ‘sensitive’ (FFS). The manufacturer of Fitbit explains that FFN counts significant movements during sleep as wake periods and is appropriate for most users, while FFS causes the tracker to record nearly all movements as awake periods. The manufacturer considers that FFS may be helpful for users with sleep disorders or for those who wear the tracker somewhere other than on the wrist while asleep. Data from the FF were uploaded to the users' online account, where the sleep measures were generated. The sleep raw data of the FF were exported after the time in bed was set as being equal to that of PSG according to the manufacturer's website. We were able to download minute-by-minute measures of each participant's sleep state during each night: categorized as ‘Sleep’, ‘Restless’ or ‘Wake’ (the restless and wake states were combined and are considered as being awake in the analyses according to the Fitbit's algorithm). The sleep metrics (TST, SE, SOL, and WASO) were calculated from the raw data of the FF. In order