SCIENTIFIC INVESTIGATIONS

Validation of Fitbit Charge 2 and Fitbit Alta HR Against Polysomnography for Assessing Sleep in Adults With Obstructive Sleep Apnea

Fernando Moreno-Pino[1,2,*]; Alejandro Porras-Segovia[3,4,*]; Pilar López-Esteban[5]; Antonio Artés[1,2]; Enrique Baca-García[3,6,7,8,9,10]

[1]Universidad Carlos III de Madrid, Leganés, Spain; [2]Gregorio Marañón Health Research Institute, Madrid, Spain; [3]Department of Psychiatry, University Hospital Jimenez Diaz Foundation, Madrid, Spain; [4]International School for Postgraduate Studies, University of Granada, Granada, Spain; [5]Neurophysiology Service, University Hospital HLA Moncloa, Madrid, Spain; [6]Department of Psychiatry, Madrid Autonomous University, Madrid, Spain; [7]Department of Psychiatry, University Hospital Rey Juan Carlos, Mostoles, Spain; [8]Department of Psychiatry, General Hospital of Villalba, Madrid, Spain; [9]Department of Psychiatry, University Hospital Infanta Elena, Valdemoro, Spain; [10]Universidad Catolica del Maule, Talca, Chile; *Contributed equally

Study Objectives: Consumer wearable devices may be a helpful method of assessing sleep, but validation is required for their use in clinical practice. Our aim was to validate two models of Fitbit sleep trackers that rely on both accelerometer and heart rate sensors against polysomnography in participants with obstructive sleep apnea (OSA).

Methods: Participants were adults presenting with symptoms of OSA and attending our outpatient sleep clinic. A polysomnography (PSG) was applied to all participants at the same time they were wearing a Fitbit sleep tracker. Using paired *t* tests and Bland-Altman plots, we compared the sleep measures provided by the wearable devices with those obtained by PSG. Since Fitbit devices' automatic detection of sleep start time can cause bias, we performed a correction using Huber loss function-based linear regression and a leave-one-out strategy.

Results: Our sample consisted of 65 patients. Diagnosis of OSA was confirmed on 55 (84.6%). There were statistically significant differences between PSG and Fitbit measures for all sleep outcomes but rapid eye movement sleep. Fitbit devices overestimated total sleep time, and underestimated wake after sleep onset and sleep onset latency. After correction of bias, Fitbit-delivered measures of sleep onset latency did not significantly differ of those provided by PSG.

Conclusions: Fitbit wearable devices showed an acceptable sensitivity but poor specificity. Consumer sleep trackers still have insufficient accuracy for clinical settings, especially in clinical populations. Solving technical issues and optimizing clinically-oriented features could make them apt for their use in clinical practice in a nondistant future.

Keywords: sleep apnea, sleep disorders, wearable technology, e-health

Citation: Moreno-Pino F, Porras-Segovia A, López-Esteban P, Artés A, Baca-García E. Validation of Fitbit Charge 2 and Fitbit Alta HR against polysomnography for assessing sleep in adults with obstructive sleep apnea. *J Clin Sleep Med.* 2019;15(11):1645–1653.

BRIEF SUMMARY

Current Knowledge/Study Rationale: Electronic wearable devices may be useful sleep assessment tools, due to their objectivity, availability to the general population and capacity to monitor sleep continuously. However, commercial wearable devices require validation in both healthy and clinical population. Here we explore the validity of two wearable devices against polysomnography in 65 patients with suspected obstructive sleep apnea (OSA).

Study Impact: This is the first study that validates Fitbit devices in patients with OSA, and the largest one performed in this population. It is also the first time Fitbit Alta HR is validated in a clinical sample. We found that this technology is still not accurate enough for clinical practice. However, the potential of these devices calls for further research, which should involve the collaboration between both developers and health care providers.

INTRODUCTION

Sleep disorders are highly prevalent in the general population and they are associated with impaired quality of life and increased mortality.[1–3] Additionally, several forms of disturbed sleep, such as insomnia, nightmares or sleep apnea, have been consistently associated with mental disorders and increased risk of suicide.[4–6]

Traditional sleep assessment methods have several limitations. Polysomnography (PSG) is considered the gold standard, as it allows for an objective and accurate measurement of sleep.[7] However, PSG is an expensive technique, which requires medical assistance for its application and needs to be carried out in a hospital or clinic, thus altering the usual sleep environment. This also means that it cannot be implemented for more than a few nights in a row, typically just one or two.[8]

On the other side of the spectrum, we find sleep questionnaires and diaries, which are inexpensive and preserve the natural sleep environment, but present with issues such as subjectivity, recall bias and high dependence on adherence.[8]

Consumer wearable devices may be a valuable addition to sleep assessment tools. Wearable devices, sometimes referred to as sleep trackers, are multisensory devices equipped

with an accelerometer and a heart monitor. They are a type of actigraphs in the sense that they record and graphically represent human activity over time. Different possible algorithms can be used to interpret data collected and provide information about sleep function. These devices aim to be objective in their measurement, and they have shown higher adherence than questionnaires since they require less effort from participants.[8] Disregarding the initial inversion of the purchase—which is rapidly decreasing as the market grows—they are much less expensive than PSG and allow for a day-to-day collection of information on a more valid ecological setting.[8–10]

Despite their potential applications, wearable devices are still commercial products, and they have been developed by nonmedical personnel. Their interpretation varies widely depending on the algorithm used, and they present with precision issues. The clinical use of wearable has been discouraged by some studies, as it was considered that they did not reach acceptable psychometric properties yet. The American Academy of Sleep Medicine (AASM), in a recent position statement about consumer sleep technology, highlights the need of validation of these devices before they can be used in clinical settings.[10]

Most previous validation studies of wearable devices are based on healthy populations.[11–15] Since assessment may be less accurate in the presence of sleep disorders and other conditions,[11,12] it is necessary to validate the different wearable devices in clinical populations before we can apply them for health care purposes. There are only a few studies that have included people with sleep disorders in their sample, mostly in pediatric populations.[16–18] Only recently, the first validation study of wearables devices in adults with sleep apnea has appeared in scientific literature.[19] In this study, Jawbone (Jawbone, San Francisco, California, United States) and Withings (Withings, Issy-les-Moulineaux, France) sleep trackers were compared with PSG in 36 patients with OSA symptoms. Additionally, a recent review exploring the use of electronical devices for the diagnosis of sleep-disordered breathing showed that most studies had quality issues, such as not reporting the sociodemographic characteristics of their samples.[20]

Most devices that have been studied employ only acceleration (actigraphy) sensors to determine the sleep status, with the notable exceptions of the Fitbit Charge 2 (Fitbit, Inc., San Francisco, California, United States)[15] and the Withings Pulse O2.[19] Both the Fitbit Charge 2 and the Fitbit Alta HR can measure heart rate using photoplethysmogram (PPG) sensors. The combination of both measurement tools drastically improves precision. Thus, the aim of this study was to validate two models of Fitbit wearable devices—Fitbit Charge 2 and the Fitbit Alta HR—against PSG for the measurement of sleep in participants with OSA.

Table 1—Characteristics of the sample.

	Total Sample (n = 65)	Charge 2 (n = 31)	Alta HR (n = 34)
Sex, n (%)			
Male	42 (64.6)	18 (58.1)	24 (70.6)
Female	23 (35.4)	13 (41.9)	10 (29.4)
Age, years, mean (SD)	58.84 (13.84)	56.87 (17.50)	60.71 (9.26)
BMI, kg/m^2, mean (SD)	29.60 (4.70)	28.27 (4.77)	31.10 (4.23)
Diagnosis, n (%)			
Normal	7 (10.8)	5 (16.1)	2 (5.9)
New OSA	42 (64.6)	20 (64.5)	22 (64.7)
Known OSA (CPAP titration)	13 (20.0)	4 (12.9)	9 (26.5)
Other	3 (4.6) *	2 (6.4)	1 (2.9)
ESS score, mean (SD)	10.05 (4.28)	9.88 (4.62)	11.00 (3.82)
AHI, events/h, mean (SD)	26.06 (21.72)	27.52 (21.56)	24.97 (22.12)
Level of accuracy, n (%)			
3-stage mode	31 (47.7)	14 (45.2)	17 (50.0)
4-stage mode	34 (52.3)	17 (54.8)	17 (50.0)

* 1 epilepsy, 1 restless leg syndrome, and 1 inconclusive. AHI = apnea-hypopnea index, BMI = body max index, CPAP = continuous positive airway pressure, ESS = Epworth Sleepiness Scale, OSA = obstructive sleep apnea, SD = standard deviation.

Table 2—PSG and Fitbit sleep outcomes and difference of means.

	PSG Mean (SD)	Fitbit Mean (SD)	Mean Diff.	SD	95% CI Lower	95% CI Upper	t	P
TST	313.63 (64.87)	373.42 (94.25)	−59.78	99.92	−84.54	−35.02	−4.824	< .001
WASO	73.72 (48.38)	37.58 (17.38)	36.14	48.60	24.09	48.18	5.995	< .001
WASO-corrected	73.72 (48.38)	633.74 (336.96)	−560.02	338.46	−643.89	−476.15	−13.340	< .001
SOL	27.56 (24.56)	4.35 (8.23)	23.22	24.47	17.15	29.28	7.648	< .001
SOL-corrected	27.56 (24.56)	157.80 (24.56)	−130.23	721.60	−309.29	48.81	−1.453	.151
Light sleep *	146.94 (60.11)	215.76 (45.61)	−68.82	76.77	−95.61	−42.04	−5.227	< .001
Deep sleep *	113.03 (48.47)	38.79 (23.08)	74.24	46.00	58.18	90.29	9.409	< .001
REM sleep *	51.37 (28.22)	54.15 (19.51)	−2.78	33.32	−14.41	8.85	−0.486	.630
Awake	73.24 (54.96)	31.31 (18.73)	41.93	55.92	28.07	55.79	6.045	< .001

* Only when working on 4-stage mode (n = 34). PSG = polysomnography, REM = rapid eye movement, SD = standard deviation, SOL = sleep onset latency, TST = total sleep time, WASO = wake after sleep onset.

Table 3—Sensitivity and specificity of Fitbit wearable devices.

	Full Sample (n = 65)	Normal PSG (n = 7)	OSA (n = 55)	OSA and CPAP (n = 13)	3-Stage Mode (n = 31)	4-Stage Mode (n = 34)	Fitbit Charge 2 (n = 31)	Fitbit Alta HR (n = 34)
Sensitivity	87.81 (79.87–95.76)	89.28 (82.69–95.87)	87.61 (85.19–90.03)	88.20 (81.78–94.62)	82.14 (68.66–95.62)	92.97 (84.38–101.56)	87.42 (75.75–99.09)	88.10 (77.22–98.98)
Specificity	43.85 (31.79–55.91)	39.53 (15.05–64.01)	44.77 (37.96–51.58)	55.08 (39.94–70.22)	34.44 (17.71–51.17)	52.44 (35.65–69.23)	35.00 (18.21–51.79)	51.90 (35.11–68.69)

Data presented as % (95% CI). CPAP = continuous positive airway pressure, OSA = obstructive sleep apnea, PSG = polysomnography.

METHODS

Setting and Design

This study was performed in agreement with the Declaration of Helsinki ethic requirements,[21] and approved by the Research Ethic Committee of the University Hospital Jimenez Diaz Foundation. All participants gave their written, informed consent to participate in the study. The Standards for the Reporting of Diagnostic accuracy (STARD) guidelines were followed where applicable.[22]

Sample

Adult patients that had been referred to our outpatient sleep clinic at the University Hospital HLA Moncloa in Madrid, Spain, were approached to participate in the study. Inclusion criteria were: being aged older than 18, presenting with known OSA or symptoms of sleep apnea assessed by a clinician, and providing informed consent to participate in the study. Exclusion criteria were: not being able to understand the inform consent and therefore to agree to it, presence of dementia or severe cognitive impairments, and being outside the age range. Sample recruitment and data collection took place between 2017 and 2018 and lasted 6 months.

PSG Measures

PSG was the gold standard against which we compared the wearable devices. Standard overnight video-EEG-PSG recordings were obtained with a computerized sleep recorder (Easy [Cadwell, Kennewick, Washington, United States] or Xltek Brain Monitor [Natus Medical Inc., Pleasanton, California, United States]). Patients underwent the following recordings: a 5-channel EEG with bipolar montages following the International 10–20 system to place electrodes in standardized scalp locations, an electrooculogram (EOG) (2 channel), a submental electromyogram (EMG), and an electrocardiogram. Thoracic and abdominal movements were recorded by inductance plethysmography, and airflow pressure by an oronasal cannula. Oxygen saturation was recorded continuously from a transcutaneous sensor (oximetry). And bilateral tibialis anterior electromyogram were also recorded. The results were later scored by a trained neurophysiologist. PSG registered five stages of sleep: awake, rapid eye movement (REM) sleep, stage N1 sleep, stage N2 sleep, and stage N3 sleep. Epochs of 30 seconds were used for scoring, following AASM guidelines.[23]

Fitbit Measures

Participants wore a Fitbit sleep tracking device in their wrist at the same time as the PSG was being performed.

Two models of Fitbit wristbands were used: the Fitbit Charge 2, released in 2016, and the Fitbit Alta HR, released in 2017.[24,25] Both devices are equipped with a three-axis accelerometer and a heart-rate monitor. They have similar operative systems and are expected to provide comparable results when used under the same conditions. The decision of using either one

Figure 1—Bland-Altman plot for total sleep time.
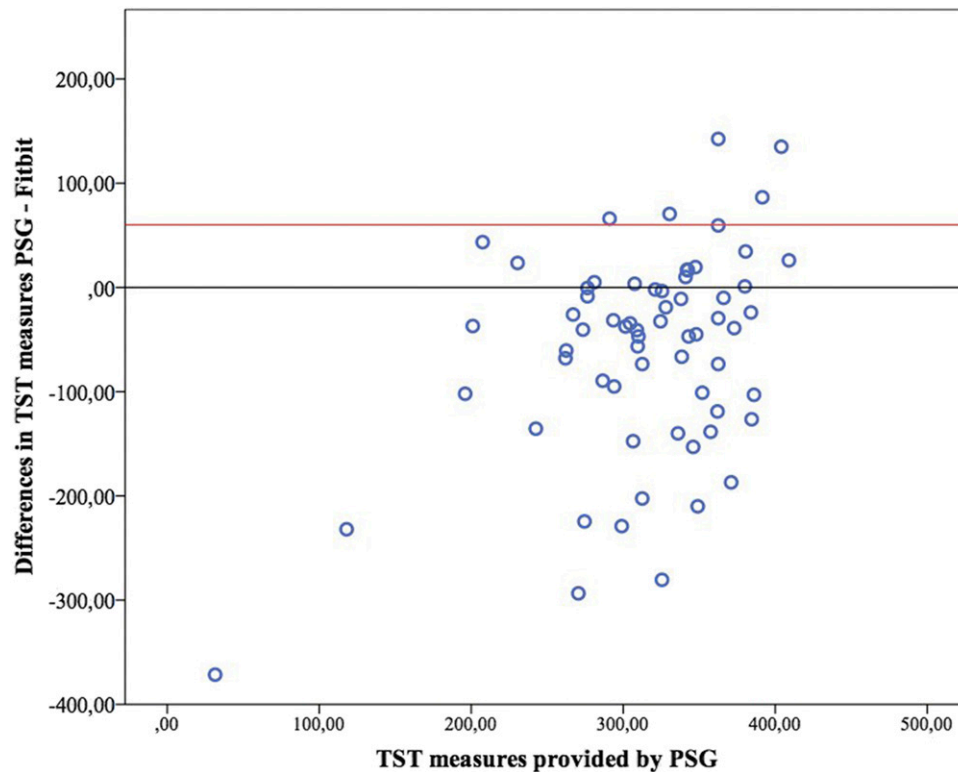
PSG = polysomnography, TST = total sleep time.

of the two models was made aleatorily and blindly by the engineers prior to meeting the participant.

Under ideal conditions, the Fitbit devices are able to record four stages of sleep: awake, light sleep, deep sleep, and REM sleep. Light sleep is referred to by the provider as the first stage of sleep, followed by deep sleep and REM sleep. In order to make comparisons between PSG and Fitbit devices we decided to equate light sleep with N1 and N2 stages measured with the PSG, while we considered deep sleep as tantamount to N3. These decisions were made after consulting with a neurophysiologist.

When a 4-stage level of precision is not reachable, the device simplifies measurement to three stages: awake, asleep, and an intermediate/indeterminate state labeled restless. Because the definition of restless provided by the developer was ambiguous, we decided to consider restless as a missing value, and only took into consideration the binary variable awake/asleep while the device was running on 3-stage mode.

As with the PSG, 30-second epochs were used. PSG and Fitbit devices were started at the same time. Engineers supervised the correct functioning of the Fitbit devices. Data were collected using the Fitbit application installed in a smartphone that was synced with the device.

All data were anonymized and uploaded to a secure web server. The algorithm used to interpret the results was the default one provided by the developers. This algorithm is not publicly available, nor it is possible to access raw data generated by the devices.

Outcomes

We obtained the following sleep outcomes: total sleep time (TST), sleep onset latency (SOL), and wake after sleep onset (WASO), and time spent in the four stages of sleep considered after grouping N1 and N2 into light sleep: awake, light sleep, deep sleep and REM sleep. Sensitivity was defined as the proportion of asleep values correctly identified as such by the wearable device over the total asleep PSG values. Specificity was the proportion of awake values correctly detected by the device. Both variables were expressed as percentages. Additionally, sociodemographic and clinical variables were collected from the patient's clinical records.
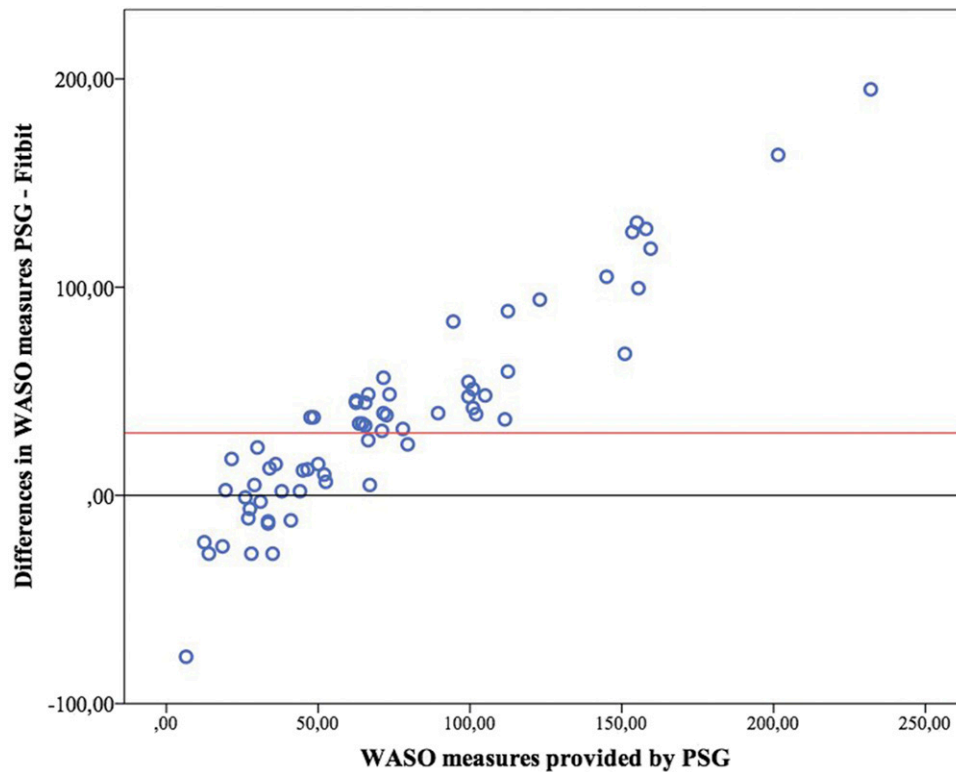
Statistical Analysis

For each sleep measure, we calculated the mean and standard deviation of each sleep measure provided by the PSG and the Fitbit devices. Differences between assessment tools were estimated using paired t tests and Bland-Altman plots.

Sensitivity and specificity were calculated with cross-tabulation, using chi square tests to determine statistical significance. We performed a subgroup analysis to test if there were any differences between the Fitbit Charge 2 and the Fitbit Alta HR models. We also explored whether there were differences between 3-stage and 4-stage measurement modes.

Statistical significance was stablished at $P < .05$, with two sided tests, and 95% confidence intervals. We used the SPSS

Figure 2—Bland-Altman plot for wake after sleep onset.

PSG = polysomnography, WASO = wake after sleep onset.

version 24.0 software (IBM Corp., Armonk, New York, United States) to perform the statistical analyses.

Bias Correction Method

Unlike PSG, both Fitbit Charge 2 and Fitbit Alta HR automatically detect when the person falls sleep, which can be a source of bias, especially when measuring SOL. To correct this bias we performed a "Leave-One-Out" strategy[26] using the formula:

$$SOL_{\text{Fitbit corrected}} = \propto SOL_{\text{Fitbit}} + \beta$$

Where, for each patient, the coefficients α and β are calculated by minimizing the regression error ($SOL_{\text{Fitbit corrected}} - SOL_{\text{Fitbit}}$) in the data from the rest of the patients. Instead the traditional square error loss function, we employ the Huber loss, less sensitive to outliers.[27,28]

RESULTS

Baseline characteristics of the sample are presented in **Table 1**. Our sample consisted of 65 participants aged 22–85 years. There were 42 (64.6%) men and 23 (35.4%) women. With the information provided by the PSG, clinical diagnosis of OSA, in conformity with DSM-5 criteria,[29] was reached in 42 (64.6%) participants. 13 participants (20.0%) had a known OSA at baseline and attended our clinic for

CPAP titration, which adds up to a total 55 participants with OSA (84.6%).

Fitbit Charge 2 was used in 31 participants, while Fitbit Alta HR was used in 34. Three-stage mode was on during 31 (47.7%) of the tests, while 4-stage mode was on during 34 (52.3%). The indeterminate restless stage was registered in 6.5% of the epochs while on 3-stage mode. Fitbit Charge 2 provided a 4-stage measurement a nonsignificant greater number of times than Fitbit Alta HR (50.0% versus 54.8%; $P = .702$).

With the exception of REM sleep ($P = .630$), t paired tests revealed statistically significant differences between PSG and Fitbit measures for all sleep outcomes ($P < .001$). The Fitbit devices overestimated TST by on average 59.78 minutes (95% CI −84.54, −35.02; $P < .001$), while WASO and SOL measures were underestimated by 36.14 minutes (95% CI 24.09, 48.18; $P < .001$) and 23.22 minutes (95% CI 17.15, 29.28; $P < .001$) respectively. Mean values and differences in sleep outcomes are presented in **Table 2**.

Overall sensitivity was 87.81% (95% CI 79.87–95.76), while specificity was 43.85% (95% CI 31.79–55.91). In patients with confirmed OSA, sensitivity and specificity were 87.61% (95% CI 85.19–90.03), and 44.77% (95% CI 37.96–51.58), respectively. In patients with OSA who were under CPAP treatment, sensitivity was 88.20% (95% CI 81.78–94.62), and specificity was 55.08% (95% CI 39.94–70.22), and in patients who tested negative on the PSG, sensitivity was 89.28% (95% CI 82.69–95.87), and specificity was 39.53% (95% CI 15.05–64.01) (**Table 3**).

Figure 3—Bland-Altman plot for sleep onset latency.
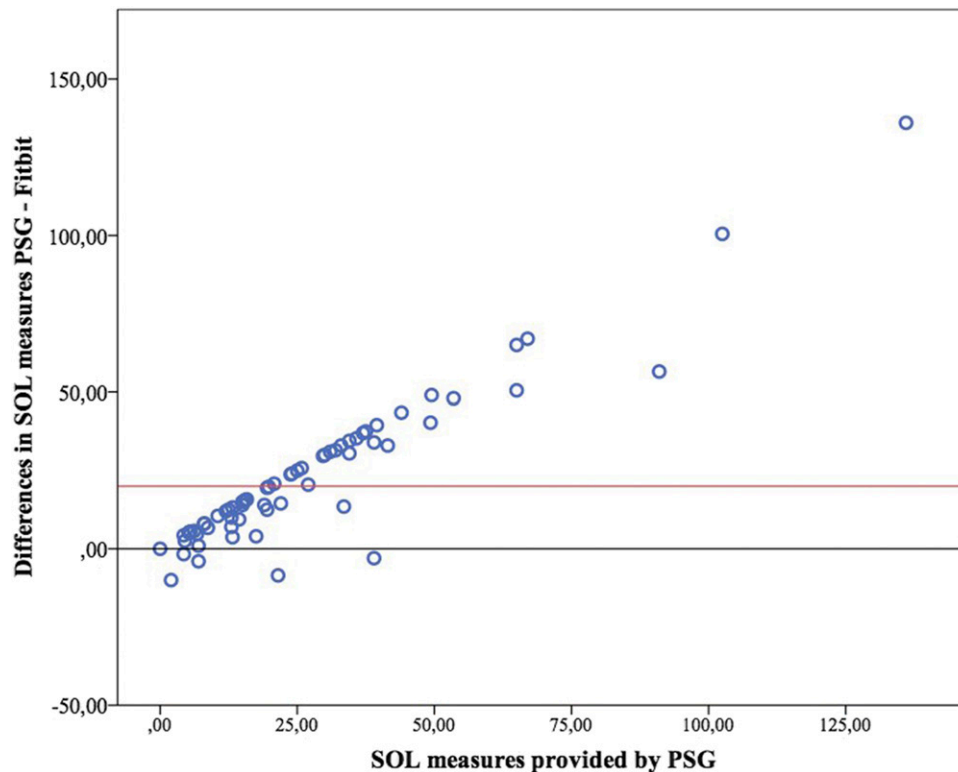
PSG = polysomnography, SOL = sleep onset latency.

There were no statistically significant differences in sensitivity between Fitbit models ($P = .743$). Specificity was significantly higher with Alta HR ($P = .06$). Sensitivity ($P < .001$) and specificity ($P = .003$) significantly increased when the devices were on 4-stage measurement mode (**Table 3**).

Bland-Altman plots for TST, WASO and SOL are presented in **Figure 1**, **Figure 2**, and **Figure 3**. **Figure 4** presents the modified Bland-Altman plots for SOL, accounting for the potential bias derived of the different initialization setup between PSG and Fitbit.

DISCUSSION

In this study, we compared two different models of Fitbit wearable devices against PSG in a sample of adults with sleep apnea. Fitbit wearable devices showed acceptable sensitivity but poor specificity, with a tendency to overestimate TST and underestimate SOL and WASO.

Previous validation studies of Fitbit devices reported a similar overestimation of TST and a high sensitivity at the expense of specificity.[11,15,30] Overall, we found a lower precision than usually reported before. These differences may be caused by our using a clinical population. Other previous studies performed in clinical populations have also found difficulties in measuring sleep with wearable devices. For instance, a 2017 study found that the accuracy of the Fitbit Flex device was significantly lower in participants with insomnia compared

with good sleepers,[18] while a later study found an overall poor accuracy in assessing sleep with the Jawbone device in individuals with suspected central hypersomnolence.[31] In our study, we did not find statistically significant differences between participants with and without OSA, but the low number of the latter makes it difficult to extract conclusions. It would be useful to compare larger samples of clinical and nonclinical populations under the same conditions of observation to ascertain this finding.

Precision increased when the 4-stage measurement mode was on. Therefore, solving the technical issues that prevent the devices from entering into this mode should be a priority for future developments. Another technical issue was the different configuration of PSG and Fitbit regarding the detection of sleep start time. The analysis we performed to minimize this bias significantly improved the precision of the SOL measurements, revealing a weakness in the devices that could be solved by the developers without excessive difficulty. Identifying and correcting this and other systematic errors could be a cost-effective way of improving precision. This also shows how the collaboration between health care professionals and developers could enhance the performance of the devices in clinical settings. As long as developers ignore the specific necessities and idiosyncrasies of clinical practice, wearable devices will not be able to reach their true potential.

It is likely that precision of sleep trackers increases with the number of measurements. The need to perform the PSG in a hospital environment makes it difficult to assess sleep several

Figure 4—Modified Bland-Altman plot for sleep onset latency (with bias correction).
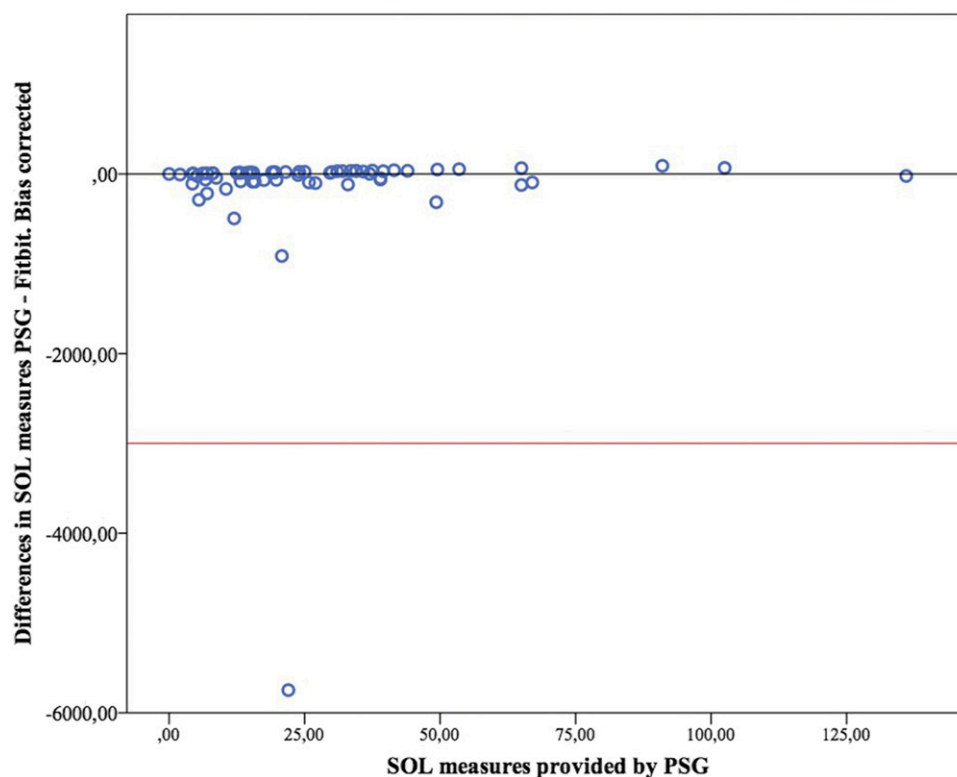
PSG = polysomnography, SOL = sleep onset latency.

nights in a row. This, in contrast, is easy to perform with wearable devices. Future studies should test if the precision of these devices increases with time of use. If this is the case, wearable devices may reach an acceptable accuracy when used for the continuous monitoring sleep. Incidentally, this is one of its more promising applications in clinical practice, since sleep is being explored as a biomarker for other conditions and risk situations, such as suicidal behavior. Sleep loss is strongly associated with suicidal behavior, and sleep disorders are highly prevalent in mentally ill patients.[4–6] Unobtrusive, passive monitoring could be especially useful in psychiatric populations, where the high rates of nonadherence and lack of cooperation make interventions in times of crisis a challenging task.[32,33] The use of e-health technologies in the prevention of suicidal behavior is already starting to be implemented in clinical settings,[34] and the role of sleep as a short-term predictor of suicidal behavior has recently been explored using actigraphy as an assessment method.[35,36]

Another potential application of this technology is the screening of sleep disturbances in the general population, which could be performed from primary care centers at a relatively low cost. Additionally, consumer sleep trackers could be used for monitoring patients with a formal diagnosis of sleep disorder to assess treatment response, as well as for increasing awareness of sleep habits in the general population.[10] However, there is still a long way before this technology can be fully applied in health care settings. The AASM recognizes the potential of commercially-available sleep monitors, and admits its usefulness to increase patient adherence to treatment, but there are several needs for the full implementation of consumer sleep technology in clinical practice and biomedical research, including widescale validation, grant of access to raw data, and FDA approval.[10]

Coordination between health care providers and technology developers is crucial in this respect. Many nonclinical factors influence the precision of sleep detection. The software used to process the data is as important as the accuracy of the sensors. Additionally, interpretation of raw data can change widely depending on the algorithms used. Fitbit wearable devices and other commercial sleep trackers use proprietary algorithms which are not made public by the developer, nor is it possible to access raw data collected by the device. It is noteworthy that one of the biggest issues occurred when trying to translate the concepts of light and deep sleep to the three stages of sleep measured by PSG. A model of wearable device more consistent with standard assessment methods, and therefore more suitable to provide directly comparable results, would greatly improve the application of this technology in research and clinical settings. Overall quality of Fitbit sleep trackers has increased over time and they expected to be more reliable as technology progresses.

E-health is opening new prospects in the diagnosis and management of several conditions. Commercially available activity monitors can be used to assess different physiological functions and monitoring is one of the most clinically relevant purposes.

To our knowledge, this is the first study to validate Fitbit wearable devices for sleep assessment in participants with sleep apnea, and the largest study to validate wearable devices in this population. It is also the first time Fitbit Alta HR is validated in a clinical sample.

Our findings should be considered in light of the study limitations. Definitions of light sleep and deep sleep are not standardized, and they may not be adequate correlates of medical terms. Similarly, the sleep stage restless has no equivalency in medical terminology, and the definition provided by the developer is ambiguous. The heterogeneity of the sample is another element to take into account. As frequently happens with clinical samples, our population was composed by dissimilar patients. Specifically, 7 of the patients had symptoms of OSA but the condition was finally ruled out during the diagnosis process, while 13 were under treatment with a CPAP machine, which can alter the PSG results. However, separate analyses for each subpopulation revealed no statistical differences regarding sensitivity and specificity between the groups. Finally, the relatively small size of our sample limits the extraction of conclusions.

ABBREVIATIONS

AASM, American Academy of Sleep Medicine
AHI, apnea-hypopnea index
AUC, area under the curve
BMI, body mass index
CPAP, continuous positive airway pressure
EES, Epworth Sleepiness Scale
OSA, obstructive sleep apnea
PSG, polysomnography
SD, standard deviation
SOL, sleep onset latency
TST, total sleep time
WASO, wake after sleep onset

REFERENCES

1. Ohayon M. Epidemiology of insomnia: what we know and what we still need to learn. *Sleep Med. Rev*. 2002;6(2):97–111.

2. Chattu VK, Manzar MD, Kumary S, Burman D, Spence DW, Pandi-Perumal SR. The global problem of insufficient sleep and its serious public health implications. *Health Care (Basel)*. 2018;7(1):E1.

3. Rosekind MR, Gregory KB. Insomnia risks and costs: health, safety, and quality of life. *Am J Manag Care*. 2010;16(8):617–626.

4. Pigeon W, Bishop T, Krueger K. Insomnia as a precipitating factor in new onset mental illness: a systematic review of recent findings. *Curr Psychiatry Rep*. 2017;19(8):44.

5. Malik S, Kanwar A, Sim L, et al. The association between sleep disturbances and suicidal behaviors in patients with psychiatric diagnoses: a systematic review and meta-analysis. *Syst Rev*. 2014;3(1):18.

6. Porras-Segovia A, Pérez-Rodríguez MM, López-Esteban P, et al. Contribution of sleep deprivation to suicidal behaviour: a systematic review. *Sleep Med Rev*. 2019;44:37–47.

7. Jafari B, Mohsenin V. Polysomnography. *Clin Chest Med*. 2010;31(2):287–297.

8. Ibáñez V, Silva J, Cauli O. A survey on sleep assessment methods. *PeerJ*. 2018;6:e4849.

9. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. *Per Med*. 2018;15(5):429–448.

10. Khosla S, Deak M, Gault D, et al. Consumer sleep technology: an American Academy of Sleep Medicine position statement. *J Clin Sleep Med*. 2018;14(05):877–880.

11. de Zambotti M, Baker F, Colrain I. Validation of sleep-tracking technology compared with polysomnography in adolescents. *Sleep*. 2015;38(9):1461–1468.

12. Evenson K, Goto M, Furberg R. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act*. 2015;12(1):159.

13. Meltzer L, Hiruma L, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep*. 2015;38(8):1323–1330.

14. Fonseca P, Weysen T, Goelema M, et al. Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. *Sleep*. 2017;40(7).

15. de Zambotti M, Goldstone A, Claudatos S, Colrain I, Baker F. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. *Chronobiol Int*. 2017;35(4):465–476.

16. Meltzer LJ, Hiruma LS, Avis K, et al. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep*. 2015;38(8):1323–1330.

17. Toon E, Davey MJ, Hollis SL, Nixon GM, Horne RSC, Biggs SN. Comparison of commercial wrist-based and smartphone accelerometers, actigraphy, and PSG in a clinical cohort of children and adolescents. *J Clin Sleep Med*. 2016;12(3):343–350.

18. Kang S, Kang J, Ko K, Park S, Mariani S, Weng J. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. *J Psychosom Res*. 2017;97:38–44.

19. Gruwez A, Bruyneel A, Bruyneel M. The validity of two commercially-available sleep trackers and actigraphy for assessment of sleep parameters in obstructive sleep apnea patients. *PLoS One*. 2019;14(1):e0210569.

20. Rosa T, Bellardi K, Viana A, Ma Y, Capasso R. Digital health and sleep-disordered breathing: a systematic review and meta-analysis. *J Clin Sleep Med*. 2018;14(09):1605–1620.

21. Shephard DA. Declaration of Helsinki and consent. *Can Med Assoc J*. 1976;115(12):1191–1192.

22. Cohen J, Korevaar D, Altman D, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.

23. Berry RB, Brooks R, Gamaldo CE, et al; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Version 2.4. Darien, IL: American Academy of Sleep Medicine; 2017.

24. Fitbit, Inc. Fitbit Charge 2 User Manual. https://staticcs.fitbit.com/content/assets/help/manuals/manual_charge_2_en_US.pdf. Accessed December 20, 2018.

25. Fitbit, Inc. Fitbit Alta HR User Manual. https://staticcs.fitbit.com/content/assets/help/manuals/manual_alta_hr_en_US.pdf. Accessed December 20, 2018.

26. Meijer R, Goeman J. Efficient approximate k-fold and leave-one-out cross-validation for ridge regression. *Biomed J*. 2013;55(2):141–155.

27. Huber PJ. Robust estimation of a location parameter. *Ann Math Stat*. 1964;35(1):73–101.

28. Wang Y-G, Lin X, Zhu M, Bai Z. Robust estimation using the huber function with a data-dependent tuning constant. *J Comput Graph Stat*. 2007;16(2):468–481.

29. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association; 2013.

30. Mantua J, Gravel N, Spencer R. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. *Sensors*. 2016;16(5):646.

31. Cook J, Prairie M, Plante D. Ability of the multisensory Jawbone UP3 to quantify and classify sleep in patients with suspected central disorders of hypersomnolence: a comparison against polysomnography and actigraphy. *J Clin Sleep Med*. 2018;14(05):841–848.

32. Lacro J, Dunn L, Dolder C, Leckband S, Jeste D. Prevalence of and risk factors for medication nonadherence in patients with schizophrenia. *J Clin Psychiatry*. 2002;63(10):892–909.

33. Bulloch A, Patten S. Non-adherence with psychotropic medications in the general population. *Soc Psychiatry Psychiatr Epidemiol*. 2009;45(1):47–56.

34. Berrouiguet S, Alavi Z, Vaiva G, et al. SIAM (Suicide intervention assisted by messages): the development of a post–acute crisis text messaging outreach for suicide prevention. *BMC Psychiatry*. 2014;14:294.

35. Littlewood DL, Kyle SD, Carter LA, Peters S, Pratt D, Gooding P. Short sleep duration and poor sleep quality predict next-day suicidal ideation: an ecological momentary assessment study. *Psychol Med*. 2019;49(3):403–411.

36. Benard V, Etain B, Vaiva G, et al. Sleep and circadian rhythms as possible trait markers of suicide attempt in bipolar disorders: an actigraphy study. *J Affect Disord*. 2019;244:1–8.

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT