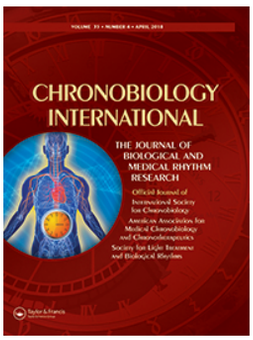

Taylor & Francis
Taylor & Francis Group

Chronobiology International

The Journal of Biological and Medical Rhythm Research

ISSN: 0742-0528 (Print) 1525-6073 (Online) Journal homepage: http://www.tandfonline.com/loi/icbi20

A validation study of Fitbit Charge 2™ compared with polysomnography in adults

Massimiliano de Zambotti, Aimee Goldstone, Stephanie Claudatos, Ian M. Colrain & Fiona C. Baker

To cite this article: Massimiliano de Zambotti, Aimee Goldstone, Stephanie Claudatos, Ian M. Colrain & Fiona C. Baker (2018) A validation study of Fitbit Charge 2™ compared with polysomnography in adults, Chronobiology International, 35:4, 465-476, DOI: 10.1080/07420528.2017.1413578

To link to this article: https://doi.org/10.1080/07420528.2017.1413578

CHRONOBIOLOGY INTERNATIONAL
2018, VOL. 35, NO. 4, 465–476
https://doi.org/10.1080/07420528.2017.1413578

Taylor & Francis
Taylor & Francis Group

A validation study of Fitbit Charge 2™ compared with polysomnography in adults

Massimiliano de Zambotti, Aimee Goldstone, Stephanie Claudatos, Ian M. Colrain, and Fiona C. Baker

Center for Health Sciences, SRI International, Menlo Park, CA, USA

ABSTRACT

We evaluated the performance of a consumer multi-sensory wristband (Fitbit Charge 2™), against polysomnography (PSG) in measuring sleep/wake state and sleep stage composition in healthy adults.

In-lab PSG and Fitbit Charge 2™ data were obtained from a single overnight recording at the SRI Human Sleep Research Laboratory in 44 adults (19—61 years; 26 women; 25 Caucasian). Participants were screened to be free from mental and medical conditions. Presence of sleep disorders was evaluated with clinical PSG. PSG findings indicated periodic limb movement of sleep (PLMS, > 15/h) in nine participants, who were analyzed separately from the main group (n = 35). PSG and Fitbit Charge 2™ sleep data were compared using paired t-tests, Bland–Altman plots, and epoch-by-epoch (EBE) analysis.

In the main group, Fitbit Charge 2™ showed 0.96 sensitivity (accuracy to detect sleep), 0.61 specificity (accuracy to detect wake), 0.81 accuracy in detecting N1+N2 sleep ("light sleep"), 0.49 accuracy in detecting N3 sleep ("deep sleep"), and 0.74 accuracy in detecting rapid-eye-movement (REM) sleep. Fitbit Charge 2™ significantly ($p < 0.05$) overestimated PSG TST by 9 min, N1+N2 sleep by 34 min, and underestimated PSG SOL by 4 min and N3 sleep by 24 min. PSG and Fitbit Charge 2™ outcomes did not differ for WASO and time spent in REM sleep. No more than two participants fell outside the Bland–Altman agreement limits for all sleep measures. Fitbit Charge 2™ correctly identified 82% of PSG-defined non-REM–REM sleep cycles across the night. Similar outcomes were found for the PLMS group.

Fitbit Charge 2™ shows promise in detecting sleep-wake states and sleep stage composition relative to gold standard PSG, particularly in the estimation of REM sleep, but with limitations in N3 detection. Fitbit Charge 2™ accuracy and reliability need to be further investigated in different settings (at-home, multiple nights) and in different populations in which sleep composition is known to vary (adolescents, elderly, patients with sleep disorders).

ARTICLE HISTORY
Received 30 October 2017
Revised 29 November 2017
Accepted 2 December 2017

KEYWORDS
Wearables; validation; multisensory; actigraphy

Introduction

Sleep is fundamental for health and wellness, and insufficient sleep is linked to severe adverse outcomes including obesity, cardiovascular diseases, diabetes, mood disorders, and mortality (Grandner 2017). Valid objective measures of sleep are a prerequisite for accurately monitoring sleep in health and pathology. The gold standard for objective sleep assessment is polysomnography (PSG), the continuous recording of electroencephalographic (EEG), electromyographic (EMG), and electrooculographic (EOG) activity via surface electrodes. Information about the relative dominance of specific EEG rhythms, muscle and eye activity across the night, is used to visually score wake and stages of sleep (N1, N2, N3, and rapid-eye-movement [REM] sleep). PSG requires trained sleep technicians, a dedicated PSG acquisition system, and technical expertise to record and visually score PSG records. PSG is usually confined to the sleep laboratory, and although portable PSG systems are also used in non-laboratory settings, it is expensive, time consuming and impractical for long-term use (Kryger, Roth, Dement 2017, for description and use of PSG).

Wrist actigraphy is used in both research and clinical sleep settings to investigate an individuals' sleep/wake patterns and is more practical and suitable than PSG for prolonged recordings in non-laboratory settings (Van De Water et al. 2011). Actigraphy uses specific algorithms (e.g., Cole–Kripke or Sadeh algorithms) to analyze an

CONTACT Massimiliano de Zambotti massimiliano.dezambotti@sri.com; maxdeze@gmail.com SRI International, 333 Ravenswood Avenue, Menlo Park, CA, 94025.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ICBI.

individuals' pattern of motion from which sleep/wake states can be estimated (motion implies wake, and no-motion implies sleep). As compared to PSG, the ability of actigraphic devices to detect sleep (sensitivity) is usually higher than the ability to accurately detect wake (specificity), which is mainly due to the intrinsic limitation of actigraphy in discriminating "*quiet wakefulness*" (no motion) from sleep. A systematic review of the literature (Van De Water et al. 2011) indicated that among studies in healthy people, the sensitivity of actigraphy ranges between 0.87 and 0.99, while specificity ranges widely between 0.28 and 0.67. Both sensitivity and specificity vary as a function of the population studied (e.g., pediatric, adults), the presence of sleep disorders (e.g., insomnia disorder, sleep related breathing disorders) and the algorithm used. In general, actigraphy devices tend to overestimate sleep and underestimate wake (Van De Water et al. 2011), with specificity decreasing as the amount of PSG wake increases (Paquet et al. 2007).

In the consumer market, the wearable industry offers a wide range of commercially available options to measure sleep. Although the wearable boom has increased awareness about the importance of sleep to the general public, accuracy and reliability of wearables in measuring sleep are still debatable, and their potential use in research and clinical sleep settings is still questionable (see de Zambotti et al. 2016b). Similar to actigraphy, the first generation of sleep-trackers used motion (an in-built accelerometer) to determine sleep/wake states via proprietary algorithms. Validation studies indicated that compared to PSG, and similarly to actigraphy, these devices overestimated sleep and underestimated wake, with high sensitivity and relatively low specificity (Kolla et al. 2016).

Few studies have simultaneously collected data from both wearables and standard actigraphy, in addition to PSG and findings are mixed. In adults, Montgomery-Downs et al. (2012) showed that both Fitbit "original" and Actiwatch-64 (Mini Mitter, Inc.) similarly overestimated total sleep time (TST) and sleep efficiency (SE) compared to PSG and both exhibited poor performance in detecting wake (specificity of 0.20 for Fitbit and 0.39 for Actiwatch-64). Similar performance (based on PSG-device discrepancies only) between standard actigraphy and wearables is also suggested by Mantua et al. (2016) who compared four different sleep-trackers (Basic Health Tracker 2014 Ed., Fitbit Flex™, Misfit Shine™, Withings Pulse O2) worn on the wrist, and Actiwatch Spectrum® (Philips Respironics, Inc.) against PSG, in young adults. Despite all devices performing similarly when estimating PSG TST, only Fitbit Flex™ and Actiwatch Spectrum® did not significantly differ from PSG when determining SE. In adults with unipolar major depressive disorder, Cook et al. (2017) found that both Fitbit Flex™ (normal setting) and Actiwatch-2, significantly overestimated TST and underestimated wake after sleep onset (WASO), compared to PSG, with similar sensitivity (0.98 and 0.97) and specificity (0.35 and 0.31) outcomes. Kang et al. (2017) investigated the accuracy of Fitbit Flex™ against PSG in a group of individuals (18–60 years) with and without insomnia, while participants also wore an Actiwatch-2. In the good sleepers, Fitbit Flex™ (normal setting) significantly overestimated PSG TST while Actiwatch-2 significantly underestimated both PSG TST and SE; poorer specificity was identified for Fitbit Flex™ (0.36) compared to that of Actiwatch-2 (0.61). In the insomnia group, both wrist-worn devices significantly underestimated WASO and had relatively low specificity (0.36 for Fitbit Flex™ and 0.45 for Actiwatch-2); Fitbit Flex™ significantly overestimated PSG TST and SE. In children and adolescents (some of whom were classified as having mild or moderate/severe obstructive sleep apnea), Meltzer et al. (2015) suggested that standard actigraphs (Actiwatch Spectrum® and AMI Motionlogger®) performed better than a commercial sleep-tracker (Fitbit Ultra™) in PSG wake estimation. Differently, Toon et al. (2015) found no differences between Actiwatch-2 (Philips Respironics, Inc.) and Jawbone UP™ in estimating PSG TST and WASO and identified similar sensitivity (0.93 and 0.92) and specificity (0.63 and 0.66), in a sample of children and adolescents with suspected sleep disordered breathing. Roane et al. (2015) found no significant differences between SenseWear® Pro3 Armband and the main PSG sleep outcomes, in a sample of adolescents; while a standard actigraphic device (AMI Motionlogger®) significantly overestimated PSG TST and underestimated WASO. However, the Bland–Altman agreement limits (reflecting the dispersion of the discrepancies) were lower for the AMI

Motionlogger® compared to that of the SenseWear® Pro3 Armband. Sensitivity and specificity were similar for SenseWear® Pro3 Armband and AMI Motionlogger® (0.947 vs. 0.97 for sensitivity, and 0.39 vs 0.43 for specificity, respectively).

Overall, the performance of commercial motion-based devices seems to share similar limitations to those of standard actigraphy. Furthermore, despite some attempts to differentiate sleep stages (e.g., "light" vs, "deep" sleep) based on motion (de Zambotti et al. 2015; Mantua et al. 2016), these devices are exclusively limited to the detection of wake and sleep states. Thanks to the advancement in sensor capability and signal processing, a new generation of wearables have now adopted a multisensory approach for sleep assessment. Such devices claim to accurately differentiate sleep stages by using sources of information (e.g., heart rate [HR] and HR variability [HRV]) in addition to motion. Despite the intriguing possibility of having easy access to a more complete picture of an individuals' sleep macrostructure, the accuracy of this new generation of devices needs to be determined. The aim of the current study is to assess the accuracy of a commercially available multisensory sleep-tracker, Fitbit Charge 2™ (Fitbit Inc.), in measuring sleep/wake states, "light" (PSG N1+N2), "deep" (PSG N3), and REM sleep against standard PSG on a single laboratory night.

Materials and methods

Participants

Forty-four adults (19–61 years; 26 women; 25 Caucasian) participated in the study. Participants were recruited from the San Francisco Bay Area. All participants were informed about the purpose of the study and gave informed consent; they were compensated for study participation. The study was approved by the SRI International IRB committee.

After a phone-screen interview, participants were scheduled for an in-person initial assessment to confirm eligibility. None of the participants reported having severe medical conditions (e.g., high blood pressure (BP), heart disease, epilepsy, diabetes) and/or current use of medications known to affect sleep (e.g., hypnotics) and/or the cardiovascular system (e.g., antihypertensives). All participants had body mass index (BMI) < 31 kg. m^{-2}, and normal BP levels (systolic BP <140 mmHg and diastolic BP <100 mmHg, as determined via digital sphygmomanometer), at the time of the PSG recording. None of the participants was a current smoker. The participants' general health status was evaluated using the 36-Item Short Form Health Survey (SF-36) (Ware and Sherbourne 1992). Self-reported sleep quality was measured using the Pittsburgh Sleep Quality Index (PSQI) (Buysse et al. 1989), and none of the participants exceeded the Insomnia Severity Index (ISI) (Bastien et al. 2001) cut-off for clinical insomnia.

All participants underwent a single clinical PSG recording (see below) at the Human Sleep Research Laboratory at SRI International in which PSG and Fitbit Charge 2™ sleep data were simultaneously collected. Participants showed no signs of breathing disorders (apnea-hypopnea index < 5 events per hour of sleep) or narcolepsy. PSG showed periodic limb movement of sleep (PLMS) > 15/h in nine participants, who were analyzed as a separate group (*PLMS* group). However, those participants did not complain of nocturnal sleep disturbance or daytime sleepiness, and thus according to the American Academy of Sleep Medicine (American Academy of Sleep Medicine, 2014) the presence of PLMS was simply recognized as a PSG finding and no diagnosis of sleep-related movement disorder was given. PLMS were determined using Compumedics Profusion PSG 3 (Compumedics, Abbotsford, Victoria, Australia) automatic analysis (significant leg movements had minimum duration of 0.5 s, maximum duration of 10 s, minimum EMG amplitude increase of 8 μV of at least 0.5 s; a minimum of 4 significant leg movements, 5–90 s apart, were required to define a periodic limb movement series). The remaining participants (N = 35) constituted the *main sample*. Sample demographics, self-reported sleep quality, and health status of the sample are provided in Table 1.

In-lab procedure

On the recording night, participants were instructed to arrive at the Human Sleep Research Laboratory

Table 1. Sample characteristics.

	Main sample	PLMS
	Mean ± SD	Mean ± SD
N Females/males	23/12	3/6
Age (years)	35 ± 12	42 ± 15
BMI (kg. m^{-2})	23.9 ± 2.8	23.7 ± 3.8
SBP (mmHg)*	113 ± 14	114 ± 8
DBP (mmHg)*	74 ± 7	74 ± 8
Self-reported sleep quality		
PSQI (total score)	3.1 ± 1.7	4.6 ± 2.2
ISI (total score)	3 ± 3	4 ± 3
SF-36 general health (*self-reported heath status with score ranging between 0 and 100. Higher values represent a more favorable health state)*	83 ± 13	84 ± 10

*Average of three readings, separated by 1 min, obtained when the participant was seated, and following 10 min of rest. **BMI**, body mass index; **DBP**, diastolic blood pressure; **ISI**, insomnia severity index; **PLMS**, periodic limb movement of sleep; **SBP**, systolic blood pressure; **SF-36**, 36-Item Short Form Health Survey.

3 h before their typical lights-off time. The beginning and the end of the PSG and Fitbit Charge 2™ data collection periods were coincident with the participants' self-selected lights-off and lights-on time. PSG and Fitbit Charge 2™ data recordings were manually synchronized. Potential bias in study outcomes due to potential errors in PSG-Fitbit Charge 2™ synchronization were tested (see analysis and results sections). Participants were instructed to refrain from consuming alcohol during the 24 h before the recording and from consuming any caffeinated products after 3 pm on the recording day. All recordings were performed in sound-attenuated and temperature-controlled rooms.

PSG sleep assessment

PSG assessment was performed using Compumedics Grael-PSG Units (Compumedics, Abbotsford, Victoria, Australia). Electroencephalographic (EEG; F3/M2, F4/M1, C3/M2, C4/M1, O1/M2, O2/M1; 256 Hz sampled and 0.03–35 Hz filtered), bilateral electrooculographic, and submental electromyographic recordings were performed and PSG sleep (wake, N1, N2, N3, and REM) scored in 30-s epochs, according to the American Academy of Sleep Medicine (AASM) rules (Iber 2007). Clinical measures of respiration, oxygen saturation, and leg movements were also collected to evaluate the presence of potential sleep disorders according to the guidelines of the AASM (Iber 2007). PSG sleep records were double scored by experienced scorers. Accepted inter-rater reliability was set at 91% concordance (scoring concordance was 91.6 ± 0.8% for the *main sample* and 91.4 ± 0.7% in the *PLMS* group).

Standard PSG outcomes were calculated according to the AASM guidelines (Iber 2007): time in bed (TIB; min), total sleep time (TST; min), sleep efficiency as TST/TIB*100 (SE; %), sleep onset latency (SOL; min), wake after sleep onset (WASO; min), time spent in N1, N2, N3 and REM sleep (min).

Fitbit charge 2™

Fitbit Charge 2™ (Fitbit Inc.) is a wearable fitness-tracker wristband able to track daily activity levels and sleep. The device automatically connects via Bluetooth and transfers data to a mobile platform via a dedicated App. Fitbit Charge 2™ allows tracking of sleep stages (minutes spent awake, in "light", "deep", and REM sleep) in addition to sleep/wake states. As disclosed on fitbit.com, Fitbit Inc. reports "*Fitbit estimates your sleep stages using a combination of your movement and heart-rate patterns. Additional data—such as the length of time your movements are indicative of sleep behavior (such as rolling over, etc.)—help confirm that you're asleep. While you're sleeping, your device tracks the beat-to-beat changes in your heart rate, known as heart rate variability (HRV), which fluctuate as you transition between light sleep, deep sleep, and REM sleep stages. When you sync your tracker in the morning, we use your movement and heart rate patterns to estimate your sleep cycles from the previous night*".

Fitbit Charge 2™ was attached to the participant's non-dominant wrist, above the wrist bone before bedtime and worn until the next morning. No action was required from participants. Sleep lab technicians assured that the devices were in close contact with the skin, without movement or excessive pressure. Sleep lab technicians assured that PSG and Fitbit Charge 2™ recordings were synchronized, with a second resolution.

For each recording, Fitbit Inc. provided sleep stage data in 30-s epochs to allow epoch-by-epoch (EBE) analysis. Fitbit Inc. was not involved in any other aspect of the study, and did not have access to participant information or PSG staging. Fitbit

Inc. was able to provide EBE data for the 44 adults included in the study. For two additional participants (not included in the study), Fitbit Inc. was unable to provide EBE data. As reported by Fitbit Inc. "*The Fitbit system does not return sleep stages under various conditions. These include cases where the heart beat signal (and hence the heart rate variability) is not cleanly detected throughout the night, if the total sleep duration is less than three hours, or if the battery runs out of power during the sleeping period*". We did not notice any potential issue with the battery charge level, and both participants had PSG total sleep duration greater than 3 h, therefore, Fitbit Charge 2™ was unable to collect data due to absence of a clean heart rate signal or some other unknown cause.

We derived the following PSG-equivalent measures from the Fitbit Charge 2™ EBE data: TST (min), SOL (min), WASO (min), time in REM, time spent in "light sleep" (PSG N1+N2, according to Fitbit Inc.) and "deep sleep" (PSG N3, according to Fitbit Inc.).

Statistical analyses

PSG and equivalent Fitbit Charge 2™ sleep outcomes were compared using paired *t*-tests. The overall agreement between PSG and equivalent Fitbit Charge 2™ sleep outcomes was analyzed using Bland–Altman plots (Bland and Altman 1986). Bland–Altman plot biases (PSG-Fitbit Charge 2™ mean differences in sleep outcomes; a positive bias indicates that Fitbit Charge 2™ underestimates PSG measures, and a negative bias indicates that Fitbit Charge 2™ overestimates them), standard deviation and ±95%CI of the biases, lower and upper agreement limits (mean difference ± 1.96*SD) and proportion of participants falling outside these limits were calculated.

From the EBE analysis we calculated sensitivity (proportion of PSG epochs identified correctly as "*Sleep*" by Fitbit Charge 2™), specificity (proportion of PSG epochs identified correctly as "*Wake*" by Fitbit Charge 2™), accuracy in detecting "*light sleep*" (proportion of PSG N1+N2 epochs identified correctly as "*light sleep*" by Fitbit Charge 2™), "*deep sleep*" (proportion of PSG N3 epochs identified correctly as "*deep sleep*" by Fitbit Charge 2™) and REM sleep (proportion of PSG REM epochs identified correctly as REM sleep by Fitbit Charge 2™). A cross-tabular representation (confusion matrix) of 4 rows containing wake, "*light sleep*", "*deep sleep*" and REM epochs classified by PSG and 4 columns containing predicted wake, "*light sleep*", "*deep sleep*" and REM epochs classified by Fitbit Charge 2™ was also created. Each cell represents mean, SD and ±95%CI. Finally, in order to assess potential biases in PSG-Fitbit Charge 2™ data synchronization, we calculate EBE specificity by sliding the PSG and Fitbit Charge 2™ epoch alignment by up to 90 s (1–3 epochs forwards and backwards).

A comparison of sleep cycles was introduced to evaluate the accuracy of Fitbit Charge 2™ in assessing NREM-REM sleep pattern distribution across the night. As previously reported (Feinberg and Floyd 1979), sleep cycles were defined using the following criteria: 1) the start of the first cycle was defined as the first epoch scored as NREM sleep; 2) the end of each NREM-REM cycle required a minimum of 15 min of NREM sleep followed by at least 5 min of REM sleep, except for the first cycle where there was no minimum duration for the REM period (but greater than zero); 3) in the final cycle, if at least 5 min of REM sleep preceded the end of the recording, the end of the cycle was defined by the last REM epoch and the final cycle was included in our analysis, however, if the final cycle did not contain at least 5 min of REM sleep (i.e., the recording ended mid-cycle), the final sleep cycle was not included in our analysis. The same rules were applied for both PSG and Fitbit Charge 2™ data.

The PSG-Fitbit Charge 2™ sleep cycle alignment was based on the overlap of the REM periods. When at least one epoch of REM sleep within a cycle overlapped between Fitbit Charge 2™ and PSG, that cycle was considered a matched cycle. When there were no overlapping REM epochs between PSG and Fitbit Charge 2™, that cycle was considered an unmatched cycle. An unmatched cycle was either the result of PSG detecting a cycle that Fitbit did not (missed cycle), or Fitbit detecting a cycle that PSG did not (extra cycle). Potential PSG-Fitbit Charge 2™ differences in the number of detected sleep cycles were tested using paired *t*-tests. The percentage of matched cycles over the total PSG cycles is provided, as well as the number of extra cycles detected by Fitbit Charge 2™. See Figure 1 for an example.

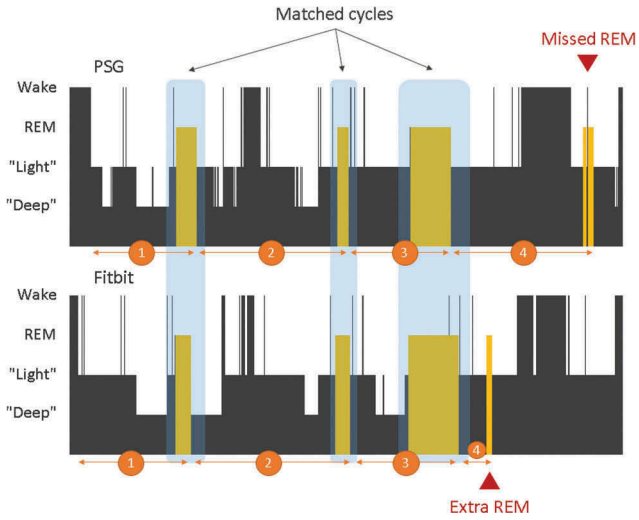

Figure 1. Hypnograms (sleep stages plotted as a function of time of the night) obtained from polysomnography (PSG; top panel) and Fitbit Charge 2™ (bottom panel) of a participant, showing PSG and Fitbit sleep cycles matching and mismatching (missed cycles or detection of extra cycles). In this participant, both PSG and Fitbit detected four sleep cycles. However, Fitbit missed the forth PSG sleep cycle and incorrectly detected an additional rapid-eye-movement (REM) episode after the third PSG non-REM-REM cycle. In this analysis, a sleep cycle is considered "matched" between PSG and Fitbit if the REM period for the two devices overlaps by at least one epoch.

All analyses were performed separately for the main group (n = 35) and the PLMS group (n = 9).

Results

PSG and Fitbit Charge 2™ sleep outcomes and Bland–Altman plots

PSG and Fitbit Charge 2™ sleep outcomes are provided in Table 2; Bland–Altman plots biases, SD and ±95%CI of the biases, upper and lower agreement limits plots are provided in Table 3; Bland–Altman plots for the main sleep outcomes are provided in Figure 2.

In the main group, Fitbit Charge 2™ significantly overestimated PSG TST by 9 min, and PSG "*light sleep*" by 34 min. It significantly underestimated PSG SOL by 4 min, and PSG "*deep sleep*" by 24 min. No significant Fitbit-PSG biases existed for WASO and time spent in REM sleep. No more than two participants fell outside the Bland–Altman agreement limits for all the considered sleep outcomes. For WASO and SOL, both participants who exceeded the agreement limits had a high amount of PSG WASO and long SOL.

Table 2. Polysomnographic (PSG) and Fitbit Charge 2™ sleep measures in the main group (n = 35), and in the periodic limb movement of sleep (PLMS) group (n = 9).

		PSG		Fitbit Charge 2™			
		Mean ± SD	±95%CI	Mean ± SD	±95%CI	t	p
Lights-off (hh:mm)	*Main group*	23:23 ± 00:46	23:07–23:39	–	––	––	–
	PLMS	23:30 ± 01:10	22:37–24:24	–			–
Lights-on (hh:mm)	*Main group*	06:39 ± 00:45	06:23–06:54	–	–	––	––
	PLMS	06:51 ± 00:56	06:08–07:33	–	–		
TIB (min)	*Main group*	439 ± 50	409–450	––	––	––	––
	PLMS	441 ± 26	421–461				
SE (%)	*Main group*	86.8 ± 8.8	83.7–89.8	––	––	––	––
	PLMS	85.5 ± 7.2	79.9–91.0				
TST (min)	*Main group*	380 ± 50	363–397	389 ± 47	373–405	2.24	0.031
	PLMS	377 ± 37	349–405	385 ± 26	365–406	0.85	0.418
SOL (min)	*Main group*	14 ± 11	10–17	9 ± 6	7–11	−2.70	0.011
	PLMS	15 ± 13	4–25	8 ± 5	5–12	−1.91	0.092
WASO (min)	*Main group*	43 ± 36	30–55	38 ± 25	29–46	−1.57	0.126
	PLMS	49 ± 26	29–69	48 ± 14	37–59	−0.17	0.866
Time in N1 (min)	*Main group*	34 ± 15	29–39	––	––	––	––
	PLMS	29 ± 14	19–40				
Time in N2 (min)	*Main group*	183 ± 31	172–194	––	––	––	––
	PLMS	200 ± 24	181–219				
Time in N1+N2 ("light sleep") (min)	*Main group*	217 ± 35	205–229	250 ± 38	237–263	5.82	<0.001
	PLMS	229 ± 31	206–253	265 ± 40	234–295	2.17	0.061
Time in N3 ("deep sleep") (min)	*Main group*	74 ± 19	67–81	50 ± 25	42–59	−4.93	<0.001
	PLMS	66 ± 31	42–90	39 ± 18	25–53	−2.37	0.045
Time in REM (min)	*Main group*	89 ± 28	79–99	88 ± 25	80–97	−0.12	0.903
	PLMS	82 ± 25	62–101	82 ± 26	66–98	0.02	0.982

REM, rapid-eye-movement; **SE**, sleep efficiency; **SOL**, sleep onset latency; **TIB**, time in bed; **TST**, total sleep time; **WASO**, wake after sleep onset

Table 3. Bland–Altman plots biases, SD and ±95%CI of the biases, upper and lower agreement limits, and number of participants exceeding these limits for polysomnographic (PSG) and equivalent Fitbit Charge 2™ sleep in the main group (n = 35), and in the periodic limb movement of sleep (PLMS) group (n = 9).

		Bias ± SD	±95%CI of the bias	Lower agreement limit	Upper agreement limit	*N participants exceeding the agreement limits*
TST (min)	*Main group*	−9 ± 24	−18–−1	−57	38	2
	PLMS	−8 ± 29	−30–14	−64	48	0
SOL (min)	*Main group*	4 ± 9	1–8	−14	23	2
	PLMS	7 ± 10	−1–15	−14	27	1
WASO (min)	*Main group*	5 ± 19	−1–11	−31	41	2
	PLMS	2 ± 26	−18–21	−49	52	0
Time in N1+N2 ("light sleep") (min)	*Main group*	−34 ± 34	−46–−22	−101	34	2
	PLMS	−35 ± 49	−73–2	−131	61	0
Time in N3 ("deep sleep") (min)	*Main group*	24 ± 28	14–33	−32	79	2
	PLMS	28 ± 35	1–54	−41	96	0
Time in REM (min)	*Main group*	1 ± 27	−9–10	−53	54	1
	PLMS	0 ± 37	−29–28	−73	73	0

REM, rapid-eye-movement; **SOL**, sleep onset latency; **TST**, total sleep time; **WASO**, wake after sleep onset

A similar trend for Fitbit-PSG discrepancies is shown in the PLMS group for whom Fitbit Charge 2™ significantly underestimated PSG "*deep sleep*" (by 28 min) and showed a trend (p = 0.061) in overestimating PSG "l*ight sleep*" (by 35 min). No more than one participant fell outside the Bland–Altman agreement limits for all the considered sleep outcomes.

EBE analysis outcomes

In the main group, Fitbit Charge 2™ had a sensitivity of 0.96 (accuracy in detecting PSG sleep epochs), a specificity of 0.61 (accuracy in detecting PSG wake), an accuracy of 0.81 for detecting "*light sleep*", 0.49 for "*deep sleep*" and 0.74 for REM sleep. When Fitbit Charge 2™ misclassified PSG wake epochs, it classified these epochs as "*light sleep*" 30% of the time. Similarly, Fitbit Charge 2™ misclassified PSG "*deep sleep*" epochs as "*light sleep*" 49% of the time and misclassified PSG REM sleep as "*light sleep*" 21% of the time.

Fitbit Charge 2™ in the PLMS group showed similar performance as in the main group. EBE analysis outcomes are provided in Table 4.

Bias in time synchronization between PSG and wearable device

Potential biases in time synchronization were checked. By shifting the PSG-Fitbit Charge 2™ epoch alignment by up to 90 s (clockwise and counterclockwise) in respect to the synchronization of the lab server and Fitbit App clocks, the chosen alignment shows the best fit and thus indicates that our manual alignment was correct. EBE specificity as a function of PSG-Fitbit Charge 2™ epoch alignment in the main group and in the PLMS participants is provided in Figure 3. This analysis was done purely to ascertain whether there was any systematic error in the alignment process. The PSG-Fitbit Charge 2™ alignment was based exclusively on the synchronization between the sleep lab server and the Fitbit App clocks and not based on the best fit determined by this analysis.

Sleep cycle analysis

Participants in the main group had, on average, 4.0 ± 1.0 NREM-REM cycles during the PSG night, whereas Fitbit Charge 2™ detected 3.7 ± 1.0 NREM-REM cycles (p = 0.086). Fitbit Charge 2™ correctly identified 82.1 ± 17.3% of the PSG cycles (percentage matched cycles by cycle number: 70.6% for the first cycle, 84.4% for the second cycle, 78.8% for the third cycle, 87.5% for the fourth cycle, 80% for the fifth cycle). Fitbit Charge 2™ identified, on average, 0.4 ± 0.7 extra cycles, compared to PSG.

Sleep cycle analysis for the PLMS group provided similar results: PSG detected, on average, 3.8 ± 0.8 NREM-REM cycles whereas Fitbit Charge 2™ detected 4.2 ± 0.8 cycles (p = 0.312). Fitbit Charge 2™ correctly identified 81.1 ± 20.7%

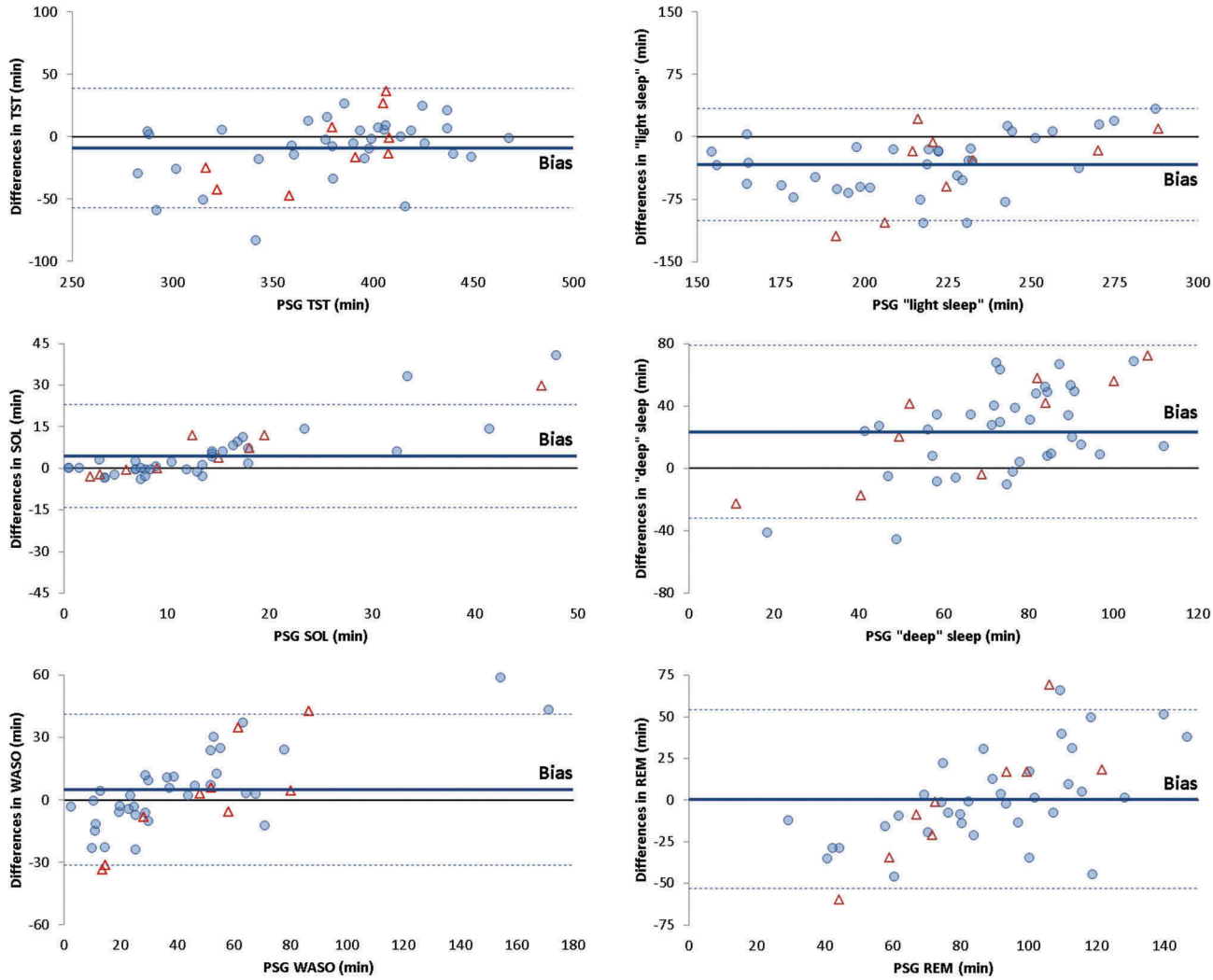

Figure 2. Bland–Altman plots for total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), time in N1+N2 ("*light*") sleep and time in N3 ("*deep*") sleep. Polysomnography (PSG)-Fitbit Charge 2™ discrepancies for sleep outcomes (*y*-axis) are plotted as a function of the PSG outcomes (*x*-axis) for each individual. Circles represent individuals in the main group and triangles represent PLMS individuals. Biases are marked; the dotted lines refer to the upper and lower Bland–Altman plots agreement limits. Biases, and upper and lower agreement limits of the biases, are displayed for the main group ($n = 35$) only for clarity in the graphical representation.

of the PSG cycles and identified 1.1 ± 1.1 extra cycles, compared to PSG.

Discussion

Fitbit Charge 2™ is among the new generation of fitness-trackers that uses a multisensory approach to sleep staging. In our study, Fitbit Charge 2™ overestimated light sleep and underestimate deep sleep but showed no bias in the estimation of REM sleep and WASO, relative to PSG. It was also able to adequately track sleep cycles across the night. While results are promising, and potentially open up the possibility of more refined measurements of sleep in large numbers of individuals using wearables, limitations exist and need to be recognized.

Only a few studies have investigated the accuracy of multisensory devices in assessing sleep stage composition against PSG. Our group previously investigated the performance of another multisensory sleep-tracker (the ŌURA ring) in a sample of healthy adolescents (de Zambotti et al. 2017). In that study, similar to our current results for Fitbit Charge 2™, the ŌURA ring had an accuracy of 0.65 for detecting "*light sleep*" and 0.61 for detecting REM sleep; however, accuracy for assessing "*deep sleep*" was poorer (0.51). For both the current study and the ŌURA ring validation, no more than 10%

Table 4. Confusion matrix for both the main group (n = 35) and the periodic limb movement of sleep (PLMS) (n = 9) group. Data are displayed as mean, SD and ±95%CI. The highlighted cells (from the top-left to the bottom-right, for each panel) represent the specificity, the accuracy in detecting N1+N2 ("light") sleep, N3 ("deep") sleep and REM sleep. PSG, polysomnography; REM, rapid-eye-movement.

		Fitbit Charge 2™			
Main group		Wake	*"light sleep"*	*"deep sleep"*	REM
PSG	Wake	**0.61(0.18) [0.67 0.55]**	0.30(0.15) [0.34 0.25]	0.01(0.02) [0.02–0.00]	0.09(0.09) [0.11 0.06]
	N1+N2	0.05(0.03) [0.06 0.04]	**0.81(0.07) [0.83 0.78]**	0.06(0.05) [0.08 0.04]	0.09(0.05) [0.10 0.07]
	N3	0.02(0.02) [0.03 0.01]	0.49(0.24) [0.57 0.41]	**0.49(0.24) [0.56 0.41]**	0.01(0.02) [0.02 0.00]
	REM	0.03(0.03) [0.04 0.02]	0.21(0.17) [0.27 0.16]	0.01(0.04) [0.03–0.00]	**0.74(0.19) [0.81 0.68]**
		Fitbit Charge 2™			
PLMS		Wake	*"light sleep"*	*"deep sleep"*	REM
PSG	Wake	**0.62(0.17) [0.73 0.51]**	0.28(0.09) [0.34 0.22]	0.02(0.03) [0.04–0.01]	0.08(0.09) [0.14 0.02]
	N1+N2	0.07(0.05) [0.10 0.03]	**0.78(0.08) [0.83 0.72]**	0.05(0.06) [0.08 0.01]	0.11(0.08) [0.16 0.06]
	N3	0.01(0.01) [0.02 0.00]	0.60(0.26) [0.77 0.43]	**0.36(0.25) [0.52 0.19]**	0.04(0.07) [0.08–0.01]
	REM	0.04(0.03) [0.06 0.02]	0.32(0.23) [0.47 0.17]	0.02(0.03) [0.05 0.00]	**0.62(0.22) [0.77 0.47]**

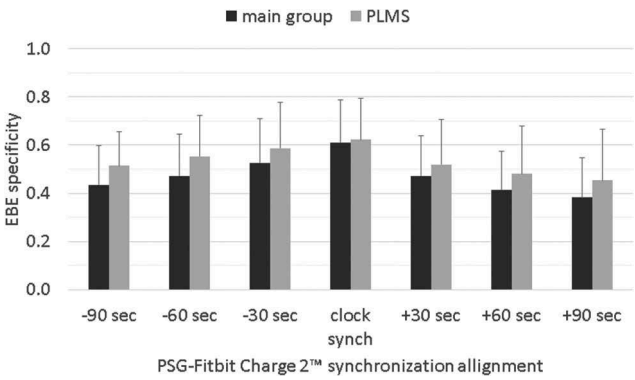

Figure 3. Epoch-by-epoch (EBE) specificity as a function of polysomnography (PSG)-Fitbit Charge 2™ epochs alignment in the main group (n = 35) and in the periodic limb movement of sleep (PLMS) group (n = 9).

of the sample fell outside the Bland–Altman agreement limits for the main sleep parameters. However, different populations (adolescents vs. adults) and experimental conditions (e.g., in the current study, participants had a single night in the laboratory compared to the ŌURA study, where an adaptation night preceded the experimental night) do not allow a direct comparison between study outcomes. The performance of Fitbit Charge 2™ in measuring sleep stage composition was presented, for the first time, at the 31st annual meeting of the Associated Professional Sleep Societies LLC (APSS) based on an internal validation from Fitbit, Inc. (Beattie et al. 2017). In that study, the authors investigated the accuracy of Fitbit Charge 2™ against PSG in "normal adult sleepers". From the published abstract (Beattie et al. 2017), authors reported an estimated Cohen's kappa coefficient of 0.52 ± 0.14 for the Fitbit Charge 2™ worn on the left wrist and 0.53 ± 0.14 if the device was worn on the right wrist, values that can be interpreted as reflecting "moderate" agreement between Fitbit Charge 2™ and PSG. They also reported an EBE accuracy (PSG epochs of wake and sleep correctly identified by Fitbit Charge 2™) of 0.69 and no biases for wake, "light sleep", "deep sleep" and REM sleep against PSG. Due to the limited details in the published abstract and the different methodologies used, a direct comparison between studies would be misleading.

The main limitation of standard actigraphs and actigraphy-based consumer wearables is the relatively low specificity (accuracy in detecting wake). In the current study, Fitbit Charge 2™ showed a specificity of 0.61. Despite no accepted standard rules to determine what can be considered "good" or "poor" performance (de Zambotti et al. 2016b), a specificity of 0.61 is higher compared to that provided in previous validation studies assessing the accuracy of older Fitbit models, which relied on only motion to determine sleep/wake states [Fitbit "original", specificity of 0.20 (Montgomery-Downs et al. 2012); Fitbit Flex, specificity of 0.35 (Cook et al. 2017) and 0.36 (Kang et al. 2017); Fitbit Ultra, specificity of 0.52 (Meltzer et al. 2015); Fitbit Charge HR™, specificity of 0.42 (de Zambotti et al. 2016a)]. A specificity of 0.61 fits within the specificity range (0.30–0.67) among studies validating standard actigraphy against PSG in healthy sleepers (see Van De Water et al. 2011). On the other hand, whether 0.61 specificity can be considered acceptable or not, is still matter of debate. In interpreting the performance of Fitbit Charge 2™ compared with previous Fitbit models and other wearables, we cannot completely

exclude that factors like the population studied (e.g., children and adolescence vs. adults, presence or absence of sleep disorders), experimental design, and methodology could have influenced the specificity outcomes. For example, in contrast to our current validation study, the majority of the previous studies used a minute-by-minute epoch comparison between PSG and wearable outputs [i.e., 30-s PSG epochs were re-coded (W + W = W; W + S or S + W = W; S + S = S)] to match the 1-min resolution of Fitbit outputs. Thus, a 1-min PSG epoch could have been misclassified as wake if one of the two single 30-s PSG epoch contained >15 s of a single wake episode. This methodological limitation may have influenced previous specificity results.

On the other hand, it is plausible to believe that the relatively greater specificity in the current study reflects a positive trend for greater accuracy for the new generation of multisensory wearable devices due to the use of multiple sources of information, in addition to motion, resulting in greater capability to differentiate between sleep/wake state as well as sleep stages. For example, motion-based wearables, other than showing lower specificity outcomes, have been previously shown to have limited success when attempting to differentiate between sleep stages ("light sleep" vs. "deep sleep") (see de Zambotti et al. 2015; Mantua et al. 2016). Although Fitbit Charge 2™ uses typical proprietary algorithms for sleep staging, it is theoretically plausible to assume that the use of HR and HRV data is a key advantage for sleep stage differentiation (particularly REM detection) as well as for the detection of "quiet wakefulness" (when individuals lie down in bed awake without moving). Indeed, sleep-stage specific shifting in autonomic activity, as detected by HRV methods, is a well-established finding (Baharav et al. 1995), with EEG and HRV measures tightly coupled across the night (Otzenberger et al. 1997). In addition, phasic sleep events such as arousals and k-complexes, which are used as markers of sleep stage transitions in standard PSG sleep scoring (see Iber 2007), are accompanied by stereotypical HR fluctuations (de Zambotti et al. 2016c; Trinder et al. 2003). Few studies have directly investigated the isolated and combined effect of motion and HRV information for automatic sleep-wake classification, with classification performance relying on the population, classification method and set of features used (Aktaruzzaman et al. 2017; Devot et al. 2010; Karlen et al. 2008). Thus, the magnitude of the isolated and combined contribution of motion, HR and HRV information in the accuracy of wake detection and sleep stage differentiation needs to be determined, as well as investigating the potential advantage of other bio-signals (e.g., skin temperature, respiration) in sleep stage classification.

With some exceptions (see Kang et al. 2017, for an example), despite recognizing the presence and absence of clinical sleep disorders, the majority of previous validation studies do not specify the presence of PSG PLMS. The presence of PSG PLMS > 15/h (as a PSG finding) has a high prevalence in the general population (Haba-Rubio et al. 2016). As reported in the literature (Haba-Rubio et al. 2016), PLMS episodes can be associated with arousals, and are accompanied by autonomic correlates such as heart rate fluctuations (acceleration followed by deceleration) (see Sforza et al. 1999). Given that Fitbit Charge 2™ relies on motion-related and HR-related measures to estimate sleep outcomes, in our study we analyzed participants with PSG PLMS > 15/h as a separate group. In this small PLMS group (n = 9), Fitbit Charge 2™ showed similar or for some parameters, slightly worse performance (13% and 12%, on average, reduced accuracy in detecting "deep sleep" and REM sleep, respectively) compared to outcomes of the main group. While a direct group comparison was not possible, due to the small sample size of the PLMS group, the presence of PLMS may have an impact on the accuracy of Fitbit Charge 2™ and requires further attention. Future studies also need to investigate the performance of Fitbit Charge 2™ in populations where the presence of PLMS reaches clinical significance, such as in periodic limb movement disorder, and in populations with other sleep disorders, such as obstructive sleep apnea syndrome, associated with frequent nocturnal arousals.

Our analysis of a particular methodological challenge when conducting validation studies, such as the synchronization of wearable devices and the PSG system (not always straightforward) shows that it is critical to accurately synchronize devices, since slight deviations in synchrony may affect performance

outcomes (drop in specificity). As evident in Figure 3, misalignment of 1–3 epochs (30–90 s) resulted in reduced specificity compared to that calculated when PSG and Fitbit were accurately aligned.

One issue related to the potential use of wearables in research and clinical settings is the amount of data loss. Mantua et al. (2016) found several hardware/software malfunctions (seven cases for Misfit Shine™, four cases for Actiwatch Spectrum®, and three cases for Basic Health Tracker 2014 Ed.) when testing multiple devices against PSG, and Baroni et al. (2016) reported that that Fitbit Flex™ is unreliable for recordings over prolonged period of time (authors reported a large amount of "missing data"). In our current study, of over 46 recordings, we recognized two cases (4.3%) in which Fitbit Charge 2™ data were unusable (see specification in the method section). The rate of failure in collecting reliable sleep data need to be further investigated in non- or less-controlled settings such as in at-home environments, over prolonged period of recordings.

A large number of US individuals use wearable technology, and millions of data points are available, but the role that wearable technology plays in the medical and research community is still not defined (see Piwek et al. 2016). Given the fast diffusion of commercial sleep-trackers and the growing adoption of these devices in research and clinical sleep settings (see https://www.fitabase.com/research-library/, for an example), we would like to emphasize the need for standard rules to evaluate and quantify the accuracy of consumer sleep-trackers. Particularly, the methods to evaluate the accuracy of a device (e.g., Bland–Altman plots, EBE analysis), specific methodological guidelines on how to conduct a validation study and what is considered an acceptable performance metric (e.g., Bland–Altman biases, Bland–Altman agreement limits, sensitivity and specificity) need to be determined.

In conclusion, Fitbit Charge 2™ shows promise in detecting sleep-wake states and sleep stage composition relative to gold standard PSG, particularly in the estimation of REM sleep, but with limitations in N3 detection. Fitbit Charge 2™ accuracy and reliability need to be further investigated in different settings and in different populations in which sleep composition is known to vary.

Acknowledgments

We would like to thank our sleep lab staff Leonardo Rosas, Vanessa Alschuler, Yun Qi Lim, and Maureen Gil for their efforts in data collection, and our participants for their valuable time.

Declaration of Interest statement

This study was funded by Fitbit, Inc. The company provided epoch-by-epoch sleep data from the Fitbit devices, blinded from PSG results; Fitbit had no involvement in data analysis, interpretation, and scientific writing. The manuscript solely reflects the views of the authors. MdZ, FB, and IC have received research funding unrelated to this work from Ebb Therapeutics Inc., and International Flavors & Fragrances Inc.

References

Aktaruzzaman M, Rivolta M, Karmacharya R, Scarabottolo N, Pugnetti L, Garegnani M, Bovi G, Scalera G, Ferrarin M, Sassi R. 2017. Performance comparison between wrist and chest actigraphy in combination with heart rate variability for sleep classification. Comput Biol Med. 89:212–21.

American Academy of Sleep Medicine. 2014. International classification of sleep disorders. *3rd ed*. Darien, IL: American Academy of Sleep Medicine.

Baharav A, Kotagal S, Gibbons V, Rubin B, Pratt G, Karin J, Akselrod S. 1995. Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability. Neurology. 45:1183–87.

Baroni A, Bruzzese J, Di Bartolo C, Shatkin J. 2016. Fitbit flex: An unreliable device for longitudinal sleep measures in a non-clinical population. Sleep Breath. 20:853–54.

Bastien C, Vallières A, Morin C. 2001. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med. 2:297–307.

Beattie Z, Pantelopoulos A, Ghoreyshi A, Oyang Y, Statan A, Heneghan C. 2017. Estimation of sleep stages using cardiac and accelerometer data from a wrist-worn device. Sleep. 40:A26.

Bland J, Altman D. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1:307–10.

Buysse D, Reynolds III C, Monk T, Berman S, Kupfer D. 1989. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. Psychiatry Res. 28:193–213.

Cook J, Prairie M, Plante D. 2017. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy. J Affect Disord. 217:299–305.

de Zambotti M, Baker F, Colrain I. 2015. Validation of sleep-tracking technology compared with polysomnography in adolescents. Sleep. 38:1461–68.

de Zambotti M, Baker F, Willoughby A, Godino J, Wing D, Patrick K, Colrain I. 2016a. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiol Behav. 158:143–49.

de Zambotti M, Godino J, Baker F, Cheung J, Patrick K, Colrain I. 2016b. The boom in wearable technology: Cause for alarm or just what is needed to better understand sleep?. Sleep. 39:1761–2.

de Zambotti M, Rosas L, Colrain I, Baker F. 2017. The Sleep of the Ring: Comparison of the ŌURA Sleep Tracker Against Polysomnography. Behav Sleep Med. 21:1–15.

de Zambotti M, Willoughby AR, Franzen PL, Clark DB, Baker FC, Colrain IM. 2016c. K-Complexes: Interaction between the Central and Autonomic Nervous Systems during Sleep. Sleep. 39:1129–37.

Devot S, Dratwa R, Naujokat E (2010). Sleep/wake detection based on cardiorespiratory signals and actigraphy. Conference of the IEEE Engineering in Medicine and Biology Society pp. 5089; Buenos Aires, Argentina.

Feinberg I, Floyd T. 1979. Systematic trends across the night in human sleep cycles. Psychophysiology. 16:283–91.

Grandner M. 2017. Sleep, health, and society. Sleep Med Clin. 12:1–22.

Haba-Rubio J, Marti-Soler H, Marques-Vidal P, Tobback N, Andries D, Preisig M, Waeber G, Vollenweider P, Kutalik Z, Tafti M. 2016. Prevalence and determinants of periodic limb movements in the general population. Ann Neurol. 79:464–74.

Iber C. 2007. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. Westchester, IL: American Academy of Sleep Medicine.

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

Karlen W, Mattiussi C, Floreano D. 2008. Improving actigraph sleep/wake classification with cardio-respiratory signals. In *Conference of the IEEE Engineering in Medicine and Biology Society*. 5262; Vancouver, BC, Canada.

Kolla BP, Mansukhani S, Mansukhani MP. 2016. Consumer sleep tracking devices: A review of mechanisms, validity and utility. Expert Rev Med Devices. 13:497–506.

Kryger MH, Roth T, Dement WC. 2017. Principles and practice of sleep medicine. Philadelphia, PA: Elsevier.

Mantua J, Gravel N, Spencer RM. 2016. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors. 16:646.

Meltzer L, Hiruma L, Avis K, Montgomery-Downs H, Valentin J. 2015. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. Sleep. 38:1323–30.

Montgomery-Downs H, Insana S, Bond J. 2012. Movement toward a novel activity monitoring device. Sleep Breath. 16:913–17.

Otzenberger H, Simon C, Gronfier C, Brandenberger G. 1997. Temporal relationship between dynamic heart rate variability and electroencephalographic activity during sleep in man. Neurosci Lett. 229:173–76.

Paquet J, Kawinska A, Carrier J. 2007. Wake detection capacity of actigraphy during sleep. Sleep. 30:1362–69.

Piwek L, Ellis DA, Andrews S, Joinson A. 2016. The rise of consumer health wearables: Promises and barriers. PLoS Med. 13:e1001953.

Roane B, Van Reen E, Hart C, Wing R, Carskadon M. 2015. Estimating sleep from multisensory armband measurements: Validity and reliability in teens. J Sleep Res. 24:714–21.

Sforza E, Nicolas A, Lavigne G, Gosselin A, Petit D, Montplaisir J. 1999. EEG and cardiac activation during periodic leg movements in sleep: Support for a hierarchy of arousal responses. Neurology. 52:786–91.

Toon E, Davey M, Hollis S, Nixon G, Horne R, Biggs S. 2015. Comparison of commercial wrist-based and smartphone accelerometers, actigraphy, and PSG in a clinical cohort of children and adolescents. J Clin Sleep Med. 12:343–50.

Trinder J, Allen N, Kleiman J, Kralevski V, Kleverlaan D, Anson K, Kim Y. 2003. On the nature of cardiovascular activation at an arousal from sleep. Sleep. 26:543–51.

Van De Water A, Holmes A, Hurley D. 2011. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography-a systematic review. J Sleep Res. 20:183–200.

Ware J Jr., Sherbourne C. 1992. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 30:473–83.