

Original Paper

Accuracy of Fitbit Wristbands in Measuring Sleep Stage Transitions and the Effect of User-Specific Factors

Zilu Liang^{1,2}, PhD; Mario Alberto Chapa-Martell³, PhD

¹School of Engineering, Kyoto University of Advanced Science, Kyoto, Japan

²Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

³Advanced Technology Division, CAC Corporation, Tokyo, Japan

Corresponding Author:

Zilu Liang, PhD

School of Engineering

Kyoto University of Advanced Science

18 Yamanouchi Gotanda-Cho

Kyoto, 6158577

Japan

Phone: 81 8040866433

Email: z.liang@csl.t.u-tokyo.ac.jp

Abstract

Background: It has become possible for the new generation of consumer wristbands to classify sleep stages based on multisensory data. Several studies have validated the accuracy of one of the latest models, that is, Fitbit Charge 2, in measuring polysomnographic parameters, including total sleep time, wake time, sleep efficiency (SE), and the ratio of each sleep stage. Nevertheless, its accuracy in measuring sleep stage transitions remains unknown.

Objective: This study aimed to examine the accuracy of Fitbit Charge 2 in measuring transition probabilities among wake, light sleep, deep sleep, and rapid eye movement (REM) sleep under free-living conditions. The secondary goal was to investigate the effect of user-specific factors, including demographic information and sleep pattern on measurement accuracy.

Methods: A Fitbit Charge 2 and a medical device were used concurrently to measure a whole night's sleep in participants' homes. Sleep stage transition probabilities were derived from sleep hypnograms. Measurement errors were obtained by comparing the data obtained by Fitbit with those obtained by the medical device. Paired 2-tailed *t* test and Bland-Altman plots were used to examine the agreement of Fitbit to the medical device. Wilcoxon signed-rank test was performed to investigate the effect of user-specific factors.

Results: Sleep data were collected from 23 participants. Sleep stage transition probabilities measured by Fitbit Charge 2 significantly deviated from those measured by the medical device, except for the transition probability from deep sleep to wake, from light sleep to REM sleep, and the probability of staying in REM sleep. Bland-Altman plots demonstrated that systematic bias ranged from 0% to 60%. Fitbit had the tendency of overestimating the probability of staying in a sleep stage while underestimating the probability of transitioning to another stage. SE > 90% (*P* = .047) was associated with significant increase in measurement error. Pittsburgh sleep quality index (PSQI) < 5 and wake after sleep onset (WASO) < 30 min could be associated to significantly decreased or increased errors, depending on the outcome sleep metrics.

Conclusions: Our analysis shows that Fitbit Charge 2 underestimated sleep stage transition dynamics compared with the medical device. Device accuracy may be significantly affected by perceived sleep quality (PSQI), WASO, and SE.

(JMIR Mhealth Uhealth 2019;7(6):e13384) doi: [10.2196/13384](https://doi.org/10.2196/13384)

KEYWORDS

wearable electronic devices; sleep; validation studies

Introduction

Importance of Consumer Sleep Tracking Devices

Having enough restorative sleep is essential for physical and mental health [1]. In recent years, consumer sleep-monitoring wristbands and associated mobile phone apps have created an effective way for individuals to understand personal sleep patterns or improve sleep quality in daily settings [2]. These devices are relatively affordable, easy to use, and ready to purchase in the consumer market. Most of the consumer wristbands rely on a similar mechanism of clinical actigraphy that infers wake and sleep cycles from limb movement [2]. Newly launched models also incorporate other streams of biosignals, such as heart rate to measure sleep stages. Users can visualize a whole night's sleep hypnogram (the temporal sequence of sleep stages) and the aggregated sleep parameters, such as total sleep time (TST) and the ratio of each sleep stage on a dashboard [3]. There is increasing evidence that consumer sleep-monitoring wristbands raise awareness of sleep health and have a positive impact on personal sleep hygiene [4-6], though the long-term impact of these technologies has not been elucidated [7]. In the meantime, researchers and clinicians are increasingly adopting consumer wristbands, such as Fitbit devices, as outcome measurement tools in research studies [6,8-14]. Compared with traditional polysomnography (PSG), Fitbit devices significantly reduce the time and monetary cost for longitudinal sleep data collection, and they could provide rich information that was not possible to collect outside sleep laboratories or clinics in the past. Participants can use the devices under free-living conditions, without the need of constant technical support. The new generation of Fitbit devices could also possibly outperform clinical actigraphy, as they leverage multiple streams of biosignals for sleep staging, whereas actigraphy is only able to detect wake and sleep on the basis of limb movement [15].

Accuracy of Consumer Sleep Tracking Devices

As consumer sleep-monitoring wristbands continue to gain popularity, their limitation in measurement accuracy raised wide concerns on the quality of data collected using these devices [7,16,17]. Data of low quality may mislead users to arrive at wrong conclusions of their sleep. In addition, data quality is of top priority for researchers who intend to use these devices in scientific studies. Therefore, understanding the validity of consumer sleep trackers has practical benefit for both individual users and for the research community. In response to this need, many studies have examined the accuracy of popular sleep trackers compared with medical devices in terms of aggregated sleep metrics, including TST, wake after sleep onset (WASO), sleep efficiency (SE), and sleep stages, that is, light sleep, deep sleep, and rapid eye movement (REM) sleep [18-24]. These studies show that the previous models of consumer wristbands have a common problem of overestimating sleep and underestimating wake [18-20]. Recent models, such as Fitbit

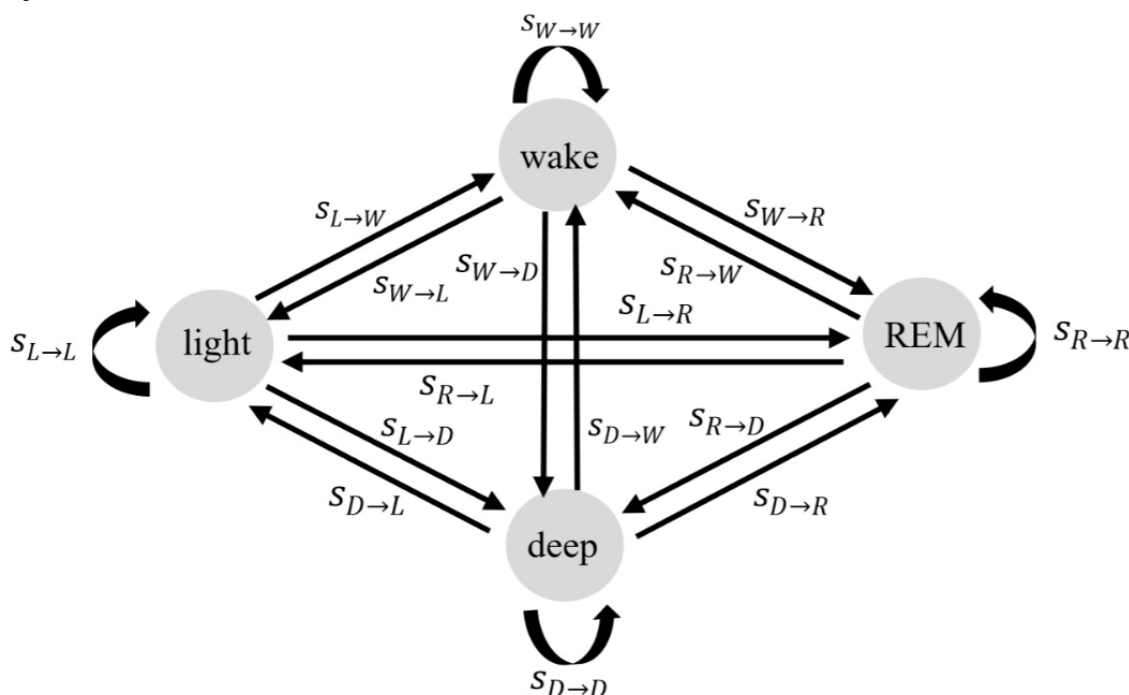
Charge 2, that rely on multistreams of biosignals have satisfying performance in measuring TST and SE but fail to produce accurate results in classifying sleep stages [21,24].

Although the main body of validation studies has been dominantly focused on polysomnographic metrics (eg, TST, WASO, sensitivity, and specificity) [2,13,24-27], the performance of consumer wristbands in measuring sleep stage transitions remains unknown. Sleep research has shown that sleep stage transition probabilities comprise rich information of sleep patterns, which have been considered more effective than polysomnographic parameters in characterizing sleep stability [28-37]. Sleep stage transition abnormality is an important indicator of sleep disorders [28,32,33,38-43]. Some studies also relied on sleep stage transition probabilities to assess the effect of treatment [44]. The clinical significance of sleep stage transition dynamics suggests the necessity of including relevant metrics (sleep stage transition probabilities) as outcome sleep parameters in validation studies. In Figure 1, a visualization of sleep stage transition dynamics is presented. The total transition probability from a single state to other states (including staying in the same state) is always 1. The $s_{X \rightarrow Y}$ represents the transition probability from sleep stage X to Y . The $\{ X, Y \}$ are derived from $\{ W, L, D, R \}$, which are abbreviations for wake, light sleep, deep sleep, and REM sleep. For example, $s_{W \rightarrow R}$ denotes the transition probability from wake to REM sleep, and $s_{W \rightarrow W}$ denotes the probability of staying in wake.

Significance of This Study

This study aimed to examine whether it would accurately measure sleep stage transitions (the transition probabilities among waking, light, deep, and REM sleep) using Fitbit Charge 2. Despite the abundant validation studies, the accuracy of consumer wristbands in measuring sleep stage transition has not been investigated. We also examined the factors that are associated with the measurement errors on sleep stage transition probabilities. Previous validation studies on other types of wearable devices found that device accuracy could vary as a function of the underlying sleep patterns, the population studied, and even how the measurand was defined [45-48]. Along the same line, we selected a set of independent variables (possible predictors), including demographic characteristics of participants, subjective sleep quality measured by Pittsburgh Sleep Quality Index (PSQI) [49], and objective sleep quality derived from medical data. The dependent variables were the absolute percent errors of Fitbit Charge 2 on sleep stage transition probabilities compared with the medical device. The outcomes of this study complement previous validation studies and contribute to the establishment of a holistic view of the capacity of consumer wristbands in measuring sleep structure under free-living conditions. This study also establishes a preliminary reference for researchers who intend to use Fitbit to measure sleep stage transitions and for individual users who rely on Fitbit sleep data to make health decisions.

Figure 1. Sleep stage transition dynamics. The W, L, D, R in the subscripts denotes the abbreviation of wake, light sleep, deep sleep, and rapid eye movement sleep.



Methods

Recruitment

We recruited participants by distributing posters around the campus of The University of Tokyo. In total, 38 people registered interest through a Web-based form, of whom 28 (74%) were eligible to participate in the study. The inclusion criteria required that the participants were adults (age > 18 years), were free of diagnosed chronic conditions, and were able to attend a briefing before the data collection phase. This research was approved by the ethical committee of the University of Tokyo. All participants provided informed consent.

Study Procedures

A face-to-face briefing was held with each participant individually before the data collection phase. In this meeting, we installed the Fitbit app on participants' mobile phones and provided verbal instructions on how to use the devices and how to synchronize the Fitbit device with its mobile phone app. Participants were provided with the following items for data collection: a Fitbit Charge 2, a medical device named Sleep Scope, electrodes, chargers, and manuals. At the end of the briefing, participants were asked to fill in a PSQI questionnaire [49] to measure their perceived sleep quality. The PSQI is a widely used instrument for assessing subjective sleep quality averaged over the past 1 month, and a $PSQI \geq 5$ is indicative of perceived poor sleep. We collected the PSQI, as it may associate to the measurement accuracy of Fitbit. More details on potential association factors of measurement accuracy will be provided in the next section.

After the briefing, participants measured their sleep using both devices for 3 consecutive nights in their homes to ensure that Fitbit Charge 2 was evaluated in an ecologically valid setting. They were asked to wear the Fitbit on the nondominant wrist

during data collection. All participants received a monetary reward when they returned the devices after data collection.

Data Collection

In this study, we collected sleep data concurrently using Fitbit Charge 2 and a medical device. Fitbit Charge 2 (Fitbit Inc) is a wearable activity wristband with an embedded triaxial accelerometer. It estimates sleep stages for each 30 second period by integrating a user's movement and heart rate data. With advances in software and hardware, Fitbit Charge 2 has overcome some problems of previous models, and it is able to measure TST and SE with good accuracy [21,24]. A medical sleep monitor named Sleep Scope (Sleep Well Co) was used to obtain the ground truth on sleep hypnograms. Sleep Scope is a clinical-grade single-channel electroencephalogram (Japanese Medical Device Certification 225ADBZX00020000), which was validated against PSG (agreement = 86.9%, average Cohen Kappa value = 0.75) [50,51]. Sleep Scope was chosen over PSG as it enabled data collection in participants' homes rather than in a sleep laboratory. This ensures that Fitbit Charge 2 was evaluated in an ecologically valid setting; this also ensures minimalizing the possible disruption of sleep by unfamiliar environment.

In the data collection phase, participants tracked their sleep for 3 consecutive nights in their homes. Following the common practice in sleep science, we analyzed the second night for each participant to remove *the first night effect* [52,53]. If the data of the second night were not valid, then the data of the third night were analyzed. The data of the first night were only selected when neither the second night nor the third night was valid.

Fitbit sleep data were retrieved through the application program interface (API) of Fitbit. Fitbit Charge 2 provides sleep data at 2 levels through public API. The *stage level* data comprise sleep

stage levels, including wake, light sleep, deep sleep, and REM sleep. These data are aggregated at 30-second granularity, which complies with the standard sleep staging in the clinical setting. If the *stage level* data are not available, the *classic level* data will be provided as an alternative. *Classic level* data comprise sleep pattern levels, including asleep, restless, and awake, and they are aggregated at a coarser granularity of 60 seconds. In this study, we were interested in the *stage level* sleep data, and the *classic level* data were discarded, as they contained no information on deep sleep, light sleep, and REM sleep.

The data of the medical device were analyzed by the Sleep Well Company, using proprietary automatic scoring algorithms, followed by epoch-by-epoch visual inspection by specialists on the basis of established standards [54], and corrections were added if needed. Fitbit data and medical data were synchronized to make sure that the start time was aligned.

To examine the effect of user-specific factors on measurement accuracy, we also collected data on the factors listed in [Table 1](#). Age and sex were based on self-report, and PSQI was measured by the PSQI questionnaire [49]. Sleep quality metrics were all derived from the medical data.

Table 1. A full list of user-specific factors.

Factors	Data type	Data collection method	Cut-off threshold
Age (years)	Ordinal	Self-reported	25
Sex	Nominal	Self-reported	Female or male
PSQI ^a	Ordinal	PSQI questionnaire	5
TST ^b (min)	Continuous	Sleep scope (medical device)	360
WASO ^c (min)	Continuous	Sleep scope	30
SOL ^d (min)	Continuous	Sleep scope	30
SE ^e , %	Continuous	Sleep scope	90.0
Light sleep, %	Continuous	Sleep scope	65.0
SWS ^f , %	Continuous	Sleep scope	20.0
REM ^g , %	Continuous	Sleep scope	20.0
T_{avg}^h (min)	Continuous	Sleep scope	90

^aPSQI: Pittsburgh Sleep Quality Index.

^bTST: total sleep time.

^cWASO: wake after sleep onset.

^dSOL: sleep onset latency.

^eSE: sleep efficiency.

^fSWS: slow wave sleep.

^gREM: rapid eye movement sleep.

^h T_{avg} : average sleep cycle.

Statistical Analysis

The overall goal of the analysis was two-fold. We aimed to examine the accuracy of Fitbit Charge 2 in measuring sleep stage transitions compared with a medical device. We were also interested in the associations of user-specific factors with the measurement accuracy of Fitbit Charge 2. All statistical significance levels reported were 2 sided, and statistical analysis was performed using R statistical software version 3.5.3 (The R Foundation)[55].

First, descriptive statistics of sleep parameters were derived from the medical data. Paired 2-tailed *t* test was used to probe if there were statistically significant differences on sleep patterns between men and women, as well as between participants below 25 years of age and above 25 years of age. Second, sleep stage

transition probabilities were calculated by dividing the number of transitions from a specific sleep state to a specific sleep state by the total number of transitions from that specific state to all sleep states (including staying in the same state). As shown in [Figure 2](#), $\{X, Y, \text{ and } B\}$ are derived from $\{W, L, D, \text{ and } R\}$ and $n_{X \rightarrow Y}$ is the number of transitions from sleep stage *X* to *Y* during a whole night's sleep. The *W*, *L*, *D*, and *R* are the abbreviations for wake, light sleep, deep sleep, and REM sleep. Sleep stage transition probabilities were calculated from Fitbit data and medical data for each participant and then averaged over the whole cohort to obtain the average sleep stage transition probabilities. Systematic difference between the 2 devices was assessed by applying paired *t* test on the sleep stage transition probabilities. A *P* value below .05 was considered statistically significant. The level of agreement between 2 devices was examined using the Bland-Altman plots [56].

Figure 2. The calculation of sleep stage transition probabilities.

$$S_{X \rightarrow Y} = \frac{n_{X \rightarrow Y}}{\sum_B n_{X \rightarrow B}} \times 100$$

Figure 3. The calculation of absolute percent error.

$$e_{X \rightarrow Y} = \left| \frac{S_{X \rightarrow Y}^F - S_{X \rightarrow Y}^M}{S_{X \rightarrow Y}^M} \right| \times 100$$

The absolute percent error $e_{X \rightarrow Y}$ was calculated using the equation in Figure 3, where $\{X, Y, \text{ and } B\}$ are derived from $\{W, L, D, \text{ and } R\}$, $s_{X \rightarrow Y}^F$ and $s_{X \rightarrow Y}^M$ are the transition probability from sleep stage X to Y , derived from Fitbit data and medical data.

To examine the effect of user-specific factors on absolute percent error, the dataset was divided into 2 subsets according to the cut-off threshold values listed in Table 1. Wilcoxon signed-rank test was conducted to examine if there were significant differences between the 2 subsets in terms of the outcome sleep metrics (sleep stage transition probabilities). The selection of cut-off threshold values was in line with literature in sleep science [49,57].

Results

Descriptive Statistics

A total of 28 young adults without chronic diseases participated in the study. A total of 5 participants were excluded from analysis because of failure to obtain *stage level* sleep data with Fitbit. That is, only *classic level* sleep data were obtained from these participants; the data had no information on light, deep, and REM sleep. Therefore, it was not possible to calculate sleep stage transition probabilities for these participants. The final dataset thus comprises sleep data from 23 participants (men:women=14:9). This number of participants is comparable with other validation studies [20,27,58-61]. All the participants were university students between 21 to 30 years old (mean 24.3, SD 2.7). A total of 8 out of the 23 participants had a PSQI higher than 5, which was indicative of unsatisfied sleep quality. Statistically significant differences were found between men

and women in terms of wake time (women: 9.7 min; men: 22.8 min; $P=.02$) and the ratio of sleep stage 1 (women: 7.7%; men: 14.3%; $P=.02$). We also compared the sleep patterns between participants below and above 25 years. Statistically significant differences were found in terms of TST (below 25 years: 308.7 min; above 25 years: 396.8 min; $P=.03$), transition probability from deep sleep to light sleep (below 25 years: 5.5%; above 25 years: 1.5%; $P=.02$), and the probability of staying in light sleep (below 25 years: 85.3%; above 25 years: 94.8%; $P=.008$).

Systematic Differences

Table 2 presents the estimated sleep stage transition probabilities derived from medical data and Fitbit data, as well as the results of paired t test. We calculated sleep stage transition probabilities individually for each participant and then averaged results across the whole cohort. It is shown that the following transitions rarely occurred: deep sleep to REM sleep and wake, light sleep to REM sleep, REM sleep to deep sleep, and REM sleep to light sleep. The t test results indicated that there were significant differences between the sleep stage transition probabilities measured by Fitbit and those measured by the medical device. Fitbit deviated from the medical device on all the transition probabilities except for the transition probability from light sleep to REM sleep ($s_{L \rightarrow R}^F = 0.9\%$; $s_{L \rightarrow R}^M = 1.7\%$), the transition probability from deep sleep to wake ($s_{D \rightarrow W}^F = s_{D \rightarrow W}^M = 0.2\%$), and the probability of staying in REM sleep stage ($s_{R \rightarrow R}^F = s_{R \rightarrow R}^M = 96.9\%$). In general, Fitbit underestimated sleep stage transition dynamics. The probabilities of staying in a specific sleep stage were significantly overestimated, whereas the probabilities of transitions from a specific stage to a different stage were mostly underestimated.

Table 2. Average sleep stage transition probabilities (%) and results of paired *t* test. Data are displayed as mean and $\pm 95\%$ CI.

Sleep stage	Wake	Light	Deep	REM ^a
Wake				
Medical	53.7 (44.0-63.3)	43.6 (33.8-53.4)	0.2 (0.0-0.4)	2.6 (1.5-3.7)
Fitbit	89.8 (81.2-98.3)	5.5 (4.3-6.7)	0.2 (0.0-0.5)	0.2 (0.0-0.5)
<i>P</i> value	<.001	<.001	.83	<.001
Light				
Medical	2.6 (2.0-3.3)	92.6 (90.9-94.4)	3.9 (2.1-5.8)	0.8 (0.7-0.9)
Fitbit	0.5 (0.3, 0.6)	97.8 (97.6-98.1)	1.1 (0.9-1.3)	0.5 (0.3-0.7)
<i>P</i> value	<.001	<.001	.005	.02
Deep				
Medical	2.5 (0.7-4.3)	57.7 (43.8-71.6)	35.5 (22.6-48.4)	0.0 (0.0-0.0)
Fitbit	0.2 (0-1.8)	3.8 (2.9-4.6)	94.9 (93.4-96.4)	1.1 (0.4-1.8)
<i>P</i> value	.02	<.001	<.001	.002
REM				
Medical	2.0 (1.6-2.4)	0.9 (0.7-1.2)	0.0 (0.0-0.0)	96.9 (96.5-97.5)
Fitbit	0.1 (0.0-0.2)	1.7 (0.7-2.6)	1.2 (0.3-2.2)	96.9 (96.0-98.0)
<i>P</i> value	<.001	.14	.01	>.99

^aREM: rapid eye movement.

Level of Agreement and Correlations

Figures 4-6 show the Bland-Altman plots comparing Fitbit Charge 2 with the medical device. Device discrepancies for sleep outcomes are plotted as a function of the medical outcomes for each individual. The mean bias ranged from 0% ($s_{R \rightarrow R}$ and $s_{D \rightarrow W}$) to approximately 60% ($s_{L \rightarrow D}$). No more than 2 participants were situated outside the lower limit of agreement or the upper limit of agreement.

In line with previous studies [62,63], we defined the acceptable error range as $e_i \leq 5\%$, as this approximates a widely acceptable standard for statistical significance in literature [64]. On the

basis of this criterion, no systematic bias was found between Fitbit and the medical device in measuring $s_{W \rightarrow L}$, $s_{W \rightarrow R}$, $s_{L \rightarrow R}$, $s_{D \rightarrow W}$, $s_{R \rightarrow L}$, $s_{R \rightarrow D}$, and $s_{R \rightarrow R}$.

Figure 4 shows that no trend was found between the difference and the mean of $s_{R \rightarrow L}$, $s_{L \rightarrow R}$ and $s_{R \rightarrow R}$. In contrast, Figure 5 and Figure 6 show clear trends that the measurement differences were greater for lower $s_{L \rightarrow L}$, $s_{D \rightarrow D}$, and $s_{W \rightarrow W}$, and the differences were greater for higher $s_{W \rightarrow L}$, $s_{W \rightarrow R}$, $s_{W \rightarrow D}$, $s_{L \rightarrow W}$, $s_{L \rightarrow D}$, $s_{D \rightarrow W}$, $s_{D \rightarrow L}$, $s_{D \rightarrow R}$, $s_{R \rightarrow W}$, and $s_{R \rightarrow D}$. These findings suggest that the accuracy of Fitbit Charge 2 in measuring sleep stage transitions could be deteriorated as sleep became more dynamic (more transitions between different sleep stages).

Figure 4. Bland-Altman plots assessing the level and limits of agreement between Fitbit Charge 2 and medical device on the transition probabilities from rapid eye movement (REM) sleep to light sleep, from light sleep to REM sleep, and the probability of staying in REM sleep. The dashed line in the middle represents the mean difference, whereas the upper and lower dashed lines represent the upper limit of agreement and the lower limit of agreement.

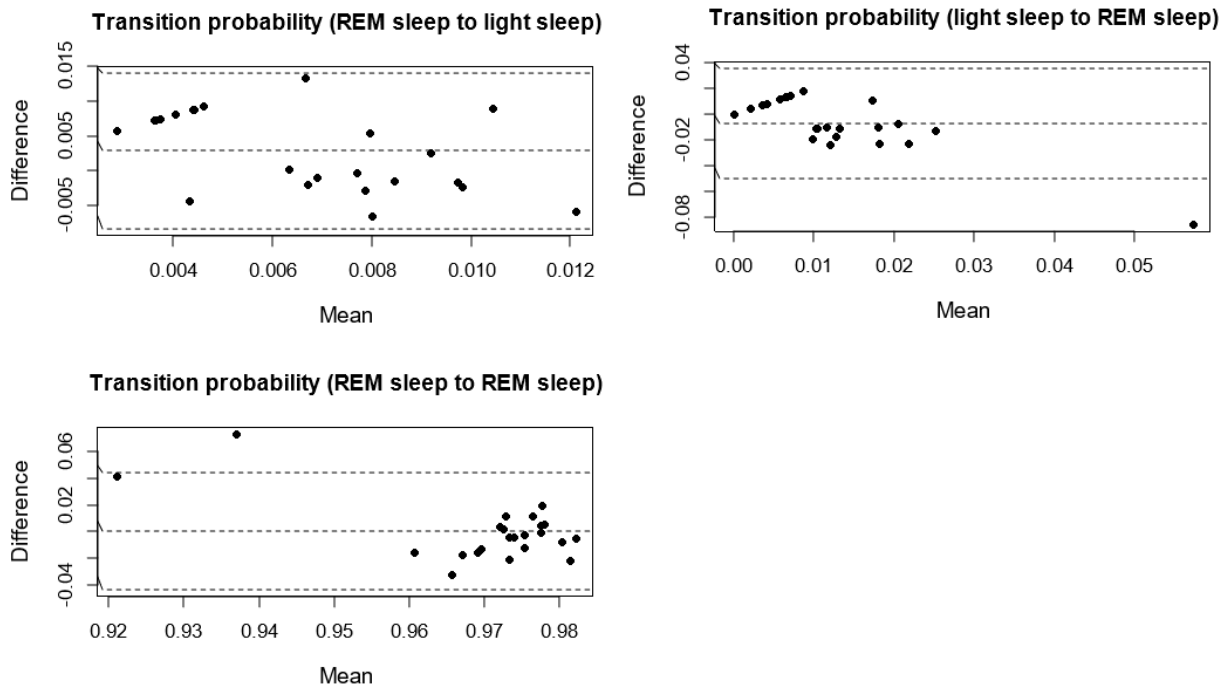


Figure 5. Bland-Altman plots assessing the level and limits of agreement between Fitbit Charge 2 and medical device on the probability of staying in light sleep, in deep sleep, and in wake.

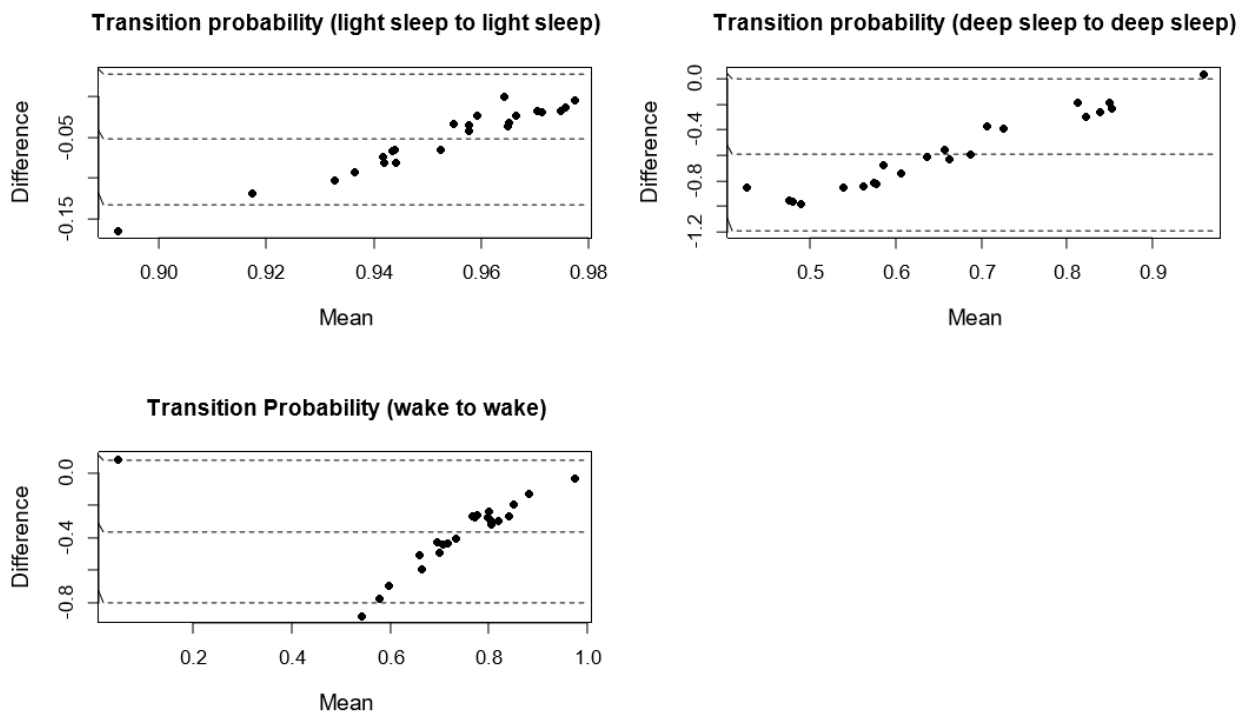
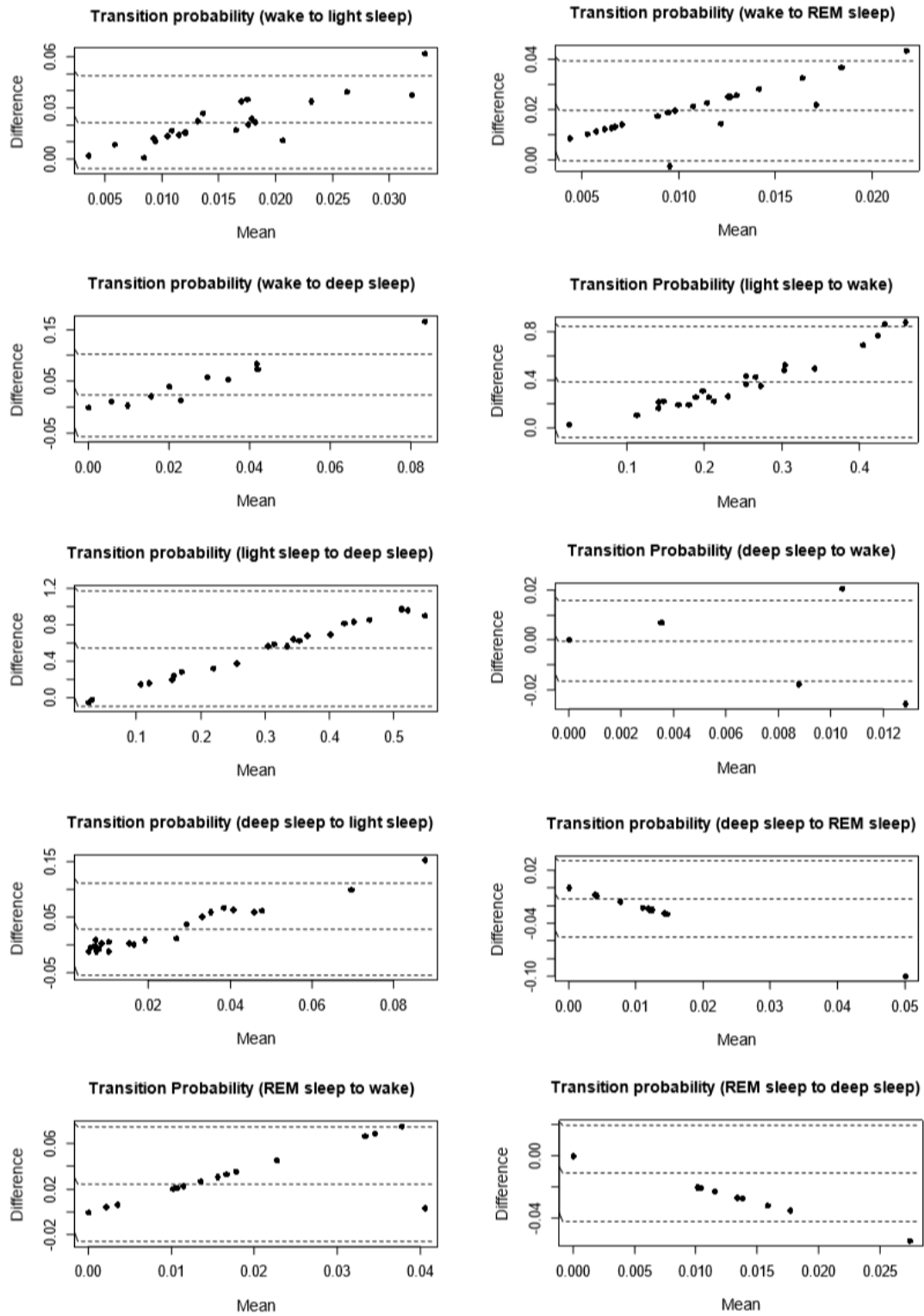


Figure 6. Bland-Altman plots assessing the level and limits of agreement between Fitbit Charge 2 and medical device on the transition probabilities from wake to light sleep, from wake to rapid eye movement (REM) sleep, from wake to deep sleep, from light sleep to wake, from light sleep to deep sleep, from deep sleep to wake, from deep sleep to light sleep, from deep sleep to REM sleep, from REM sleep to wake, and from REM sleep to deep sleep.



Effect of User-Specific Factors

The results of Wilcoxon signed-rank test showed that good subjective sleep quality indicated by PSQI as lower than 5 was associated with decreased errors in the probability of staying in deep sleep stage (PSQI<5, 132.1±173.1%; PSQI≥5, 346.8±250.0%; *P*=.04), but it was associated with increased

errors in transition probability from waking to REM sleep (PSQI<5, 100.0±0.0%; PSQI≥5, 85.1±25.5%; *P*=.02).

Wake time longer than 30 min was associated with increased errors in transition probability from light sleep to REM sleep (WASO≥30, 265.8±176.5; WASO<30, 103.9±49.1%; *P*=.02), but it was associated with decreased errors in transition

probability from light sleep to wake ($WASO \geq 30$, $78.6 \pm 10.2\%$; $WASO < 30$, $86.7 \pm 8.6\%$; $P = .049$), as well as the probability of staying in wake ($WASO \geq 30$, $117.3 \pm 269.5\%$; $WASO < 30$, $125.2 \pm 103.6\%$; $P = .006$).

SE above 90% was associated with increased measurement errors in transition probability from REM sleep to light sleep ($SE > 90$, $107.1 \pm 53.2\%$; $SE \leq 90\%$, $55.9 \pm 40.4\%$; $P = .047$).

In addition, age below 25 years (age < 25 , $7.9 \pm 5.4\%$; age ≥ 25 , $3.1 \pm 2.3\%$; $P = .01$), sleep onset latency (SOL) shorter than 30 min (SOL < 30 , $8.6 \pm 5.8\%$; SOL ≥ 30 , $4.1 \pm 3.4\%$; $P = .02$), and deep sleep ratio above 20% (slow wave sleep; SWS $< 20\%$, $3.9 \pm 3.5\%$; SWS ≥ 20 , $9.5 \pm 5.2\%$; $P = .007$) were associated with slight increased measurement error in the probability of staying in light sleep stage. Nevertheless, the average errors were no more than 10% in all the corresponding cases.

No significant associations were found between measurement errors of Fitbit and other factors, including sex, TST, SOL, light sleep ratio, REM sleep ratio, and T_{avg} .

Discussion

Principal Findings

We have demonstrated a numerical comparison on sleep stage transition probabilities between Fitbit Charge 2 and the medical device. The level and limits of agreement between the 2 types of devices were illustrated using Bland-Altman plots. The results of Wilcoxon signed-rank test were presented to demonstrate the associations between user-specific factors and measurement errors. This study generated 2 main findings. First, we found that Fitbit Charge 2 underestimated sleep stage transition dynamics compared with the medical device. Second, device accuracy was mainly associated with 3 user-specific factors: subjective sleep quality measured by PSQI, WASO, and SE.

Sleep stage transition analysis has been used to characterize sleep continuity and the temporal stability of non-REM and REM bouts in sleep science [28-30,32,40,44]. In this study, the sleep stage transition probabilities derived from the medical data demonstrated interesting patterns. As expected, the probability for any sleep stage to stay in the same stage was constantly higher than that for this stage to change to a different stage. Direct transition between deep sleep and REM sleep rarely happened. The probability of transitions from wake to deep sleep or from wake to REM sleep was low. Similarly, the probability of transition from deep sleep to wake was also low. These characteristics were consistent with findings reported in previous sleep studies on sleep stage transition patterns in healthy people [31,44].

Sleep stage transition is the result of complex interactions among many brain regions. Not being able to detect markers in brainwaves, such as k-complexes [54], consumer wristbands have limited performance in classifying sleep stages. Previous studies show that Fitbit Charge 2 devices significantly overestimated light sleep and underestimated deep sleep when validated in lab settings [21], whereas they underestimated deep sleep and overestimated light and REM sleep when validated under free-living conditions [24]. This study complements

previous findings and contributes new insights into Fitbit's capacity in capturing sleep stage transitions. Overall, we observed that Fitbit Charge 2 significantly deviated from the medical device in measuring sleep stage transition dynamics. Notably, the average probabilities of staying in wake stage and deep stage measured by Fitbit were significantly higher than those measured by the medical device. In contrast, Fitbit underestimated the probabilities of stage transitions from light sleep to wake and from light sleep to deep sleep. This is probably because of the misclassification of wake and deep sleep epochs to light sleep [21]. Systematic bias (between 40% and 60%) was illustrated in the Bland-Altman plots on these sleep stage transition probabilities. On the other hand, no systematic bias and mean difference were observed in measuring the probability of staying in REM sleep stage. This result provides complementary evidence to the finding in the study by De Zambotti et al [21] that Fitbit Charge 2 agreed well to medical devices in detecting REM sleep.

A unique aspect of this study is that we also examined the effect of user-specific factors and found multiple associations. Our analysis showed that subjective sleep quality measured by PSQI, wake after WASO, and SE were significantly strong predictors of measurement errors in sleep stage transition probabilities. Age, SOL, and deep sleep ratio were significant but weak predictors, whereas sex, TST, light sleep ratio, REM sleep ratio, and average sleep cycle were not associated with the measurement errors of Fitbit.

Despite the finding from previous validation studies that poor sleep quality is associated with deteriorated performance of sleep monitoring devices in measuring polysomnographic sleep metrics [21,25,65], this study reveals that the relationship is more complicated between sleep quality and device accuracy in measuring sleep stage transitions. Indeed, we found that good subjective sleep quality (PSQI < 5) was associated with decreased measurement error in the probability of staying in deep sleep stage, and less fragmented sleep ($WASO < 30$ min) was associated with decreased errors in transition probability from light sleep to REM sleep. Nevertheless, it is also found that good sleep characterized by quick sleep onset (SOL < 30 min), high ratio of deep sleep (SWS $> 20\%$), good subjective feeling (PSQI < 5), short awakenings ($WASO < 30$ min), and high SE ($SE > 90\%$) were associated with increased measurement errors in different outcome transition probabilities. This result contradicts previous findings on actigraphy that deteriorated sleep (eg, long WASO and SOL) increased measurement errors [21,25,65]. This disparity suggests that findings related to clinical actigraphy should not be generalized to consumer wristbands without further validation.

In addition, age was found to be a significant but weak predictor of measurement errors. Participants in the age range of 25 to 30 had decreased measurement errors in the probability of staying in light sleep stage compared with those younger than the age of 25. As age has been widely recognized as a significant factor that alters sleep patterns [43,57], the effect of age may also be traced back to the difference in underlying sleep patterns. The medical sleep data showed that younger participants generally had shorter sleep and higher sleep stage transition dynamics (transition from deep sleep to light sleep), which may

account for the increase in measurement errors. Nevertheless, this finding should not be generalized to a wide range of age groups because of the restricted sampling of age in this study. Further studies are needed to systematically examine the effect of age on device accuracy.

Our findings complement those of previous validation studies on consumer wristbands for sleep tracking in general. Fitbit Charge 2 has demonstrated satisfying performance in measuring TST and SE, but it remains incapable of classifying sleep stages with good accuracy [21,24]. Our findings show that Fitbit Charge 2 may also underestimate sleep transition dynamics, and it should thus be used with caution. This study establishes a preliminary reference for researchers who intend to use the Fitbit device to measure sleep stage transitions in scientific studies, and this study suggests that both perceived and objective sleep patterns may need to be considered when choosing sleep monitoring tools.

Limitations

This study is subject to the following limitations. First, the participants represent a young healthy population that was free of sleep disorders or chronic diseases. Therefore, the results cannot be generalized to older or clinical populations. Second,

the data collection phase was not longitudinal in nature, and only 1 night of sleep from each participant was analyzed. Thus, the results may fail to count intrapersonal variations. Third, the list of potential affecting factors investigated in this study was not exhaustive, and it may be affected by restricted sampling. Further research should address these limitations by including a diverse population, extending data collection duration, and examining the effect of other potential predictors of device accuracy.

Conclusions

We have demonstrated that Fitbit Charge 2 significantly underestimated sleep stage transition dynamics compared with the medical device and that measurement accuracy could be mainly affected by perceived sleep quality, sleep continuity, and SE. Despite the positive trend of enhanced accuracy for the latest consumer wearable sleep trackers, the limitation of these devices in detecting sleep stage transition dynamics needs to be recognized. As an outcome measurement tool, Fitbit Charge 2 may not be suited for research studies related to sleep stage transitions or for health care decision making. Further research should focus on enhancing the accuracy of these consumer wristbands in measuring not only polysomnographic parameters but also sleep stage transition dynamics.

Acknowledgments

This study was sponsored by a JSPS KAKENHI Grant-in-Aid for Research Activity Start-up (Grant Number 16H07469) and a JSPS KAKENHI Grant-in-Aid for Early Career Scientists (Grant Number 19K20141).

Conflicts of Interest

None declared.

References

1. Buysse DJ. Sleep health: can we define it? Does it matter? *Sleep* 2014 Jan 01;37(1):9-17 [FREE Full text] [doi: [10.5665/sleep.3298](https://doi.org/10.5665/sleep.3298)] [Medline: [24470692](https://pubmed.ncbi.nlm.nih.gov/24470692/)]
2. Kolla BP, Mansukhani S, Mansukhani MP. Consumer sleep tracking devices: a review of mechanisms, validity and utility. *Expert Rev Med Devices* 2016 May;13(5):497-506. [doi: [10.1586/17434440.2016.1171708](https://doi.org/10.1586/17434440.2016.1171708)] [Medline: [27043070](https://pubmed.ncbi.nlm.nih.gov/27043070/)]
3. Duncan M, Murawski B, Short CE, Rebar AL, Schoeppe S, Alley S, et al. Activity trackers implement different behavior change techniques for activity, sleep, and sedentary behaviors. *Interact J Med Res* 2017 Aug 14;6(2):e13 [FREE Full text] [doi: [10.2196/ijmr.6685](https://doi.org/10.2196/ijmr.6685)] [Medline: [28807889](https://pubmed.ncbi.nlm.nih.gov/28807889/)]
4. Shelgikar AV, Anderson PF, Stephens MR. Sleep tracking, wearable technology, and opportunities for research and clinical care. *Chest* 2016 Dec;150(3):732-743. [doi: [10.1016/j.chest.2016.04.016](https://doi.org/10.1016/j.chest.2016.04.016)] [Medline: [27132701](https://pubmed.ncbi.nlm.nih.gov/27132701/)]
5. Liu W, Ploderer B, Hoang T. In bed with technology: challenges and opportunities for sleep tracking. New York, NY: ACM; 2015 Presented at: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction ; , Parkville, Australia; December 07-10, 2015; Parkville, VIC, Australia p. 142-151 URL: <https://dl.acm.org/citation.cfm?id=2838739.2838742> [doi: [10.1145/2838739.2838742](https://doi.org/10.1145/2838739.2838742)]
6. Liang Z, Ploderer B, Liu W, Nagata Y, Bailey J, Kulik L, et al. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. *Pers Ubiquit Comput* 2016 Sep 16;20(6):985-1000. [doi: [10.1007/s00779-016-0960-6](https://doi.org/10.1007/s00779-016-0960-6)]
7. Liang Z, Ploderer B. Sleep tracking in the real world: a qualitative study into barriers for improving sleep. New York, NY: ACM; 2016 Presented at: Proceedings of the 28th Australian Conference on Computer-Human Interaction ; , Tasmania, Australia; 2016; Launceston, Tasmania, Australia p. 537-541 URL: <https://dl.acm.org/citation.cfm?id=3010988&dl=ACM&coll=DL> [doi: [10.1145/3010915.3010988](https://doi.org/10.1145/3010915.3010988)]
8. Yang R, Shin E, Newman M, Ackerman M. When fitness trackers don't 'fit'nd-user difficulties in the assessment of personal tracking device accuracy. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, NY: ACM; 2015 Presented at: International Joint Conference on Pervasive and Ubiquitous Computing; September 07-11, 2015; Osaka, Japan p. 623-634 URL: <https://dl.acm.org/citation.cfm?id=2804269> [doi: [10.1145/2750858.2804269](https://doi.org/10.1145/2750858.2804269)]

9. Liang Z, Chapa-Martell M, Nishimura T. A personalized approach for detecting unusual sleep from time series sleep-tracking data. In: Proceedings of the IEEE International Conference on Health Informatics (ICHI).: IEEE; 2016 Presented at: IEEE International Conference on Health Informatics (ICHI); 2016; Chicago, US URL: <https://ieeexplore.ieee.org/document/7776322> [doi: [10.1109/ICHI.2016.99](https://doi.org/10.1109/ICHI.2016.99)]
10. Cook JD, Prairie ML, Plante DT. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: a comparison against polysomnography and wrist-worn actigraphy. *J Affect Disord* 2017 Aug 01;217:299-305. [doi: [10.1016/j.jad.2017.04.030](https://doi.org/10.1016/j.jad.2017.04.030)] [Medline: [28448949](https://pubmed.ncbi.nlm.nih.gov/28448949/)]
11. Weatherall J, Paprocki Y, Meyer T, Kudel I, Witt E. Sleep tracking and exercise in patients with type 2 diabetes mellitus (Step-D): pilot study to determine correlations between Fitbit data and patient-reported outcomes. *JMIR Mhealth Uhealth* 2018 Jun 05;6(6):e131 [FREE Full text] [doi: [10.2196/mhealth.8122](https://doi.org/10.2196/mhealth.8122)] [Medline: [29871856](https://pubmed.ncbi.nlm.nih.gov/29871856/)]
12. Bian J, Guo Y, Xie M, Parish AE, Wardlaw I, Brown R, et al. Exploring the association between self-reported asthma impact and Fitbit-derived sleep quality and physical activity measures in adolescents. *JMIR Mhealth Uhealth* 2017 Jul 25;5(7):e105 [FREE Full text] [doi: [10.2196/mhealth.7346](https://doi.org/10.2196/mhealth.7346)] [Medline: [28743679](https://pubmed.ncbi.nlm.nih.gov/28743679/)]
13. Baron K, Duffecy J, Berendsen M, Cheung Mason I, Lattie E, Manalo N. Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. *Sleep Med Rev* 2018 Dec;40:151-159. [doi: [10.1016/j.smrv.2017.12.002](https://doi.org/10.1016/j.smrv.2017.12.002)] [Medline: [29395985](https://pubmed.ncbi.nlm.nih.gov/29395985/)]
14. Kelly JM, Strecker RE, Bianchi MT. Recent developments in home sleep-monitoring devices. *ISRN Neurol* 2012;2012:768794-768710 [FREE Full text] [doi: [10.5402/2012/768794](https://doi.org/10.5402/2012/768794)] [Medline: [23097718](https://pubmed.ncbi.nlm.nih.gov/23097718/)]
15. Goldstone A, Baker F, de Zambotti M. Actigraphy in the digital health revolution: still asleep? *Sleep* 2018 Sep 01;41(9):- [doi: [10.1093/sleep/zsy120](https://doi.org/10.1093/sleep/zsy120)] [Medline: [29982834](https://pubmed.ncbi.nlm.nih.gov/29982834/)]
16. West P, Van Kleek M, Giordano R, Weal M, Shadbolt N. Information quality challenges of patient-generated data in clinical practice. *Front Public Health* 2017;5:284 [FREE Full text] [doi: [10.3389/fpubh.2017.00284](https://doi.org/10.3389/fpubh.2017.00284)] [Medline: [29209601](https://pubmed.ncbi.nlm.nih.gov/29209601/)]
17. Liang Z, Ploderer B, Chapa-Martell M. Is Fitbit fit for sleep-tracking? Sources of measurement errors and proposed countermeasures. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. New York, NY: ACM; 2017 Presented at: International Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2017; Barcelona, Spain p. 476-479 URL: <https://dl.acm.org/citation.cfm?id=3154897> [doi: [10.1145/3154862.3154897](https://doi.org/10.1145/3154862.3154897)]
18. de Zambotti M, Claudatos S, Inkelis S, Colrain IM, Baker FC. Evaluation of a consumer fitness-tracking device to assess sleep in adults. *Chronobiol Int* 2015;32(7):1024-1028 [FREE Full text] [doi: [10.3109/07420528.2015.1054395](https://doi.org/10.3109/07420528.2015.1054395)] [Medline: [26158542](https://pubmed.ncbi.nlm.nih.gov/26158542/)]
19. de Zambotti M, Baker FC, Willoughby AR, Godino JG, Wing D, Patrick K, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiol Behav* 2016 May 01;158:143-149 [FREE Full text] [doi: [10.1016/j.physbeh.2016.03.006](https://doi.org/10.1016/j.physbeh.2016.03.006)] [Medline: [26969518](https://pubmed.ncbi.nlm.nih.gov/26969518/)]
20. de Zambotti M, Baker FC, Colrain IM. Validation of sleep-tracking technology compared with polysomnography in adolescents. *Sleep* 2015 Sep 01;38(9):1461-1468 [FREE Full text] [doi: [10.5665/sleep.4990](https://doi.org/10.5665/sleep.4990)] [Medline: [26158896](https://pubmed.ncbi.nlm.nih.gov/26158896/)]
21. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. *Chronobiol Int* 2018 Apr;35(4):465-476. [doi: [10.1080/07420528.2017.1413578](https://doi.org/10.1080/07420528.2017.1413578)] [Medline: [29235907](https://pubmed.ncbi.nlm.nih.gov/29235907/)]
22. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep* 2015 Aug;38(8):1323-1330 [FREE Full text] [doi: [10.5665/sleep.4918](https://doi.org/10.5665/sleep.4918)] [Medline: [26118555](https://pubmed.ncbi.nlm.nih.gov/26118555/)]
23. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. *Sleep Breath* 2012 Sep;16(3):913-917. [doi: [10.1007/s11325-011-0585-y](https://doi.org/10.1007/s11325-011-0585-y)] [Medline: [21971963](https://pubmed.ncbi.nlm.nih.gov/21971963/)]
24. Liang Z, Chapa Martell MA. Validity of consumer activity wristbands and wearable EEG for measuring overall sleep parameters and sleep structure in free-living conditions. *J Healthc Inform Res* 2018 Apr 20;2(1-2):152-178. [doi: [10.1007/s41666-018-0013-1](https://doi.org/10.1007/s41666-018-0013-1)]
25. Sánchez-Ortuño MM, Edinger JD, Means MK, Almirall D. Home is where sleep is: an ecological approach to test the validity of actigraphy for the assessment of insomnia. *J Clin Sleep Med* 2010 Feb 15;6(1):21-29 [FREE Full text] [Medline: [20191934](https://pubmed.ncbi.nlm.nih.gov/20191934/)]
26. Kang S, Kang JM, 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 Jun;97:38-44. [doi: [10.1016/j.jpsychores.2017.03.009](https://doi.org/10.1016/j.jpsychores.2017.03.009)] [Medline: [28606497](https://pubmed.ncbi.nlm.nih.gov/28606497/)]
27. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int J Behav Nutr Phys Act* 2015;12:42 [FREE Full text] [doi: [10.1186/s12966-015-0201-9](https://doi.org/10.1186/s12966-015-0201-9)] [Medline: [25890168](https://pubmed.ncbi.nlm.nih.gov/25890168/)]
28. Kishi A, Struzik ZR, Natelson BH, Togo F, Yamamoto Y. Dynamics of sleep stage transitions in healthy humans and patients with chronic fatigue syndrome. *Am J Physiol Regul Integr Comp Physiol* 2008 Jun;294(6):R1980-R1987 [FREE Full text] [doi: [10.1152/ajpregu.00925.2007](https://doi.org/10.1152/ajpregu.00925.2007)] [Medline: [18417644](https://pubmed.ncbi.nlm.nih.gov/18417644/)]

29. Kishi A, Yamaguchi I, Togo F, Yamamoto Y. Markov modeling of sleep stage transitions and ultradian REM sleep rhythm. *Physiol Meas* 2018 Dec 31;39(8):084005. [doi: [10.1088/1361-6579/aad900](https://doi.org/10.1088/1361-6579/aad900)] [Medline: [30089099](https://pubmed.ncbi.nlm.nih.gov/30089099/)]
30. Kishi A, Yasuda H, Matsumoto T, Inami Y, Horiguchi J, Tamaki M, et al. NREM sleep stage transitions control ultradian REM sleep rhythm. *Sleep* 2011 Oct 01;34(10):1423-1432 [FREE Full text] [doi: [10.5665/SLEEP.1292](https://doi.org/10.5665/SLEEP.1292)] [Medline: [21966074](https://pubmed.ncbi.nlm.nih.gov/21966074/)]
31. Lo C, Bartsch RP, Ivanov PC. Asymmetry and basic pathways in sleep-stage transitions. *Europhys Lett* 2013 Apr 01;102(1):10008 [FREE Full text] [doi: [10.1209/0295-5075/102/10008](https://doi.org/10.1209/0295-5075/102/10008)] [Medline: [24653582](https://pubmed.ncbi.nlm.nih.gov/24653582/)]
32. Bianchi MT, Cash SS, Mietus J, Peng C, Thomas R. Obstructive sleep apnea alters sleep stage transition dynamics. *PLoS One* 2010 Jun 28;5(6):e11356 [FREE Full text] [doi: [10.1371/journal.pone.0011356](https://doi.org/10.1371/journal.pone.0011356)] [Medline: [20596541](https://pubmed.ncbi.nlm.nih.gov/20596541/)]
33. Burns JW, Crofford LJ, Chervin RD. Sleep stage dynamics in fibromyalgia patients and controls. *Sleep Med* 2008 Aug;9(6):689-696. [doi: [10.1016/j.sleep.2007.10.022](https://doi.org/10.1016/j.sleep.2007.10.022)] [Medline: [18314389](https://pubmed.ncbi.nlm.nih.gov/18314389/)]
34. Kemp B, Kamphuisen HA. Simulation of human hypnograms using a Markov chain model. *Sleep* 1986;9(3):405-414. [doi: [10.1093/sleep/9.3.405](https://doi.org/10.1093/sleep/9.3.405)] [Medline: [3764288](https://pubmed.ncbi.nlm.nih.gov/3764288/)]
35. Yang MC, Hirsch CJ. The use of a semi-Markov model for describing sleep patterns. *Biometrics* 1973 Dec;29(4):667-676. [Medline: [4785232](https://pubmed.ncbi.nlm.nih.gov/4785232/)]
36. Zung W, Naylor T, Gianturco D, Wilson W. Computer simulation of sleep EEG patterns with a Markov chain model. *Recent Adv Biol Psychiatry* 1965;8:335-355. [Medline: [5871725](https://pubmed.ncbi.nlm.nih.gov/5871725/)]
37. Comte JC, Ravassard P, Salin PA. Sleep dynamics: a self-organized critical system. *Phys Rev E Stat Nonlin Soft Matter Phys* 2006 May;73(5 Pt 2):056127. [doi: [10.1103/PhysRevE.73.056127](https://doi.org/10.1103/PhysRevE.73.056127)] [Medline: [16803018](https://pubmed.ncbi.nlm.nih.gov/16803018/)]
38. Penzel T, Kantelhardt JW, Lo C, Voigt K, Vogelmeier C. Dynamics of heart rate and sleep stages in normals and patients with sleep apnea. *Neuropsychopharmacology* 2003 Jul;28 Suppl 1:S48-S53 [FREE Full text] [doi: [10.1038/sj.npp.1300146](https://doi.org/10.1038/sj.npp.1300146)] [Medline: [12827144](https://pubmed.ncbi.nlm.nih.gov/12827144/)]
39. Klerman EB, Davis JB, Duffy JF, Dijk D, Kronauer RE. Older people awaken more frequently but fall back asleep at the same rate as younger people. *Sleep* 2004 Jun 15;27(4):793-798. [Medline: [15283016](https://pubmed.ncbi.nlm.nih.gov/15283016/)]
40. Langrock R, Swihart BJ, Caffo BS, Punjabi NM, Crainiceanu CM. Combining hidden Markov models for comparing the dynamics of multiple sleep electroencephalograms. *Stat Med* 2013 Aug 30;32(19):3342-3356 [FREE Full text] [doi: [10.1002/sim.5747](https://doi.org/10.1002/sim.5747)] [Medline: [23348835](https://pubmed.ncbi.nlm.nih.gov/23348835/)]
41. Schlemmer A, Parlitz U, Luther S, Wessel N, Penzel T. Changes of sleep-stage transitions due to ageing and sleep disorder. *Philos Trans A Math Phys Eng Sci* 2015 Feb 13;373(2034):- [doi: [10.1098/rsta.2014.0093](https://doi.org/10.1098/rsta.2014.0093)] [Medline: [25548271](https://pubmed.ncbi.nlm.nih.gov/25548271/)]
42. Ferri R, Pizza F, Vandi S, Ilti M, Plazzi G. Decreased sleep stage transition pattern complexity in narcolepsy type 1. *Clin Neurophysiol* 2016 Dec;127(8):2812-2819. [doi: [10.1016/j.clinph.2016.05.364](https://doi.org/10.1016/j.clinph.2016.05.364)] [Medline: [27417057](https://pubmed.ncbi.nlm.nih.gov/27417057/)]
43. Zhang X, Kantelhardt JW, Dong XS, Krefting D, Li J, Yan H, et al. Nocturnal dynamics of sleep-wake transitions in patients with narcolepsy. *Sleep* 2017 Feb 01;40(2). [doi: [10.1093/sleep/zsw050](https://doi.org/10.1093/sleep/zsw050)] [Medline: [28364512](https://pubmed.ncbi.nlm.nih.gov/28364512/)]
44. Kim JW, Lee J, Robinson PA, Jeong D. Markov analysis of sleep dynamics. *Phys Rev Lett* 2009 May 01;102(17):178104. [doi: [10.1103/PhysRevLett.102.178104](https://doi.org/10.1103/PhysRevLett.102.178104)] [Medline: [19518839](https://pubmed.ncbi.nlm.nih.gov/19518839/)]
45. Van de Water AT, Holmes A, Hurley D. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography: a systematic review. *J Sleep Res* 2011 Mar;20(1 Pt 2):183-200 [FREE Full text] [doi: [10.1111/j.1365-2869.2009.00814.x](https://doi.org/10.1111/j.1365-2869.2009.00814.x)] [Medline: [20374444](https://pubmed.ncbi.nlm.nih.gov/20374444/)]
46. Paquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. *Sleep* 2007 Oct;30(10):1362-1369 [FREE Full text] [Medline: [17969470](https://pubmed.ncbi.nlm.nih.gov/17969470/)]
47. Modave F, Guo Y, Bian J, Gurka MJ, Parish A, Smith MD, et al. Mobile device accuracy for step counting across age groups. *JMIR Mhealth Uhealth* 2017 Jun 28;5(6):e88 [FREE Full text] [doi: [10.2196/mhealth.7870](https://doi.org/10.2196/mhealth.7870)] [Medline: [28659255](https://pubmed.ncbi.nlm.nih.gov/28659255/)]
48. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep* 1992 Oct;15(5):461-469. [doi: [10.1093/sleep/15.5.461](https://doi.org/10.1093/sleep/15.5.461)] [Medline: [1455130](https://pubmed.ncbi.nlm.nih.gov/1455130/)]
49. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989 May;28(2):193-213. [Medline: [2748771](https://pubmed.ncbi.nlm.nih.gov/2748771/)]
50. Matsuo M, Masuda F, Sumi Y, Takahashi M, Yamada N, Ohira MH, et al. Comparisons of portable sleep monitors of different modalities: potential as naturalistic sleep recorders. *Front Neurol* 2016;7:110 [FREE Full text] [doi: [10.3389/fneur.2016.00110](https://doi.org/10.3389/fneur.2016.00110)] [Medline: [27471489](https://pubmed.ncbi.nlm.nih.gov/27471489/)]
51. Yoshida M, Shinohara H, Kodama H. Assessment of nocturnal sleep architecture by actigraphy and one-channel electroencephalography in early infancy. *Early Hum Dev* 2015 Sep;91(9):519-526. [doi: [10.1016/j.earlhumdev.2015.06.005](https://doi.org/10.1016/j.earlhumdev.2015.06.005)] [Medline: [26140905](https://pubmed.ncbi.nlm.nih.gov/26140905/)]
52. McCall C, McCall WV. Objective vs subjective measurements of sleep in depressed insomniacs: first night effect or reverse first night effect? *J Clin Sleep Med* 2012 Feb 15;8(1):59-65 [FREE Full text] [doi: [10.5664/jcsm.1664](https://doi.org/10.5664/jcsm.1664)] [Medline: [22334811](https://pubmed.ncbi.nlm.nih.gov/22334811/)]
53. Ahmadi N, Shapiro G, Chung S, Shapiro C. Clinical diagnosis of sleep apnea based on single night of polysomnography vs two nights of polysomnography. *Sleep Breath* 2009 Aug;13(3):221-226. [doi: [10.1007/s11325-008-0234-2](https://doi.org/10.1007/s11325-008-0234-2)] [Medline: [19067010](https://pubmed.ncbi.nlm.nih.gov/19067010/)]
54. Berry RB, Brooks R, Gamaldo C, Harding SM, Lloyd RM, Quan SF, et al. AASM Scoring Manual Updates for 2017 (Version 2.4). *J Clin Sleep Med* 2017 Dec 15;13(5):665-666 [FREE Full text] [doi: [10.5664/jcsm.6576](https://doi.org/10.5664/jcsm.6576)] [Medline: [28416048](https://pubmed.ncbi.nlm.nih.gov/28416048/)]

55. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
56. Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. *Br Med J* 1994 Jun 11;308(6943):1552 [[FREE Full text](#)] [doi: [10.1136/bmj.308.6943.1552](https://doi.org/10.1136/bmj.308.6943.1552)] [Medline: [8019315](https://pubmed.ncbi.nlm.nih.gov/8019315/)]
57. Ohayon M, Wickwire E, Hirshkowitz M, Albert S, Avidan A, Daly F, et al. National Sleep Foundation's sleep quality recommendations: first report. *Sleep Health* 2017 Dec;3(1):6-19. [doi: [10.1016/j.sleh.2016.11.006](https://doi.org/10.1016/j.sleh.2016.11.006)] [Medline: [28346153](https://pubmed.ncbi.nlm.nih.gov/28346153/)]
58. Taïbi DM, Landis CA, Vitiello MV. Concordance of polysomnographic and actigraphic measurement of sleep and wake in older women with insomnia. *J Clin Sleep Med* 2013 Mar 15;9(3):217-225 [[FREE Full text](#)] [doi: [10.5664/jcsm.2482](https://doi.org/10.5664/jcsm.2482)] [Medline: [23493815](https://pubmed.ncbi.nlm.nih.gov/23493815/)]
59. O'Hare E, Flanagan D, Penzel T, Garcia C, Frohberg D, Heneghan C. A comparison of radio-frequency biomotion sensors and actigraphy versus polysomnography for the assessment of sleep in normal subjects. *Sleep Breath* 2015 Mar;19(1):91-98. [doi: [10.1007/s11325-014-0967-z](https://doi.org/10.1007/s11325-014-0967-z)] [Medline: [24614968](https://pubmed.ncbi.nlm.nih.gov/24614968/)]
60. Bellone GJ, Plano SA, Cardinali DP, Chada DP, Vigo DE, Golombek DA. Comparative analysis of actigraphy performance in healthy young subjects. *Sleep Sci* 2016;9(4):272-279 [[FREE Full text](#)] [doi: [10.1016/j.slsci.2016.05.004](https://doi.org/10.1016/j.slsci.2016.05.004)] [Medline: [28154740](https://pubmed.ncbi.nlm.nih.gov/28154740/)]
61. Lucey B, Mclelland JS, Toedebusch C, Boyd J, Morris J, Landsness E, et al. Comparison of a single-channel EEG sleep study to polysomnography. *J Sleep Res* 2016 Dec;25(6):625-635 [[FREE Full text](#)] [doi: [10.1111/jsr.12417](https://doi.org/10.1111/jsr.12417)] [Medline: [27252090](https://pubmed.ncbi.nlm.nih.gov/27252090/)]
62. Meltzer L, Walsh C, Traylor J, Westin A. Direct comparison of two new actigraphs and polysomnography in children and adolescents. *Sleep* 2012 Jan 01;35(1):159-166 [[FREE Full text](#)] [doi: [10.5665/sleep.1608](https://doi.org/10.5665/sleep.1608)] [Medline: [22215930](https://pubmed.ncbi.nlm.nih.gov/22215930/)]
63. Werner H, Molinari L, Guyer C, Jenni O. Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns. *Arch Pediatr Adolesc Med* 2008 Apr;162(4):350-358. [doi: [10.1001/archpedi.162.4.350](https://doi.org/10.1001/archpedi.162.4.350)] [Medline: [18391144](https://pubmed.ncbi.nlm.nih.gov/18391144/)]
64. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. 24 hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Med Sci Sports Exerc* 2015 Oct 17. [doi: [10.1249/MSS.0000000000000778](https://doi.org/10.1249/MSS.0000000000000778)] [Medline: [26484953](https://pubmed.ncbi.nlm.nih.gov/26484953/)]
65. Kushida C, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement W. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med* 2001 Sep;2(5):389-396. [Medline: [14592388](https://pubmed.ncbi.nlm.nih.gov/14592388/)]

Abbreviations

- API:** application programming interface
- PSG:** polysomnography
- PSQI:** Pittsburgh Sleep Quality Index
- REM:** rapid eye movement
- SE:** sleep efficiency
- SOL:** sleep onset latency
- SWS:** slow wave sleep
- TST:** total sleep time
- WASO:** wake after sleep onset

Edited by G Eysenbach; submitted 19.01.19; peer-reviewed by F Modave, K Ng, S Berrouiguet; comments to author 28.03.19; revised version received 04.04.19; accepted 23.04.19; published 06.06.19

Please cite as:

Liang Z, Chapa-Martell MA

Accuracy of Fitbit Wristbands in Measuring Sleep Stage Transitions and the Effect of User-Specific Factors

JMIR Mhealth Uhealth 2019;7(6):e13384

URL: <https://mhealth.jmir.org/2019/6/e13384/>

doi: [10.2196/13384](https://doi.org/10.2196/13384)

PMID: [31172956](https://pubmed.ncbi.nlm.nih.gov/31172956/)

©Zilu Liang, Mario Alberto Chapa-Martell. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 06.06.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.