

## Paradox in Validating Consumer Wearable Activity Trackers on Measuring Sleep Quality

Zilu Liang<sup>†</sup> Mario Alberto Chapa-Martell<sup>‡</sup>

We present a paradox that we discovered in validating a consumer activity tracker Fitbit on measuring sleep quality. The sleep measurements by Fitbit were compared to both subjective and objective sleep quality. The results reveal that Fitbit may not correlate to objective sleep measures, but rather agree well to subjective ones. Depending on whether subjective or objective sleep quality is taken as the ground truth, Fitbit may both be valid and invalid, leading to a paradox that could be explained by the discrepancy between objective and subjective sleep quality. We argue that future validation studies should properly address this paradox.

Key words: validation, Fitbit, sleep, wearable, pervasive computing

## 1. Introduction

Several studies have investigated the validity of wearable activity trackers on measuring sleep quality in comparison to more sophisticated devices such as PSG or actigraphy that measure objective sleep quality. However, a revisit to the definition of sleep quality in clinical sleep research show that sleep quality can be quantified not only through the sleep measures estimated from sleep monitoring devices but also based on people's subjective feeling upon waking up or during the day, and that subjective sleep quality and objective sleep quality are only moderately correlated [1-3]. In this study, we validate a popular wearable activity tracker, Fitbit, in comparison to subjective sleep quality measured using a standard sleep diary as well as to objective sleep measures detected using a clinical device SLEEP SCOPE. The results reveal that Fitbit may not correlate to objective sleep measures but agree well to subjective sleep quality. Depending on whether subjective or objective sleep quality is taken as the ground truth, contradictory conclusions may be drawn on the validity of Fitbit, leading to an interesting paradox.

We further delve deeper into the discrepancy between subjective and objective sleep quality on healthy people. Such discrepancy has been well studied in sleep research community on clinical populations such as old people with cognitive impairment [4], depression patients [5] and insomnia patients [6] but not on healthy populations. Our study shows that the discrepancy between subjective and objective sleep quality also exist in healthy people, which may account for the paradox in validating wearable sleep trackers. Since objective and subjective sleep quality measures sleep from different perspective and thus are both important in helping us understand people's sleep quality, we argue that both aspects should be considered in future validation studies of consumer sleep tracking devices.

## 2. Related Work

### 2.1 Fundamental Concepts of Human Sleep

Human sleep can be measured along multiple dimensions such as quantity, continuity and timing [7]. Focusing on the most measurable characteristics of sleep that are closely related to physical and mental wellbeing, sleep health was quantified in the following five dimensions in [1]: sleep duration, sleep continuity or efficiency, timing, alertness/sleepiness, satisfaction/quality.

The multiple dimensions of human sleep can be measured both objectively and subjectively [1]. Subjective methods include

Pittsburgh Sleep Quality Index (PSQI) [8] and sleep diary [9]. The PSQI questionnaire is widely used in clinical settings for rough evaluation of sleep quality over the past one month. Since human sleep is dynamic, PSQI may fail to capture sleep variations in long time period. In contrast, sleep diary is widely used to collect longitudinal sleep data for understanding long-term trends and patterns of sleep [9]. The objective methods for measuring human sleep focus on analyzing a set of metrics called sleep structure [10]. The components of sleep structure include total sleep time (TST), wake after sleep onset (WASO), number of awakenings (NAWK), sleep onset latency (SOL), sleep efficiency index (SE), REM sleep latency, total time in each sleep stage, and sleep stage ratio [11]. Human sleep that satisfies the following range is considered as abnormal: SOL  $\geq$  46 minutes, WASO  $\geq$  41 minutes, NAWK (for awakenings longer than 5 minutes)  $\geq$  4, SE  $<$  75%, REM ratio  $\geq$  41%, Stage 1 sleep  $\geq$  21%, Stage 2 sleep  $\geq$  81%, Stage 3+4 sleep  $<$  10% [12]. However, since people's sleep needs varies significantly, there may be a wide spectrum of acceptable sleep structures in addition to the recommended standards. PSG and actigraphy are widely used tools for objectively measuring sleep in clinical settings. A polysomnography (PSG) sleep test measures parameters of the whole sleep structure over one night, and is mainly used for diagnosing sleep diseases [11]. Actigraphy is a wristband-like device that is widely used for diagnosing circadian-related disorders [13]. A common problem of actigraphy is that it tends to underestimate awakenings and overestimate total sleep time [13].

Despite measuring the same phenomenon, subjective sleep quality and objective sleep quality characterize different aspects of human sleep and are only modestly correlated [1-3]. We therefore adopted both objective and subjective sleep assessments as the ground truth in this study to compare the consumer devices with.

### 2.2 Validity of Consumer Sleep-Tracking Technology

Consumer sleep-tracking technologies help people monitor and reflect on sleep in home settings at low cost. This field is expanding very rapidly and new devices are entering the market every year. Based on the mechanisms of the technologies, consumer sleep-tracking tools can be divided into two categories: accelerometer-based (e.g. mobile apps, activity wristbands, smart mattress) and EEG-based (e.g. ZEO headband, Neuroon eye mask, Sleep Shepherd headband). Comprehensive reviews on recent developments in home sleep-tracking devices and mobile apps can be found in [14-15]

A number of studies have evaluated the validity of some of the consumer devices both quantitatively and qualitatively. Quantitative validation studies compared the data obtained from consumer sleep tracking tools to measurements by clinical devices or instruments including PSG, actigraphy, and PSQI [14].

<sup>†</sup> The University of Tokyo.

<sup>‡</sup> CAC Corporation. The opinions expressed in this paper are solely of the author's and do not represent the view of the company.

These studies mainly validated older models of popular activity trackers for home sleep tracking, including Fitbit Tracker [16], Fitbit Ultra [17], and Jawbone [18]. A few studies also investigated the accuracy of some mobile apps such as Sleep Time [19]. These studies found that activity trackers had the common problem of overestimating sleep duration and underestimating awakening.

Researchers in the human-computer interaction research community have also investigated the validity issues of consumer sleep tracking technologies from user's perspective. The long-term impact of sleep-tracking was studied in [20] and the measurement accuracy was highlighted as one of the main obstacles for improving sleep health using consumer sleep trackers. Following the same line, [21] investigated the source of measurement errors and proposed countermeasures in the aspects of both technologies and human factors.

### 3. Methods

#### 3.1 Sleep Metrics

TABLE I. Definition of Sleep Metrics

Sleep Metrics		Definition
Total Sleep Time (TST)		Time in minutes from sleep onset to sleep offset.
Wake after Sleep Onset (WASO)		Periods of wakefulness occurring after defined sleep onset.
Number of Awakenings (NAWK)		The number of awakenings occurring after defined sleep onset.
Sleep Onset Latency (SOL)	Objective	Time in minutes from "light out" to the first epoch scored as sleep.
	Subjective	Individual's sense of how easy it was to fall asleep (1~5; 1=immediately fell asleep, 5=it was difficult to fall asleep).
Sleep Efficiency (SE)	Objective	Percentage of total time in bed actually spent in sleep. $TST/(TST+WASO)$
	Subjective	Individual's sense of whether the sleep was good or poor (1~5; 1=very poor, 5=very good).

The sleep metrics that we measured using consumer devices, clinical device, and sleep diary are summarized in Table 1. It is worth noting that the objective and subjective assessment on the SOL and the SE used different scales. Using consumer devices and clinical device, the SOL was automatically measured in minutes from "light out" to the first epoch scored as sleep, and the SE was calculated as the percentage of total time in bed actually spent in sleep. In comparison, the subjective evaluation of SOL and SE was based on an individual's feeling on how easy it was to fall asleep and how well the sleep was according to an ordinal scale in the range of 1 and 5.

#### 3.2 Devices and Instruments

##### 3.2.1 Fitbit

The Fitbit Charge 2 is a wearable activity wristband that tracks the frequency and intensity of a user's movements with an embedded 3-dimensional accelerometer. It tracks sleep in addition to physical activity, workout, calorie consumption, etc. The normal sleep-recording mode was used during the data collection process, which accounts "significant movements (such as rolling over) as being awake, and is appropriate for most users" according to the manufacturer's website [22]. A Fitbit Charge 2 can automatically detect the start of sleep if a user has not moved

for approximately one hour. The movement data were collected in 1-min epochs by default. After being synced to the Fitbit database, these data are mapped to aggregated sleep metrics such as total minutes asleep and minutes awake using proprietary software and algorithms. The final results are then shown to the end users on the Fitbit dashboard. A screenshot of one night's sleep data on the Fitbit dashboard is shown in Figure 1.

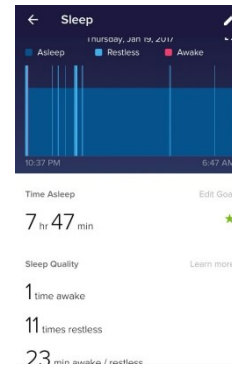


Fig.1. A screenshot of sleep data on Fitbit dashboard.

##### 3.2.2 Sleep Scope

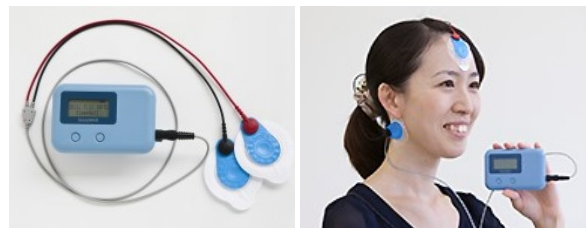


Fig.2. Single-channel EEG clinical sleep monitor Sleep Scope (pictures credits: <https://sleepwell.co.jp/>).

As shown in Figure 2, Sleep Scope is a clinical 1-channel portable EEG device developed by the Sleep Well company based in Osaka, Japan. Sleep Scope has been previously validated against PSG and achieved 86.9% agreement (average Cohen's Kappa value = 0.753) [23]. Given that the average inter-scoring reliability on sleep stage scoring is 82.6% [24], the agreement was high between Sleep Scope and PSG. We chose Sleep Scope as an alternative of PSG because the purpose of this study was not to diagnose sleep problems. Sleep Scope is more portable and less obtrusive in comparison to PSG, and it can be used to monitor sleep at a user's home, though still not as convenient as wearable devices. A Sleep Scope device uses two electrodes to be placed on the forehead and behind an ear. In this study we used gel type electrodes to improve the accuracy of EEG measurement. The data (i.e., raw EEG signals) from Sleep Scope need to be sent to the company for analysis. A sleep report will be generated based on the analysis and a sample report is shown in Figure 4.

##### 3.2.3 Sleep Diary

Tracking sleep on a night-by-night basis with a sleep diary has been considered as a useful methodology for accessing sleep [25]. In this study, we created a sleep diary for each participant using Google Forms based on the core version of the standard sleep diary [9]. The standard sleep diary was developed primarily for insomnia research, but it is also helpful for good sleepers to understand personal sleep patterns. There are in total 9 questions to answer and it takes less than 1 minute to complete the diary.

All questions are consistent with the wording in the standard sleep diary, except the question on sleep onset latency. We changed the question “how long did it take you to fall asleep (in minutes)?” to “how easy was it for you to fall asleep (1-5; 1 = immediately fell asleep, 5 = it was difficult to fall asleep)”, as previous studies suggested that the subjective estimation on the time to fall asleep could be extremely inaccurate [26]. A sample sleep diary can be accessed at <https://goo.gl/forms/3BMLJfE2WeMBT8Zn1>.



Fig.3. A sample report of Sleep Scope (in Japanese only).

### 3.3 Study Procedure

Sleep data were collected from two healthy young adults (1 female and 1 male; age range: 25-35 years old) without sleep problems. Ethics approval was obtained from the Ethic Committee of the University of Tokyo. All participants provided informed consent.

Each participant tracked their sleep over multiple nights during January and March 2017 using a Fitbit Charge 2 and a Sleep Scope simultaneously. The Fitbit Charge 2 was worn on the non-dominant wrist. The two electrodes of Sleep Scope were attached to the upper forehead and behind ear, while the main body of the device was placed beside the participant’s pillow. In the morning following each night, the participants received a reminder to fill in the sleep diary online.

When participants completed the tracking, we retrieved data from all devices as well as sleep diary. The aggregated sleep data of Fitbit were retrieved through Fitbit public API using a web application SleepExplorer [26]. The EEG data from the Sleep Scope were extracted from the SD card of the device and were forwarded to the Sleep Well Company for analysis. At the Sleep Well company, raw EEG data were routinely analyzed at 30-second epoch. The sleep stages were determined using proprietary automatic scoring system. The validity of the sleep staging was then visually assessed epoch-by-epoch by specialists according to sleep scoring standards [27] and corrections were added when necessary.

### 3.4 Data Analysis

#### 3.4.1 Data Preprocessing

As summarized in Table 2, Fitbit and SLEEP SCOPE use different names for the sleep metrics of interest. The initial step in data preprocessing is to unify the naming of the sleep metrics measured by Fitbit in line with Sleep Scope. In addition, missing data was removed pair-wisely.

TABLE II. Naming of Sleep Metrics by Fitbit in Comparison to Sleep Scope (Clinical Terminology)

SLEEP SCOPE	Fitbit
Total Sleep Time (TST)	Minutes Asleep
Wake After Sleep Onset (WASO)	Minutes Awake
Number of Awakenings (NAWK)	Number of Awakenings
Sleep Onset Latency (SOL)	Time to Fall Asleep
Sleep Efficiency	Sleep Efficiency

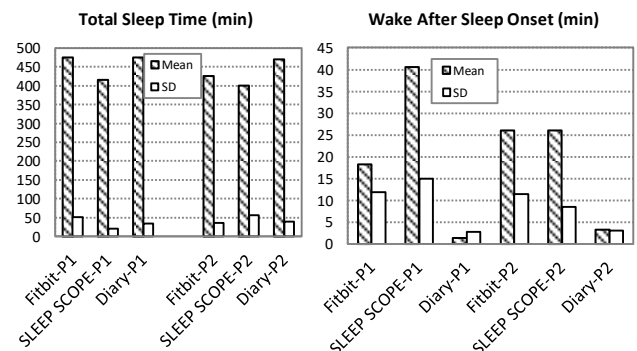


Fig.4. Mean and standard deviation of (left) total sleep time (TST) and (right) wake after sleep onset (WASO) measured by Fitbit, SLEEP SCOPE and sleep diary for each participant.

#### 3.4.2 Statistical Analysis

The validity of Fitbit refers to how well a Fitbit device actually measures the underlying sleep phenomenon in comparison to the ground truth [28]. Taken the measurements by the clinical device Sleep Scope and the standard sleep diary as the ground truth, we use several statistical tests to evaluate the validity of the

consumer trackers. Scatter plots and Bland-Altman plots [29] were used to examine relationship and the level of agreement between Fitbit and the ground truth. In clinical settings, if the within-mean differences equal  $\pm 1.96$  SD are not clinically important, then the two devices are equivalent and may be used interchangeably [28]. In addition, Pearson correlation was used to assess the relationships between measurements by consumer devices and those by the ground truth. The results are described in detail in the next section.

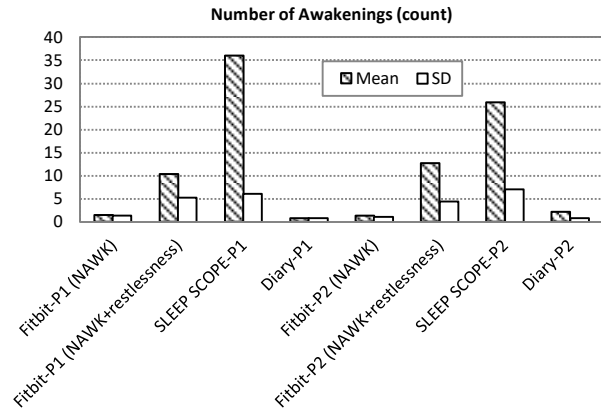


Fig.5. Mean and standard deviation of number of awakenings measured by Fitbit, SLEEP SCOPE and sleep diary.

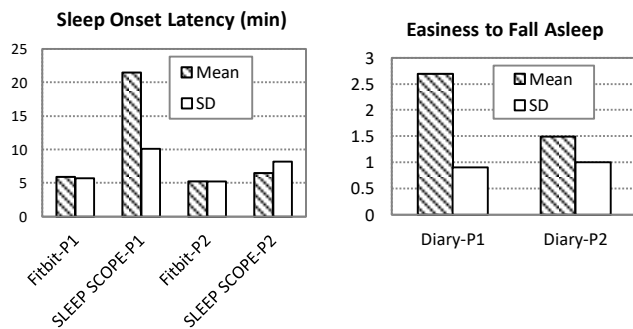


Fig.6. Mean and standard deviation of (left) sleep onset latency (SOL) measured by Fitbit and SLEEP SCOPE, and (right) easiness to fall asleep measured by sleep diary (1=immediately fell asleep; 5=it was difficult to fall asleep).

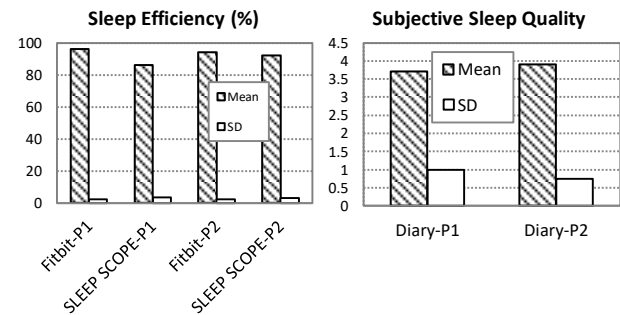


Fig.7. Mean and standard deviation of (left) sleep efficiency (SE) measured by Fitbit and SLEEP SCOPE, and (right) subjective sleep quality measured by sleep diary (1=very poor; 5=very good).

## 4. Results

### 4.1 Sleep Metrics Means and Standard Deviation

The average and standard deviation of sleep metrics are shown in Figure 4-7. The results of participant P1 are indicated by the left six bars, while those of participant P2 are indicated by the right six bars. The sleep patterns of the two participants are different in both objective and subjective sleep measures.

### 4.2 Fitbit in Comparison to subjective and objective sleep quality

The Pearson correlation coefficients shown in Table III suggest that the relationship between Fitbit and subjective/objective sleep quality was distinct for each participant. For participant 1, Fitbit has better correlation to sleep diary in terms of TST but is not correlated to SLEEP SCOPE. For participant 2, Fitbit has strong correlation to both sleep diary and SLEEP SCOPE in the dimension of TST. In addition, Fitbit is moderately correlated to sleep diary in term of NAWK, and is moderately correlated to SLEEP SCOPE in terms of WASO, both with statistical significance.

TABLE III. Fitbit in Comparison to Subjective (Sleep Diary) and Objective Sleep Quality (SLEEP SCOPE) for each participant.

Fitbit v.s.	TST	WASO	NAWK	SOL	SE
Diary (P1)	<b>0.56*</b>	0.13	0.18	0.02	0.26
Diary (P2)	<b>0.75</b>	0.20	<b>0.46</b>	0.07	0.14
SLEEP SCOPE (P1)	0.25	0.25	0.08	0.04	0.23
SLEEP SCOPE (P2)	<b>0.72</b>	<b>0.43</b>	0.02	0.22	0.23

\*bold indicate statistically significant correlations.

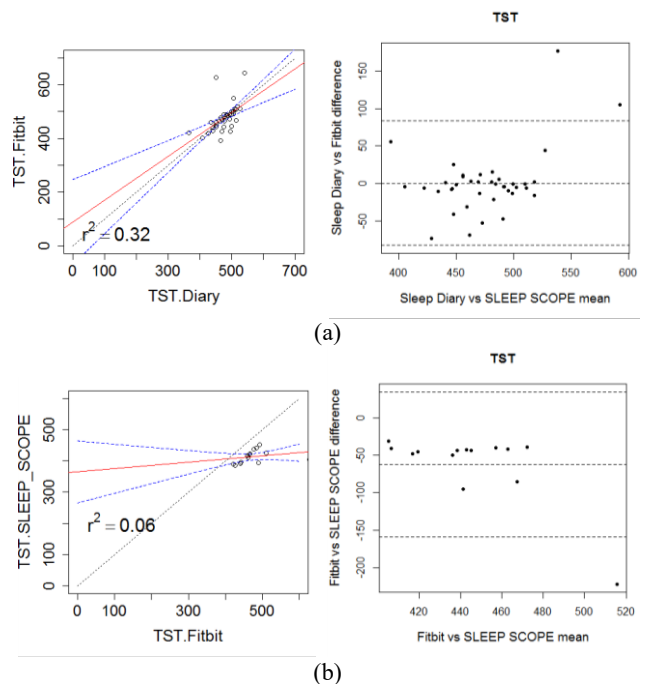


Fig.5. Scatterplots and Bland-Altman plots on TST (a) between Fitbit and sleep diary, and (b) between Fitbit and SLEEP SCOPE for Participant1.

Due to page limit, we only selectively present the agreement on TST for participant 1 (Figure 8) and that on NAWK for participant 2 (Figure 9). The scatterplots and Bland-Altman plots of TST for participant 1 show that Fitbit agreed well to sleep diary on TST, while Fitbit overestimated TST in comparison to SLEEP SCOPE. The scatterplots and Bland-Altman plots of NAWK for participant 2 show Fitbit agreed well to sleep diary, but underestimated NAWK in comparison to SLEEP SCOPE.

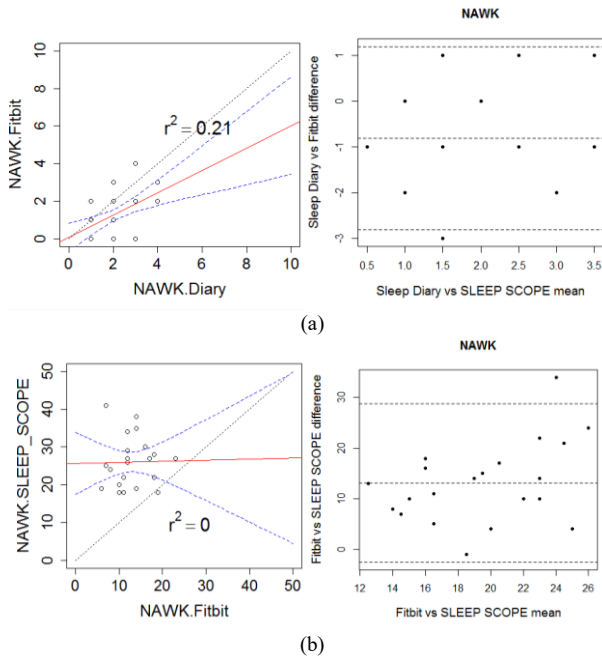


Fig.9. Scatterplots and Bland-Altman plots on NAWK (a) between Fitbit and sleep diary, and (b) between Fitbit and SLEEP SCOPE for Participant2.

### 4.3 Discrepancy between subjective and objective sleep quality

Since the validity of Fitbit Charge 2 manifested distinct characteristics when compared to objective sleep quality (measured by Sleep Scope) and to subjective sleep quality (measured by sleep diary), we delve deeper into the discrepancy in sleep estimates between Sleep Scope and sleep diary. As shown in Table IV, objective and subjective sleep quality was only modestly correlated on SOL and SE for participant 1, and modestly correlated on TST and WASO for participant 2. Sleep diary and SLEEP SCOPE was not correlated on NAWK for either of the participant. The Bland-Altman plots on TST, WASO and NAWK are shown in Figure 10. Since SOL and SE were measured on different magnitude by subjective sleep quality (SOL: 1~5; SE: 1~5) and objective sleep quality (SOL: 0~∞ min; SE: 0 ~ 100%), we did not conduct Bland-Altman analysis on these two sleep metrics. The plots demonstrate significant discrepancy between subjective sleep quality and objective sleep quality on all sleep metrics for both participants.

TABLE IV. Comparison between Subjective (Sleep Diary) and Objective Sleep Quality (SLEEP SCOPE).

	TST	WASO	NAWK	SOL	SE
P1	0.52	0.09	0.06	<b>0.69*</b>	<b>0.56</b>
P2	<b>0.60</b>	<b>0.48</b>	0.10	0.20	0.28

\*bold indicate statistically significant correlations.

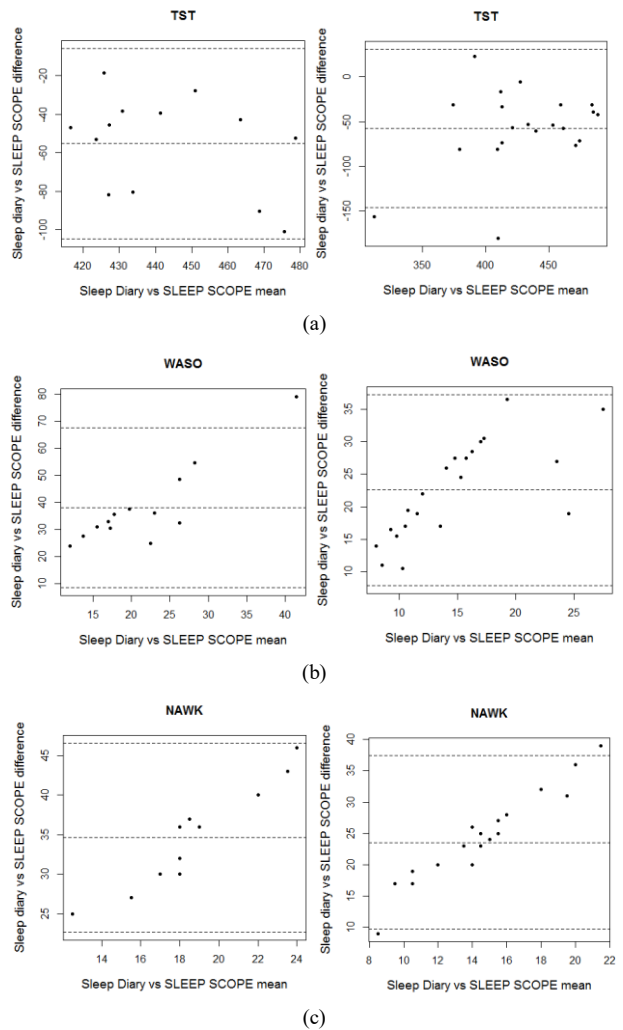


Fig.10. Bland-Altman plots between subjective sleep quality (sleep diary) and objective sleep quality (SLEEP SCOPE) on (a) TST, (b) WASO, and (c) NAWK. The left column shows the plots for Participant 1 and the right column shows the plots for Participant 2.

## 5. Discussions

The sleep research community has studied the discrepancy between objective and subjective sleep quality on specific populations such as insomnia patients [6,30], aged population [4], and people with mental disorder [5]. Following the same line, our study showed that systematic discrepancy exist between objective and subjective assessment on TST, WASO, and NAWK for healthy people, and subjective sleep quality was characterized by longer sleep and shorter awakenings compared to objective sleep quality. Our study showed that the objective and subjective assessment on TST, SOL and SE were only modestly correlated and the subjective perception on awakenings significantly deviated from objective measurements. Such discrepancies may account for the distinct characteristics of Fitbit when compared to objective and subjective sleep quality respectively. In addition, we also found that such discrepancy became more significant with more disrupted sleep (characterized by longer WASO and higher NAWK).

In our previous study, we found that people's definition of sleep quality large depended on their subjective perception [26], which may not necessarily be consistent with objective measurements. As a result, the perceived errors (which are the

discrepancies between consumer sleep trackers and subjective sleep experience) may deviate from the objective errors (which are the discrepancies between consumer sleep trackers and clinical sleep monitors). Therefore, different conclusions may be drawn on the validity of consumer sleep trackers depending on whether objective sleep quality or subjective sleep quality is taken as the ground truth. Although existing validation studies mostly adopted objective sleep quality as the ground truth, we argue that subjective sleep quality should also be considered in future validation studies, as subjective sleep reports “reflect an individual’s perspective or state of mind in addition to some component of their objective sleep patterns” [31].

## 6. Conclusions

This study investigated the validity of Fitbit Charge 2 for sleep tracking in comparison to a clinical device SLEEP SCOPE and to sleep diary. We found that regardless of interpersonal difference, Fitbit agreed well to the subjective sleep quality measured by sleep diary on TST and NAWK. However, Fitbit measures significantly deviated from the objective sleep quality measured by SLEEP SCOPE. We further examined the agreement between subjective and objective sleep quality and found significant discrepancy between the two, which may explain the contradictory results on Fitbit validity. Since subjective and objective sleep quality reflect different aspects of a person’s sleep experience, we argue that both modalities should be considered in future validation studies.

### Acknowledgement

This study was supported by JSPS KAKENHI Grant-in-Aid for Research Activity Start-up (Grant Number 16H07469).

### References

- [1] Daniel J Buysse, “Sleep health: can we refine it? Does it matter?” *Sleep*, vol.37, no.1, 9-17 (2014).
- [2] Vitiello MV, Larsen LH, Drolet G, et al. “Gender differences in subjective-objective sleep relationships in non-complaining healthy older adults,” *Sleep*, vol. 25S, no.A61 (2002).
- [3] Coates TJ, Killen JD, George J, et al. “Estimating sleep parameters: a multitrait-multimethod analysis,” *J Consult Clin Psychol*, vol.50, no. 3, p.345-52 (1982).
- [4] Landry GJ, Best JR, Liu-Ambrose T, “Measuring sleep quality in older adults: a comparison using subjective and objective methods,” *Frontiers in Aging Neuroscience*, vol.7, article 166, (2015).
- [5] Van Raveesteyn LM, Tulen JHM, Kamperman, AM, et al. “Perceived sleep quality is worse than objective parameters of sleep in pregnant women with a mental disorder,” *Journal of Clinical Sleep Medicine*, vol.10, no.10, p.1137-1141 (2014).
- [6] Lund HG, Rybarczyk BD, Perrin PB, et al. “The discrepancy between subjective and objective measures of sleep in older adults receiving CBT for comorbid insomnia,” *Journal of Clinical Psychology*, vol.69, no.10, p.1108-1120, (2013).
- [7] Carskadon MA, Dement WC. “Normal human sleep: an overview”. In: Kryger MH, Roth T, Dement WC, eds. *Principle and practice of sleep medicine*. 4th ed. Philadelphia, PA: Elsevier Saunders, 13-23 (2015).
- [8] Daniel J Buysse, Charles F Reynolds, Timothy H Monk, et al. *The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research*. *Psychiatry research* 28, 2 (1989), 193-213.
- [9] Colleen E. Carney, Daniel J. Buysse, Sonia Ancoli-Israel, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. *SLEEP*, 35:2, 287-302 (2012).
- [10] Christer Hublin, Markku Partinen, Markku Koskenvuo and Jaakko Kaprio. Sleep and Mortality: A Population-Based 22-Year Follow-up Study. *Sleep* 30, 10 (2007), 1245.
- [11] Clete A Kushida, Michael R Littner, Timothy Morgenthaler, et al. (2005). Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep*, 28:4, 499-519.
- [12] Maurice Ohayon, Emerson M Wickwire, Max Hirshkowitz, et al. (2017). National sleep foundation’s sleep quality recommendations: first report. *Sleep Health*, 3, 6-19
- [13] Avi Sadeh. (2011). The role and validity of actigraphy in sleep medicine: an update. *Sleep Medicine Reviews*, 15, 259-267.
- [14] Bhanu Prakash Kolla, Subir Mansukhani, and Meghna P. Mansukhani. (2016). Consumer sleep tracking devices: a review of mechanisms, validity and utility. *Expert Review of Medical Devices*, 13:5, 497-506.
- [15] Anita Valanju Shelgikar, Patricia F Anderson, and Marc R Stephens. (2016). Sleep tracking, wearable technology, and opportunities for research and clinical care. *Chest*, 150:3, 732-743.
- [16] Montgomery-Downs HE, Insana SP, Bond JA. (2012). Movement toward a novel activity monitoring device. *Sleep Breath*, 16:3, 913-917.
- [17] Meltzer LJ, Hiruma LS, Avis K, et al. (2015). Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep*, 38:8, 1323-1330.
- [18] De Zambotti M, Claudatos S, Inkelis S, et al. (2015). Evaluation of a consumer fitness-tracking device to access sleep in adults. *Chronobiol Int.*, 32:7, 1024-1028.
- [19] Bhat S, Ferraris A, Gupta D, et al. “Is there a clinical role for smartphone sleep apps? Comparison of sleep cycle detection by a smartphone application to polysomnography,” *J Clin Sleep Med*, vol.11, no.7, p.709-715 (2015)
- [20] Liang, Z., Ploderer, B. (2016). Sleep tracking in the real world: a qualitative study into barriers for improving sleep. *OZCHI 2016*: 537-541.
- [21] Liang, Z., Ploderer, B., and Chapa-Martell, MA. (2017) Is Fitbit Fit for Sleep-Tracking? Sources of Measurement Errors and Proposed Countermeasures. In *PervasiveHealth 2017*, Spain.
- [22] Fitbit Charge 2: How do I track my sleep? [https://help.fitbit.com/articles/en\\_US/Help\\_article/1314/?l=en\\_US&c=Topics%3ASleep&p=charge\\_2&fs=Search&pn=1#Whatisthedifference](https://help.fitbit.com/articles/en_US/Help_article/1314/?l=en_US&c=Topics%3ASleep&p=charge_2&fs=Search&pn=1#Whatisthedifference).
- [23] Neuroon sleep analytics. <https://neuroon.jp/features/sleep-analytics/>
- [24] Richard S. Rosenberg and Steven Van Hout. The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring. *Journal of Clinical Sleep Medicine*, 9:1, 81-87, 2013.
- [25] Buysse DJ, Ancoli-Israel S, Edinger JD, et al. Recommendations for a standard research assessment of insomnia. *Sleep*, 29:1155-73, 2006.
- [26] Zilu Liang, Bernd Ploderer, Wanyu Liu, et al. “SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors,” *Personal and Ubiquitous Computing* 20(6): 985-1000 (2016).
- [27] Iber C, Ancoli-Israel S, Chesson A, et al. (2007). *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*, 1st ed. Westchester, IL: American Academy of Sleep Medicine.
- [28] Higgins, PA, Straub, AJ. Understanding the error of our ways: mapping the concepts of validity and reliability. *Nurs Outlook* 2006; 54:23-29.
- [29] Bland JM, Altman DG, “Statistical methods for assessing agreement between two methods of clinical measurement,” *Lancet*, vol.1, p.307-10, (1986).
- [30] Rosa RR, Bonnet MH. 2000. Reported chronic insomnia is independent of poor sleep as measured by electroencephalography. *Psychosom Med*, 62, 474-482.
- [31] Tworoger, SS, Davis, S, Vitiello, MV, et al. (2005). Factors associated with objective (actigraphic) and subjective sleep quality in young adult women. *Journal of Psychosomatic Research*, 59:1, 11-19.
- [32] Cellini N, Buman MP, McDevitt, EA, et al. (2013). Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults. *Chronobiol. Int.*, 30, 691-698.