



Consumer-grade sleep trackers are still not up to par compared to polysomnography

Kyubo Kim¹ · Do-Yang Park² · Yong Jae Song² · Sujin Han¹ · Hyun Jun Kim² 

Received: 7 June 2021 / Revised: 20 August 2021 / Accepted: 9 September 2021 / Published online: 5 November 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

Purpose Although polysomnography (PSG) is the gold standard for monitoring sleep, it has many limitations. We aimed to prospectively determine the validity of wearable sleep-tracking devices and smartphone applications by comparing the data to that of PSGs.

Methods Patients who underwent one night of attended PSG at a single institution from January, 2015 to July 2019 were recruited. Either a sleep application or wearable device was used simultaneously while undergoing PSG. Nine smartphone applications and three wearable devices were assessed.

Results We analyzed the results of 495 cases of smartphone applications and 170 cases of wearables by comparing each against PSG. None of the tested applications were able to show a statistically significant correlation between sleep efficiency, durations of wake time, light sleep or deep sleep with PSG. Snore time correlated well in both of the two applications which provided such information. Deep sleep duration and WASO measured by two of the three wearable devices correlated significantly with PSG. Even after controlling for transition count and moving count, the correlation indices of the wearables did not increase, suggesting that the algorithms used by the wearables were not largely affected by tossing and turning.

Conclusions Most of the applications tested in this study showed poor validity, while wearable devices mildly correlated with PSG. An effective use for these devices may be as a tool to identify the change seen in an individual's sleep patterns on a day-to-day basis, instead of as a method of detecting absolute measurements.

Keywords Sleep · Smartphone · Actigraphy · Polysomnography

Introduction

Good quality sleep is essential for optimal health [1, 2]. However, a considerable portion of the population is not able to enjoy the benefits of healthy sleep. According to the Centers for Disease Control and Prevention, about 70 million Americans suffer from chronic sleep problems. Sleep disorders generally pertain to problems with the quality, duration, and timing of sleep.

Overnight in-laboratory polysomnography (PSG) is currently the gold standard for monitoring sleep. However, a

level I PSG requires various equipment as it consists of at least eight channels (electroencephalography, electrooculography, electromyography of the chin and bilateral anterior tibialis, electrocardiography, abdominal and thoracic respiratory effort, airflow, and pulse oximetry) and is labor-intensive as it has to be attended by a sleep technologist [3]. Furthermore, PSG is not a tool that can be used on a nightly basis and thus is not a viable method for long-term monitoring of sleep. Therefore, due to the inherent limitations of PSG, alternative sleep monitoring tools are attracting interest for clinical and personal use. Wrist-worn sleep and activity trackers with wearable sensors (wearables) and mobile sleep applications (apps) are most commonly used because of their convenience [4]. However, it is not known how accurately they measure sleep and sleep disturbance events. In 2018, the American Academy of Sleep Medicine (AASM) published a position statement cautioning against the use of consumer sleep technology as a replacement for validated diagnostic tools as there are limited data regarding

✉ Hyun Jun Kim
entkhj@ajou.ac.kr

¹ Department of Otorhinolaryngology-Head and Neck Surgery, Kangdong Sacred Heart Hospital, Hallym University College of Medicine, Seoul, Republic of Korea

² Department of Otolaryngology, Ajou University School of Medicine, Suwon, Republic of Korea

the scientific accuracy of such devices or apps [5]. To forward the understanding of consumer sleep technology, we aimed to determine prospectively the validity of wearable sleep-tracking devices and smartphone apps by comparing the data to that of PSGs.

Subjects and methods

Patients who underwent one night of attended-PSG at the sleep clinic of a single tertiary institution from January 2015 to July 2019 were recruited for this study. Informed consent was obtained from all individual participants included in the study. This study was conducted in accordance with the Declaration of Helsinki and was approved by the institutional IRB (AJIRB-MED-MDB-17-328, AJIRB-DEV-DEO-14-293).

Polysomnography

Polysomnography was performed in the sleep laboratory of Ajou University (Embla N7000, ResMed, Amsterdam, Netherlands) and was used to record the following items: airflow using oronasal thermal airflow sensors, thoracic and abdominal movements using plethysmography belts, body position sensors, pulse oximetry, one submental and one anterior tibialis electromyogram, an electroencephalogram using six channels (C3/A2, C4/A1, F3/A2, F4/A1 O1/A2, and O2/A1), right and left electrooculograms, and an electrocardiogram. Scoring was conducted manually by a trained sleep specialist in accordance with the American Academy of Sleep Medicine scoring guidelines [6]. Sleep was scored in sleep stages: N1, N2, N3, and rapid eye movement (REM) sleep.

Smartphone applications

Nine smartphone applications were chosen based on their popularity at Google Store and Apple App Store at the time of study initiation (in alphabetical order); Good Sleep!, MotionX, Sleep Analyzer, Sleep Cycle, Sleep Time, Smart Sleep, SnoreClock, SnoreLab, and WakeApp Pro.

Three smartphones were used during the study—Samsung Galaxy K (SHW-M130K, version 2.3.3), iPhone 4, and later, iPhone 6 (ios 10.2.1). The smartphones were placed on the bed or on the pillow as recommended by the application developer. To accurately determine the start and end time, we synchronizing the app start time with the Embla® Rem-Logic™ Software “Lights Off” time and the app end time with the “Lights On” time. The subject was told not to touch the smartphone and sleep as normally as possible.

Wearables

Mi Band 2 (Xiaomi Inc., China; Mi Fit 4.0.15), Gearfit 2 (Samsung Inc., Korea; Samsung Health 6.8.5.009), and Fitbit Alta HR (Fitbit Inc., USA; Fitbit 6.4.2) were used as wearable wrist-worn sleep and activity trackers. The actigraph was strapped on the non-dominant wrist. At the end of the sleep period, data were downloaded from the wrist device to the application on the smartphone.

Total sleep time (TST: number of minutes scored as sleep between start and end of sleep period time) was recorded from all three wearables, while deep sleep, light sleep, REM sleep, and wake time after sleep onset (WASO: number of minutes scored as wake between start and end of SPT) were analyzed when applicable. Light sleep was defined as the sum of stage N1 and N2 sleep, while deep sleep was identified as stage N3.

Statistical analyses

Statistical analyses were performed using Student’s *t* test, Spearman’s rank correlation coefficient (R_s) test and intra-class correlation (ICC) test. Bland-Altman plots were used to assess the limit of agreement. All statistical analyses were performed with R software (version 3.5.1). $P < 0.05$ was considered to be statistically significant.

Results

Demographic data

We analyzed the results of 495 cases of smartphone applications and 170 cases of wearables by comparing each against a full night PSG. The demographic data and PSG data are shown in Table 1, and all parameters were not statistically different between the two groups.

Smartphone applications

Table 2 shows the correlation indices between the consumer-grade sleep apps and PSG, and Figure 1 assesses the limit of agreement. Sleep efficiency, wake time, duration of light sleep, duration of deep sleep, sleep onset, percentage of wake time, percentage of light sleep time, percentage of deep sleep time, and snore time detected by the apps were compared to that determined by PSG. “Light sleep” detected by the apps was compared to the sum of stages N1 and N2 measured by PSG. None of the apps claimed they could detect REM sleep; MotionX, Sleep Time, and Wakeapp Pro stated that their “deep sleep” was

Table 1 Characteristics of subjects and polysomnography (PSG) results

	PSG data of app group	PSG data of wearable group
Number of cases	495	170
Sex ratio (M:F)	446:49	149:21
Age (years)	41.9 ± 12.9	41.6 ± 13.4
BMI (kg/m ²)	26.5 ± 4.4	26.7 ± 3.8
AHI (/h)	28.1 ± 26.2	30.2 ± 24.5
RDI (/h)	37.7 ± 30.1	36.8 ± 22.8
ODI (/h)	27.1 ± 25.5	28.4 ± 23.7
TRT (min)	458.8 ± 21.9	459.9 ± 23.6
TST (min)	381.2 ± 56.6	388.3 ± 52.6
Sleep efficiency (%)	83.2 ± 12.1	84.5 ± 10.6

BMI body mass index, *AHI* apnea-hypopnea index, *RDI* respiratory disturbance index, *ODI* oxygen desaturation index, *TRT* total recording time, *TST* total sleep time

the sum of stage N3 and REM sleep, while the other apps could not be verified.

None of the tested apps were able to show a statistically significant correlation between sleep efficiency, duration of wake time, light sleep duration, and deep sleep duration with PSG. Sleep time paradoxically correlated inversely with PSG when comparing percent of light sleep and percent of deep sleep, but the absolute light sleep time and deep sleep time showed no statistical significance. Although sleep analyzer’s wake time percentage significantly correlated with PSG, the Spearman’s rho was only 0.27 ($p = 0.03$), and as wake time duration and sleep efficiency did not correlate with PSG, it would be difficult to argue that sleep analyzer’s wake time percentage has clinical significance. Furthermore, the limits of agreement seen on the Bland-Altman

plot (Fig. 1a) were wide, with the difference between 54.1 and -56.3%. Sleep onset was only assessed by WakeApp Pro, and it significantly but weakly corresponded with PSG ($R_s = 0.30$, $p = 0.021$), but the Bland-Altman plot (Fig. 1B) suggested that the app underestimated sleep onset with the bias being 10.85 min. Snore time correlated well in both of the two apps which provided such information (SnoreClock and SnoreLab.), but both Bland-Altman plots (Fig. 1c, d) had wide limits of agreement and the mean bias was 71.3 min and 106.1 min, respectively.

Wearables

Table 3 depicts the correlation and agreement indices between the wearables and full-night PSG, while Figure 2 and 3 assess the limit of agreement. The TST, light sleep duration, deep sleep duration, REM sleep duration, WASO, and sleep efficiency outcomes were compared. As with the apps data, ‘light sleep’ detected by the wearables was compared to the sum of stages N1 and N2 measured by PSG. According to the manufacturer, Fitbit Alta HR is capable of detecting REM sleep, so “deep sleep” measured by Fitbit was compared to only N3 stage measured by PSG. On the other hand, Mi Band 2 did not provide any information regarding REM sleep.

TST results (Spearman’s rho) significantly correlated with PSG in all three wearables, although only the ICC of Mi Band 2 was statistically significant ($ICC = 0.297$, $p = 0.013$). On the other hand, the Bland-Altman plot of Fitbit Alta HR (Fig. 2C) had the narrowest limits of agreement of the three plots. WASO measured by Mi Band 2 correlated significantly with PSG, but ICC was not significant and the Bland-Altman plot did not show good agreement levels (Fig. 3a). Sleep efficiency checked by Gear Fit2 did not correlate with PSG.

Table 2 Correlation indices (Spearman Rs) between polysomnography and apps’ sleep variables

Application	n	Polysomnography								
		Sleep efficiency	Wake	Light sleep	Deep sleep	Sleep onset	Wake time %	Light sleep %	Deep sleep %	Snore time
Good Sleep!	45	0.058					0.034			
MotionX	64	-0.004	-0.04	0.186	-0.01					
Sleep Analyzer	64	0.222	0.153	0.129	0.121		0.268*	0.021	0.055	
Sleep Cycle	52	-0.262								
Sleep Time	65	0.006	-0.11	0.07	0.000		-0.131	-0.371*	-0.302*	
Smart Sleep	48				-0.12				-0.195	
SnoreClock	47									0.483***
SnoreLab	54									0.475**
WakeApp Pro	56	0.02	0.018	0.058	-0.06	0.307*				

* $p < 0.05$

** $p \leq 0.01$

*** $p \leq 0.001$

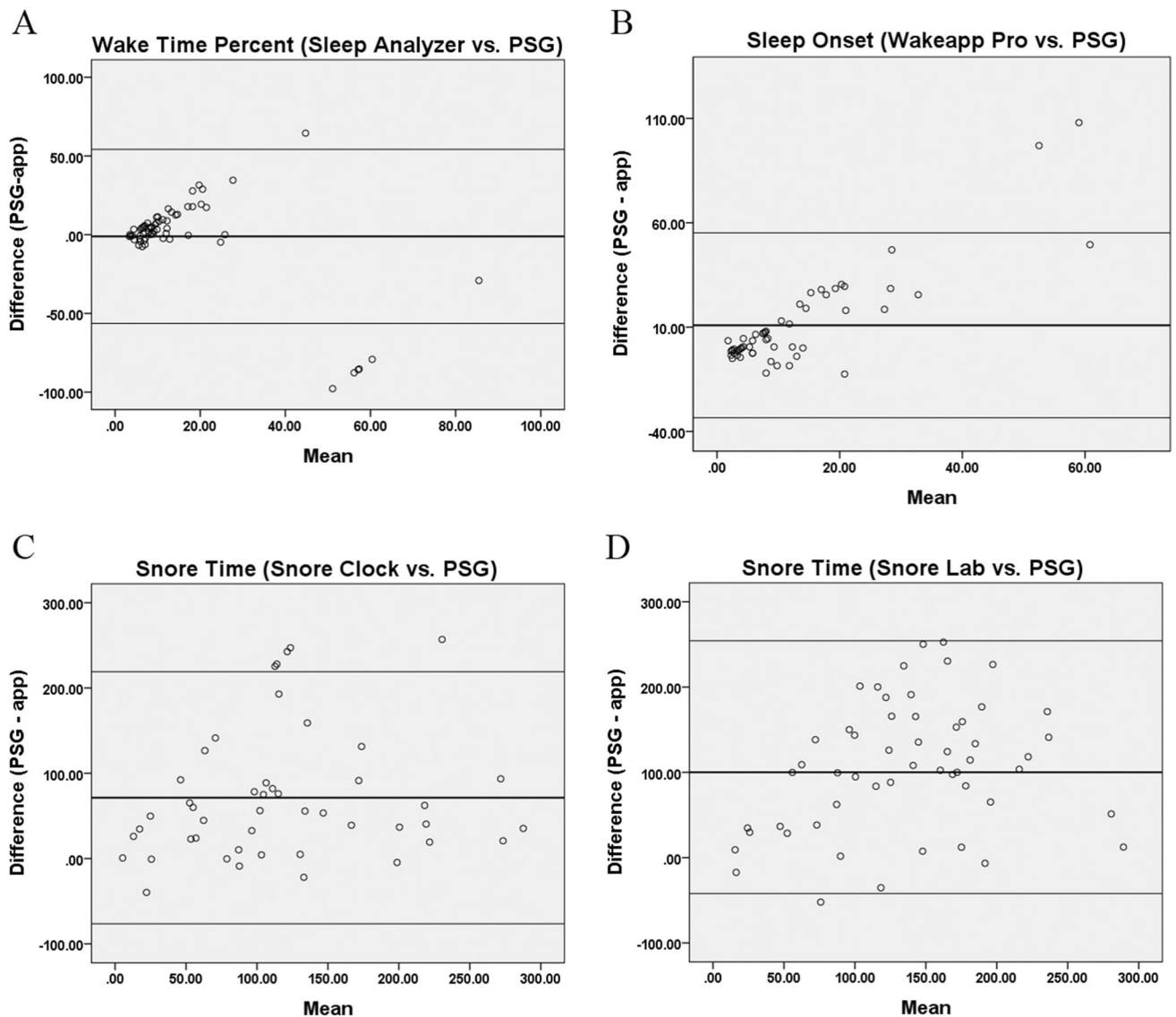


Fig. 1 Bland-Altman plots comparing sleep variables between applications and polysomnography (PSG). (**A**) variable: wake time percent; app: Sleep Analyzer, (**B**) variable: sleep onset; app: Wakeapp

Pro, (**C**) variable: snore time; app: Snore Clock, (**D**) variable: snore time, app: Snore Lab

Interestingly, the deep sleep duration, REM sleep duration, and WASO measured by the Fitbit Alta HR significantly correlated with PSG results, although Fitbit Alta HR underestimated the duration of light sleep when compared to PSG (253 min vs 287 min). However, intra-class correlations revealed only mild validity for the three parameters (deep sleep duration, $R_s = 0.372$, $ICC = 0.301$; REM duration, $R_s = 0.310$, $ICC = 0.323$; and WASO, $R_s = 0.425$, $ICC = 0.213$). The limits of agreement on the three Bland-Altman plots (Figs. 3b-d) for Fitbit Alta HR regarding deep sleep duration, REM sleep duration, and WASO

were wide and did not show good agreement between the wearable and PSG.

Controlling for transition count, transition index, and moving count

As consumer-grade apps and wearables generally use an accelerometer to measure sleep parameters, we hypothesized that the transition count/index and moving count may possibly be confounding factors when determining the correlation between apps/wearables and PSG. Transition

Table 3 Correlation and agreement indices between polysomnography (PSG) and wearable sleep trackers' variables.

Wearable	n	PSG					SE
		TST	Light sleep	Deep sleep	REM sleep	WASO	
Mi Band2	55	r = 0.367 ** ICC = 0.297*	r = 0.032 ICC = 0.024	r = 0.116 ICC = 0.095		r = 0.383 ** ICC = 0.148	
Gear Fit2	54	r = 0.307 * ICC = 0.209					r = 0.18 ICC = 0.048
Fitbit Alta HR	61	r = 0.466 *** ICC = 0.205	r = 0.179 ICC = -0.19	r = 0.372 ** ICC = 0.301 **	r = 0.310 * ICC = 0.323 **	r = 0.425 *** ICC = 0.213 *	

Light sleep: sum of N1 and N2 sleep

Deep sleep: N3 sleep

TST: Total sleep time

REM: Rapid eye movement

WASO: Wake time after sleep onset

SE: Sleep efficiency

r: Spearman's rho

ICC: Intraclass correlation coefficient

* $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

count was defined as the total number of times the patient changed sleeping positions, while transition index was calculated as transition count per hour. Movement was identified as the maximum change of gravity per second detected by the gravity body position sensors of the PSG and moving count was defined as the total number of movements.

Table 4 depicts the correlation indices for the apps/wearables after controlling individually for transition count, transition index or moving count, calculated by using partial correlation analyses. After controlling for each of the parameters, the correlation indices of Fitbit Alta HR unexpectedly decreased, while on the other hand, there was not a large gap between the correlation indices of Mi Band 2. The statistical insignificance of sleep efficiency measured by Gear Fit2 was not altered after controlling for the three movement parameters.

Light sleep duration detected by the MotionX app did not originally significantly correlate with PSG ($R_s = 0.186$, $p = 0.14$) but became significant after controlling for transition index ($R_s = 0.262$, $p = 0.04$) or moving count ($R_s = 0.312$, $p = 0.01$). This might indicate that the algorithm used in the MotionX app attached more importance to movement when compared to other apps.

Wake time percent measured by sleep analyzer and sleep onset measured by WakeApp Pro remained significant even after controlling for the three parameters.

Discussion

The present study assessed the reliability of smartphone applications and wearables for estimating various sleep-related outcomes against PSG, the gold-standard method.

Regrettably, we were unable to show evidence that commercially available apps or wearables provide comparable data regarding sleep parameters when matched with PSG. We determined that while the TST measured by all three wearables tested correlated well with PSG, none of the nine tested smartphone applications were able to appraise TST. This is in line with a previous study where Fino et al. tested four smartphone apps and reported that there was a high correspondence between all the apps and "time in bed (TIB)" measured by PSG, but only one app reliably detected TST [7]. Multiple studies have shown that measurement of sleep parameters using algorithms based on movement detected by the smartphone's accelerometer cannot accurately assess sleep stages [8–10]. However, one of the limitations of previous studies were that the study populations were very small, with about 10–20 cases tested per app. One strength of this study is that we tested nine apps, with an average of about 55 cases per app. To our knowledge, this is the most number of cases tested in the literature up to date. Even with the larger study population, the results were similar to the previous studies.

In our study performed on patients with suspicion of sleep disorders, compared to PSG, the sleep time app was not able to effectively measure sleep efficiency, wake time, light sleep duration, or deep sleep duration. Bhat et al. compared the sleep time app to PSG in twenty healthy volunteers and also found that there was no correlation between PSG and app sleep efficiency, light sleep duration percentage, and deep sleep duration percentage [11]. According to their report, the sleep time app had high sensitivity (detecting sleep) but low specificity (detecting wake).

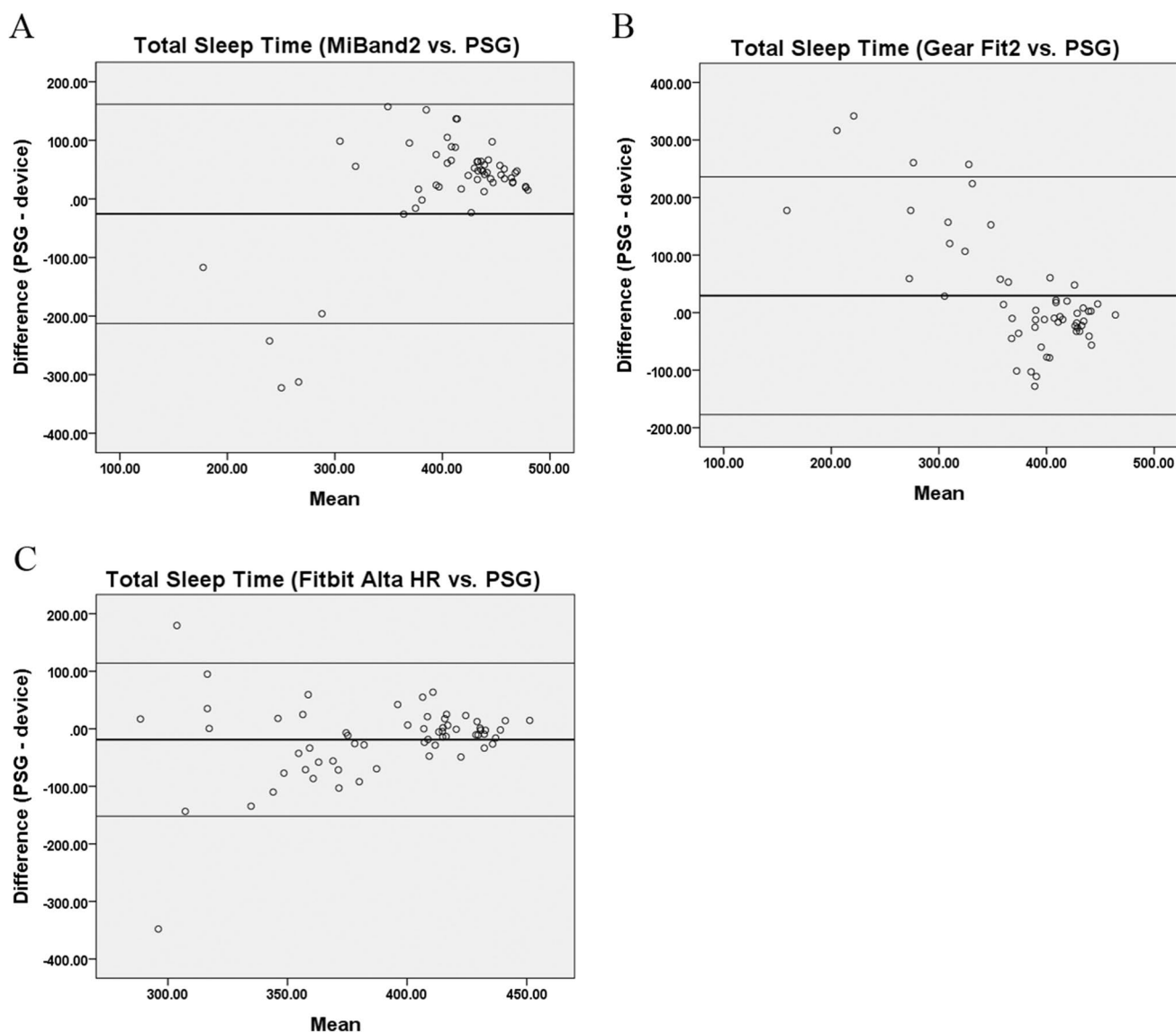


Fig. 2 Bland-Altman plots comparing total sleep time (min) measured by polysomnography (PSG) and wearables (A) Mi Band2, (B) Gear Fit2, and (C) Fitbit Alta HR

In contrast to sleep stage measurements, the detection of snoring time and sleep apnea based on sounds recorded by the smartphone's built-in microphone significantly correlated to that of PSG. The two apps that could measure snoring time (SnoreClock and SnoreLab) showed that snoring time correlated well with that of PSG even though the Bland-Altman plots suggested that both apps generally underestimated the snoring time. These apps may still be able to help monitor snoring patients, especially those who undergo surgery or conservative treatment such as weight loss or lifestyle changes [12].

In regard to wearables, all three devices (Mi Band 2, Gear Fit 2, and Fitbit Alta HR) were able to reliably measure TST when compared to PSG. The Fitbit data is in agreement with

previous studies where it was reported that assessment of sleep outcomes by Fitbit was valid in insomniacs and good sleepers [13, 14]. A recent study in healthy adults also suggested that the Fitbit Charge 2 model showed high sensitivity in detecting sleep and precisely identified 82% of non-REM and REM sleep cycles [15]. Our study demonstrated that the Fitbit data also mostly correlated with PSG in patients suspicious of OSA. Another study showed that two different sleep trackers revealed poor validity when compared to PSG [16]. Fitbit was reliably able to determine deep sleep, REM stage sleep, and WASO when compared to PSG. The Fitbit model used in this study, Fitbit Alta HR, is equipped with a heart rate sensor, which supposedly enables it to estimate

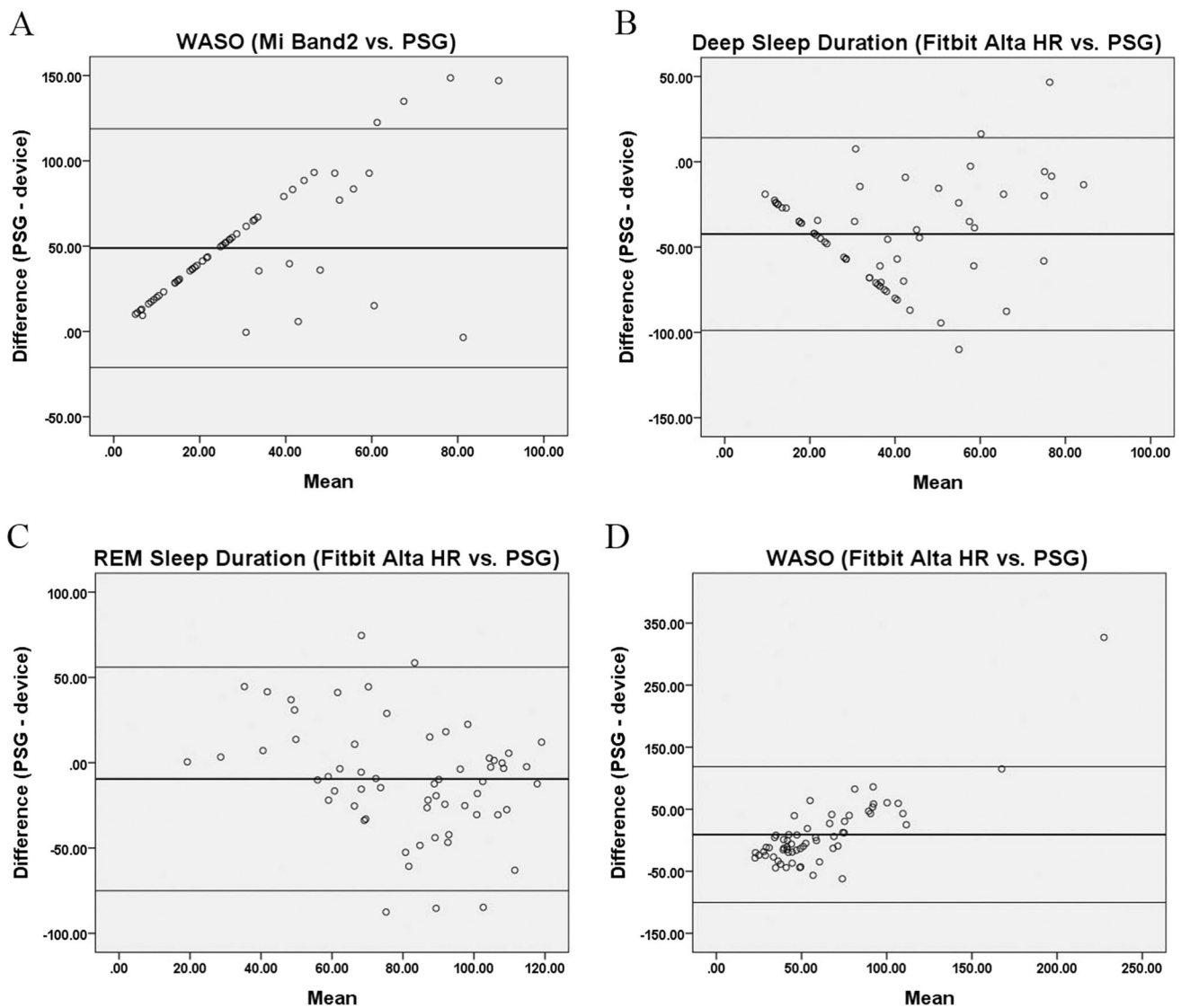


Fig. 3 Bland-Altman plots comparing sleep variables between wearables and polysomnography (PSG). **(A)** variable: WASO (wake time after sleep onset); Mi Band2, **(B)** variable: deep sleep duration; Fitbit

Alta HR, **(C)** variable: REM sleep duration; Fitbit Alta HR, **(D)** variable: WASO, Fitbit Alta HR.

light, deep, and REM sleep stages by matching the heart rate data with the body movement data from the accelerometer.

Mi Band differentiates light sleep from deep sleep by combining body movement data detected by a triaxial accelerometer and heart rate data monitored by a photoplethysmography sensor [17]. In our study, Mi Band 2 was also capable of detecting TST and WASO, when compared with PSG, albeit weakly and with low agreement. In a recent validation study conducted to determine whether low-cost trackers were able to measure TST in adults in a free-living environment, Mi Band 2 was also shown to have weak correlation and low agreement with the free-living golden standard used in the study [18]. Although Mi Band 2’s manual did not provide any information regarding REM

sleep, correlation between the “deep sleep” measured by Mi Band 2 was higher when compared with N3+REM (data not shown), instead of just N3 only, suggesting that Mi Band 2 is not able to distinguish REM sleep from N3 sleep.

Even though some of the tested apps and wearables were able to show a statistically significant correlation with PSG regarding several sleep parameters, the indices were low and strong correlation was not observed. As consumer-grade apps and wearables generally use an accelerometer to measure sleep parameters, we hypothesized that a high transition count/index and/or moving count may possibly be detrimental for the devices to accurately measure sleep, especially in patients who toss and turn frequently while sleeping. When controlling for either transition index or

Table 4 Correlation indices (Spearman's rho) of app/wearable variables compared to polysomnography when controlling for transition count, transition index, and moving count

Variable	Controlled parameter	Fitbit	Gear Fit	Mi Band	MotionX	Sleep Analyzer	Wakeapp Pro
WASO	None	0.425***		0.383**			
	Transition count	0.185		0.364**			
	Transition index	0.083		0.336*			
	Moving count	0.342**		0.372**			
REM sleep	None	0.310*					
	Transition count	0.256*					
	Transition index	0.204					
	Moving count	0.292*					
Light sleep (N1+N2)	None	0.179		0.032	0.186		
	Transition count	0.140		0.021	0.236		
	Transition index	0.159		0.010	0.262*		
	Moving count	0.145		0.029	0.312*		
Deep sleep (N3)	None	0.333**		0.116			
	Transition count	0.330**		0.160			
	Transition index	0.347**		0.162			
	Moving count	0.337**		0.097			
Wake time percent	None					0.268*	
	Transition count					0.292*	
	Transition index					0.265*	
	Moving count					0.273*	
Sleep Efficiency	None		0.179				
	Transition count		0.135				
	Transition index		0.103				
	Moving count		0.144				
Sleep Onset	None						0.307*
	Transition count						0.293*
	Transition index						0.289*
	Moving count						0.304*

REM rapid eye movement, WASO wake time after sleep onset

* $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

moving count, correlation of the light sleep detected by the MotionX app with PSG data became statistically significant, indicating that the more the patient changed positions or moved considerably, the MotionX app was not able to accurately count light sleep.

However, surprisingly, the correlation of WASO measured by Fitbit Alta HR with PSG became insignificant when controlled for transition count and index. The correlation of Fitbit's WASO with PSG remained significant, but the correlation index was decreased when controlled for moving count. The correlation of deep sleep (N3) did not significantly differ when controlling for transition or movement, probably due to the fact that there is ordinarily little movement during deep sleep. The correlation of WASO and deep sleep detected by Mi Band 2 was similar before and after controlling for transition count/index and

moving count. Variable wearables typically use sensors with similar capacities, so the differences seen here may be attributed to the different algorithms used to detect sleep for each wearable program. Even though the Fitbit Alta HR and Mi Band 2 both have actimetric sensors and can monitor heart rate, our results seem to suggest that Fitbit uses an algorithm based more on body movement, while Mi Band's algorithm may be more focused on heart rate. Furthermore, we were able to determine that the algorithms used by the wearables was not largely affected by tossing and turning, and were more accurate than previously perceived.

Although we did not directly compare the sleep parameters measured by the smartphone apps to that of the wearables, in general, even though the PSG data of the two groups were comparable, the wearables seem to be more reliable

as sleep-trackers when compared to the results of the sleep apps. Spearman's correlation between PSG and two wearables (Fitbit and Mi Band 2) was statistically significant but the ICC's indicated low agreement. Based on these findings, although the absolute numbers measured may not be accurate, an effective use for these wearables may be to use them as a tool to identify the change seen in an individual's sleep patterns on a day-to-day basis.

Most of the apps tested in this study showed poor validity when compared to the wearable devices. One way to increase the accuracy of apps may be to use the apps in conjunction with other devices, such as wristbands or a micromovement-sensitive mattress with a sleep monitoring system [19]. One disadvantage of wearable devices, compared to apps, is that wearables do not contain microphones and thus are not able to measure snoring time. Future studies using three-way comparisons (app-wearable-PSG) are warranted to further our knowledge on various sleep-monitoring products. The object of this study was not to make judgements on any one specific app or device but to evaluate the clinical usefulness of these more easily accessible tools. We acknowledge the fact that these sleep monitoring wearables and applications were developed as commercial products and not as medical devices to be used for formal purposes. However, data regarding the validity of such products is needed to lower the risk of misleading consumers. Clinicians need to be vigilant as the adoption of such sleep-monitoring wearables and apps by the general public is on the rise.

Funding This research was supported by a National Research Foundation of Korea grant funded by the Korean Government (2017R1E1A1A01074543) and the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (HC15C3415).

Declarations

Ethics approval This study was approved by the institutional IRB (AJIRB-MED-MDB-17-328, AJIRB-DEV-DEO-14-293).

Conflict of interest The authors do not have any financial or nonfinancial conflicts of interest to disclose.

References

- Luyster FS, Strollo PJ, Jr., Zee PC, Walsh JK, Boards of Directors of the American Academy of Sleep Medicine, the Sleep Research Society (2012) Sleep: a health imperative. *Sleep* 35(6):727–734
- Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Buysse D, Dinges DF, Gangwisch J, Grandner MA, Kushida C, Malhotra RK, Martin JL, Patel SR, Quan SF, Tasali E (2015) Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society. *Sleep* 38(6):843–844
- Kryger MH, Roth T, Dement WC (2017) Principles and practice of sleep medicine, Sixth, edition. Elsevier, Philadelphia, PA
- Baron KG, Duffecy J, Berendsen MA, Cheung Mason I, Lattie EG, Manalo NC (2018) Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. *Sleep Med Rev* 40:151–159
- Khosla S, Deak MC, Gault D, Goldstein CA, Hwang D, Kwon Y, O'Hearn D, Schutte-Rodin S, Yurcheshen M, Rosen IM, Kirsch DB, Chervin RD, Carden KA, Ramar K, Aurora RN, Kristo DA, Malhotra RK, Martin JL, Olson EJ, Rosen CL, Rowley JA, Academy American, of Sleep Medicine Board of D, (2018) Consumer Sleep Technology: An American Academy of Sleep Medicine Position Statement. *J Clin Sleep Med* 14(5):877–880
- Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, Marcus CL, Mehra R, Parthasarathy S, Quan SF, Redline S, Strohl KP, Davidson Ward SL, Tangredi MM, American Academy of Sleep Medicine (2012) Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med* 8(5):597–619
- Fino E, Plazzi G, Filardi M, Marzocchi M, Pizza F, Vandi S, Mazzetti M (2020) (Not so) Smart sleep tracking through the phone: Findings from a polysomnography study testing the reliability of four sleep applications. *J Sleep Res* 29(1):e12935
- Choi YK, Demiris G, Lin SY, Iribarren SJ, Landis CA, Thompson HJ, McCurry SM, Heitkemper MM, Ward TM (2018) Smartphone Applications to Support Sleep Self-Management: Review and Evaluation. *J Clin Sleep Med* 14(10):1783–1790
- Fino E, Mazzetti M (2019) Monitoring healthy and disturbed sleep through smartphone applications: a review of experimental evidence. *Sleep Breath* 23(1):13–24
- Lee-Tobin PA, Ogeil RP, Savic M, Lubman DI (2017) Rate My Sleep: Examining the Information, Function, and Basis in Empirical Evidence Within Sleep Applications for Mobile Devices. *J Clin Sleep Med* 13(11):1349–1354
- Bhat S, Ferraris A, Gupta D, Mozafarian M, DeBari VA, Gushway-Henry N, Gowda SP, Polos PG, Rubinstein M, Seidu H, Chokroverty S (2015) Is There a Clinical Role For Smartphone Sleep Apps? Comparison of Sleep Cycle Detection by a Smartphone Application to Polysomnography. *J Clin Sleep Med* 11(7):709–715
- Kazikdas KC (2019) The Perioperative Utilization of Sleep Monitoring Applications on Smart Phones in Habitual Snorers With Isolated Inferior Turbinate Hypertrophy. *Ear Nose Throat J* 98(1):32–36
- Kang SG, Kang JM, Ko KP, Park SC, 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
- Kahawage P, Jumabhoy R, Hamill K, de Zambotti M, Drummond SPA (2020) Validity, potential clinical utility, and comparison of consumer and research-grade activity trackers in Insomnia Disorder I: In-lab validation against polysomnography. *J Sleep Res* 29(1):e12931
- de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC (2018) A validation study of Fitbit Charge 2 compared with polysomnography in adults. *Chronobiol Int* 35(4):465–476
- Gruwez A, Bruyneel AV, Bruyneel M (2019) The validity of two commercially-available sleep trackers and actigraphy for assessment of sleep parameters in obstructive sleep apnea patients. *PLoS One* 14(1):e0210569

17. Ameen MS, Cheung LM, Hauser T, Hahn MA, Schabus M (2019) About the Accuracy and Problems of Consumer Devices in the Assessment of Sleep. *Sensors (Basel)* 19(19)
18. Degroote L, Hamerlinck G, Poels K, Maher C, Crombez G, De Bourdeaudhuij I, Vandendriessche A, Curtis RG, DeSmet A (2020) Low-Cost Consumer-Based Trackers to Measure Physical Activity and Sleep Duration Among Adults in Free-Living Conditions: Validation Study. *JMIR Mhealth Uhealth* 8(5):e16674
19. Xu ZF, Luo X, Shi J, Lai Y (2019) Quality analysis of smart phone sleep apps in China: can apps be used to conveniently screen for obstructive sleep apnea at home? *BMC Med Inform Decis Mak* 19(1):224

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.