JCSM
Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

The Validity of a New Consumer-Targeted Wrist Device in Sleep Measurement: An Overnight Comparison Against Polysomnography in Children and Adolescents

Anu-Katriina Pesonen, PhD; Liisa Kuula, MA

Department of Psychology and Logopedics, University of Helsinki, Finland

Study Objectives: The validity of consumer-targeted wrist-worn sleep measurement systems has been little studied in children and adolescents. We examined the validity of a new fitness tracker (PFT) manufactured by Polar Electro Oy and the previously validated Actiwatch 2 (AW2) from Philips Respironics against polysomnography (PSG) in children and adolescents.

Methods: Seventeen children (age 11.0 ± 0.8 years) and 17 adolescents (age 17.8 ± 1.8 years) wore the PFT and AW2 concurrently with an ambulatory PSG in their own home for 1 night. We compared sleep onset, offset, sleep interval (time from sleep on to offset), actual sleep time (time scored as sleep between sleep on and offset), and wake after sleep onset (WASO) between accelerometers and PSG. Sensitivity, specificity, and accuracy were calculated from the epoch-by-epoch data.

Results: Both devices performed adequately against PSG, with excellent sensitivity for both age groups (> 0.91). In terms of specificity, the PFT was adequate in both groups (> 0.77), and AW2 adequate in children (0.68) and poor in adolescents (0.58). In the younger group, the PFT underestimated actual sleep time by 29.9 minutes and AW2 underestimated actual sleep time by 43.6 minutes. Both overestimated WASO, PFT by 24.4 minutes and AW2 by 20.9 minutes. In the older group, both devices underestimated actual sleep time (PFT by 20.6 minutes and AW2 by 26.8 minutes) and overestimated WASO (PFT by 12.5 minutes and AW2 by 14.3 minutes). Both devices were accurate in defining sleep onset.

Conclusions: This study suggests that this consumer-targeted wrist-worn device performs as well as, or even better than, the previously validated AW2 against PSG in children and adolescents. Both devices underestimated sleep but to a lesser extent than seen in many previous validation studies on research-targeted accelerometers.

Keywords: accelerometer, actigraphy, adolescent, child, sleep, polysomnography, validation

Citation: Pesonen AK, Kuula L. The validity of a new consumer-targeted wrist device in sleep measurement: an overnight comparison against polysomnography in children and adolescents. *J Clin Sleep Med.* 2018;14(4):585–591.

BRIEF SUMMARY

Current Knowledge/Study Rationale: The validity of consumer-targeted accelerometers in measuring sleep is little studied. Most validation studies focus only on devices designed for research use, though sleep-tracking devices are widely used among consumers. This study compared two wrist-worn devices (Polar Electro Oy and Philips Respironics Actiwatch 2) against polysomnography in children and adolescents.

Study Impact: Both wrist-worn devices had high sensitivity, but specificity was generally higher for the Polar fitness tracker than for the Actiwatch 2. The Polar fitness tracker measures sleep at a level generally accepted in clinical and research contexts.

INTRODUCTION

As insufficient sleep[1] and/or societal pressure that delays sleep-wake behavior[2] affect an increasing number of children and adolescents, the question of how to validly measure sleep in these populations is relevant and timely. As argued before,[3] the number of night awakenings and the actual sleep onset times may not be captured by a standard sleep diary approach. In the case of younger children, parents may not be aware of the actual sleep duration,[4] limiting the use of sleep diaries as a reliable source of sleep data. During adolescence, a commitment to a sleep diary may be challenging without a strong personal motivation. Keeping a paper-and-pencil sleep log to document bedtimes is also required to extract valid sleep data from the good-quality accelerometers.[5] Although polysomnography (PSG) is the gold standard to assess clinical sleep disorders, it is not practical in examining sleep rhythm and quantity over an extended time period. Therefore, there is a need to find reliable and feasible methods to measure sleep in pediatric populations, to assess quantity and quality of sleep, and/or to track and adjust very late circadian rhythms that are common in adolescence.

The accelerometer devices that have been validated against PSG for children and adolescents are usually targeted to research and clinical contexts and are rarely affordable to the general public. To our knowledge, there is only one study that compared a commercially available wrist accelerometer against PSG in children and adolescents.[3] The report showed that the accelerometer

Table 1—Sample characteristics.

	Younger Age Group	Older Age Group
Valid data, n	17	17
Age, years, mean ± SD	11.0 ± 0.8	17.8 ± 1.8
Sex, female, n (%)	9 (53)	8 (47)
Postpubertal, n (%) *	0 (0)	17 (100)
Sleep quality †		
Excellent, %	47	0
Quite good, %	53	94
Quite poor, %	0	6

* = measured with the Pubertal Development Scale. † = self-reported for the polysomnography night. SD = standard deviation.

(Fitbit Ultra) demonstrated good sensitivity (0.86) and accuracy (0.84), but poor specificity (0.52) in the normal measuring mode, and, in the sensitive measuring mode, adequate specificity (0.79), but inadequate sensitivity (0.70) and accuracy (0.71).

In this study, we set out to assess the validity of a new, commercially available Polar fitness tracker (PFT; Polar Electro Oy, Kempele, Finland) and the previously validated Actiwatch 2 (AW2; Philips Respironics, Murrysville, Pennsylvania, United States) accelerometer against PSG in 11-year-old children and 18-year-old adolescents over 1 night slept at home. In contrast to the accelerometer devices usually used for research purposes, the PFT functions without registering the lights off and lights on times (eg, by pressing an event marker or by using a paper-and-pencil sleep diary). If working correctly, the automatic bedtime detection increases the feasibility of the measurement, as self-reported bedtimes often are not reliable. Manually inserting the analysis windows and selecting correct event markers (often resulting from several markers for one night) also takes considerable time and effort.

METHODS

Participants

We recruited the participants from local basketball teams to participate in a double-purpose study on accelerometer validation and the effect of physical activity and screen time on sleep the following night. We invited 20 males and 20 females from 4 teams targeted to 9- to 11- and 17- to 19-year-olds. Thirty-four of the 40 invited subjects had successful measurements. This sample was 85% of those invited, 50% boys, and contained 17 participants in each age group. The actualized age in the younger group was 11.0 ± 0.8 (range 9.9–12.6) years and 17.8 ± 1.8 (range 14.4–19.8) years in the older group. The characteristics of the participants are given in **Table 1**. Each team received an incentive of 2000 euros to be used for the participating team's benefit, and the participants themselves received two movie tickets each. The research protocol was approved by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioural Sciences. Each parent for those in the younger age group and all the participating children and adolescents provided informed consent.

Polysomnography

A single-night, unattended, ambulatory PSG was performed for all participants. An experienced research nurse visited the participants' homes in the evenings to set up the PSG (SOMNOscreen plus, SOMNOmedics GmbH, Germany) and in the morning to remove the leads. The measurement time was fitted to the usual sleep rhythm of the participants, except in the adolescent group, some having very late usual bedtimes (eg, 2:00 AM). Electroencephalography (EEG) measurements were recorded with gold cup electrodes at 6 EEG locations (F3, F4, C3, C4, O1 and O2) and 2 channels for the mastoids (A1, A2) according to the standardized 10/20 system. The electrocardiogram (ECG), electrooculogram (EOG) and the electromyogram (EMG) were measured by using disposable adhesive electrodes (Ambu Cardiology Blue Sensor M; Ambu Neuroline 715, Ambu A/S, Denmark), two locations for ECG and EOG, and three locations for EMG. In addition, an online reference Cz and a ground electrode in the middle of the forehead were used. The sampling rate was 256 Hz. All signals were filtered with pass band of 0.5–40 Hz (Hamming windowed sinc zero-phase FIR filter, cutoff [−6 dB] 0.25 Hz and 44.3 Hz respectively) and re-referenced to the average signal of A1 and A2 electrodes. Sleep stages from PSG data were scored manually with the DOMINO program version 2.7 (SOMNOmedics GmbH, Germany) by two experienced researchers in 30-second epochs, in accordance with the rules published by American Academy of Sleep Medicine in The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.2 (2015).

Accelerometers

All the participants wore the PFT and AW2 on their nondominant wrist during the PSG night, such that the PFT was positioned closer to the wrist, with a 1- to 2-cm gap between the devices. Wrist devices and PSG were initialized with the same computer with an atomic time clock.

Polar Fitness Tracker (PFT)

We used a prototype PFT that corresponds to the commercially available model A370 fitness tracker, with an embedded microelectromechanical systems (MEMS) 3D sensor (LIS3DH, STMicroelectronics, Geneva, Switzerland), and a sampling rate of 50 Hz. Data were uploaded to a prototype version of a web service that generated sleep-wake detection in 30-second epochs. The calculation of the sleep variables is grounded on Polar's proprietary algorithm and is uniform with the "sleep plus" feature that is available in the model A370 fitness tracker and model M430 GPS running watch (Polar Electro Oy, Kempele, Finland). The epoch-specific sleep-wake classification was exported from the service.

Actiwatch 2 (AW2)

The AW2 (Philips Respironics, Murrysville, Pennsylvania, United States) is solid-state piezo-electric sensor with a sampling rate of 32 Hz. Each 30-second epoch was scored as either "wake" or "sleep" based on the Philips Respironics default medium sensitivity threshold (40 counts per epoch). The epoch-specific sleep-wake classification was exported from the Actiware version 6.0.8 software (Philips Respironics, Murrysville, Pennsylvania, United States).

Table 2—Definitions of the sleep parameters used in the current study.

	Philips Actiware Software	Polar Sleep Plus Software	Current Study
Time scored as sleep between sleep onset and offset	Total sleep time	Actual sleep in minutes	Actual sleep time
Time between sleep onset and offset	Sleep interval	Sleep time	Sleep interval
Number of minutes awake between sleep onset and offset	Wake after sleep onset	Interruptions	Wake after sleep onset
The percentage of sleep between sleep onset and offset	Not directly available*	Actual sleep in percentages	Actual sleep percent

* = calculated for the current study from other variables: (total sleep time / sleep interval) × 100.

Other Variables

All participants were asked to report their sleep quality the morning following the PSG measurement using a four-point scale: (1) excellent, (2) quite good, (3) quite poor, or, (4) extremely poor. Additionally, all participants were instructed to complete an online questionnaire that included questions regarding sleep, health, pubertal development derived from the Pubertal Development Scale (PDS),[6] frequency and intensity of physical activity, frequency and duration of both sedentary and screen time, and highest attained education of parents. Scores from PDS were used to ensure the two age groups differed regarding their pubertal status in order to make comparisons between prepubertal and postpubertal participants.

Statistical Analyses

We used 30-second epochs in all devices to facilitate matching of the signals epochs, and PSG lights off and lights on times defined the analysis window for the accelerometers. We defined PSG sleep onset as the first epoch of sleep scored on the PSG. Sleep onset in AW2 is defined as the first minute of a consecutive 10 minutes of scored sleep, with 1 minute of activity allowed within the 10 minutes. Sleep onset definition in the PFT is based on their proprietary algorithm and is not publicly available. Please see **Table 2** for the measured sleep variables and their definitions. For the PFT, there is no option for the consumer or researcher to register lights-off time, excluding sleep efficiency from these analyses.

The first paired sample *t* tests compared sleep metrics to evaluate differences between PSG and the accelerometers in sleep onset, sleep offset, sleep interval, actual sleep time, WASO, and actual sleep percentage separately in both participating age groups. Second, we conducted epoch-by-epoch comparisons between the accelerometers and PSG in order to calculate sensitivity (the ability of an accelerometer to detect true sleep), specificity (the ability of accelerometer to detect true wake), and accuracy (the ability of the accelerometer to detect both sleep and wake),[7,8] in females, males, and in a pooled sample in both age groups. Finally, we followed the lead of the previous validation studies[3,9] using the Bland-Altman method[10] and counted the share of participants who were within 30 minutes from each other in actual sleep time and in terms of WASO as measured with the different devices. We used the Bland-Altman method to compare the two accelerometers and to estimate the percentage of individuals measured correctly with the accelerometer in relation to the gold standard measure, PSG. We illustrated the concordance between the devices with Bland-Altman mean difference plots. All analyses were conducted using SPSS version 24.0 (IBM Corp, Armonk, New York, United States).

RESULTS

Initial Analyses

Table 1 shows that all participants were prepubertal in the younger age group and postpubertal in the older group. No significant sex distribution bias was observed. When subjectively assessing the sleep quality in the PSG assessment night, 5.9% of the adolescent group rated the quality as quite poor, and the rest as quite good, whereas none of the children in the younger age group reported poor sleep quality. The mean time difference from device placement to bedtime was (hours:minutes) 2:07, 95% confidence interval (CI) 1:37 to 2:37) in the younger group, and 5:47, 95% CI 4:13 to 7:22 in the older group.

Sleep Characteristics in Accelerometers and PSG

Table 3 shows the mean sleep metrics for each measurement device in each age group and the mean differences of both wrist-worn devices compared to PSG from the paired *t* tests.

Younger Age Group

The sleep onset times in the accelerometers were all within 5 minutes from the PSG (mean difference [MD] = 5 minutes for PFT and 1 minute for AW2). Sleep offset time was correctly identified by the PFT (MD = 0 minutes), but the AW2 resulted in an underestimation of 24 minutes (P = .01). This accumulated in 22.7 minutes shorter sleep interval in AW2 in contrast to PSG (P = .04), whereas the sleep interval measured by PFT resulted in 5 minutes shorter time (P = .10). In terms of the actual sleep time, both devices underestimated the sleep, AW2 by 43.6 minutes and PFT by 28.9 minutes (both P < .001), with both accelerometers showing a comparable overestimation in WASO (24.4 minutes for PFT, 20.9 minutes for AW2; all P < .001). In the end, the actual sleep percent was rather similar for both devices, with an underestimation of 4.5% and 4.2% for PFT and AW2, respectively.

Older Age Group

The sleep onset and offset times in the accelerometers were all within 1 to 13 minutes from the PSG (sleep onset MD = 13 minutes and MD = 2 minutes; sleep offset MD = −1 minutes and MD = 10 minutes for PFT and AW2, respectively, all P > .17 except sleep offset for AW2, P = .001). Both devices underestimated sleep intervals (PFT −8.6, P = .46; AW2 −12.4,

Table 3—Summary of the sleep metrics in PSG and accelerometers and their mean differences.

	PSG	PFT	AW2	Difference PSG − PFT		Difference PSG − AW2	
	M (SD)	M (SD)	M (SD)	MD (SD)	*P*	MD (SD)	*P*
Younger Age Group							
Sleep onset, hours:minutes	21:52 (00:56)	21.57 (1:00)	21:51 (00:56)	00:05 (00:10)	.07	−00:01 (00:14)	.78
Sleep offset, hours:minutes	7:09 (00:34)	7:09 (00:34)	6:45 (00:55)	−00:00 (00:12)	.94	−00:24 (00:34)	.01
Sleep interval, minutes	557.0 (50.1)	552.1 (53.8)	534.4 (69.5)	−5.0 (11.8)	.10	−22.7 (42.2)	.04
Actual sleep, minutes	535.8 (45.0)	506.9 (52.5	492.2 (67.4)	−28.9 (9.3)	.001	−43.6 (43.3)	.001
WASO, minutes	21.2 (17.6)	45.7 (17.3)	42.1 (15.8)	24.4 (11.5)	.001	20.9 (13.0)	.001
Actual sleep %	96.3 (2.8)	91.7 (2.9)	92.1 (2.9)	−4.5 (2.2)	.001	−4.2 (2.5)	.001
Older Age Group							
Sleep onset, hours:minutes	24:35 (01:43)	24:48 (01:49)	24:37 (01:44)	00:13 (0:38)	.17	00:02 (00:11)	.48
Sleep offset, hours:minutes	7:27 (00:41)	7:25 (00:39)	7:16 (00:35)	−0:01 (0:08)	.48	00:10 (00:13)	.001
Sleep interval, min	411.7 (89.0)	403.1 (89.0)	399.3 (92.8)	−8.6 (47.2)	.46	−12.4 (19.8)	.001
Actual sleep, min	394.8 (89.3)	374.1 (76.3)	368.0 (86.2)	−20.6 (9.3)	.04	−26.8 (22.9)	.01
WASO, min	16.9 (17.7)	29.4 (14.1)	31.3 (17.5)	12.5 (23.6)	.04	14.3 (13.4)	.001
Actual sleep %	95.8 (4.5)	92.9 (2.7)	92.3 (3.9)	−2.9 (5.2)	.03	−3.5 (2.9)	.001

Actual sleep % = percentage of sleep between sleep onset and offset, AW2 = Actiwatch 2, M = mean, MD = mean difference, PFT = Polar fitness tracker, PSG = polysomnography, SD = standard deviation, WASO = wake after sleep onset.

Table 4—Sensitivity, specificity, and accuracy of the accelerometers in the younger age group.

	Females		Males		All		
	M (SD)	Range	M (SD)	Range	M (SD)	Range	Sex Difference, *P*
Sensitivity							
PFT	0.94 (0.02)	0.90–0.97	0.92 (0.04)	0.88–1.00	0.93 (0.03)	0.88–1.00	.30
AW2	0.94 (0.02)	0.88–0.96	0.91 (0.06)	0.79–0.97	0.93 (0.04)	0.79–0.97	.19
Specificity							
PFT	0.71 (0.22)	0.31–0.94	0.85 (0.19)	0.42–0.98	0.77 (0.21)	0.31–0.98	.19
AW2	0.60 (0.21)	0.25–0.86	0.76 (0.13)	0.33–0.95	0.68 (0.22)	0.25–0.95	.13
Accuracy							
PFT	0.92 (0.04)	0.85–0.95	0.91 (0.02)	0.88–0.94	0.91 (0.04)	0.85–0.95	.48
AW2	0.91 (0.03)	0.85–0.95	0.89 (0.04)	0.72–0.96	0.90 (0.05)	0.72–0.96	.42

The variables were formed as follows: true positive (TP) = the number of epochs correctly identified as sleep; false positive (FP) = the number of epochs incorrectly identified as sleep; true negative (TN) = the number of cases correctly identified as wake; false negative (FN) = the number of cases incorrectly identified as wake. Specificity = (TN / [TN + FP]); Sensitivity = (TP / [TP + FN]); Accuracy = ([TP + TN] / [TP + TN + FP + FN]). AW2 = Actiwatch 2, M = mean, PFT = Polar fitness tracker, SD = standard deviation.

P = .001). In terms of the actual sleep time, both devices resulted in an underestimation, AW2 by 26.8 minutes and PFT by 20.6 minutes (*P* < .01 and .04, respectively), with both accelerometers showing a very comparable overestimation of WASO (12.5 minutes for PFT, 14.3 minutes for AW2) (*P* < .04 and .001, respectively). The actual sleep percent was also rather similar for both devices, with an underestimation of 2.9% and 3.5% for PFT and AW2, respectively.

Epoch-by-Epoch Analyses

Table 4 shows the summary of the sensitivity, specificity, and accuracy from PFT and AW2 separately for females and males in the younger age group. There were no significant sex differences, although specificity was somewhat lower in females compared to males. With regard to sensitivity, PFT and AW2 were at equal estimation (0.93), but in specificity, PFT scored higher (0.77) compared to AW2 (0.68), both being at an adequate level. The overall accuracy was very similar in both PFT and AW2 devices (0.91 and 0.90, respectively).

Table 5 shows the summary of the sensitivity, specificity, and accuracy from PFT and AW2 separately for females and males in the older age group. No significant sex differences were observed. The sensitivity was 0.91 for PFT and 0.93 for AW2. However, there was a large difference in specificity between the devices, with PFT scoring 0.83 and AW2 0.58. In terms of overall accuracy, both devices fared well (0.90 and 0.89).

Percent of Observations Within 30 Minutes From Each Other

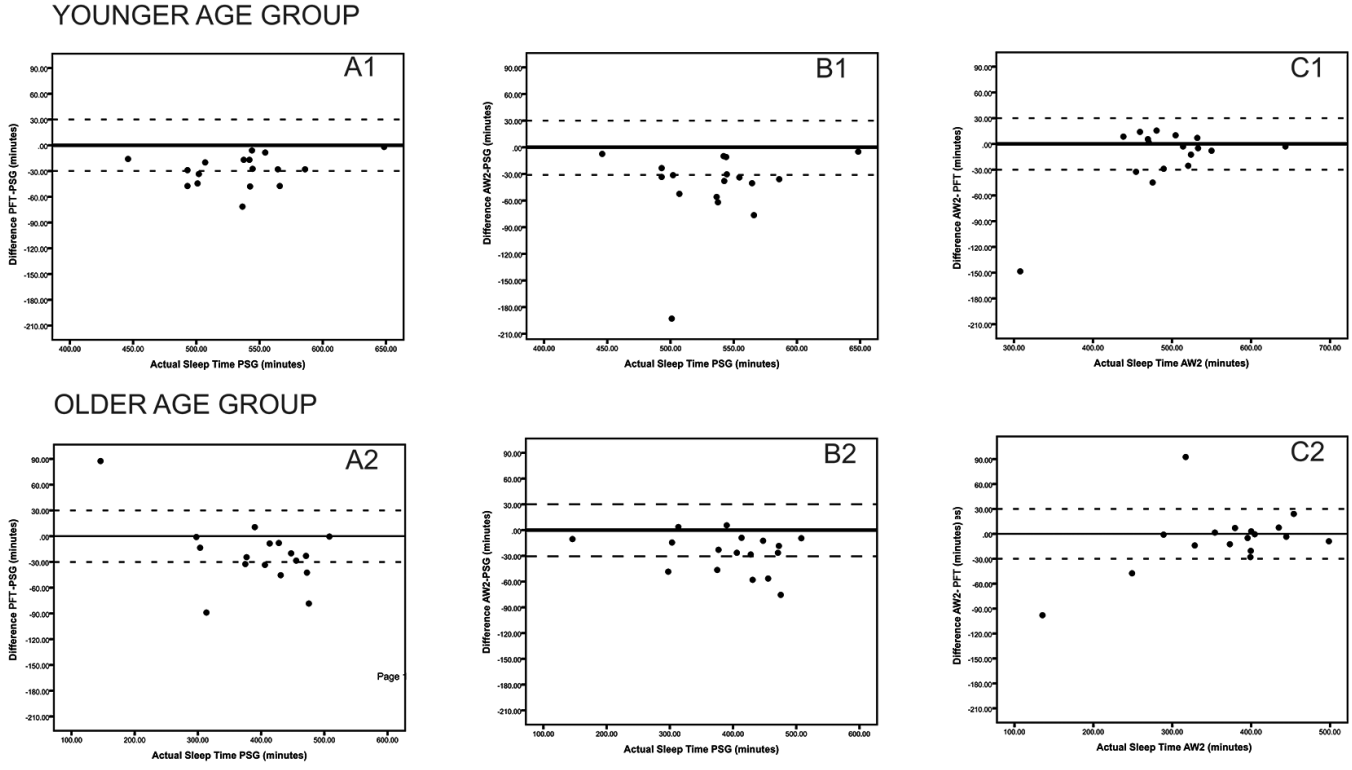
Figure 1 shows Bland-Altman plots picturing the concordance between the two accelerometers and between the accelerometers and PSG. In the younger age group, for 64.7% (PFT) and 29.4% (AW2) of the participants the difference in the actual sleep time was smaller than 30 minutes compared to PSG

Table 5—Sensitivity, specificity, and accuracy of the accelerometers in the older age group.

	Females		Males		All		
	M (SD)	Range	M (SD)	Range	M (SD)	Range	Sex Difference, *P*
Sensitivity							
PFT	0.89 (0.09)	0.68–0.96	0.92 (0.04)	0.89–0.96	0.91 (0.03)	0.68–0.96	.37
AW2	0.93 (0.03)	0.88–0.96	0.92 (0.01)	0.90–0.94	0.93 (0.02)	0.88–0.96	.71
Specificity							
PFT	0.83 (0.12)	0.67–0.95	0.83 (0.20)	0.40–1.00	0.83 (0.17)	0.40–1.00	.98
AW2	0.57 (0.19)	0.33–0.88	0.60 (0.17)	0.27–0.77	0.58 (0.22)	0.27–0.88	.71
Accuracy							
PFT	0.89 (0.07)	0.73–0.94	0.91 (0.02)	0.88–0.94	0.90 (0.05)	0.73–0.94	.38
AW2	0.90 (0.03)	0.85–0.94	0.88 (0.04)	0.79–0.92	0.89 (0.04)	0.79–0.94	.32

The variables were formed as follows: true positive (TP) = the number of epochs correctly identified as sleep; false positive (FP) = the number of epochs incorrectly identified as sleep; true negative (TN) = the number of cases correctly identified as wake; false negative (FN) = the number of cases incorrectly identified as wake. Specificity = (TN / [TN + FP]); Sensitivity = (TP / [TP + FN]); Accuracy = ([TP + TN] / [TP + TN + FP + FN]). AW2 = Actiwatch 2, M = mean, PFT = Polar fitness tracker, SD = standard deviation.

Figure 1— Bland-Altman plots.



Bland-Altman plots of the correspondence of the actual sleep time between the devices in the younger (upper panels) and older (bottom panels) age groups. **(A1,A2)** PSG and PFT, **(B1,B2)** PSG and AW2, **(C1,C2)** AW2 and PFT. AW2 = Actiwatch 2, PFT = Polar fitness tracker, PSG = polysomnography.

(**Figure 1**, panels A1 and B1). When comparing the two accelerometers, for 82.4% of the individuals the actual sleep time was within 30 minutes (**Figure 1**, panel C1). In terms of WASO, the corresponding proportions were 70.6 % (PFT versus PSG), 88.2% (AW2 versus PSG) and 100% (AW2 versus PFT).

In the older age group for 58.8% (PFT) and 70.6% (AW2) of the participants the difference in the actual sleep time was smaller than 30 minutes compared to PSG (**Figure 2**, panels A2 and B2). When comparing the two accelerometers, for 82.4% of the individuals the actual sleep time was within 30 minutes (**Figure 1**, panel C2). In terms of WASO, the corresponding proportions were 82.4% (PFT versus PSG), 88.2% (AW2 versus PSG), and 100% (AW2 versus PFT).

DISCUSSION

User-friendly wrist-worn activity and sleep trackers are increasingly available for all consumers, including children and adolescents.[11] This reflects also on the demand to understand

their true measurement capacity, as users, patients, and parents are increasingly aware and concerned about the questions of sufficient sleep. Home-based measurements using commercial wrist devices may also lead to clinical consultations, with a growing demand for the clinicians to understand the validity of these measurements.

We studied a new PFT, equivalent to their commercially available model A370 fitness tracker, that operates through their proprietary web service, and the previously validated AW2, a device mainly targeted to clinicians and researchers. We studied the validity against PSG in two age groups, one prepubertal and the other postpubertal. In this study, we found an adequate and comparable measurement validity for both accelerometers over 1 night of measurement. At some points the PFT outperformed the AW2. Both devices were adequately accurate in measuring sleep onset. When measuring sleep offset, the AW2 had a statistically significant difference from PSG—underestimating sleep offset in the younger group, and overestimating sleep offset in the older group. Both devices underestimated actual sleep time and overestimated WASO in both age groups.

With regard to the epoch-by-epoch analyses, both devices resulted in excellent and equal sensitivity. In terms of specificity, PFT was somewhat stronger compared to AW2 both in the younger and older age groups. Specificity was at an adequate level for PFT in both age groups and for AW2 in the younger age group, but poor for AW2 in the older age group. Testing whether the devices stay within a 30-minute range from the PSG in the actual sleep time revealed that the PFT clearly excelled AW2 in the younger age group, but the AW2 fared better in the older age group. A comparison between the two accelerometers showed that 82.4% participants from both age groups were scored within a 30-minute range from each other in actual sleep time. Compared to previous studies using AW2 with medium sensitivity but longer, 1-minute, epoch length,[12] we detected better sensitivity, specificity, and accuracy for AW2. However, in a study by Meltzer et al. including an age range of 3 to 18 years using high-sensitivity settings, the corresponding device to AW2, MiniMitter, also reached better specificity than observed in this study (0.68) in the younger age group.[7] Compared to the study by Meltzer et al., the current study had more specific age ranges.

Limitations

This study was conducted with great care to ensure a reliable comparison of two different accelerometers and PSG. However, some limitations are present within this protocol. First, the small sample size including only healthy participants restricts the external validity of our results. Thus, sleep-wake-measurements from children or adolescents with health issues or sleep disorders might not be estimated as accurately as in this sample. Second, the consumer products do not usually give epoch-specific data to the user, so a collaboration with the manufacturer is a necessity for potential replication studies. In these cases, it is crucial to design the data flow such that the manufacturer does not have access to the PSG data at any stages to maintain impartiality, as was done in our study. It should be also emphasized that we used a prototype of the commercially available model A370 fitness tracker, but it was similar in size and appearance to the final product. Third, our comparison covered only 1 night without any adjustment period; hence, it is possible that we did not capture the full scope of typical sleep behavior. Fourth, we did not balance the position of the accelerometers in the same wrists, but all measurements were conducted with PFT being on top of the pisiform bone, and AW2 1–2 cm above (closer to the torso). This may have been biased either way for estimations. Finally, the sample size was rather small for detecting statistical significance of any potential sex differences.

Strengths

Our study had several strengths. First, we had a highly ecologically valid sample in adolescents, with sleep rhythms varying in time and duration (eg, one participant slept for less than three hours). This represents the typically irregular sleep behavior seen in adolescents. Second, we used a 30-second epoch length in all devices allowing precise comparisons. Third, by simultaneously using two accelerometers we were able to compare the actual differences between their abilities to detect sleep and wake.

CONCLUSIONS

Wrist-worn consumer-aimed devices with integrated sleep estimations are becoming increasingly common,[11] but their role in research and clinical use is still under debate. As consumers are likely to evaluate their sleep behavior based on the estimations their device reports, it is important to critically assess the reliability of available solutions and the data their algorithms produce. As noted before, the lack of standard scoring rules and the use of manufacturer-specific sleep variables represents a problem when comparing the validity across products and age groups, making it more difficult to make use of the products in pediatric sleep medicine.[13] In addition, commercial products may have proprietary algorithms for sleep detection.

In this study we compared the output of one novel commercial PFT to the sleep parameters derived from PSG epoch to epoch. Based on this comparison we detected excellent sensitivity, adequate specificity, and excellent accuracy in the PFT in both prepubertal and postpubertal children and adolescents.

In our epoch-to-epoch comparisons we also included the previously validated AW2, and tested its sensitivity, specificity, and accuracy against PSG. These comparisons yielded similarly excellent sensitivity, considerable variation in specificity from poor (children) to adequate (adolescents), and excellent accuracy. When comparing the accelerometers, PFT had a higher specificity within both age groups, and AW2 had minimally higher sensitivity within the older age group. These results suggest PFT slightly outperformed AW2 in estimating sleep in epoch-to-epoch comparisons. We conclude that consumer-targeted accelerometers may be able to provide information on sleep that is as accurate as the current, commonly used devices in research and clinical practice. However, more validation studies are needed,[11] and the findings must be replicated in larger samples with more varying sleep profiles, including children with sleep disorders or chronic health problems.

ABBREVIATIONS

AW2, Philips Respironics Actiwatch 2
EEG, electroencephalography
ECG, electrocardiogram
EMG, electromyogram
EOG, electrooculogram
MD, mean difference
MEMS, microelectromechanical systems
PFT, Polar fitness tracker
PSG, polysomnography
SD, standard deviation
WASO, wake after sleep onset

REFERENCES

1. Owens J. Insufficient sleep in adolescents and young adults: an update on causes and consequences. *Pediatrics*. 2014;134(3):e921-e932.
2. Gradisar M, Gardner G, Dohnt H. Recent worldwide sleep patterns and problems during adolescence: a review and meta-analysis of age, region, and sleep. *Sleep Med*. 2011;12(2):110–118.
3. 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;38(8):1323–1330.
4. Belanger ME, Simard V, Bernier A, Carrier J. Investigating the convergence between actigraphy, maternal sleep diaries, and the child behavior checklist as measures of sleep in toddlers. *Front Psychiatry*. 2014;5:158.
5. Galland B, Meredith-Jones K, Terrill P, Taylor R. Challenges and emerging technologies within the field of pediatric actigraphy. *Front Psychiatry*. 2014;5:99.
6. Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: Reliability, validity, and initial norms. *J Youth Adolesc*. 1988;17(2):117–133.
7. Meltzer LJ, Walsh CM, Traylor J, Westin AM. Direct comparison of two new actigraphs and polysomnography in children and adolescents. *Sleep*. 2012;35(1):159–166.
8. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med*. 2001;2(5):389–396.
9. Werner H, Molinari L, Guyer C, Jenni OG. Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns. *Arch Pediatr Adolesc Med*. 2008;162(4):350–358.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.
11. de Zambotti M, Godino JG, Baker FC, Cheung J, Patrick K, Colrain IM. The boom in wearable technology: cause for alarm or just what is needed to better understand sleep? *Sleep*. 2016;39(9):1761–1762.
12. Meltzer LJ, Wong P, Biggs SN, et al. Validation of actigraphy in middle childhood. *Sleep*. 2016;39(6):1219–1224.
13. Meltzer LJ, Montgomery-Downs HE, Insana SP, Walsh CM. Use of actigraphy for assessment in pediatric sleep research. *Sleep Med Rev*. 2012;16(5):463–475.

ACKNOWLEDGMENTS

The authors are thankful to research assistant Teo Piitulainen for the device management, and research nurse Helena Alfthan for the sleep measurements at participants' homes.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication October 4, 2017
Submitted in final revised form November 28, 2017
Accepted for publication December 29, 2017
Address correspondence to: Anu-Katriina Pesonen, PO Box 9, Department of Psychology and Logopedics, Faculty of Medicine, 00014 University of Helsinki, Finland; Email: anukatriina.pesonen@helsinki.fi

DISCLOSURE STATEMENT

Work for this study was performed at the Department of Psychology and Logopedics, University of Helsinki. All authors have seen and approved the final version of the manuscript. The research work (conducting the study, analyzing the data, writing the manuscript) was not covered by any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Polar Electro sponsored the immediate expenses from the research nurse, research assistant, and the participant incentives. The sponsor provided the Polar prototype devices to the study, agreed on the study design by the authors, and provided the data summary (epoch-by-epoch data are not openly available through the consumer product) and technical details of the device. Polar Electro had no involvement in the analysis and interpretation of data; nor in the writing of the report; nor in the decision to submit the article for publication. Acknowledging the role of the funding source as explained, the authors take complete responsibility for the integrity of the data and the accuracy of the data analysis and report no individual conflicts of interest.