http://dx.doi.org/10.5665/sleep.1608

Direct Comparison of Two New Actigraphs and Polysomnography in Children and Adolescents

Lisa J. Meltzer, PhD[1]; Colleen M. Walsh, BS[2]; Joel Traylor[3]; Anna M. L. Westin, MS[4]

[1]Division of Pulmonary Medicine and Department of Pediatrics, Children's Hospital of Philadelphia and University of Pennsylvania School of Medicine; [2]Department of Psychology, Drexel University; [3]Division of Pulmonary Medicine, Children's Hospital of Philadelphia; [4]Department of Psychology, University of Maryland, Baltimore County

Study Objectives: To evaluate the validity and reliability of 2 new models of commercially available actigraphs compared to polysomnography for children and adolescents.

Design and Setting: Subjects concurrently wore the Ambulatory Monitoring Inc. Motionlogger Sleep Watch (AMI) and the Phillips Respironics Mini-Mitter Actiwatch-2 (PRMM) while undergoing overnight polysomnography (PSG) in a pediatric sleep laboratory housed in a tertiary care children's hospital.

Participants: 115 youth (59 girls, 56 boys), ages 3-18 years (mean 8.8 years, SD 4.4 years).

Measurements: Outcome variables were total sleep time (TST), wake after sleep onset (WASO), and sleep efficiency (SE). Epoch-by-epoch comparisons were made between the 2 devices and PSG to determine sensitivity, specificity, and accuracy. Agreement between the 2 devices was determined with *t*-tests and the Bland-Altman concordance technique. Different algorithms/sensitivities, developmental age groups, and sleep disordered breathing (SDB) status were also examined.

Results: For both device brands, sensitivity (0.89-0.97), specificity (0.54-0.77), and accuracy (0.87-0.90) were similar to previous reports. Notably, compared to PSG, both device brands significantly overestimated WASO, while the AMI device also significantly underestimated TST. Inter-device comparison of the 2 brands found poor agreement for TST, WASO, and SE. Agreement with PSG differed depending on the scoring algorithm (AMI) or sensitivity setting (PRMM), as well as across developmental age group and sleep disordered breathing (SDB) status.

Conclusions: Similar to previous reports, both new actigraph brands were found to have good sensitivity (to detect sleep), but poorer specificity (to detect wake). Study results also suggest that researchers should adjust the scoring algorithm/sensitivity depending on a study's design (e.g., young children vs. adolescents, healthy children vs. youth with SDB). Further, inter-device reliability was poor, suggesting the need for caution when comparing results across studies that use different brands of actigraphic devices.

Keywords: Actigraphy, accelerometer, polysomnography, children, adolescents, validation, sensitivity, specificity, Bland-Altman

Citation: Meltzer LJ; Walsh CM; Traylor J; Westin AML. Direct comparison of two new actigraphs and polysomnography in children and adolescents. *SLEEP* 2012;35(1):159-166.

INTRODUCTION

The question of how to accurately measure sleep patterns in children and adolescents has recently received increased attention in the literature,[1-3] with strengths and weaknesses reported for both objective and subjective measures. Polysomnography (PSG) is considered the "gold standard" for measuring sleep architecture and for detecting underlying sleep disruptors such as obstructive sleep apnea. However, as a measure of sleep patterns, PSG is limited by: (1) typically providing only 1-2 nights of information about a child's sleep, (2) an insufficient number of sleep labs prepared to test children, and (3) significant financial cost. Self- or parent-reported questionnaires are the most commonly used measure of children's sleep patterns. Yet questionnaires are limited by several aspects of reporter bias. First, parents may not always be aware of prolonged sleep onset, frequent and/or prolonged night wakings, or early morning wake times, especially in older children and adolescents who do not alert a parent when they are awake. Second, there is a higher rate of reported sleep problems when children have another behavioral or developmental disorder (e.g., autism, ADHD), even though these differences have not been supported by objective measures of sleep.[4-7] In addition, parent sleep is often disturbed by child sleep problems, which can also influence parental report. Finally, most questionnaires require respondents to estimate their sleep patterns over an extended interval, typically the past week or month, not allowing for night-to-night variability.

Actigraphy has been shown to provide a valid estimate of sleep patterns in children and adolescents.[1,8-10] An actigraph is a small, wrist-watch sized activity monitor that provides an estimate of sleep patterns based on data collected by an internal accelerometer. The collected data are then translated into epochs (typically 30 sec or 1 min) of activity. Using validated algorithms, epochs are then scored as sleep or wake.

These activity monitors are commercially available from a number of companies. However, the two brands of actigraphs most commonly utilized and reported on in the sleep literature are Ambulatory Monitoring Inc. (AMI) and Philips Respironics Mini-Mitter (PRMM). Although the devices from these companies have different measurement mechanisms and scoring algorithms, results are used interchangeably and interpreted equally across studies.[11]

Each brand of actigraph has been shown to be valid compared to PSG, yet only a handful of studies have directly compared the two with each other,[12-15] with one additional study reporting on the comparison of the PRMM to another brand less commonly used.[16] Of these, only 2 studies compared both devices

Submitted for publication January, 2011
Submitted in final revised form July, 2011
Accepted for publication July, 2011
Address correspondence to: Lisa J. Meltzer, PhD, National Jewish Health, 1400 Jackson Street, G311, Denver, CO 80206; Tel: (303) 398-1837, E-mail: meltzerl@njhealth.org

Table 1—Epoch-by-epoch comparisons

	Polysomnography		
Actigraph	**Sleep**	**Wake**	**Total**
Sleep	True sleep (TS)	False Wake (FW)	TS + FW
Wake	False sleep (FS)	True Wake (TW)	FS + TW
Total	TS + FS	FW +TW	TS + FS +TW + FW

Sensitivity = TS / (TS + FS); Specificity = TW / (FW + TW);
Accuracy = (TS + TW) / (TS + FS + TW + FW).

simultaneously to PSG.[14,15] Overall, studies comparing 2 devices have included small samples, with none including children < 12 years of age. Another consideration is that both companies released new devices in 2008 (Motionlogger Sleep Watch, Ambulatory Monitoring Inc., Ardsley, NY and Actiwatch-2, Philips Respironics Mini-Mitter, Bend, OR), with validity determined by the companies through comparison with older models of the devices (T. Kazluskly and M. Reed, personal communications).

The 2 primary aims of this study were: (1) to evaluate the validity of 2 new models of commercially available actigraphs compared to PSG for children and adolescents; and (2) to examine the inter-device reliability of the 2 new actigraph devices. The secondary aim of this study was to compare the impact of different scoring algorithms/sensitivity settings of these 2 new devices on total sleep time, wake after sleep onset, and sleep efficiency across developmental age groups (preschool, school-age, adolescent) and sleep disordered breathing status (no obstructive sleep apnea [OSA], mild OSA, moderate/severe OSA).

METHODS

Subjects

Participants were 115 youth (ages 3-18 years) who were scheduled for overnight PSG at the Children's Hospital of Philadelphia (CHOP) Sleep Lab for either a clinical evaluation (n = 104) or as part of a research protocol for youth with sickle cell disease (n = 11). All participants wore 2 actigraphs during the PSG. The study was approved by the hospital's institutional review board and informed consent and assent (when appropriate) was obtained for all participants.

Polysomnography

Overnight PSG was performed in the CHOP Sleep Lab using a Rembrandt polysomnography system (Embla, Broomfield, CO). Recorded parameters included: electroencephalography (F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1); left and right electrooculogram; submental electromyogram; bilateral tibial electromyogram; electrocardiogram; oronasal airflow with 3-pronged thermistor; nasal pressure with pressure transducer; rib cage and abdominal wall motion via respiratory impedance plethysmography; and end-tidal capnometry. Arterial oxygen saturation with pulse waveform was also recorded, as well as digital video and audio. Studies were scored based on American Academy of Sleep Medicine (AASM) pediatric criteria.[17] The sleep period was scored from "lights out" to "lights on," with lights out scheduled as close as possible with participants normal sleep schedule. Average "lights off" time was 21:17, and average "lights on" time was 06:04. All participants had ≥ 7.6 h of PSG recording completed. The apnea-hypopnea index (AHI) was used to determine sleep disordered breathing status using the following criteria: No OSA: AHI < 1.5; Mild OSA: AHI ≥ 1.5 and ≤ 5; Moderate/Severe OSA: AHI > 5.[18]

Actigraphy

Participants wore the Motionlogger Sleep Watch (AMI) and the Mini-Mitter Actiwatch-2 (PRMM) on the non-dominant wrist. Seventeen randomly selected subjects wore 2 devices of the same brand (AMI-AMI or PRMM-PRMM) to examine intra-device reliability. Placement of the actigraphs in relation to the wrist was randomly assigned (AMI-PRMM or PRMM-AMI), with devices placed on the wrist by a member of the research team. Actigraphs were removed by a sleep technician in the morning.

Data for both devices were collected in 1-min epochs. AMI data (collected in the Zero-Crossing Mode) were scored using Action W-2 version 2.6.9905 software (Ambulatory Monitoring Inc., Ardsley NY). The Sadeh algorithm is the most commonly reported analysis for children and adolescents, thus it was examined as the primary algorithm. The Cole-Kripke algorithm[19] was applied separately for secondary analyses. PRMM data were scored using Actiware software version 5.59.0015 (Phillips Respironics, Bend, OR). The medium sensitivity threshold (40 counts per epoch) is the default setting for this program, thus it was used as the primary threshold. Secondary analyses included both the low (80 counts per epoch) and high (20 counts per epoch) wake sensitivity thresholds.

Data Analysis

Data were synchronized by initializing the actigraphs on the same computer used for PSG, with the PSG "lights off" and "lights on" time applied to both actigraph devices as the start and end points. Sleep onset was determined by the time of the first epoch of sleep as scored by PSG. In order to match the 30-sec PSG epochs with the 1-min actigraphy epochs, each minute of PSG data was scored as wake if either one or both 30-sec epochs were scored as wake.[20,24] Thus a minute of sleep on PSG required both 30-sec epochs to be scored as sleep. Only 4.7% of the combined 1-min PSG epochs were scored as wake when one of the 30-sec PSG epochs was scored as sleep.

The outcome sleep variables for this study were defined as total sleep time (TST: number of minutes scored as sleep between lights off and lights on), wake after sleep onset (WASO: number of minutes scored as wake between PSG scored sleep onset and lights on), and sleep efficiency (SE: TST divided by sleep period or number of total minutes from lights off to lights on). These 3 variables are commonly used to provide overall summary data in sleep research.

All analyses were conducted using SPSS 15.0 (SPSS, Inc., Chicago, IL). Preliminary analyses examined intra-device reliability with Pearson correlations and paired *t*-tests. Two approaches were used to evaluate the validity of each device compared to PSG. First, repeated-measures ANCOVAs (controlling for age and sleep disordered breathing status) were used to evaluate differences in sleep outcome variables between each device and PSG. Second, epoch-by-epoch (EBE) comparisons were used to determine agreement for each device with PSG. Table 1 shows how the variables for the EBE comparison were determined.

To determine inter-device reliability, *t*-tests were used to evaluate differences in sleep outcome variables for the 2 devices, and the Bland-Altman concordance technique was used to examine degree of agreement between the 2 devices for TST, WASO, and SE.[21] Similar to Werner et al., we defined a priori a difference between the 2 devices of ≤ 30 min satisfactory for TST, with a difference < 5% for SE satisfactory.[1] Default settings (Sadeh algorithm for the AMI device and medium sensitivity for the PRMM device) were used for all primary aim analyses.

Secondary aims utilizing alternative scoring algorithms/sensitivity settings were approached in a similar fashion: (1) repeated-measures ANCOVAs (controlling for age and sleep disordered breathing status) were used to compare each scoring algorithm/sensitivity setting with PSG, and (2) repeated-measures ANOVA was used to compare each scoring algorithm/sensitivity setting with PSG by developmental age group and sleep disordered breathing status. EBE comparisons were also used to determine agreement across scoring algorithm/sensitivity for each age group and sleep disordered breathing status.

Table 2—Differences between actigraphy and PSG using standard settings (controlling for age and SDB status)

	Actigraphy Mean (SD)	PSG Mean (SD)	*F*	P	Mean Difference (95% CI)
AMI - Sadeh					
TST (min)	412.1 (73.3)	435.7 (64.6)	47.2	< 0.001	-23.6 (-30.8 to -16.4)
WASO (min)	87.0 (60.4)	61.9 (39.3)	41.5	< 0.001	25.1 (18.2 to 32.0)
SE (%)	78.2 (13.2)	82.5 (10.5)	44.0	< 0.001	-4.4 (-5.8 to -3.0)
PRMM – Medium Sensitivity					
TST (min)	434.1 (60.9)	436.7 (65.1)	51.1	0.33	-2.7 (-8.2 to 2.9)
WASO (min)	69.2 (40.7)	59.7 (38.2)	59.4	< 0.001	9.5 (4.4 to 14.5)
SE (%)	82.3 (10.3)	82.7 (10.6)	50.2	< 0.001	-0.4 (-1.5 to 0.7)

AMI df = (1,104), PRMM df = (1,103).

Table 3—Sensitivity, specificity, and accuracy for full-night and for period after PSG-scored sleep onset

	Full Night Mean (SD)	After PSG Scored Sleep Onset Mean (SD)	*t*	P
AMI - Sadeh				
Sensitivity	0.89 (0.09)	0.89 (0.09)	-0.64	0.52
Specificity	0.73 (0.20)	0.63 (0.24)	7.30	< 0.001
Accuracy	0.87 (0.06)	0.88 (0.15)	-0.77	0.44
AMI – Cole-Kripke				
Sensitivity	0.92 (0.08)	0.92 (0.08)	-0.50	0.62
Specificity	0.65 (0.23)	0.54 (0.26)	6.04	< 0.001
Accuracy	0.89 (0.07)	0.90 (0.15)	-1.17	0.24
PRMM – Medium Sensitivity				
Sensitivity	0.93 (0.05)	0.94 (0.05)	-1.48	0.14
Specificity	0.69 (0.21)	0.69 (0.22)	0.44	0.66
Accuracy	0.89 (0.04)	0.92 (0.14)	-1.97	0.05
PRMM – Low Sensitivity				
Sensitivity	0.89 (0.06)	0.89 (0.06)	-0.84	0.40
Specificity	0.77 (0.20)	0.77 (0.20)	-0.39	0.70
Accuracy	0.87 (0.05)	0.88 (0.13)	-1.26	0.21
PRMM – High Sensitivity				
Sensitivity	0.97 (0.03)	0.96 (0.10)	0.86	0.39
Specificity	0.54 (0.24)	0.54 (0.26)	0.26	0.80
Accuracy	0.90 (0.05)	0.93 (0.14)	-2.48	0.02

AMI df = (1,106), PRMM df = (1,105).

RESULTS

Participants included 59 girls and 56 boys, with a mean age of 8.8 years (SD 4.4 years). Self-identified race was 46% Caucasian, 37.4% African American, 6.1% Hispanic, and 10.4% other. In terms of SDB, 54.8% of participants had no OSA (AHI mean = 0.5), 25.2% had mild OSA (AHI mean = 3.1), and 20.0% had moderate/severe OSA (AHI mean = 11.0). Five participants (4.3%) had periodic limb movement disorder (PLM index mean = 14.0).

Intra-Device Comparison

Although 17 participants wore 2 devices of the same brand, one device had a technical failure for one participant, resulting in 8 subjects for each brand of actigraph. Statistically significant correlations were found for both devices for TST (AMI = 0.995, PRMM = 0.999), WASO (AMI = 0.997, PRMM = 0.999), and SE (AMI = 0.996, PRMM = 0.999). Paired *t*-tests found no significant differences between devices of the same brand for TST (AMI: 438.0 vs. 438.8; PRMM: 441.4 vs. 438.4), WASO (AMI: 72.9 vs. 72.6; PRMM: 68.4 vs. 70.8), or SE (AMI: 82.7% vs. 82.8%; PRMM: 83.1% vs. 82.5%).

Actigraph versus PSG

For the AMI Sleep Watch, significant differences were found for all sleep outcome variables controlling for age and SDB status (Table 2), with the AMI device underestimating TST by almost 24 min and overestimating wake by 25 min, resulting in a lower SE (underestimated by 4.4%). EBE comparisons for the full recording period found sensitivity, specificity, and accuracy to be comparable or better than previous reports (Table 3).[22,23] After PSG sleep onset, sensitivity and accuracy did not change, but specificity was significantly lower (Table 3).

For the PRMM Actiwatch-2, no significant differences were found for TST when controlling for age and SDB status; however, the PRMM device overestimated WASO by 10 minutes. A statistically significant but non-clinically meaningful difference was found for SE (underestimated SE by 0.4%; Table 2). EBE comparisons found sensitivity, specificity, and accuracy to be comparable or better than previous reports (Table 3).[22,23] EBE comparisons did not differ when examining only the period after PSG sleep onset.

AMI Sleep Watch versus PRMM Actiwatch-2

Significant differences were found between the AMI and PRMM devices for all sleep variables. Specifically, the AMI

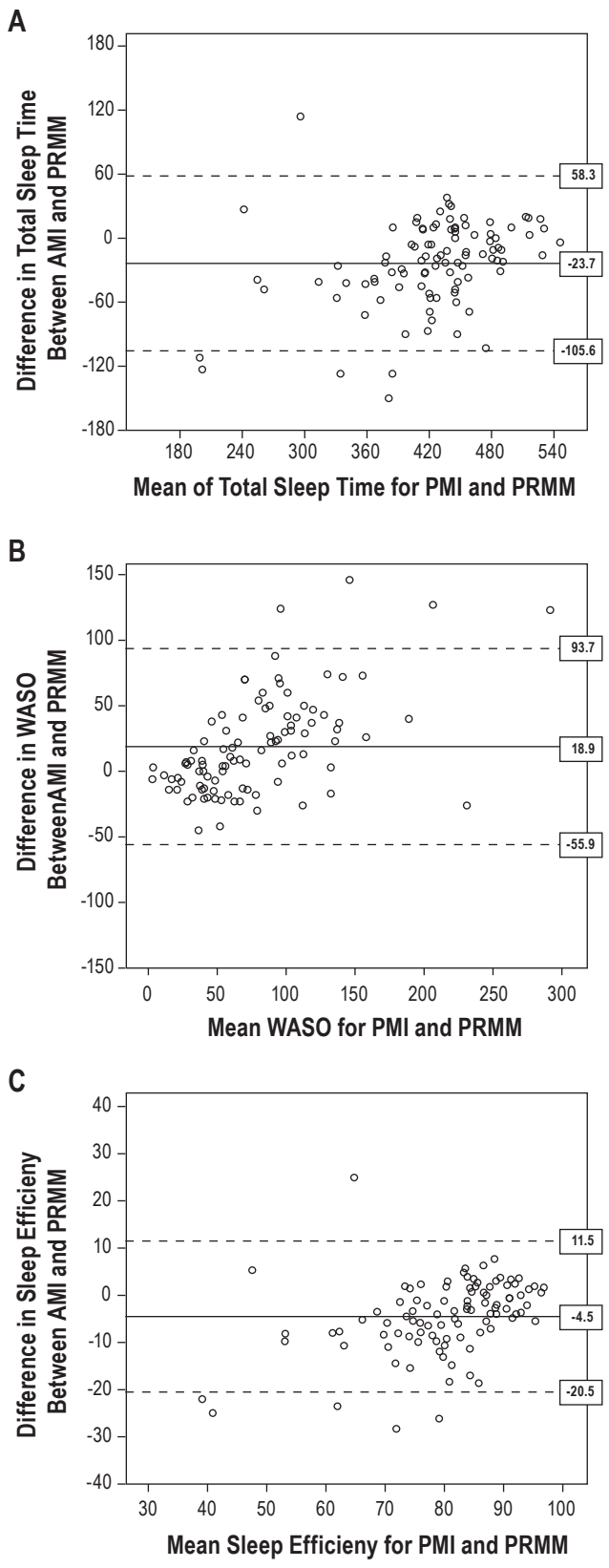

Figure 1—Bland Altman plots for AMI Sleep Watch vs. PRMM Actiwatch-2 comparing **(A)** total sleep time, **(B)** wake after sleep onset, and **(C)** sleep efficiency. The solid lines indicate the means of the differences, and the dashed lines 2 standard deviations above and below those means. See text for further details.

device found a shorter TST (AMI = 409.8 min, PRMM = 433.5 min, $t_{97} = -5.72$, $P < 0.001$), more WASO (AMI = 88.1 min, PRMM = 69.3 min, $t_{97} = 5.0$, $P < 0.001$), and lower SE (AMI = 77.8%, PRMM = 82.2%, $t_{97} = -5.55$, $P < 0.001$) compared to the PRMM device. The differences in agreement are also seen in the Bland-Altman plots (Figure 1), which show a range of ± 82 min for TST, a range of ± 75 min for WASO, and a range of ± 16% for SE.

Alternative Scoring Algorithms/Sensitivity Settings versus PSG

EBE comparisons between the alternative scoring algorithms and sensitivity settings (AMI-Cole-Kripke, PRMM-Low and PRMM-High) can be found in Table 3, while comparisons of the summary variables (TST, WASO, SE) can be found in Table 4.

For AMI-Cole-Kripke, statistically significant differences were found for all three sleep variables (controlling for age and SDB) compared to PSG, yet none of these differences were clinically meaningful (2 min for TST, 5 min for WASO, and 0.4% for SE). EBE comparisons found sensitivity, specificity, and accuracy to be comparable or better than previous reports.[22,23] However, for the period after PSG sleep onset, specificity was significantly lower for AMI-Cole-Kripke.

Compared to PSG, significant differences were found for both PRMM-Low and PRMM-High. The Low Sensitivity setting underestimated TST by 31 min, overestimated WASO by 36 min, and underestimated SE by almost 6%. The High Sensitivity setting overestimated TST by 26 min, underestimated WASO by 15 min, and overestimated SE by 5%. EBE comparisons for the Low Sensitivity settings found sensitivity, specificity and accuracy to be comparable or better than previous reports.[22,23] For the High Sensitivity setting, sensitivity and accuracy were high, while specificity was very low. No differences were found for the full-night recording period versus after PSG sleep onset.

Developmental Age Group

To examine differences by child age (Table 5), participants were divided into three groups: Preschool (ages 3-5 years), School-Age (ages 6-12 years), and Adolescent (ages 13-18 years). A significant main effect was found for both devices compared to PSG for TST, WASO, and SE (all $P < 0.001$).

Post hoc analyses show that for preschoolers, both AMI-Sadeh and AMI-Cole-Kripke underestimated TST and SE and overestimated WASO compared to PSG. Both the PRMM-Low and PRMM-Medium significantly underestimated TST and SE and overestimated WASO compared to PSG, while PRMM-High overestimated TST and SE. For school-age children, post hoc analyses show that AMI-Sadeh underestimated TST and SE compared to PSG. No other differences were found for AMI-Sadeh or AMI-Cole-Kripke. Compared to PSG, PRMM-Low underestimated sleep, overestimated WASO, and underestimated SE; while PRMM-High overestimated TST, underestimated WASO, and overestimated SE. Post hoc analyses for adolescents show AMI-Cole-Kripke, PRMM-Medium and PRMM-High all overestimated TST and SE and underestimated WASO.

As seen in Table 6, similar to the full sample results, specificity for both AMI-Sadeh and AMI-Cole-Kripke was significantly lower for the period after PSG sleep onset. Across age

Table 4—Differences between actigraphy algorithms/sensitivity settings and PSG (controlling for age and SDB status)

	Actigraphy Mean (SD)	PSG Mean (SD)	F	P	Mean Difference (95% CI)
AMI – Cole-Kripke					
TST (min)	433.3 (71.7)	435.7 (64.6)	19.7	< 0.001	-2.4 (-9.1 to 4.3)
WASO (min)	66.8 (56.2)	61.9 (39.3)	13.6	< 0.001	4.8 (-1.7 to 11.4)
SE (%)	82.2 (12.6)	82.5 (10.5)	18.9	< 0.001	-0.4 (-1.7 to 0.9)
PRMM – Low Sensitivity					
TST (min)	406.2 (63.0)	436.7 (65.1)	108.2	< 0.001	-30.5 (-36.5 to -24.5)
WASO (min)	95.9 (44.5)	59.7 (38.2)	142.9	< 0.001	36.2 (30.9 to 41.4)
SE (%)	77.0 (11.0)	82.7 (10.6)	104.7	< 0.001	-5.7 (-6.8 to -4.5)
PRMM – High Sensitivity					
TST (min)	462.7 (56.9)	436.7 (65.1)	8.1	0.005	25.9 (20.3 to 31.6)
WASO (min)	45.1 (35.3)	59.7 (38.2)	14.1	< 0.001	-14.7 (-19.6 to -9.8)
SE (%)	87.7 (9.1)	82.7 (10.6)	9.0	0.003	5.0 (3.9 to 6.1)

AMI df = (1,104), PRMM df = (1,103).

Table 5—Mean (SD) values for polysomnography and over-/underestimation (SD) of total sleep time, wake after sleep onset, and sleep efficiency for actigraphy settings by developmental age group

	Total Sleep Time (minutes)		
	Preschool (n = 31)	**School-Age (n = 48)**	**Adolescent (n = 28)**
PSG	452.6 (35.7)	450.1 (57.0)	392.4 (81.7)
AMI: Sadeh	-51.0 (35.9)*	-27.2 (37.9)*	12.9 (41.7)
AMI: Cole-Kripke	-24.2 (28.7)*	-6.9 (36.7)	29.5 (40.1)*
	Preschool (n = 32)	**School-Age (n = 50)**	**Adolescent (n = 24)**
PSG	452.2 (36.8)	449.9 (56.1)	388.6 (87.6)
PRMM: Medium	-23.0 (23.4)*	-7.8 (29.6)	35.2 (31.3)*
PRMM: Low	-54.4 (27.0)*	-34.9 (33.1)*	10.6 (30.3)
PRMM: High	10.2 (23.1)†	19.7 (27.2)*	59.9 (37.0)*
	Wake After Sleep Onset (minutes)		
	Preschool (n = 31)	**School-Age (n = 48)**	**Adolescent (n = 28)**
PSG	57.2 (34.0)	55.4 (35.6)	78.4 (46.8)
AMI: Sadeh	49.8 (38.2)*	26.4 (35.9)	-4.5 (37.4)
AMI: Cole-Kripke	22.9 (30.8)*	7.0 (36.3)	-18.9 (36.2)†
	Preschool (n = 32)	**School-Age (n = 50)**	**Adolescent (n = 24)**
PSG	52.2 (28.6)	56.0 (34.6)	77.7 (50.4)
PRMM: Medium	29.4 (23.5)*	11.0 (29.2)†	-20.4 (22.7)*
PRMM: Low	60.9 (26.7)*	37.2 (29.5)*	1.2 (22.6)
PRMM: High	0.0 (22.0)	-11.2 (26.9)†	-41.6 (25.0)*
	Sleep Efficiency (percent)		
	Preschool (n = 31)	**School-Age (n = 48)**	**Adolescent (n = 28)**
PSG	84.8 (5.5)	84.4 (8.8)	76.8 (14.8)
AMI: Sadeh	-9.5 (6.7)*	-5.1 (7.1)*	-2.6 (8.5)
AMI: Cole-Kripke	-4.5 (5.4)*	-1.3 (6.9)	-5.8 (8.1)*
	Preschool (n = 32)	**School-Age (n = 50)**	**Adolescent (n = 24)**
PSG	85.0 (5.5)	84.4 (8.7)	76.4 (15.9)
PRMM: Medium	-4.3 (4.4)*	-1.5 (5.6)	6.9 (6.2)*
PRMM: Low	-10.2 (5.0)*	-6.5 (6.2)*	2.1 (6.0)
PRMM: High	1.9 (4.3)†	3.7 (5.2)*	11.8 (7.2)*

Negative values indicate an underestimate of TST/WASO/SE compared to PSG; positive values indicate an overestimate of TST/WASO/SE compared to PSG. AMI, Ambulatory Monitoring Inc., Motionlogger Sleep Watch; PRMM, Phillips Respironics Mini-Mitter Actiwatch-2; PSG, Polysomnography. *Post hoc comparison vs. PSG, P ≤ 0.001. †Post hoc comparison vs. PSG, P ≤ 0.01

Table 6—Sensitivity, specificity, and accuracy for full-night (FN) and for period after PSG-scored sleep onset (APSO) by age and SDB status

	Preschool		School-Aged		Adolescent		No OSA		Mild OSA		Mod/Severe OSA	
	FN	APSO	FN	APSO	FN	APSO	FN	APSO	FN	APSO	FN	APSO
AMI - Sadeh												
Sensitivity	0.86	0.86	0.90	0.90	0.92	0.92	0.91	0.92	0.88	0.88	0.85	0.85
Specificity	0.81	0.72*	0.75	0.64*	0.60	0.50	0.70	0.59*	0.77	0.69	0.75	0.63
Accuracy	0.86	0.89	0.88	0.88	0.87	0.88	0.88	0.90	0.87	0.86	0.84	0.83
AMI – Cole-Kripke												
Sensitivity	0.90	0.91	0.93	0.93	0.93	0.93	0.95	0.95	0.90	0.91	0.89	0.89
Specificity	0.76	0.64*	0.68	0.55*	0.49	0.42	0.63	0.50*	0.69	0.61	0.67	0.55
Accuracy	0.88	0.93	0.90	0.90	0.86	0.87	0.90	0.93	0.87	0.88	0.86	0.86
PRMM – Medium												
Sensitivity	0.91	0.91	0.93	0.94	0.97	0.97	0.94	0.94	0.94	0.94	0.92	0.92
Specificity	0.78	0.79	0.71	0.69	0.54	0.53	0.71	0.71	0.68	0.66	0.67	0.67
Accuracy	0.89	0.95	0.90	0.91	0.88	0.90	0.90	0.94	0.89	0.90	0.88	0.89
PRMM – Low												
Sensitivity	0.85	0.85	0.89	0.89	0.93	0.93	0.89	0.89	0.89	0.89	0.87	0.87
Specificity	0.83	0.86	0.79	0.77	0.63	0.67	0.78	0.79	0.74	0.76	0.75	0.75
Accuracy	0.85	0.89	0.88	0.87	0.88	0.90	0.87	0.90	0.87	0.87	0.86	0.87
PRMM – High												
Sensitivity	0.96	0.96	0.97	0.95	0.99	0.99	0.97	0.97	0.97	0.94	0.96	0.96
Specificity	0.63	0.65	0.57	0.56	0.38	0.36	0.56	0.55	0.50	0.51	0.54	0.53
Accuracy	0.91	0.97	0.91	0.93*	0.86	0.89*	0.91	0.95	0.89	0.92*	0.89	0.90

*Full-night vs. after PSG sleep onset, P ≤ 0.001

groups, AMI-Cole-Kripke had a slightly better sensitivity, yet AMI-Sadeh had a much higher specificity. For the PRMM device, PRMM-Low had the best specificity, although a slightly lower sensitivity (compared to PRMM-Medium and PRMM-High). Notably, specificity declined across age groups for both device brands, with adolescents having the poorest specificity (across devices/settings).

Sleep Disordered Breathing Status

Age was equally distributed across SDB status, $\chi^2(4) = 7.0$, $P = 0.14$. A significant main effect was found for both devices for TST, WASO, and SE (all $P < 0.001$). As seen in Table 7, post hoc comparisons between youth with and without sleep disordered breathing were also made between the different actigraphic algorithms/sensitivities and PSG. For youth without OSA, AMI-Sadeh and PRMM-Medium provided good estimates of both TST and SE. For youth with SDB (mild and moderate/severe), AMI-Cole-Kripke and PRMM-Medium provided good estimates of TST and SE. Across SDB groups, EBE comparisons (Table 6) found that AMI-Sadeh had better specificity than AMI-Cole-Kripke, but slightly lower sensitivity to detect sleep. Similarly, PRMM-Low had the best specificity, but a slightly lower sensitivity than PRMM-Medium.

DISCUSSION

This study is one of the first to directly compare two different brands of commercially available actigraphs with polysomnography in a large sample of children and adolescents, and it addresses several of the recommendations from the AASM's 2007 practice parameters for actigraphy.[24] Specifically this study compared results from different types of actigraphs and across different algorithms. In addition, polysomnography was used as a reference standard for the comparison of actigraphy to PSG, while the Bland-Altman concordance technique was used to compare the two brands of devices to one another (since neither brand is considered a gold standard). The results of this study are limited by the single night of assessment in the sleep lab. In addition, although the overall sample was larger than previous studies, some of the group comparisons (e.g., developmental age group, SDB status) may have been limited by smaller sample sizes. Finally, the specificity results may have been influenced by the need to collapse the 30-second PSG epochs into 1-minute epochs in order to do an epoch-by-epoch comparison with actigraphy.

Despite these limitations, there are a number of strengths to this study. The inclusion of a wide age-range allowed us to examine the validity of different devices for both children and adolescents. The side-by-side comparison of both devices with PSG allowed us to examine the validity of these two new devices compared to the "gold standard." By examining the reliability of these devices compared to one another, we were able to determine whether different brands of actigraphs can be used in the same study, as well as whether results using different brands of actigraphs can be compared across studies. Finally, since the use of only one statistical approach (e.g., correlation, epoch-by-epoch comparison) can provide misleading agreement between devices,[1,15,21] multiple statistical approaches were used in this study, including repeated-measures ANCOVA for summary sleep variables, epoch-by-epoch agreement rates, and the Bland-Altman concordance technique.

As expected, the correlation between devices of the same brand (intra-device reliability) was excellent with no significant differences found when the devices were compared side by side. This suggests that researchers can use multiple devices of the same brand and model within the same study and obtain comparable results. However, this study did not compare these two new models with older actigraph models (e.g., AMI Basic Mini-Motionlogger or PRMM AW-64). Further validation is needed to determine whether different models within the same brand provide similar data in children and adolescents.

For both devices, epoch-by-epoch comparisons with the "gold standard" PSG were similar to or better than previous reports.[22,23,25-30] For the full study sample, both devices provided good sensitivity to measure sleep (89% to 97%) and overall good accuracy (87% to 90%). Specificity for measuring wake was similar to or slightly better than previous reports (54% to 77%). That said, the ability of actigraphy to identify wake after sleep onset remains a significant limitation of these devices.[22,28] In particular, the AMI device had significantly poorer specificity after PSG sleep onset, suggesting the ability to detect wake prior to sleep onset is better than the ability to detect wake after sleep onset.

Table 7—Mean (SD) values for polysomnography and over-/underestimation (SD) of total sleep time, wake after sleep onset, and sleep efficiency for actigraphy settings by sleep disordered breathing status

	Total Sleep Time (minutes)		
	No OSA (n = 59)	**Mild OSA (n = 27)**	**Mod/Severe OSA (n = 21)**
PSG	439.3 (58.9)	438.9 (75.3)	421.7 (66.5)
AMI: Sadeh	-13.7 (44.2)	-32.3 (35.0)*	-40.0 (52.6)†
AMI: Cole-Kripke	7.0 (39.2)	-12.1 (30.2)	-16.1 (50.6)
	No OSA (n = 58)	**Mild OSA (n = 26)**	**Mod/Severe OSA (n = 22)**
PSG	442.9 (59.5)	436.5 (76.4)	420.7 (65.2)
PRMM: Medium	-5.5 (32.6)	3.1 (39.8)	-1.9 (37.7)
PRMM: Low	-33.6 (37.7)*	-23.6 (39.5)†	-30.6 (40.9)†
PRMM: High	21.6 (29.0)*	34.9 (41.4)*	26.7 (36.3)†
	Wake After Sleep Onset (minutes)		
	No OSA (n = 59)	**Mild OSA (n = 27)**	**Mod/Severe OSA (n = 21)**
PSG	56.4 (32.8)	63.4 (34.0)	75.7 (57.2)
AMI: Sadeh	15.6 (39.0)†	33.5 (34.6)	40.8 (51.9)†
AMI: Cole-Kripke	3.8 (34.7)	14.4 (29.2)	17.0 (50.0)
	No OSA (n = 58)	**Mild OSA (n = 26)**	**Mod/Severe OSA (n = 22)**
PSG	53.1 (28.4)	60.7 (34.6)	76.2 (57.1)
PRMM: Medium	11.4 (31.5)†	5.7 (30.4)	8.8 (34.4)
PRMM: Low	37.9 (34.6)*	32.4 (34.0)*	36.0 (36.3)*
PRMM: High	-11.8 (26.7)*	-19.5 (33.0)†	16.7 (32.8)
	Sleep Efficiency (percent)		
	No OSA (n = 59)	**Mild OSA (n = 27)**	**Mod/Severe OSA (n = 21)**
PSG	83.1 (9.2)	82.7 (12.3)	80.9 (11.7)
AMI: Sadeh	-2.4 (8.4)	-6.1 (6.7)*	-7.6 (10.0)†
AMI: Cole-Kripke	1.5 (7.6)	-2.3 (5.7)	-3.0 (9.7)
	No OSA (n = 58)	**Mild OSA (n = 26)**	**Mod/Severe OSA (n = 22)**
PSG	83.6 (9.1)	82.7 (12.6)	80.5 (11.8)
PRMM: Medium	-1.0 (6.2)	0.7 (7.8)	-0.3 (7.3)
PRMM: Low	-6.2 (7.0)*	-4.4 (7.6)†	-5.8 (7.9)†
PRMM: High	4.2 (5.6)*	6.7 (8.2)*	5.2 (7.0)†

Negative values indicate an underestimate of TST/WASO/SE compared to PSG, positive values indicate an overestimate of TST/WASO/SE compared to PSG. AMI, Ambulatory Monitoring Inc., Motionlogger Sleep Watch; PRMM, Phillips Respironics Mini-Mitter Actiwatch-2; PSG, Polysomnography. *Post hoc comparison vs. PSG, $P < 0.001$. †Post hoc comparison vs. PSG, $P \leq 0.01$. ‡Post hoc comparison vs. PSG, $P \leq 0.05$.

A comparison of outcome variables (TST and SE) told a slightly different story than the EBE comparisons. For the full sample, AMI-Sadeh underestimated TST and SE by clinically meaningful amounts (24 min and 4%, respectively), and this bias was stronger among younger and (to a lesser degree) more sleep disordered participants. Based on previous reports, however, this result is not surprising as the Sadeh algorithm was developed and validated on a healthy sample of adolescents/young adults.[20] In fact, when examining the results by developmental group and SDB status, AMI-Sadeh performed best for adolescents, as well as for youth of all ages without SDB.

The need to consider developmental stage and SDB status when selecting an actigraph device has also been found in other studies. For preschoolers, sensitivity and accuracy were found to be similar to Sitnick et al.[27]; yet in the current study, actigraphy was found to underestimate sleep compared to PSG, whereas the Sitnick study found that actigraphy overestimated sleep compared to videosomnography. For children and adolescents with sleep disordered breathing, actigraphy can provide a valid estimate of total sleep time and sleep efficiency. However, as in previous studies, this is not true for all settings on either brand of actigraph.[29,30] In addition, actigraphy alone should not be considered a valid way to assess sleep disordered breathing.[24]

Finally, the side-by-side comparison of the two brands of actigraphs was less encouraging in terms of validity. Similar to Werner et al.,[1] the a priori defined limits of agreement between the two devices was not found, with the two brands providing very different data for total sleep time, wake after sleep onset, and sleep efficiency. Thus, as suggested by LeBourgeois et al., the two brands cannot be used within the same study.[13] Further, caution should be used when comparing results from studies that have used different brands of actigraphs.

In sum, the purpose of this study was not to demonstrate the superiority of one brand of actigraphic device over another, but rather to look at the validity of two new devices on the market, as well as the inter-device reliability. Similar to previous

models of actigraphs, epoch-by-epoch comparisons showed that the AMI Motionlogger Sleep Watch and the PRMM Actiwatch-2 were both valid devices compared to PSG, although inter-device reliability was poor. For those who choose to use actigraphy in research, the selection of a device and scoring algorithm should be made based on study population (e.g., young children vs. adolescents, healthy youth vs. youth with SDB). Further, these selections need to be outlined in detail when reporting results.[11,24] Finally, the field of sleep would benefit from the development of standard procedures for the use, scoring, and interpretation of actigraphy.

ACKNOWLEDGMENTS

The authors thank the children and families who participated in this study; Kelly Ann Davis, Ashley Peightal, and the sleep laboratory technicians for their assistance; Avi Sadeh for his mentorship with this project; and Hawley Montgomery-Downs and Carole Marcus for their helpful review and comments. This study was funded in part by MH077662.

DISCLOSURE STATEMENT

This was not an industry supported study. The authors have indicated no financial conflicts of interest.

REFERENCES

1. 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;162:350-8.
2. Spruyt K, Gozal D, Dayyat E, Roman A, Molfese DL. Sleep assessments in healthy school-aged children using actigraphy: concordance with polysomnography. J Sleep Res 2011;20:223-32.
3. Holley S, Hill CM, Stevenson J. A comparison of actigraphy and parental report of sleep habits in typically developing children aged 6 to 11 years. Behav Sleep Med 2010;8:16-27.
4. Hering E, Epstein R, Elroy S, Iancu DR, Zelnik N. Sleep patterns in autistic children. J Autism Dev Disord 1999;29:143-7.
5. Wiggs L, Montgomery P, Stores G. Actigraphic and parent reports of sleep patterns and sleep disorders in children with subtypes of attention-deficit hyperactivity disorder. Sleep 2005;28:1437-45.
6. Corkum P, Tannock R, Moldofsky H, Hogg-Johnson S, Humphries T. Actigraphy and parental ratings of sleep in children with attention deficit hyperactivity disorder (ADHD). Sleep 2001;24:303-12.
7. Goodlin-Jones BL, Tang K, Liu J, Anders TF. Sleep patterns in preschool-age children with autism, developmental delay, and typical development. J Am Acad Child Psychiatry 2008;47:930-8.
8. Acebo C, Sadeh A, Seifer R, et al. Estimating sleep patterns with activity monitoring in children and adolescents: How many nights are necessary for reliable measures? Sleep 1999;22:95-103.
9. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003;26:342-92.
10. Acebo C, Sadeh A, Seifer R, Tzischinsky O, Hafer A, Carskadon MA. Sleep/wake patterns derived from activity monitoring and maternal report for healthy 1- to 5-year-old children. Sleep 2005;28:1568-77.
11. Berger AM, Wielgus KK, Young-McCaughan S, Fischer P, Farr L, Lee KA. Methodological challenges when using actigraphy in research. J Pain Symptom Manage 2008;1-9.
12. Benson K, Friedman L, Noda A, Wicks D, Wakabayashi E, Yesavage J. The measurement of sleep by actigraphy: Direct comparison of 2 commercially available actigraphs in a nonclinical population. Sleep 2004;27:986-9.
13. LeBourgeois MK, Acebo C, Seifer R, Carskadon MA. Comparing estimates of adolescent sleep and wake from two activity monitoring systems. Sleep 2002;25:A273-4.
14. Tonetti L, Pasquini F, Fabbri M, Belluzzi M, Natale V. Comparison of two different actigraphs with polysomnography in healthy young subjects. Chronobiol Int 2008;25:145-53.
15. Weiss AR, Johnson NL, Berger NA, Redline S. Validity of activity-based devices to estimate sleep. J Clin Sleep Med 2010;6:336-42.
16. Pollak CP, Stokes PE, Wagner DR. Direct comparison of two widely used activity recorders. Sleep 1998;21:207-12.
17. Iber C, Ancoli-Israel S, Chesson AL Jr., Quan SF, for the American Academy of Sleep Medicine. The AASM manual for the scoring of sleep and associated events. Westchester, IL: American Academy of Sleep Medicine, 2007.
18. Katz ES, Marcus CL. Diagnosis of obstructive sleep apnea syndrome in infants and children. In: Sheldon SH, Ferber R, Kryger MH, eds. Principles and practice of pediatric sleep medicine. Philadelphia: Elsevier Saunders, 2005:197-210.
19. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992;15:461-9.
20. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: An empirical test of methodological issues. Sleep 1994;17:201-7.
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
22. Meltzer LJ, Montgomery-Downs HE, Insana SP, Walsh CM. Use of actigraphy for the assessment of pediatric sleep. Sleep Med Rev (under review).
23. Martin JL, Hakim AD. Wrist actigraphy. Chest 2011;139:1514-27.
24. Morgenthaler TI, Alessi C, Friedman L, et al. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. Sleep 2007;30:519-29.
25. de Souza L, Benedito-Silva AA, Pires ML, Poyares D, Tufik S, Calil HM. Further validation of actigraphy for sleep studies. Sleep 2003;26:81-5.
26. 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;2:389-96.
27. Sitnick SL, Goodlin-Jones BL, Anders TF. The use of actigraphy to study sleep disorders in preschoolers: some concerns about detection of nighttime awakenings. Sleep 2008;31:395-401.
28. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 2011;15:259-67.
29. Johnson NL, Kirchner HL, Rosen CL, et al. Sleep estimation using wrist actigraphy in adolescents with and without sleep disordered breathing: A comparison of three data modes. Sleep 2007;30:899-905.
30. Hyde M, O'Driscoll DM, Binette S, et al. Validation of actigraphy for determining sleep and wake in children with sleep disordered breathing. J Sleep Res 2007;16:213-6.