pii: sp-00501-14 http://dx.doi.org/10.5665/sleep.4918

Comparison of a Commercial Accelerometer with Polysomnography and Actigraphy in Children and Adolescents

Lisa J. Meltzer, PhD[1]; Laura S. Hiruma, PhD[2]; Kristin Avis, PhD[3]; Hawley Montgomery-Downs, PhD[4]; Judith Valentin, RPSGT[5]

[1]National Jewish Health, Denver, CO; [2]University at Albany, State University of New York, Albany, NY; [3]University of Alabama at Birmingham, Birmingham, AL; [4]West Virginia University, Morgantown, WV; [5]Children's of Alabama, Birmingham, AL

Study Objectives: To evaluate the reliability and validity of the commercially available Fitbit Ultra (2012) accelerometer compared to polysomnography (PSG) and two different actigraphs in a pediatric sample.

Design and Setting: All subjects wore the Fitbit Ultra while undergoing overnight clinical polysomnography in a sleep laboratory; a randomly selected subset of participants also wore either the Ambulatory Monitoring Inc. Motionlogger Sleep Watch (AMI) or Phillips-Respironics Mini-Mitter Spectrum (PRMM).

Participants: 63 youth (32 females, 31 males), ages 3–17 years (mean 9.7 years, SD 4.6 years).

Measurements: Both "Normal" and "Sensitive" sleep-recording Fitbit Ultra modes were examined. Outcome variables included total sleep time (TST), wake after sleep onset (WASO), and sleep efficiency (SE). Primary analyses examined the differences between Fitbit Ultra and PSG using repeated-measures ANCOVA, with epoch-by-epoch comparisons between Fitbit Ultra and PSG used to determine sensitivity, specificity, and accuracy. Intra-device reliability, differences between Fitbit Ultra and actigraphy, and differences by both developmental age group and sleep disordered breathing (SDB) status were also examined.

Results: Compared to PSG, the Normal Fitbit Ultra mode demonstrated good sensitivity (0.86) and accuracy (0.84), but poor specificity (0.52); conversely, the Sensitive Fitbit Ultra mode demonstrated adequate specificity (0.79), but inadequate sensitivity (0.70) and accuracy (0.71). Compared to PSG, the Fitbit Ultra significantly overestimated TST (41 min) and SE (8%) in Normal mode, and underestimated TST (105 min) and SE (21%) in Sensitive mode. Similar differences were found between Fitbit Ultra (both modes) and both brands of actigraphs.

Conclusions: Despite its low cost and ease of use for consumers, neither sleep-recording mode of the Fitbit Ultra accelerometer provided clinically comparable results to PSG. Further, pediatric sleep researchers and clinicians should be cautious about substituting these devices for validated actigraphs, with a significant risk of either overestimating or underestimating outcome data including total sleep time and sleep efficiency.

Keywords: actigraphy, Fitbit, accelerometer, polysomnography, validation, pediatric, sensitivity, specificity

Citation: 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.

INTRODUCTION

Measuring sleep in pediatric populations is challenging. Parent- or self-report measures, including questionnaires, sleep diaries, and a clinical history, are most commonly used. Yet there are a number of well-known limitations, including the inability of younger children or those with developmental delays to report their sleep patterns.[1] In addition, as children get older, parents become less involved with bedtime routines and may not be aware of their child's sleep onset latency or nighttime awakenings.[2–4] For the diagnosis of sleep disorders, polysomnography (PSG) is considered the "gold standard" sleep measurement, using multiple channels to assess sleep physiology and underlying sleep disorders. However, in addition to the cost and required in-lab setting, PSG is limited in its ability to provide information on sleep patterns over multiple nights. Because sleep disturbances are relatively common in childhood, it is important to have affordable and accurate ways to measure sleep in children, capturing multiple nights of sleep patterns in the child's natural sleep environment.[5,6]

Submitted for publication August, 2014
Submitted in final revised form November, 2014
Accepted for publication December, 2014
Address correspondence to: Lisa J. Meltzer, PhD, National Jewish Health, 1400 Jackson Street, G311, Denver, CO 80206; Tel: (303) 398-1837; Email: meltzerl@njhealth.org

Wrist-worn accelerometers, or actigraphy, have become increasingly popular for measuring sleep patterns in pediatric populations.[7,8] Actigraphs are small, portable devices with accelerometers that detect physical movement, with activity translated into estimates of sleep and wake. Although actigraphy is limited by its ability to accurately capture wake after sleep onset, past research has demonstrated that actigraphy provides a valid estimate of sleep among children and adolescents.[6,9–11]

Over the past few years, there has been a rapid growth in commercially available, consumer-friendly devices marketed as being able to measure sleep. The Fitbit Ultra (Fitbit Inc., San Francisco, CA, 2012), which costs approximately $100 and has no fee to use the online software, is one such commercially available accelerometer that is marketed for the measurement of both daytime activity and sleep. The Fitbit Ultra, as well as other similar commercially available devices, are attractive to sleep clinicians and researchers as potentially accessible and affordable alternative to traditional actigraphs, such as the Ambulatory Monitoring Inc. (AMI) Motionlogger Sleep Watch ($1,850 for one watch, interface, and software) and the Philips Respironics Mini-Mitter (PRMM) Actiwatch Spectrum ($1,225 for one watch, interface, and software). However, these more traditional devices have been shown to provide a reliable estimate of sleep-wake patterns, while the validity of newer commercial devices like the Fitbit Ultra remains undetermined in pediatrics. Montgomery-Downs and colleagues conducted one of the only published studies examining the reliability and

validity of the Fitbit compared to PSG and actigraphy for measuring sleep among adults.[12] Results indicated that the Fitbit (using the default Normal mode) demonstrated good intra-device reliability; however, while both the Fitbit and actigraphy overestimated sleep compared to PSG, the Fitbit overestimated total sleep time by 24 minutes more than actigraphy.[12] While the authors concluded that the Fitbit could possibly serve in a limited capacity as an additional tool to provide information about activity levels during sleep for normal adult populations, further validation was recommended in specific populations such as pediatrics.

Thus, the purpose of this study was to examine the utility of the Fitbit Ultra as a measure of sleep-wake patterns in pediatric populations. Study aims were to: (1) examine the validity of the Fitbit Ultra compared to overnight PSG in children and adolescents; (2) compare the Fitbit Ultra to two commonly used actigraphs to determine whether the Fitbit Ultra provides comparable outcome data for clinical or research use; and (3) to compare sleep outcomes as measured by the Fitbit Ultra between different age groups and sleep disordered breathing (SDB) status.

METHODS

Subjects

Sixty-three youth, ages 3–17 years, who were scheduled for an overnight clinical PSG in the pediatric sleep laboratory at Children's of Alabama between June 2012 and June 2013 participated in the study. The study was approved by the hospital's institutional review board, and informed consent and assent (when appropriate) were obtained for all participants.

Polysomnography

Overnight PSG was conducted using the Sandman 10.0 platform (Embla, Broomfield, CO). PSG-recorded measurements included electroencephalography (F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1), bilateral electrooculogram, submental and bilateral electromyogram, electrocardiogram, oronasal airflow with 3-pronged thermistor, nasal pressure with pressure transducer, rib cage and abdominal wall motion via respiratory impedance plethysmography, end-tidal capnometry, and arterial oxygen saturation via pulse waveform, as well as video and audio recordings. Sleep studies were scored based on the pediatric criteria outlined by the American Academy of Sleep Medicine (AASM).[13] The total sleep period was scored from lights out to lights on, with lights out time scheduled as close as possible to the child's typical bedtime.

All participants had ≥ 7.5 h of PSG recording completed (mean 8.4 h, SD 0.44). The average lights out time was 21:48 (SD 0:27) and the average lights on time was 06:12 (0:12). The apnea-hypopnea index (AHI) was used to determine sleep disordered breathing (SDB) status using the following criteria: No OSA: AHI < 1.5; Mild OSA: AHI ≥ 1.5 and ≤ 5; Moderate/Severe OSA: AHI > 5.[14]

Fitbit Ultra

All participants wore a Fitbit Ultra activity-monitoring device on their non-dominant wrist (using the Fitbit Velcro cuff). Stratified randomization (for age and gender) was used to randomly select 9 subjects to wear 2 Fitbit Ultra devices on the same wrist to examine intra-device reliability. The Fitbit Ultra tracks frequency and intensity of movement with a 3-dimensional accelerometer system. All devices were placed on subjects' wrists by a member of the research team and were removed by a sleep technician in the morning. The proprietary software algorithms for both the Normal and Sensitive sleep-recording modes were assessed. According to the manufacturer's website, the Normal mode counts "significant movements such as rolling over as being awake, and is appropriate for most users."[15] The Sensitive mode is the recommended setting for individuals with suspected sleep disorders.[15] Data were collected in 1-min epochs (default) and downloaded onto the Fitbit Inc. data-tracking website, using the device's USB docking port. Minute-by-minute data were extracted manually from the Fitbit website separately for the Normal and Sensitive modes, with 2 research assistants (RA) viewing and recording each subject's minute-by-minute data. If the 2 RAs extractions did not have 100% agreement, the first author (LJM) also viewed the file to resolve the discrepancy.

Actigraphy

Stratified randomization (by age and gender) was used to randomly assign participants to wear a second device on the non-dominant wrist, which included either the AMI Motionlogger Sleep Watch (18 subjects), the PRMM Actiwatch Spectrum (18 subjects), or the IM Systems ActiTrac (Individual Monitoring Systems, Inc. Arnold, MD, n = 18). While the AMI and PRMM devices have yet to be validated in pediatrics, previous versions by these manufacturers have been shown to be reliable and valid.[6,9–11,16,17] However, the ActiTrac has not been validated for use in pediatrics; thus, it was not included in this study. Placement of the actigraphs and the Fitbit Ultra in relation to the wrist was randomly assigned (Fitbit Ultra-AMI or AMI-Fitbit Ultra; Fitbit Ultra-PRMM or PRMM-Fitbit Ultra). Data for both devices were collected in 1-min epochs. AMI data were collected in the Zero-Crossing Mode and scored using the Sadeh algorithm with Motionlogger WatchWare 1.94.1.3 software (Ambulatory Monitoring Inc. Ardsley, NY). PRMM data were scored using the default medium sensitivity threshold (40 counts per epoch) with Actiware version 6.0.0 software (Phillips Respironics, Bend, OR).

Data Analysis

Fitbit Ultra data were synchronized by initializing the devices on the same computer for PSG and actigraphs. The sleep period for the Fitbit Ultra and the actigraphs were identified using the PSG lights off and lights on times, and sleep onset was based on the time of the first PSG-scored sleep epoch. In order to match the 30-sec PSG epochs with the 1-min Fitbit Ultra/actigraphy epochs, each minute of PSG data was scored as wake if either one or both 30-sec epochs were scored as awake.[6,10,18]

Outcome variables of interest for sleep in this study included 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 after PSG sleep onset) and sleep efficiency (SE: TST divided by minutes between lights off and lights on, expressed as a percent). Additionally,

Table 1—Comparison of sleep outcomes for PSG, Fitbit Ultra Normal and Fitbit Ultra Sensitive modes (n = 49, controlling for age and sleep disordered breathing status).

Sleep Outcome	PSG Mean (SE)	Fitbit Ultra Normal Mean (SE)	Fitbit Ultra Sensitive Mean (SE)	F (df = 2, 45)	P
TST (min)	420.4 (9.5)	461.6 (5.3)	315.7 (8.9)	39.9	< 0.001
WASO (min)	65.3 (6.7)	33.1 (3.8)	171.7 (8.5)	28.8	< 0.001
Sleep Efficiency (%)	83.4 (1.7)	91.8 (0.8)	62.7 (0.2)	39.3	< 0.001

Post hoc analyses indicate Fitbit Ultra Normal and Fitbit Ultra Sensitive modes were significantly different from PSG for all three outcomes, P < 0.001. PSG, polysomnography; TST, total sleep time; WASO, wake after sleep onset; SE, standard error; min, minutes.

epoch-by-epoch (EBE) comparisons between Fitbit Ultra and PSG were used to determine sensitivity (i.e., detection of true sleep), specificity (i.e., detection of true wake), and accuracy (i.e., ability to detect both true sleep and true wake).[6,19]

Analyses were conducted using SPSS 20.0 (SPSS, Inc., Chicago, IL). Preliminary analyses examined the intra-device reliability for both Fitbit Ultra modes using paired *t*-tests. Primary outcome analyses were used to compare the Fitbit Ultra to PSG by (1) repeated-measures ANCOVAs (controlling for developmental age category and sleep disordered breathing status) to examine differences in sleep outcomes, with paired t-tests for post hoc analyses, and (2) epoch-by-epoch analyses to examine the sensitivity, specificity, and accuracy of the Fitbit Ultra. To examine the reliability of the Fitbit Ultra as a measure of sleep outcomes compared to actigraphy, *t*-tests were used to compare means, and the Bland-Altman concordance technique was utilized.[20,21] We used *a priori* standards for determining satisfactory clinical agreement between sleep measures based on previous research. Satisfactory agreement was defined as ≤ 30-min difference between devices for TST and WASO, and ≤ 5% difference for sleep efficiency.[6,22] Secondary analyses examined differences between sleep outcomes as measured by PSG and Fitbit Ultra by developmental age group (preschool, school-aged, or adolescent) and sleep disordered breathing (SDB) status.

RESULTS

Sixty-three children and adolescents participated (31 boys and 32 girls), with a mean age of 9.7 years (SD 4.6 years). Parent-identified race of the participants was 51% Caucasian, 44% African American, 3% Asian, and 2% multiracial. For SDB, 61% of participants had no OSA (AHI mean = 0.4), 23% had mild OSA (AHI mean = 2.4), and 16% had moderate/severe OSA (AHI mean = 27.9). No participants had periodic limb movement disorder. Because of technical issues with the Fitbit Ultra (data were not recorded, n = 12) and PSG (corrupt file, n = 2) the final sample included 49 participants. Due to technical issues with the actigraph devices, an additional 2 subjects were excluded from the PRMM analyses (data were not recorded), and 2 subjects were excluded from the AMI analyses (unexplained artifact), resulting in 12 subjects per actigraph. There were no demographic differences between participants who were included and excluded due to technical issues.

Intra-Device Comparison

In order to determine whether the Fitbit Ultra devices were interchangeable (i.e., provide similar outcome data), 9 participants each wore 2 Fitbit Ultra devices side-by-side on their non-dominant wrist to assess intra-device reliability; however, data were not recorded for one pair, and there was a significant discrepancy between the PSG times and the Fitbit Ultra times for another device. We used paired *t*-tests to compare the remaining outcome data (n = 7) provided by 2 devices (mean 1 derived from device closest to wrist, mean 2 from device furthest from wrist). When examining the Normal mode, no statistically significant differences were found for intra-device TST or SE (TST: 468.7 vs. 471.1; SE: 92.9% vs. 93.3%). Additionally, there were no statistically significant differences for intra-device TST or SE for Sensitive mode (TST: 300.4 vs. 289.9; SE: 59.4% vs. 57.4%). The difference between devices fell within the acceptable range for the *a priori* determined values for satisfactory clinical agreement.

Fitbit Ultra versus PSG

Significant differences on all three sleep outcome variables were found between PSG and both Fitbit Ultra modes (Table 1). Post hoc analyses indicated that Normal mode significantly overestimated TST by 41 min and underestimated WASO by 32 min, resulting in a significant overestimate of SE by 8% (Table 1). Significant differences were also found between PSG and Fitbit Ultra in Sensitive mode, with the Fitbit Ultra underestimating sleep TST by 105 min and overestimating WASO by 106 min, resulting in an underestimation of SE by 21% (Table 1).

Based on EBE comparisons between Fitbit Ultra Normal mode and PSG, sensitivity (0.87) and accuracy (0.84) were adequate, but specificity was poor (0.52). For Sensitive mode, specificity was adequate (0.79), but sensitivity (0.70) and accuracy (0.71) were low.

Fitbit Ultra versus Actigraphy

AMI Motionlogger Sleep Watch

Significant differences were found between both Fitbit Ultra modes and AMI across outcome measures (Table 2). Compared to the AMI device, Fitbit Ultra Normal mode significantly overestimated TST by 37 min and underestimated WASO by 32 min, resulting in an overestimation of SE by 7% (Table 2). The differences in agreement can also be seen in the Bland-Altman plots (Figure 1A), which highlight > 30-min discrepancy in TST for 8% of participants and > 5% discrepancy in sleep efficiency for 17% of participants.

Table 2—Comparison of sleep outcomes for Actigraphy, Fitbit Ultra Normal, and Fitbit Ultra Sensitive modes (n = 12 per device, controlling for age and sleep disordered breathing status).

Sleep Outcome	PSG Mean (SE)	Actigraphy Mean (SE)	Fitbit Ultra Normal Mean (SE)	Fitbit Ultra Sensitive Mean (SE)	F (df = 2,10)	P
AMI Motionlogger Sleep Watch[a]						
TST (min)	451.8 (16.1)	447.1 (12.0)	484.2 (9.2)	349.6 (10.4)	17.2	0.001
WASO (min)	48.0 (12.2)	52.8 (5.9)	20.4 (5.4)	153.8 (11.6)	18.4	0.001
Sleep Efficiency (%)	88.6 (2.8)	87.8 (1.5)	95.1 (1.2)	68.7 (2.2)	19.2	0.001
PRMM Actiwatch Spectrum[b]						
TST (min)	428.0 (14.5)	427.8 (15.2)	464.3 (10.3)	310.8 (11.9)	13.9	0.002
WASO (min)	62.7 (10.9)	63.3 (15.3)	31.3 (8.7)	172.3 (17.5)	8.50	0.01
Sleep Efficiency (%)	85.1 (2.9)	84.9 (2.8)	92.1 (1.7)	61.7 (2.6)	16.6	0.001

[a]Post hoc analyses indicate Fitbit Ultra Normal and Fitbit Ultra Sensitive modes were significantly different from PSG and the AMI Motionlogger Sleep Watch for all three outcomes. [b]Post hoc analyses indicate Fitbit Ultra Sensitive mode was significantly different from PSG and the PRMM Actiwatch Spectrum for all three outcomes. PSG, polysomnography; TST, total sleep time; WASO, wake after sleep onset; AMI, Ambulatory-Monitoring Inc.; PRMM, Philips Respironics Mini-Mitter; SE, standard error; min, minutes.

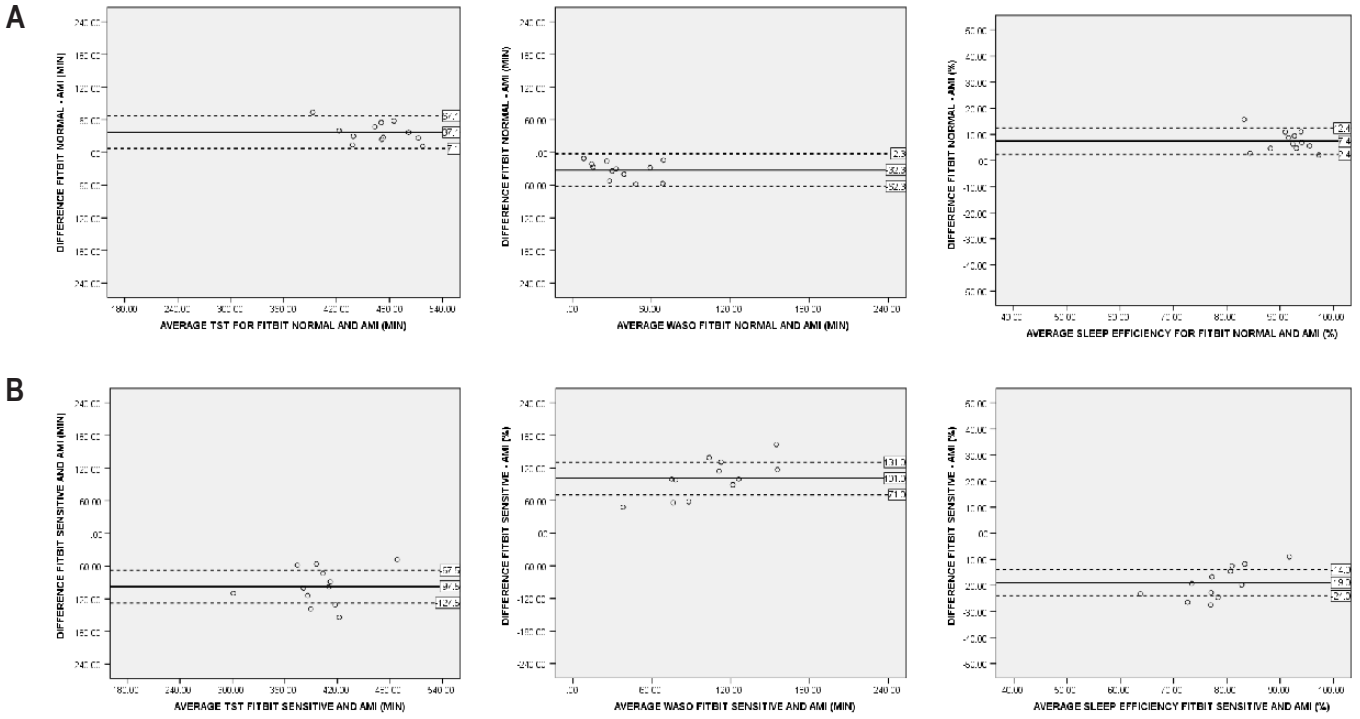

Figure 1—Bland-Altman plots. Fitbit Ultra Normal vs. AMI **(A)**. Fitbit Ultra Sensitive vs. AMI **(B)**.The solid lines indicate the mean of the differences. Based on a priori limits of clinical agreement, the dashed lines are 30 minutes above and below the mean difference for total sleep time (TST) and wake after sleep onset (WASO) and 5% above and below the mean difference for sleep efficiency.

Compared to the AMI device, Fitbit Ultra Sensitive mode significantly underestimated TST by 98 min and overestimated WASO by 101 min, resulting in an underestimation of SE by 19% (Table 2). The Bland-Altman plots (Figure 1B) show > 30-min discrepancy in TST for 50% of participants, > 30-min discrepancy in WASO for 42% of participants, and > 5% discrepancy in sleep efficiency for 50% of participants.

PRMM Actiwatch Spectrum

Significant differences were found between both Fitbit Ultra modes and PRMM across outcome measures (Table 2). However, post hoc analyses indicated that compared to the PRMM device, Fitbit Ultra Normal mode was not statistically different. However, clinical significance was noted with Fitbit Ultra Normal overestimating TST by 36 min, underestimating WASO by 32 min, and overestimating SE by 7% (Table 2). Further, as seen in the Bland-Altman Plots (Figure 2A), 42% of participants had both > 30-min discrepancy in TST, > 30-min discrepancy in WASO for 33% of participants, and > 5% discrepancy in sleep efficiency.

Compared to the PRMM device, the Fitbit Ultra Sensitive mode underestimated TST by 117 min and overestimated WASO by 109 min, resulting in an underestimation of SE by

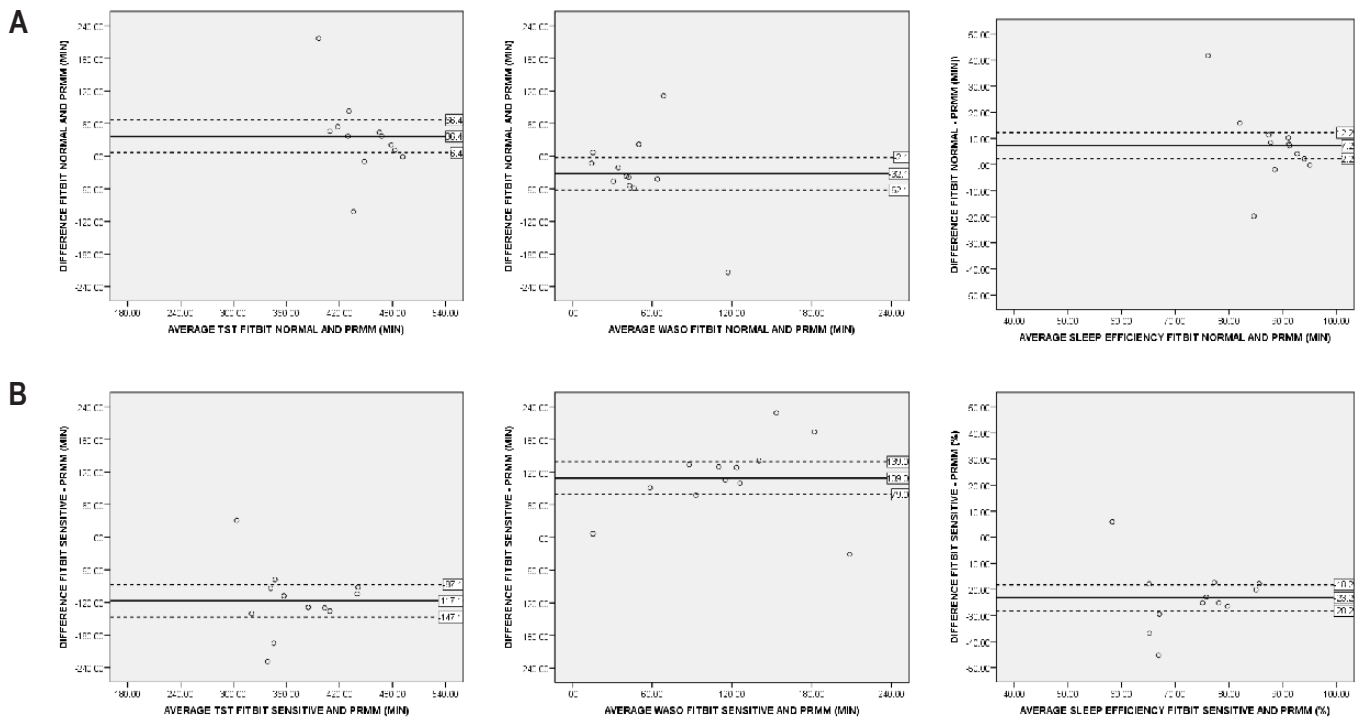

Figure 2—Bland-Altman plots. Fitbit Ultra Normal vs. PRMM **(A)**. Fitbit Ultra Sensitive vs. PRMM **(B)**. The solid lines indicate the mean of the differences. Based on a priori limits of clinical agreement, the dashed lines are 30 minutes above and below the mean difference for total sleep time (TST) and wake after sleep onset (WASO) and 5% above and below the mean difference for sleep efficiency.

Table 3—Comparison of sleep outcomes for PSG, Fitbit Ultra Normal and Fitbit Ultra Sensitive modes for each developmental age group (controlling for sleep disordered breathing status).

Sleep Outcome	PSG Mean (SE)	Fitbit Ultra Normal Mean (SE)	Fitbit Ultra Sensitive Mean (SE)	F[a]	P
Preschool (n = 14)					
TST (min)	448.6 (16.2)	469.6 (11.6)[c]	308.6 (15.2)[b]	62.9	< 0.001
WASO (min)	45.5 (10.0)	26.6 (4.6)	187.0 (13.7)[b]	66.7	< 0.001
Sleep Efficiency (%)	88.8 (2.4)	93.0 (1.1)[c]	61.1 (2.8)[b]	71.8	< 0.001
School-Age (n = 17)					
TST (min)	436.4 (12.6)	463.1 (8.7)[c]	327.5 (14.0)[b]	59.0	< 0.001
WASO (min)	54.9 (9.6)	35.2 (6.8)	158.6 (14.7)[b]	25.0	< 0.001
Sleep Efficiency (%)	86.3 (2.2)	91.7 (1.5)[c]	64.8 (2.6)[b]	53.1	< 0.001
Adolescent (n = 18)					
TST (min)	383.6 (19.8)	454.0 (8.7)[c]	309.9 (17.3)[b]	21.3	< 0.001
WASO (min)	90.4 (14.1)	36.3 (7.2)[c]	172.0 (15.6)[b]	17.9	< 0.001
Sleep Efficiency (%)	76.7 (3.6)	90.9 (1.5)[c]	61.9 (3.1)[b]	21.7	< 0.001

[a]Preschool df = (2,11), School-Age df = (2,14), Adolescent df = (2, 15). [b]Post hoc analyses indicate significantly different from PSG, P ≤ 0.001. [c]Post hoc analyses indicate significantly different from PSG, P < 0.05. PSG, polysomnography; TST, total sleep time; WASO, wake after sleep onset; SE, standard error; min, minutes.

23% (Table 2). The Bland-Altman plots (Figure 2B) identify 33% of participants having > 30-min discrepancy in TST, > 30-min discrepancy in WASO for 50% of participants, and 58% of participants having > 5% discrepancy in sleep efficiency.

Developmental Age Group

In order to examine differences in Fitbit Ultra sleep measurements by developmental age, participants were identified as belonging to one of three groups: Preschool (ages 3–5 years; n = 14); School-Age (ages 6–12 years; n = 17); or Adolescent (ages 13–18 years; n = 18). Significant differences for all three outcomes were found between PSG and both Fitbit Ultra modes (Table 3).

For both preschoolers and school-aged children, post hoc analyses found that Fitbit Ultra Normal significantly overestimated total sleep time and SE, while the Fitbit Ultra Sensitive

Table 4—Comparison of sleep outcomes for PSG, Fitbit Ultra Normal and Fitbit Ultra Sensitive modes for each sleep disordered breathing group (controlling for developmental age group).

Sleep Outcome	PSG Mean (SE)	Fitbit Ultra Normal Mean (SE)	Fitbit Ultra Sensitive Mean (SE)	F[a]	P
No OSA (n = 30)					
TST (min)	446.8 (9.9)	472.0 (5.6)[c]	331.6 (8.5)[b]	27.5	< 0.001
WASO (min)	46.7 (7.4)	26.9 (4.4)[d]	163.2 (7.6)[b]	22.9	< 0.001
Sleep Efficiency (%)	88.4 (1.8)	93.5 (0.8)[c]	65.7 (1.6)[b]	27.5	< 0.001
Mild OSA (n = 12)					
TST (min)	393.5 (23.1)	452.4 (11.6)[c]	293.9 (12.8)[b]	43.0	< 0.001
WASO (min)	75.7 (15.4)	32.1 (4.9)[c]	181.0 (11.2)[b]	33.4	< 0.001
Sleep Efficiency (%)	79.3 (4.0)	91.5 (1.3)[c]	59.5 (2.0)[b]	49.4	< 0.001
Moderate/Severe OSA (n = 7)					
TST (min)	353.6 (39.0)	433.0 (24.4)	284.4 (52.1)[d]	0.67	0.56
WASO (min)	127.1 (26.6)	61.6 (18.0)	192.0 (50.4)[d]	0.04	0.96
Sleep Efficiency (%)	69.1 (6.9)	84.9 (4.3)	55.3 (9.6)[d]	0.73	0.54

[a]No OSA df = (2,27), Mild OSA df = (2,9), Adolescent df = (2,4). [b]Post hoc analyses indicate significantly different from PSG, P ≤ 0.001. [c]Post hoc analyses indicate significantly different from PSG, P < 0.01. [d]Post hoc analyses indicate significantly different from PSG, P < 0.05. OSA, obstructive sleep apnea; PSG, polysomnography; TST, total sleep time; WASO, wake after sleep onset; SE, standard error; min, minutes.

mode significantly underestimated TST, overestimated WASO, and underestimated SE (Table 3). A significant difference in WASO was not found for either age group using the Normal mode, with differences under 20 minutes (Table 3). For adolescents, significant post hoc differences were found for all sleep outcomes, with the Fitbit Ultra Normal significantly overestimating TST, underestimating WASO, and overestimating SE, while the Fitbit Ultra Sensitive mode significantly underestimated TST, overestimated WASO, and underestimated SE (Table 3).

Sleep Disordered Breathing Status

Participants were divided into 3 SDB groups based on their AHI, resulting in 30 youth with No OSA; 12 with Mild OSA; and 7 with Moderate/Severe OSA. Significant differences were found between PSG and both Fitbit Ultra modes for all 3 sleep outcomes for both the No OSA and Mild OSA groups (Table 4). Although a significant difference was not found between PSG and Fitbit Ultra for those with Moderate/Severe OSA, this was likely due to the small number of youth in this group. For exploratory purposes, post hoc *t*-tests were still performed.

For both youth with No OSA and those with Mild OSA, significant post hoc differences were found for all sleep outcomes, with the Fitbit Ultra Normal significantly overestimating TST, underestimating WASO, and overestimating SE, while the Fitbit Ultra Sensitive mode significantly underestimated TST, overestimated WASO, and underestimated SE (Table 4). For youth with Moderate/Severe SDB, a significant difference was found only between PSG and the Fitbit Ultra Sensitive mode for all sleep outcomes, with the Fitbit Ultra Sensitive mode significantly underestimating TST, overestimating WASO, and underestimating SE (Table 4). Notable, although not statistically significant, the Fitbit Ultra Normal mode overestimated TST by 79 min, underestimated WASO by 66 min, and overestimated SE by 16% (Table 4).

DISCUSSION

To our knowledge, this study is the first to examine the validity of the Fitbit Ultra for the measurement of sleep patterns in children and adolescents. Inexpensive and commercially available accelerometers, such as the Fitbit Ultra, have become increasingly popular. That said, it is important to note that Fitbit Ultra "is not directed at persons under the age of 13" (see http://www.fitbit.com/privacy). Furthermore, Fitbit's terms and conditions state "the Fitbit Service is not intended to diagnose, treat, cure, or prevent any disease" (see http://www.fitbit.com/terms). However, it is becoming more common in clinical practice to have patients and parents reporting on a child or adolescent's sleep based on data obtained with commercially available devices. Thus it is of value for clinicians to understand the strengths and limitations of these devices.

In addition, as a measure of sleep, these devices offer an appealing and consumer-friendly alternative for sleep researchers. However, the results of this study highlight the inability of the Fitbit Ultra to accurately capture sleep duration and sleep efficiency in pediatric populations. The strengths of this study include direct comparison of the Fitbit Ultra with both polysomnography and previously validated actigraphs in pediatric populations, as well as the inclusion of youth of varying ages and SDB status. Limitations of the present study include the relatively small sample size (especially for the age group and SDB status comparisons), a single night of assessment completed in the sleep lab (which does not capture potential movement caused by bed sharing or pets), and the need to collapse the 30-sec PSG into 1-min epochs in order to make EBE comparisons with the Fitbit Ultra.

Side-by-side Fitbit Ultra comparisons indicated that intra-device reliability was acceptable in both modes; however, intra-device reliability was much lower for both settings compared to previously published reports of the intra-device reliability of actigraphy.[6] In other words, one Fitbit Ultra device may not be substituted for another Fitbit Ultra device in the

middle of a research protocol, as sleep outcomes may differ between devices. That said, we only looked at a single night of measurement; thus, we cannot draw conclusions about the ability of a single Fitbit Ultra device to reliably measure one person's sleep over multiple nights.

Our results also demonstrated that neither sleep-recording mode of the Fitbit Ultra device provided consistent or comparable results to PSG or actigraphy. Compared to PSG and more traditional actigraphy devices, the Fitbit Ultra Normal mode significantly overestimated sleep duration and underestimated wake after sleep onset, resulting in an overestimation of sleep efficiency. This may lead a patient, clinician, or researcher to conclude that sleep is much better than PSG or actigraphy would indicate.

The reverse was true for the Sensitive mode, with the Fitbit Ultra significantly underestimating sleep duration and overestimating wake after sleep onset, resulting in an underestimation of sleep efficiency. Use of the Sensitive mode may lead a patient, clinician, or researcher to conclude that sleep is much worse than PSG or actigraphy would indicate.

It is important to again note that the utility of Fitbit Ultra as a multi-night measure was not examined in this study, so there may be benefit if the clinical question is whether or not a person's sleep patterns change, for example, pre/post treatment. However, it is notable that most of the overestimates and underestimates of sleep fell outside the *a priori* range of satisfactory agreement for clinical outcomes (TST < 30 minutes and SE < 5% different). Thus, clinicians should use caution when interpreting Fitbit Ultra data provided by patients who purchased the device, as these data may be significant overestimate or underestimates of the patient's actual sleep (depending on the mode use).

The accuracy of clinical sleep outcomes differed for both modes depending on the child's developmental age and SDB status. Most notably for age, there was no significant difference in WASO for both preschool and school-aged children, while for adolescents there was an underestimation by 54 minutes. This highlights the fact that adolescents are more likely to exhibit motionless wakefulness (not moving but still awake) than younger children. Differences between PSG and Fitbit Ultra were greater for youth with mild OSA—and especially moderate or severe OSA—than for youth with no SDB. This suggests that children and adolescents with SDB likely experience more movement during sleep. These findings do not support the manufacturer's recommendation that the Sensitive mode may be better for those with suspected sleep disorders.[15]

The clinically meaningful differences in sleep outcome measures found between the Fitbit Ultra and both PSG and actigraphy highlight the limitations of the Fitbit Ultra as a sleep research measure. The utility of the Fitbit Ultra as a research measure is further diminished by the difficulty in extracting raw sleep data from the Fitbit website, as well as the use of scoring algorithms based on unknown proprietary algorithms. Not only must the minute-by-minute data be extracted manually (as was done in this study and the Montgomery-Downs et al. study[12]), summary variables (e.g., total sleep time) cannot be exported directly into an analyzable format (e.g., a .txt or .csv format). Further, when extracting data (including sleep outcome variables) by hand, one must view each day individually, with no option to examine daily data for longer periods (e.g., 1–2 weeks) on a single screen.

While this study highlights the significant limitations of the Fitbit Ultra as a research device, there may be a role for commercially available devices within a clinical setting. For example, the Fitbit Ultra may provide within-subject changes to sleep patterns (e.g., bedtimes, wake times) over multiple nights that is more objective than self- or parent-report measures. However, it should also be noted that data were not recorded for 12 participants in this study (19%) due to a technical failure of the Fitbit Ultra. Further, Fitbit Ultra users should carefully consider their selection of one sleep-recording mode over another because the two modes yield vastly different data. Finally, additional studies are needed to assess the Fitbit Ultra's utility as a longer-term sleep measurement compared to traditional self- and parent-report sleep diary methods.

ACKNOWLEDGMENTS

The authors thank the children and families who participated in this study, and Devon Ambler and Genery Booster for their assistance with scoring.

DISCLOSURE STATEMENT

This was not an industry supported study. This study was funded in part by NIH K23 MH066772 and the UAB Health Services Foundation General Endowment Fund. The authors have indicated no financial conflicts of interest. Data collection for this research project took place at Children's of Alabama hospital in Birmingham, Alabama. Research analyses took place at National Jewish Health in Denver, Colorado.

REFERENCES

1. Acebo C, Sadeh A, Seifer R, et al. Sleep/wake patterns derived from activity monitoring and maternal report for healthy 1- to 5-year-old children. Sleep 2005;28:1568–77.
2. Paavonen EJ, Aronen ET, Moilanen I, et al. Sleep problems of school-aged children: a complementary view. Acta Paediatr 2000;89:223–8.
3. Owens JA, Spirito A, McGuinn M, et al. Sleep habits and sleep disturbance in elementary school-aged children. J Dev Behav Pediatr 2000;21:27–36.
4. Meltzer LJ, Avis KT, Biggs S, et al. The Children's Report of Sleep Patterns (CRSP): a self-report measure of sleep for school-aged children. J Clin Sleep Med 2013;9:235–45.
5. Malow BA, Marzec ML, McGrew SG, et al. Characterizing sleep in children with autism spectrum disorders: a multidimensional approach. Sleep 2006;29:1563–71.
6. Meltzer LJ, Walsh CM, Traylor J, et al. Direct comparison of two new actigraphs and polysomnography in children and adolescents. Sleep 2012;35:159–66.
7. Meltzer LJ, Montgomery-Downs HE, Insana SP, et al. Use of actigraphy for assessment in pediatric sleep research. Sleep Med Rev 2012;16:463–75.
8. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 2011;15:259–67.
9. Weiss AR, Johnson NL, Berger NA, et al. Validity of activity-based devices to estimate sleep. J Clin Sleep Med 2010;6:336–42.
10. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep 1994;17:201–7.
11. Sadeh A, Lavie P, Scher A, et al. Actigraphic home-monitoring of sleep-disturbed and control infants and young children: a new method for pediatric assessment of sleep-wake patterns. Pediatrics 1991;87:494–9.
12. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. Sleep Breath 2012;16:913–7.

13. Iber C, Ancoli-Israel S, Chesson AL Jr., Quan S. The AASM manual for the scoring of sleep and associated events. Westchester, IL: American Academy of Sleep Medicine, 2007.
14. 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.
15. Sleep Tracking FAQs. Fitbit.com; https://help.fitbit.com/customer/portal/articles/1230068-about-sleep-tracking.
16. O'Driscoll DM, Foster AM, Davey MJ, et al. Can actigraphy measure sleep fragmentation in children? Arch Dis Child 2010;95:1031–3.
17. 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.
18. 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.
19. Kushida CA, Chang A, Gadkary C, et al. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med 2001;2:389–96.
20. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. Statistician 1983;32:307–17.
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
22. Werner H, Molinari L, Guyer C, et al. Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns. Arch Pediatr Adolesc Med 2008;162:350–8.