ORIGINAL RESEARCH

Actigraphy-based sleep estimation in adolescents and adults: a comparison with polysomnography using two scoring algorithms

Mirja Quante[1–3]
Emily R Kaplan[2]
Michael Cailler[2]
Michael Rueschman[2]
Rui Wang[2–5]
Jia Weng[2]
Elsie M Taveras[3,5,6]
Susan Redline[2,3,7]

[1]Department of Neonatology, University of Tuebingen, Tuebingen, Germany; [2]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, USA; [3]Harvard Medical School, Boston, MA, USA; [4]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA; [5]Department of Population Medicine, Harvard Medical School and The Harvard Pilgrim Health Care Institute, Boston, MA, USA; [6]Division of General Academic Pediatrics, Department of Pediatrics, MassGeneral Hospital for Children, Boston, MA, USA; [7]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

Correspondence: Mirja Quante
Department of Neonatology, University of Tuebingen, Calwerstr, 7, 72076 Tuebingen, Germany
Tel +49 7071 80877
Fax +49 7071 4356
Email mirja.quante@med.uni-tuebingen.de

Objectives: Actigraphy is widely used to estimate sleep–wake time, despite limited information regarding the comparability of different devices and algorithms. We compared estimates of sleep–wake times determined by two wrist actigraphs (GT3X+ versus Actiwatch Spectrum [AWS]) to in-home polysomnography (PSG), using two algorithms (Sadeh and Cole–Kripke) for the GT3X+ recordings.

Subjects and methods: Participants included a sample of 35 healthy volunteers (13 school children and 22 adults, 46% male) from Boston, MA, USA. Twenty-two adults wore the GT3X+ and AWS simultaneously for at least five consecutive days and nights. In addition, actigraphy and PSG were concurrently measured in 12 of these adults and another 13 children over a single night. We used intraclass correlation coefficients (ICCs), epoch-by-epoch comparisons, paired *t*-tests, and Bland–Altman plots to determine the level of agreement between actigraphy and PSG, and differences between devices and algorithms.

Results: Each actigraph showed comparable accuracy (0.81–0.86) for sleep–wake estimation compared to PSG. When analyzing data from the GT3X+, the Cole–Kripke algorithm was more sensitive (0.88–0.96) to detect sleep, but less specific (0.35–0.64) to detect wake than the Sadeh algorithm (sensitivity: 0.82–0.91, specificity: 0.47–0.68). Total sleep time measured using the GT3X+ with both algorithms was similar to that obtained by PSG (ICC=0.64–0.88). In contrast, agreement between the GT3X+ and PSG wake after sleep onset was poor (ICC=0.00–0.10). In adults, the GT3X+ using the Cole–Kripke algorithm provided data comparable to the AWS (mean bias=3.7±19.7 minutes for total sleep time and 8.0±14.2 minutes for wake after sleep onset).

Conclusion: The two actigraphs provided comparable and accurate data compared to PSG, although both poorly identified wake episodes (i.e., had low specificity). Use of actigraphy scoring algorithm influenced the mean bias and level of agreement in sleep–wake times estimates. The GT3X+, when analyzed by the Cole–Kripke, but not the Sadeh algorithm, provided comparable data to the AWS.

Keywords: validation, actigraphy, polysomnography, scoring algorithm

Introduction

Actigraphy is increasingly used to provide objective measurements of sleep.[1–3] A PubMed search done in April 2017 revealed 2,503 scientific publications that include actigraphy and sleep. Nonetheless, there are limited data that directly assess agreement with polysomnography (PSG), and evaluate the extent to which findings change as a result of the use of different actigraphs and alternative scoring algorithms and the variation across age groups. Such data are timely, given the active use of actigraphs in large population-based initiatives, including the UK Biobank. Understanding the

comparability of data across studies is important in interpreting and generalizing findings.

The most commonly used algorithms for sleep–wake scoring in children, adolescents, and adults have been developed by Sadeh et al[4] and Cole et al[5]. Sadeh's (S) algorithm was originally validated on a healthy sample of adolescents and young adults (age range 10–25 years), whereas Cole–Kripke's (CK) algorithm was validated in an adult sample (age range 35–65 years). Few studies have validated these two scoring algorithms against the gold standard of PSG; these studies were limited by not assessing both children and adults concurrently,[6–9] studying patients referred for PSG,[7–9] using only a single actigraph,[6,8,9] and not synchronizing clock times for PSG and actigraphy.[8]

In particular, little research is available that addresses the validity of a newer actigraph, the GT3X+ (ActiGraph™; ActiGraph, Pensacola, FL, USA), compared to the "gold standard" PSG. The GT3X+ is in use in large-scale epidemiologic studies, such as the National Health and Nutrition Examination Survey, the Transdisciplinary Research on Energetics initiative, and a prebirth cohort study, Project Viva. Little information is available regarding how the GT3X+ device performs in comparison to a device commonly used in sleep research, the Actiwatch Spectrum (AWS; Philips/Respironics, Murrysville, PA, USA).[10–12] To the best of our knowledge, no study has yet validated the GT3X+ for sleep in school children and young teens (age 10–14 years). It is very important to conduct actigraphy validation studies in pediatric populations, given the significant sleep problems that children suffer from and the emerging data from small- as well as large-scale research studies.[2]

Accordingly, in this study, our aims were to: 1) assess the accuracy of sleep–wake estimation of the GT3X+ actigraph in children in addition to adults; 2) examine the validity of GT3X+ with the CK scoring algorithm; and 3) evaluate the agreement between sleep parameters from concurrently used GT3X+ and AWS wrist actigraph devices.

Subjects and methods
Study sample
Participants consisted of healthy volunteers, 22 adults and 13 children (age 10–14 years). We identified volunteers through local advertisement at gyms, workplaces, and schools. Exclusion criteria included self-reported severe untreated diagnosed sleep disorder (e.g., untreated severe sleep apnea), serious or acute illness, use of sedating medications, and disabilities that might interfere or restrict mobility. This study was conducted in accordance with "The Code of Ethics of

the World Medical Association" (Declaration of Helsinki), and the Institutional Review Board of Brigham and Women's Hospital approved the study. All participants, or legal guardians of the children under the legal age, provided written informed consent for this study. In addition, we obtained assent from children under 14 years of age.

Actigraphy and sleep diary
We asked all adults and children to wear a GT3X+ device on the nondominant wrist for at least 5 days. Participants completed a diary using a smartphone app before going to bed and upon awakening each morning, in which they noted time in bed, time out of bed, nocturnal awakenings, and any naps taken. The app was programmed with diary questions similar to those used for the consensus sleep diary by Ginger.io (California Medical P.C., San Francisco, CA, USA).[13] All logs and participants' answers were sent to a secure web-based dashboard. Electronic sleep diaries have been shown to provide valid information on sleep patterns.[14] To compare the GT3X+ with the AWS, adults also wore an AWS concurrently on the nondominant wrist. The type of device worn nearer to the wrist was randomly determined. Children only wore the GT3X+ to minimize burden.

Activity counts for the GT3X+ were collected in 1-minute epochs, whereas data for the AWS were collected in 30-s epochs, reflecting the default modes of these devices. We analyzed GT3X+ data using ActiLife 6 software (Version 6.13.1, ActiGraph). The software provides the S and CK algorithms to score sleep, and we applied each independently to derive sleep–wake estimates from the GT3X+ devices. The S algorithm was originally validated in adolescents and young adults wearing an AMA-32 actigraph (Ambulatory Monitoring, Inc., Ardsley, NY, USA), whereas the CK algorithm was validated in adults wearing a Motionlogger Actigraph (Ambulatory Monitoring, Inc.). In order to match the output of the more sensitive GT3X+ device with the Motionlogger Actigraph and AMA-32, the ActiGraph company adapted the original S and CK algorithms to the ActiGraph devices by performing a side-by-side test using devices from both companies worn together (D Judge, ActiGraph, personal communication, April, 2016). For the CK algorithm, the epoch data were adjusted through scaling the count values by 100 and setting scaled values over 300 to 300. For the S algorithm, the company changed the original scoring algorithm threshold for "probability of sleep" from ≥0 counts to ≥−4 counts. We scored the AWS data using Respironics Actiware 5 (Version 5.59, Philips/Respironics), which is the only algorithm provided by this device for analysis. The

algorithm has three threshold settings. We used a medium threshold of 40 counts per epoch for wake detection with the AWS. The medium-threshold setting is thought to give the best compromise between detecting sleep and wake in terms of sensitivity and specificity compared to PSG.[10] We manually identified the night-time sleep period based on the self-completed Ginger.io smartphone app sleep diaries and observation of a sharp decrease/increase in activity.

Polysomnography

For a single night during the actigraphy recording period, 12 of the 22 enrolled adults and all 13 children underwent an unattended overnight PSG at their homes (Somté; Compumedics, Abbotsford, Australia), collecting the following channels for sleep staging: electroencephalogram (Fz-Cz, C4-M1, Oz-Cz), bilateral electrooculograms, and submental electromyogram. The Compumedics system has been proven to be feasible for children in an unattended setting.[15] Trained research staff went to the homes 1–3 hours before usual bedtime to place all sensors. PSG and actigraphs were precisely synchronized by initializing all devices on the same computer. PSG studies were scored manually by a registered polysomnologist, blinded to the actigraphy and all other data, using established criteria for sleep staging in 30-s epochs.[16] We identified "lights off" and "lights on" from information recorded by the participant using the Ginger.io smartphone app diary. The PSG was considered satisfactory if there were at least 5 hours of electroencephalogram that could be staged for sleep and sleep efficiency exceeded 80%.

Data analysis

We performed two sets of analyses to assess the agreement between actigraphy and PSG: an analysis of sleep summary statistics derived from data across the sleep period and an epoch-by-epoch analysis. The PSG to actigraphy comparison was based on a single concurrent night of monitoring. Total sleep time (TST) was calculated as the number of minutes scored as sleep between "lights off" and "lights on". Wake after sleep onset (WASO) was defined as the number of minutes scored as WASO during the sleep period, and sleep efficiency (SE) as the ratio of TST to the total time between "lights off" and "lights on" and reported as percentage. Paired t-tests or Wilcoxon rank-sum tests were used to evaluate differences in sleep variables between each device, specific for each algorithm, and PSG. We calculated the mean differences between actigraphy settings and PSG and its corresponding 95% CI. Agreement between PSG and actigraphy was also examined using intraclass correlation coefficients

(ICCs). The Bland–Altman concordance technique was applied to examine degree of agreement between the two actigraphy devices, GT3X+ and AWS, for TST and WASO. For the epoch-by-epoch analysis, we calculated accuracy, sensitivity, and specificity. In order to match the 30-s PSG sleep epochs with the 1-minute actigraphy epochs from the GT3X+, we converted the 1-minute GT3X+ epochs to 30-s epochs by dividing into half the GT3X+ epochs. PSG data were reduced to binary form (0=wakefulness and 1=any sleep state). Accuracy was defined as the total number of epochs of sleep defined by PSG that were correctly classified by actigraphy divided by the total number of scored epochs (true sleep categorized by actigraphy/total scored epochs). Sensitivity represents the proportion of epochs identified by PSG that were correctly classified as sleep by actigraphy (true sleep categorized by actigraphy/sleep categorized by PSG), whereas specificity is the proportion of epochs identified correctly as wake by the actigraph (true wake categorized by actigraphy/wake categorized by PSG). We performed separate analyses for children and adults. Analyses were conducted using SAS 9.4 (SAS Institute, Inc., Cary, NC, USA).

Results

Descriptive characteristics of the study sample are shown in Table 1. Participants were 22 adults (mean age of 32.3±11.4 years) and 13 children (mean age of 13.0±1.4 years). The sample consisted of 19 females and 16 males, and 46% were non-Hispanic white participants. Participants provided a mean of 7.3±1.5 nights of valid actigraphy data. Of the initial 25 overnight PSG studies, 13 adolescent studies and nine adult studies were satisfactory, but three adult studies failed (two with general technical failure and one with misplaced

Table 1 Descriptive characteristics of the study population

	Children, n=13	Adults, n=22
	n	n
Male	7	9
Race/ethnicity		
Non-Hispanic white	2	14
Hispanic		
Black	2	0
White	9	3
Other	0	5
	Mean ± SD, median (interquartile range)	
Age, years	13.0±1.4, 13.0 (2.9)	32.3±11.4, 29.3 (9.7)
Weight, kg	59.9±25.6	72.2±16.7
BMI, kg/m^2	24.2±10.6, 20.9 (16.3)	24.6±4.4, 23.5 (4.7)

Abbreviation: BMI, body mass index.

leads). A failure rate of about 10% has been described for in-home PSG.[17]

Agreement between sleep summary statistics using the actigraphs and PSG

Mean values, mean differences, and ICCs for sleep variables comparing PSG with actigraphy for the sample of children and adults are presented in Tables 2 and 3. In adults, average TST by PSG was 435.6±36.9 minutes. Compared to this, the GT3X+ analyzed using the CK algorithm overestimated TST by an average of 14 minutes, whereas the GT3X+ analyzed using the S algorithm underestimated sleep by an average of 14 minutes. In children, the corresponding average TST by PSG was 483.4±62.8 minutes. The GT3X+ analyzed using the CK algorithm and the GT3X+ analyzed using the S algorithm underestimated TST by an average of 18 and 49 minutes, respectively. The ICC for TST between the GT3X+ with the CK algorithm and PSG was 0.81 in adults and 0.88 in children. The ICC for TST between PSG and the GT3X+ using the S algorithm was somewhat lower (ICC=0.64 in adults and ICC=0.79 in children).

The mean PSG WASO was 36.9±23.6 minutes in adults and 18.1±14.6 minutes in children. For average WASO derived from GT3X+, the S algorithm overestimated wake by 21 minutes (57%) in adults and by 59 minutes (327%) in children, resulting in a 4.5% (adults) and 11.6% (children) lower SE.

Table 2 mean ± SD/median (interquartile range) values for PSG and actigraphy of total sleep time, wake after sleep onset, and sleep efficiency

	Children, n=13	Adults, n=9
Total sleep time, minutes		
PSG	483.4 ± 62.8	435.6 ± 36.9
GT3X+: Sadeh	434.8 ± 55.9*	421.6 ± 48.5
GT3X+: Cole–Kripke	465.2 ± 56.6	449.7 ± 39.5
Actiwatch Spectrum: medium threshold	NA	444.6 ± 53.7
Wake after sleep onset, minutes		
PSG	18.1 ± 14.6, 14.0 (15.0)	36.9 ± 23.6
GT3X+: Sadeh	77.3 ± 37.5, 70.0 (72.0)**	58.1 ± 6.7
GT3X+: Cole–Kripke	49.1 ± 27.3, 45.0 (38.0)**	35.6 ± 9.5
Actiwatch Spectrum: medium threshold	NA	30.2 ± 7.0
Sleep efficiency, %		
PSG	95.2 ± 3.5, 95.1 (2.8)	90.4 ± 4.3
GT3X+: Sadeh	83.6 ± 6.7, 84.0 (14.1)**	85.9 ± 6.4
GT3X+: Cole–Kripke	89.3 ± 4.7, 89.3 (6.4)**	91.6 ± 2.5
Actiwatch Spectrum: medium threshold	NA	92.4 ± 2.1

Notes: Comparison versus PSG, *p<0.05, **p<0.01.
Abbreviations: NA, not applicable; PSG, polysomnography.

The CK algorithm underestimated wake by 1 minute (4%) in adults and overestimated wake by 31 minutes (63%) in children. This resulted in a 1.3% higher (adults) and 5.8% lower (children) SE, respectively. However, there was little overall agreement between the GT3X+ and PSG for WASO (CK: ICC=0.00 in children and ICC=0.10 in adults; S: ICC=0.00 in children and ICC=0.00 in adults) or for SE (CK: ICC=0.00 in children and ICC=0.33 in adults; S: ICC=0.00 in children and ICC=0.13 in adults).

Compared to PSG, TST, WASO, and SE derived from AWS differed by 9 minutes, 7 minutes, and 2%, respectively. The ICCs between PSG and AWS were high for TST (ICC=0.84) and low for WASO (ICC=0.00) and SE (0.18).

Device agreement analysis

We constructed Bland–Altman plots to visualize the level of agreement between GT3X+ and AWS for TST and WASO in 22 adults. The mean biases were estimated at −81.9±19.7 minutes (TST) and 34.2±18.0 minutes (WASO) for the S algorithm and 3.7±19.7 minutes (TST) and 8.0±14.2 minutes (WASO) for the CK algorithm, compared to the AWS (Figure 1).

Epoch-by-epoch agreement analysis

Sensitivity, specificity, and accuracy of epoch-by-epoch comparisons between the actigraphy and PSG data in the adolescent and adult samples are shown in Table 4. Overall sensitivity (i.e., ability to accurately identify sleep) was high for all actigraphs and algorithms, ranging from 0.82 to 0.88 in children and from 0.91 to 0.96 in adults. However, specificity values (i.e., ability to accurately identify wake) were moderate to low, ranging from 0.64 to 0.68 in children and from 0.34 to 0.47 in adults. The GT3X+ with the CK algorithm was not only more sensitive (0.96 in adults, 0.88 in children), but also less specific (0.35 in adults, 0.64 in children), compared to the GT3X+ with the S algorithm.

Discussion

Although the GT3X+ and similar actigraph devices are increasingly used across a range of research, there is a paucity of research that addresses the comparability of information from these devices to other actigraphs, or validity as compared to PSG. Specifically, the validity of the GT3X+ in children using the CK algorithm has not yet been examined. To address this gap, we evaluated the accuracy of the wrist-placed GT3X+ to detect sleep–wake epochs and summary statistics compared to overnight concurrent PSG, using both the S and CK sleep algorithms in healthy children and adults.

Table 3 Mean differences and intraclass correlation coefficients between PSG and actigraphy

	Children, n=13		Adults, n=9	
	Mean difference (95% CI)	ICC	Mean difference (95% CI)	ICC
TST – GT3X+: Sadeh (minutes)	−48.6 (−96.7 to −0.5)	0.79	−14.0 (−57.1 to 29.1)	0.64
TST – GT3X+: Cole–Kripke (minutes)	−18.2 (−66.5 to 30.2)	0.88	14.1 (−24.1 to 52.3)	0.81
TST – Actiwatch Spectrum: medium threshold (minutes)	NA	NA	9.1 (−37.0 to 55.1)	0.84
WASO – GT3X+: Sadeh (minutes)	59.2 (35.5–83.0)	0.00	21.2 (−0.7 to 43.1)	0.00
WASO – GT3X+: Cole–Kripke (minutes)	31.0 (13.0–49.0)	0.00	−1.4 (−20.1 to 17.4)	0.10
WASO – Actiwatch Spectrum: medium threshold (minutes)	NA	NA	−6.8 (−25.2 to 11.6)	0.00
SE – GT3X+: Sadeh (%)	−11.6 (−15.9 to −7.3)	0.00	−4.5 (−9.9 to 1.0)	0.13
SE – GT3X+: Cole–Kripke (%)	−5.8 (−9.2 to −2.4)	0.00	1.3 (−2.2 to 4.8)	0.33
SE – Actiwatch Spectrum: medium threshold (%)	NA	NA	2.0 (−1.4 to 5.4)	0.18

Abbreviations: ICC, intraclass correlation coefficient; NA, not applicable; PSG, polysomnography; SE, sleep efficiency; TST, total sleep time; WASO, wake after sleep onset.
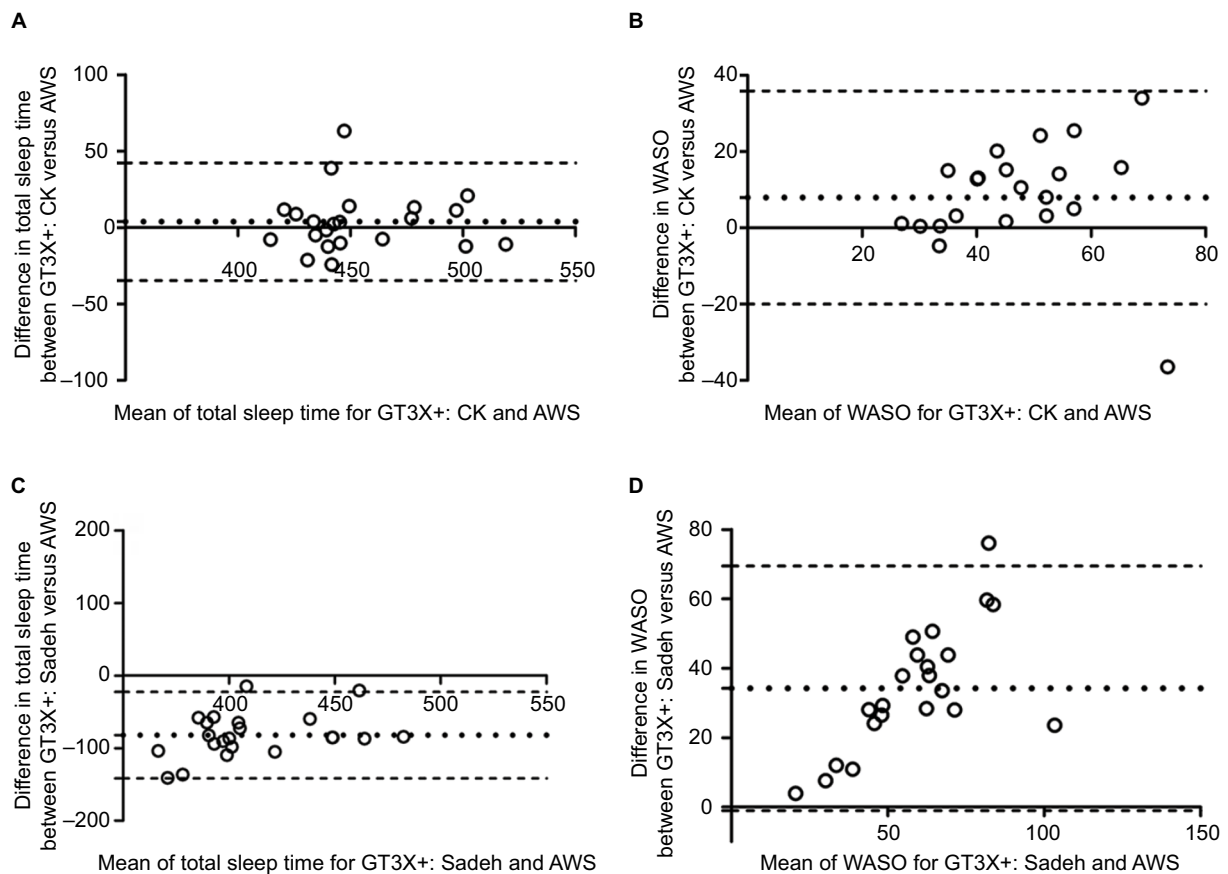

Figure 1 Bland–Altman plots of GT3X+ and AWS in 22 adults for (**A**) total sleep time: GT3X+: CK versus AWS, (**B**) WASO: GT3X+: CK versus AWS, (**C**) total sleep time: GT3X+: Sadeh versus AWS, and (**D**) WASO: GT3X+: Sadeh versus AWS. The dotted line indicates the mean of the differences or bias, and the dashed lines indicate the lower and upper 95% limits of agreement.
Abbreviations: AWS, Actiwatch Spectrum; CK, Cole–Kripke; WASO, wake after sleep onset.

We also examined the agreement between the GT3X+ and AWS over at least five nights of monitoring in adults. Each actigraphy device had comparable moderate to high accuracy compared to PSG. Consistent with previous studies, the specificity for both actigraphs and algorithms to detect wakefulness within sleep periods was moderate to low.[1] Overall,

the CK algorithm was more sensitive, but less specific than the S algorithm. For both algorithms, sleep misclassification was higher in children, whereas wake misclassification was higher in adults. Both the S and CK algorithms overestimated wake and underestimated sleep in children. A different picture emerged for the CK algorithm and the AWS in adults:

Table 4 Sensitivity, specificity, and accuracy of epoch-by-epoch comparisons with PSG

	Children, n=13	Adults, n=9
GT3X+: Sadeh		
Sensitivity	0.82	0.91
Specificity	0.68	0.47
Accuracy	0.81	0.83
GT3X+: Cole–Kripke		
Sensitivity	0.88	0.96
Specificity	0.64	0.35
Accuracy	0.86	0.85
Actiwatch Spectrum: medium threshold		
Sensitivity	NA	0.95
Specificity	NA	0.34
Accuracy	NA	0.84

Abbreviations: NA, not applicable; PSG, polysomnography.

as compared to PSG, sleep was slightly overestimated and wake underestimated. When used with the CK algorithm, the GT3X+ presented data comparable to the AWS in adults.

So far, only two previous studies have investigated the performance of the wrist-placed GT3X+ compared to PSG and to our knowledge, ours is the first study in school children and young teens.[10,12] Cellini et al compared the GT3X+ with the S algorithm against PSG-recorded naps in 30 healthy young adults and found that the GT3X+ overestimated SE (by 14.5%) and TST (by 8.8 minutes) and underestimated WASO (by 5.9 minutes). In contrast, we found that the S algorithm underestimated sleep and overestimated wake both in children and adults. This is likely explained by the fact that Cellini et al recorded naps and not overnight sleep.[10] Naps and overnight sleep differ in length, and the amount of deep sleep and rapid eye movement sleep, which can be associated with different movements.[18]

Slater et al evaluated the GT3X+ with the S algorithm in 108 young adults and compared it to an overnight PSG.[12] The authors reported similar values for TST and SE (mean differences of <1 minutes and <2%, respectively), whereas WASO was systematically overestimated by about 16 minutes, resulting in a high sensitivity of 0.90 and a moderate specificity of 0.60. Our results are in line with these findings by showing a significant overestimation of WASO using the GT3X+ with the S algorithm in both children and adults.

To the best of our knowledge, ours is the first study also to evaluate the GT3X+ with the CK scoring algorithm against PSG. Overall, there was stronger agreement with PSG using the CK algorithm, but the specificity in detecting wakefulness was better for the S algorithm. These results are consistent with four other studies that evaluated various actigraph models produced by Ambulatory Monitoring, Inc.

Meltzer et al examined the S and CK scoring algorithms in 115 youths, aged 3–18 years, wearing the Motionlogger Sleep Watch compared to PSG. The researchers found that the S algorithms underestimated TST and SE by clinically meaningful amounts (23.6 minutes and 4.4%, respectively), whereas the corresponding differences for the CK algorithm were only 2.4 minutes and 0.4%.[7] The authors reported high sensitivities of 0.89 and 0.92 and moderate specificities of 0.73 and 0.65 for the S and CK algorithms, respectively. Kim et al also compared both the S and CK algorithms with PSG using the Action-W actigraph in patients with sleep apnea. The intraclass correlation analysis for TST estimated by PSG and actigraphy revealed that the CK algorithm performed better than the S algorithm (CK, ICC=0.39 and S, ICC=0.27).[8] However, their ICCs for TST were much lower than ours. These studies, including ours, show that the CK algorithm is more accurate not only in adults but also in children, which is striking since the CK was originally validated in an adult sample. Kim et al also performed a subgroup analysis revealing that actigraphy was more reliable in patients with less severe sleep apnea.[8] Several other studies have shown that actigraphy misclassifies sleep and wake in patients with sleep apnea, in particular, tending to underestimate sleep.[9,19,20] de Souza et al assessed the Mini Motionlogger Actigraphy in 21 healthy volunteers and also reported a slightly higher sensitivity (0.97 versus 0.99) and lower specificity (0.44 versus 0.34) for the S compared to the CK algorithm, when assessing epoch-by-epoch agreement data with PSG in 21 adult volunteers.[6] In an earlier study, Johnson et al found high agreement for TST estimation when comparing the S versus CK algorithms (i.e., ICC of 0.98) with the Octagonal Sleep Watch 2.01 in a sample of 181 adolescents participating in the Cleveland TeenZzz Study.[9] There is always a trade-off between sensitivity and specificity, and further analysis in exploring the effect of sleep algorithm thresholds might, therefore, be useful.

We also evaluated the agreement between sleep parameters from concurrently used GT3X+ and AWS devices over 1 week of recording. In adults, the GT3X+ with the CK algorithm provided data that was comparable to the ASW with average differences in sleep and wake time of 5 minutes. In contrast, larger differences were observed between the GT3X+ analyzed with the S and the AWS (23 minutes for TST and 28 minutes for WASO). Of note, the difference in WASO between the two devices increased with increasing WASO, meaning that the comparability decreases with more wake.

Our study has a number of significant strengths, including inclusion of both children and adults, performance of

a side-by-side comparison of two devices with PSG, use of rigorous statistical testing, and assessments conducted under free-living conditions. We assessed agreement using both summary variables and epoch-by-epoch measurements. However, the study also has several limitations. Our sample size was small (13 children and nine adults for aims 1 and 2, and 22 adults for aim 3) and we only performed a single night of PSG. But PSG was performed on the same night as actigraphy, directly informing on sleep–wake agreement. We used an alternative electroencephalography montage, equally acceptable for use, which often results in "cleaner" electroencephalography signals.[21] The sample consisted of healthy volunteers, and therefore, the results cannot be generalized to patients with sleep disorders or other comorbidities. Future research should, therefore, repeat this validation study in populations with sleep disorders such as obstructive sleep apnea or insomnia.

Conclusion

In conclusion, our study suggests that the GT3X+, a device of increasing use in a wide variety of research settings in children and adults, is a valid device to estimate TST, WASO, and SE in school children and adults. However, the poor specificity suggests that actigraphy should be used more cautiously in studies where periods of wake detection are important. For example, evaluation of wake time will be underestimated, which can reduce clinical utility in disorders such as insomnia, where there is interest in identifying wake periods. For both the S and CK algorithms, sensitivity tended to be higher in adults, while specificity tended to be higher in children. When analyzing data from the GT3X+, the CK algorithm tended to be more sensitive (0.88–0.96) to detect sleep, but less specific (0.35–0.64) to detect wake than the S algorithm (sensitivity: 0.82–0.91, specificity: 0.47–0.68). The GT3X+ data analyzed using the CK algorithm provided comparable sleep/wake estimation compared to the AWS, suggesting that information derived from each device provides similar data in healthy adults.

Acknowledgments

Disclosure

References

1. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev.* 2011;15(4):259–267.
2. 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.
3. Amy W, Hawley MD, Rosemary SCH, Sarah NB. Actigraphy and Sleep/Wake Diaries. Oxford: Oxford University Press; 2013.
4. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. *Sleep.* 1994;17(3):201–207.
5. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep.* 1992;15(5):461–469.
6. de Souza L, Benedito-Silva AA, Pires ML, Poyares D, Tufik S, Calil HM. Further validation of actigraphy for sleep studies. *Sleep.* 2003;26(1):81–85.
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. Kim MJ, Lee GH, Kim CS, et al. Comparison of three actigraphic algorithms used to evaluate sleep in patients with obstructive sleep apnea. *Sleep Breath.* 2013;17(1):297–304.
9. 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(7):899–905.
10. Cellini N, Buman MP, McDevitt EA, Ricker AA, Mednick SC. Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults. *Chronobiol Int.* 2013; 30(5):691–698.
11. Zinkhan M, Berger K, Hense S, et al. Agreement of different methods for assessing sleep characteristics: a comparison of two actigraphs, wrist and hip placement, and self-report with polysomnography. *Sleep Med.* 2014;15(9):1107–1114.
12. Slater JA, Botsis T, Walsh J, King S, Straker LM, Eastwood PR. Assessing sleep using hip and wrist actigraphy. *Sleep Biol Rhythms.* 2015;13(2):172–180.
13. Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep.* 2012;35(2):287–302.
14. Muller S, Hemmi MH, Wilhelm FH, Barr RG, Schneider S. Parental report of infant sleep behavior by electronic versus paper-and-pencil diaries, and their relationship to actigraphic sleep measurement. *J Sleep Res.* 2011;20(4):598–605.
15. Goodwin JL, Enright PL, Kaemingk KL, et al. Feasibility of using unattended polysomnography in children for research--report of the Tucson Children's Assessment of Sleep Apnea study (TuCASA). *Sleep.* 2001;24(8):937–944.
16. Berry RB, Brooks R, Gamaldo CE, et al; The American Academy of Sleep Medicine. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.2. Darien, IL: American Academy of Sleep Medicine; 2015.
17. Tan HL, Kheirandish-Gozal L, Gozal D. Pediatric home sleep apnea testing: slowly getting there! *Chest.* 2015;148(6):1382–1395.

18. Kawada T, Suzuki H, Shimizu T, Katsumata M. Agreement in regard to total sleep time during a nap obtained via a sleep polygraph and accelerometer: a comparison of different sensitivity thresholds of the accelerometer. *Int J Behav Med.* 2012;19(3):398–401.

19. Elbaz M, Roue GM, Lofaso F, Quera Salva MA. Utility of actigraphy in the diagnosis of obstructive sleep apnea. *Sleep.* 2002;25(5):527–531.

20. Hedner J, Pillar G, Pittman SD, Zou D, Grote L, White DP. A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. *Sleep.* 2004;27(8):1560–1566.

21. Duce B, Rego C, Milosavljevic J, Hukins C. The AASM recommended and acceptable EEG montages are comparable for the staging of sleep and scoring of EEG arousals. *J Clin Sleep Med.* 2014;10(7):803–809.