HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Sleep Science

journal homepage: www.elsevier.com/locate/ssci

Sleep Science

Full length article

Inter-device reliability of an automatic-scoring actigraph for measuring sleep in healthy adults

CrossMark

Matthew Driller *, Joseph McQuillan, Shannon O'Donnell

University of Waikato, Hamilton, New Zealand

A B S T R A C T

Actigraphy has become a common method of measuring sleep due to its non-invasive, cost-effective nature. An actigraph (Readiband™) that utilizes automatic scoring algorithms has been used in the research, but is yet to be evaluated for its inter-device reliability. A total of 77 nights of sleep data from 11 healthy adult participants was collected while participants were concomitantly wearing two Readiband™ actigraphs attached together (ACT1 and ACT2). Sleep indices including total sleep time (TST), sleep latency (SL), sleep efficiency (SE%), wake after sleep onset (WASO), total time in bed (TTB), wake episodes per night (WE), sleep onset variance (SOV) and wake variance (WV) were assessed between the two devices using mean differences, 95% levels of agreement, intraclass correlation coefficients (ICC), typical error of measurement (TEM) and coefficient of variation (CV%) analysis. There were no significant differences between devices for any of the measured sleep variables (p > 0.05). TST, SE, SL, TTB, SOV and WV all resulted in *very high* ICC's (> 0.90), with WASO and WE resulting in *high* ICC's between devices (0.85 and 0.80, respectively). Mean differences of −2.1 and 0.2 min for TST and SL were associated with a low TEM between devices (9.5 and 3.8 min, respectively). SE resulted in a 0.3% mean difference between devices. The Readiband™ is a reliable tool for researchers using multiple devices of this brand in sleep studies to assess basic measures of sleep quality and quantity in healthy adult populations.

© 2016 Brazilian Association of Sleep. Production and Hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The quantification and measurement of sleep amongst various interventional and population research studies and clinical settings is of increasing importance. Different methods of monitoring sleep have been extensively researched and validated in the literature, with little focus on the inter-device reliability of such tools. Indeed, the precision required to determine changes in sleep patterns amongst different individuals and populations is of critical importance to understanding and interpreting the results of any sleep research studies.

Although considered the gold standard method of sleep measurement, polysomnography (PSG) requires a somewhat intrusive and expensive assessment of sleep indices [1]. Moreover, PSG monitoring typically requires attendance at a sleep laboratory with specialist staff, in a foreign environment, which may be inconvenient and unnatural for most individuals. Because of this, attempts have been made to measure sleep using less-invasive methods. Such methods include sleep-logs/questionnaires and

wristwatch actigraphy. The use of sleep-logs and questionnaires are common as they are in-expensive and easy to administer. However, these have been shown to have a poor relationship with objective measures of sleep [2], therefore questioning their efficacy. Wristwatch actigraphy is a non-intrusive, cost-effective tool used to estimate sleep quantity and quality which has been compared to PSG, showing accuracies of ~90% in some studies for total sleep time and sleep efficiency [3–5] and as such, are widely used in the sleep literature [3]. Actigraphy involves the use of a device housed in a wristwatch that contains a small accelerometer capable of sensing movement along any one of three axes. The accelerometer samples multiple times per second and with each limb movement, the accelerometer registers this information and stores it in an adjacent memory chip. Once the recording period has finished, the actigraph is downloaded and manually scored for sleep indices by a trained sleep technician [6]. While the process of manually scoring actigraph data has been described for its inter and intra-scorer reliability, it is difficult to make conclusions on the overall reliability of actigraphy given the variation of scoring methods, brands of actigraphs and researchers themselves [3].

Given the limitations of manually scoring actigraph files, a plethora of new actigraphy devices designed to automatically score sleep have emerged. One such device, the Readiband™ (Fatigue Science, Honolulu, USA), is gaining popularity for its use in

sleep research studies [7–10]. The Readiband™ records data at a sample rate of 16 Hz and uses a patented algorithm to automatically score sleep data via download to the companies software. The Readiband™ has been validated against PSG, with levels of accuracy 93% being reported [11]. However, while the device has been shown to be a valid sleep measurement tool, the inter-device reliability of the Readiband™ is yet to be evaluated. Assessing the inter-device reliability for multiple devices of the same brand and model is important for researchers to have confidence that separate devices are reading in a similar and reliable manner. Indeed, this type of assessment has become commonplace in evaluating the reliability of physical activity trackers [12]. However, this type of assessment is not yet standard procedure for new actigraphs that measure sleep indices. Therefore, the purpose of the current study was to investigate the inter-device reliability of the Readiband™ by evaluating 77 nights of data from healthy adult participants concomitantly wearing two Readiband™ devices attached together.

2. Materials and methods

2.1. Participants

A total of 11 healthy adults (4 male/7 female, mean $\pm$ SD; age: 33 $\pm$ 7 years) volunteered to participate in the current study. All participants provided informed written consent before taking part in the study and were free of any diagnosed sleep disorders. Ethical approval for the study was obtained through the institutions Human Research Ethics Committee.

2.2. Methodology

Participants were required to wear two wrist actigraphs (SBV2 Readiband™, Fatigue Science, Honolulu, USA), attached together over a 7-day period to assess inter-device reliability between the two devices (ACT1 and ACT2). The Readiband™ devices have been shown to have good validity (overall accuracy of 93%) when compared to the gold standard of PSG in 50 participants undergoing overnight sleep monitoring at a sleep centre [11] and have been accepted as an approved device by the Federal Drug Administration (FDA) based on this validation. The Readiband™ has also been assessed in a mini-validation study against another actigraph (Micro Mini-Motion Loggers, Ambulatory Monitoring Inc., Ardsley, USA) [10] where the two brands of actigraph were attached together for 3 nights in 8 participants, resulting in acceptable levels of agreement for sleep duration and rest duration ($r=0.84$ and 0.94, respectively). In the current study, both devices were tightly secured together using electrical tape so that they could not move independently of each other and were worn on the participants' non-dominant wrist before initialization of the two devices to record data in 1-minute epochs [13]. This method of determining inter-device reliability of actigraphy monitors has been used previously [14]. Participants were required to wear the actigraph continuously for the 7-day period, with the exception of time spent in water, bathing or showering. Participants were instructed to maintain their usual sleep habits and general daily activity patterns during the monitoring period. At the conclusion of the recording period, actigraph data were wirelessly downloaded to a study computer using a Nordic 2.4 GHz ANT transceiver, which was then analyzed using Fatigue Science software (16 Hz sampling rate: Readiband™, Fatigue Science, Vancouver). The raw activity scores were translated to sleep-wake scores based on computerized scoring algorithms. The five measures obtained from the actigraphy device and software that were used as sleep indices are described in Table 1.

Table 1

Definitions of each sleep variable measures using the Fatigue Science, Readiband™ actigraph.

Sleep indices	Units	Description
Total Sleep Time (TST)	Minutes	Total time spent asleep
Sleep Efficiency (SE)	%	Total time in bed divided by total sleep time
Total Time in Bed (TTB)	Minutes	Total time spent in bed
Sleep Latency (SL)	Minutes	Time taken for sleep onset
Wake Episodes per Night (WE)	Number count	Total number of awakenings per night
Wake After Sleep Onset (WASO)	Minutes	Time spent awake after sleep onset per night
Sleep Onset Variance (SOV)	Minutes	Variation in sleep onset time
Wake Variance (WV)	Minutes	Variation in wake time
Sleep Onset Time (SOT)	Time of day (p.m.)	Time fell asleep at night
Wake Time (WT)	Time of day (a.m.)	Time woken in morning

2.3. Statistical analysis

Simple group statistics are shown as means $\pm$ standard deviations unless stated otherwise. A students paired *t*-test was used to compare ACT1 and ACT2 using a Statistical Package for Social Science (V. 22.0, SPSS Inc., Chicago, IL), with statistical significance set at p $\leq$ 0.05. Inter-device agreements for ACT1 and ACT2 were examined using intraclass correlation coefficients (ICC) with 95% confidence intervals (95% CI) and interpreted as 0.90–1.00=*very high* correlation, 0.70–0.89=*high* correlation, 0.50–0.69=*moderate* correlation, 0.26–0.49=*low* correlation and 0.00–0.25=*little*, *if any* correlation [15]. The mean differences and upper and lower limits of agreement (1.96 standard deviations or 95% of a normally distributed population) between devices were determined in absolute values for TST, SL and SE. Between-device typical error of measurement (TEM) was determined using an excel spreadsheet [16] and are presented as a coefficient of variation percentage (CV%) and as absolute values. Similar to Werner et al. [17], we defined an apriori difference between the 2 devices of $\leq$ 30 min satisfactory for TST, with a difference < 5% for SE satisfactory.

3. Results

There were no significant differences between devices (ACT1 and ACT2) for any of the measured sleep variables (p > 0.05, Table 2). There was a mean difference between devices of -2.1 ± 13.4 min over the 77 nights of data for TST. This difference was associated with a *very high* correlation and a low TEM (9.5 min) and CV (2.3%) between devices (Table 3).

Table 2

Mean $\pm$ SD values for both devices (ACT1 and ACT2) for all measured sleep variables and p-values for each comparison.

	ACT1	ACT2	P-Value
Total Sleep Time (min)	461.6 $\pm$ 86.6	459.5 $\pm$ 87.9	0.20
Sleep Efficiency (%)	83.0 $\pm$ 8.9	83.2 $\pm$ 8.9	0.73
Sleep Latency (min)	21.9 $\pm$ 20.0	21.7 $\pm$ 19.6	0.79
Total Time in Bed (min)	564.1 $\pm$ 98.7	563.2 $\pm$ 99.0	0.55
Wake After Sleep Onset (min)	11.7 $\pm$ 8.0	12.2 $\pm$ 8.4	0.32
Wake Episodes (No. per night)	3.5 $\pm$ 2.5	3.6 $\pm$ 2.6	0.72
Sleep Onset Variance (min)	-0.8 ± 74.0	-2.3 ± 75.1	0.13
Wake Variance (min)	-3.1 ± 48.4	-2.6 ± 47.3	0.37
Sleep Onset Time (time of day)	22:47 $\pm$ 0:49	22:48 $\pm$ 0:49	0.76
Wake Time (time of day)	7:03 $\pm$ 0:52	7:02 $\pm$ 0:50	0.41

Table 3
Typical error of measurement (TEM) expressed in raw values and as a coefficient of variation (CV%), mean difference, range of mean difference and intra-class correlation (ICC) for each sleep variable between ACT1 and ACT2.

	TEM (95% CL)	CV% (95% CL)	Mean difference (± SD)	Range of mean difference (1.96xSD)	ICC (± 95% CL)
Total Sleep Time (min)	9.5 8.2–11.3	2.3 2.0–2.8	−2.1 + 13.4	−28.8 to 24.7	0.99 0.98–0.99 *very high*
Sleep Efficiency (%)	2.4 2.0–2.9	NA	0.2 + 3.4	−6.2 to 6.6	0.93 0.89–0.96 *very high*
Sleep Latency (min)	3.8 3.2–4.6	32.5 26.9–41.1	−0.2 + 5.4	−10.8 to 10.4	0.97 0.94–0.98 *very high*
Total Time in Bed (min)	8.5 7.3–10.3	1.5 1.3–1.8	−0.9 + 12.1	−25.0 to 23.3	0.99 0.99–1.00 *very high*
Wake After Sleep Onset (min)	3.3 2.8–3.9	37.8 31.5–47.4	0.6 + 4.6	−8.6 to 9.7	0.85 0.76–0.90 *high*
Wake Episodes (No. per night)	1.2 1.0–1.4	41.8 34.6–52.6	0.1 + 1.6	−3.2 to 3.3	0.80 0.70–0.87 *high*
Sleep Onset Variance (min)	5.4 4.6–6.6	31.8 24.7–44.7	−1.5 + 7.7	−16.8 to 13.9	0.99 0.99–1.00 *very high*
Wake Variance (min)	3.6 3.1–4.3	15.0 11.8–20.5	0.5 + 5.1	−21.9 to 25.5	0.99 0.99–1.00 *very high*

SE resulted in a TEM between devices of 2.4%, which was associated with an ICC of 0.93–*very high* (Table 3). SL, TTB, SOV and WV also resulted in *very high* correlations between devices and a mean difference of < 1.5 min (Table 3). Comparison between these devices for these variables also resulted in TEM's of < 8.5 min (Table 3).

The remaining variables; WASO and WE, resulted in *high* correlations between devices, with TEM values of 3.3 min and 1.2 (wake episodes), respectively (Table 3).

Level of agreement plots showing ± 95% limits of agreement between ACT1 and ACT2 for TST, SL and SE are displayed in Fig. 1.

4. Discussion

The current study was the first to determine the inter-device reliability of a commercially available automatic-scoring actigraph in healthy adult participants by wearing two Readiband™ devices simultaneously. The correlation between devices of the same brand (Readiband™) was *very high* for the most important sleep variables of total sleep time, sleep latency and sleep efficiency, with no significant differences in any of the measured sleep variables between devices. All differences between the two devices in the current study are deemed to be acceptable according to Werner et al. [17], who stated that a difference between 2 devices of ≤ 30 min can be deemed satisfactory for total sleep time, with a difference < 5% for sleep efficiency satisfactory. This suggests that researchers can use multiple devices of the same brand and model within the same study and obtain comparable results. The results in the current study are similar to those described by Dennis et al. [13], who studied the agreement between the Readiband™ and the Micro Mini-Motion Logger actigraphs and reported correlation coefficients of > 0.80 for total sleep time and total time in bed.

A major limitation of actigraphy methods that require manual sleep scoring, is that it introduces human error, as opposed to the automatic scoring device used in the current study. Indeed, proposed limitations of the use of actigraphy in sleep research are the inter-scorer reliability or the potential for intra-scorer bias. The use of a computerized scoring algorithm helps to account for both of these factors. Furthermore, the inter-scorer reliability of sleep data using the 'gold-standard' PSG, for determining sleep-wake has been studied extensively, with agreements between scorers ranging from 65% to 85% [18,19]. In a large comparative study investigating inter-scorer agreement between sleep laboratories, Norman et al. [20] reported that the level of agreement in sleep indices varies between scorers and between laboratories. Results showed that the level of agreement between laboratories is lower than what can be maintained between scorers within the same laboratory. The authors expressed caution when comparing sleep data scored by experts from separate laboratories [20]. This would suggest that even PSG for detecting sleep-wake, may have reliability issues as well as lacking ecological validity.
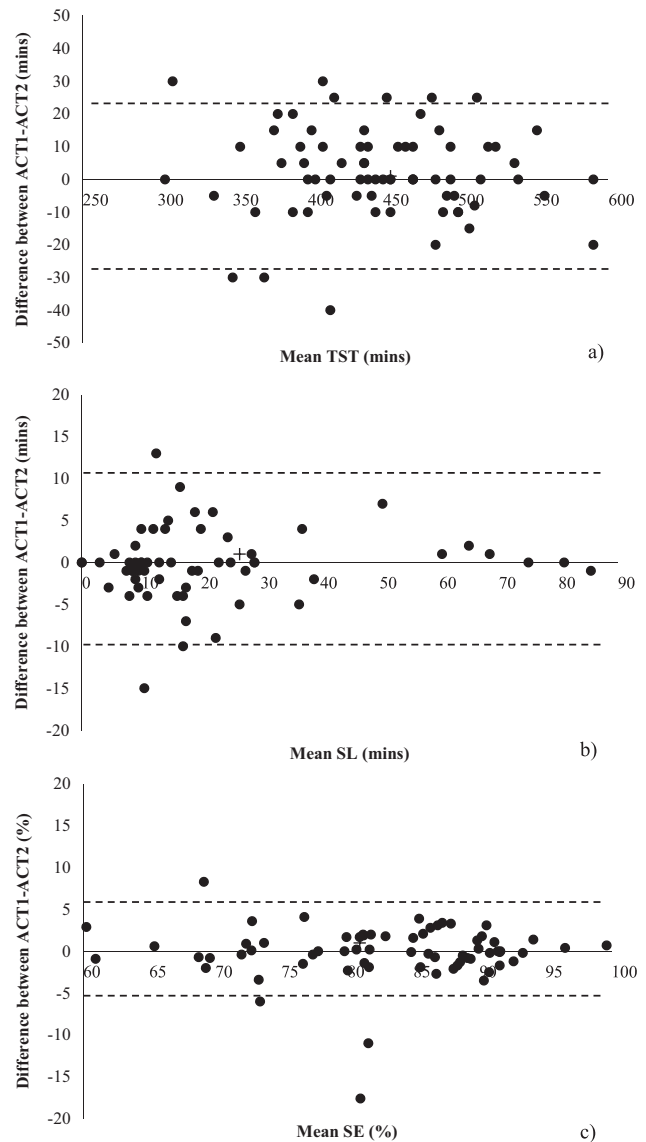
Fig. 1. Level of agreement plots showing ± 95% limits of agreement between ACT1 and ACT2 for a) total sleep time (TST); b) sleep latency (SL); c) sleep efficiency (SE).

The foreign environment experienced during PSG monitoring in a sleep laboratory may alter the normal sleeping patterns of an individual [21]. Indeed, differences between at-home and laboratory PSG monitoring have been shown to produce different results [22,23]. The un-natural laboratory environment, combined with the cost of assessment, accessibility of the laboratory and technicians, makes it difficult to attain for healthy sleepers wanting to monitor their sleep. For this reason, it has been suggested that

sleep monitoring at home, in a familiar environment may be the most appropriate for monitoring normal sleep patterns [23]. Even with at-home PSG monitoring, the comfort of sleeping with multiple electrodes and attachments must be questioned. While PSG monitoring in a laboratory may be important for diagnosing sleep disorders, the basic determination of sleep-wake cycles and sleep efficiency may be adequate for individuals wanting to know more about their sleep hygiene. Therefore, the importance of valid and reliable tools, such as actigraphy, may serve this purpose.

5. Conclusion

In summary, the results from the current study would suggest that the Readiband™ is a reliable tool for researchers aiming to use multiple devices of the same brand in sleep studies. These findings, along with the previously reported validity of the Readiband™ device, make it an easy to use, practical tool in both the clinical and research setting, without the need for qualified sleep scorers. The automatic scoring make this a novel device that can be used in many different fields to assess basic measures of sleep quality and quantity.

References

[1] van de Water A, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography–a systematic review. J Sleep Res 2011;20:183–200.

[2] Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? Epidemiology 2008;19:838–45.

[3] Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 2011;15:259–67.

[4] 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.

[5] Babin L, Lee S, Halko S, Boudreau A, George C. Determining sleep-wake activity using actiwatch. Sleep Res 1997;26:640.

[6] Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak C. The role of actigraphy in the study of sleep and circadian rhythms. American Academy of Sleep Medicine Review Paper. Sleep; 2003, 26. p. 342–92.

[7] Fowler P, Duffield R, Vaile J. Effects of simulated domestic and international air travel on sleep, performance, and recovery for team sports. Scand J Med Sci Sports 2015;25:441–51.

[8] Fullagar HH, Duffield R, Skorski S, White D, Bloomfield J, Kölling S, et al. Sleep, travel and recovery responses of national footballers during and following long-haul international travel. Int J Sports Physiol Perform 2016;11:86–95.

[9] Noor ZM, Smith AJ, Smith SS, Nissen LM. Feasibility and acceptability of wrist actigraph in assessing sleep quality and sleep quantity: a home-based pilot study in healthy volunteers. Health 2013;2013.

[10] Dennis J, Dawson B, Heasman J, Rogalski B, Robey E. Sleep patterns and injury occurrence in elite australian footballers. J Sci Med Sport 2016;19:113–6.

[11] Russell C, Caldwell J, Arand D, Myers L, Wubbels P, Downs H. Validation of the Fatigue Science Readiband™ Actigraph and Associated Sleep/WakeClassification Algorithms; 2011.

[12] Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12:1.

[13] Dennis J, Dawson B, Heasman J, Rogalski B, Robey E. Sleep patterns and injury occurrence in elite Australian footballers. J Sci Med Sport 2016:1–4.

[14] Meltzer LJ, Walsh CM, Traylor J, Westin A. Direct comparison of two new actigraphs and polysomnography in children and adolescents. Sleep 2012;35:159–66.

[15] Munro BH. Statistical Methods for Health Care Research, vol. 1. Philadelphia, USA: Lippincott Williams & Wilkins; 2005.

[16] Hopkins W. Analysis of reliability with a spreadsheet. A new view of statistics, Internet Society for Sport Science. Available online: ⟨http://sportsci.org/resource/stats/xrely.xls⟩; 2010 [accessed 10.06.15].

[17] 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.

[18] Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring. J Clin Sleep Med 2013;9:81–7.

[19] Penzel T, Zhang X, Fietze I. Inter-scorer reliability between sleep centers can teach us what to improve in the scoring rules. J Clin Sleep Med 2013;9:81–7.

[20] Norman RG, Pal I, Stewart C, Walsleben JA, Rapoport DM. Interobserver agreement among sleep scorers from different centers in a large dataset. Sleep 2000;23:901–8.

[21] Campbell AJ, Neill AM. Home set-up polysomnography in the assessment of suspected obstructive sleep apnea. J Sleep Res 2011;20:207–13.

[22] Edinger JD, Fins AI, SullivanJr RJ, Marsh GR, Dailey DS, Hope TV, et al. Sleep in the Laboratory and Sleep at home: Comparisons of older insomniacs and Normal sleepers. Sleep 1997.

[23] Iber C, Redline S, Gilpin AK, Quan SF, Zhang L, Gottlieb DJ, et al. Polysomnography performed in the unattended home versus the attended laboratory setting-Sleep Heart Health Study methodology. Sleep 2004;27:536–40.