Journal of Science and Medicine in Sport xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Journal of Science and Medicine in Sport

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

Original research

The validity of activity monitors for measuring sleep in elite athletes

Charli Sargent [a,*], Michele Lastella [a], Shona L. Halson [b], Gregory D. Roach [a]

[a] Appleton Institute for Behavioural Science, Central Queensland University, Australia
[b] Department of Physiology, Australian Institute of Sport, Australia

ARTICLE INFO

Article history:
Received 1 June 2015
Received in revised form
26 November 2015
Accepted 6 December 2015
Available online xxx

Keywords:
Sleep monitoring
Wrist activity monitor
Polysomnography
Agreement
Sensitivity
Specificity

ABSTRACT

Objectives: There is a growing interest in monitoring the sleep of elite athletes. Polysomnography is considered the gold standard for measuring sleep, however this technique is impractical if the aim is to collect data simultaneously with multiple athletes over consecutive nights. Activity monitors may be a suitable alternative for monitoring sleep, but these devices have not been validated against polysomnography in a population of elite athletes.
Design: Participants ($n = 16$) were endurance-trained cyclists participating in a 6-week training camp.
Methods: A total of 122 nights of sleep were recorded with polysomnography and activity monitors simultaneously. Agreement, sensitivity, and specificity were calculated from epoch-for-epoch comparisons of polysomnography and activity monitor data. Sleep variables derived from polysomnography and activity monitors were compared using paired *t*-tests. Activity monitor data were analysed using low, medium, and high sleep–wake thresholds.
Results: Epoch-for-epoch comparisons showed good agreement between activity monitors and polysomnography for each sleep–wake threshold (81–90%). Activity monitors were sensitive to sleep (81–92%), but specificity differed depending on the threshold applied (67–82%). Activity monitors underestimated sleep duration (18–90 min) and overestimated wake duration (4–77 min) depending on the threshold applied.
Conclusions: Applying the correct sleep–wake threshold is important when using activity monitors to measure the sleep of elite athletes. For example, the default sleep–wake threshold (>40 activity counts = wake) underestimates sleep duration by ~50 min and overestimates wake duration by ~40 min. In contrast, sleep–wake thresholds that have a high sensitivity to sleep (>80 activity counts = wake) yield the best combination of agreement, sensitivity, and specificity.

1. Introduction

The ability to monitor the amount and quality of sleep obtained by an elite athlete can be a useful tool for evaluating recovery from training and competition. The gold standard for measuring sleep is polysomnography (PSG) because it provides superior information about the structure and depth of sleep.[1] However, it is a complex procedure that involves the recording, analysis, and interpretation of multiple physiological parameters, including brain activity, eye movement, and muscle tone.[2] In addition, PSG is expensive, it can be invasive for participants, and it is impractical if the aim is to collect data simultaneously with multiple participants over consecutive nights. Recently, activity monitors have emerged as a potential alternative to PSG for monitoring the sleep of elite athletes.[3–5]

An activity monitor is a device—typically worn on the wrist—that contains an accelerometer. The device is based on the principle that people move more when they are awake compared with when they are asleep.[6] Unlike PSG, an activity monitor can be worn with little inconvenience as it is small and light and requires no intervention from the wearer. A number of studies have assessed the validity of activity monitors for measuring sleep by comparing sleep/wake durations determined by PSG with sleep/wake durations determined by activity monitors. In healthy adults, high correlations have been reported for sleep duration (i.e., 0.89–0.98)[6–8] and moderate-to-high correlations have been reported for wake time within sleep (i.e., 0.36–0.85).[7–9] More recently, validation of activity monitors has moved beyond comparing sleep/wake durations, toward considering epoch-for-epoch matches with PSG. Agreement between PSG and activity monitors is high, with rates ranging from 77 to 91%.[6,10,11] The ability of

* Corresponding author.
E-mail address: charli.sargent@cqu.edu.au (C. Sargent).

http://dx.doi.org/10.1016/j.jsams.2015.12.007
1440-2440/

activity monitors to detect sleep is also high (91–96%), however their capacity to detect wake is low (38–54%).[6,10,11]

In general, wrist activity monitors appear to be less accurate in populations exhibiting short or fragmented sleep.[9,11,12] For example, activity monitors tend to overestimate sleep in people who lie quietly in bed for long periods of time.[13] Conversely, activity monitors tend to underestimate sleep in people who exhibit high levels of movement during light sleep.[13] This raises a potential issue with the use of activity monitors for measuring sleep in elite athletes. Several studies indicate that elite athletes obtain less sleep than members of the general population[5,14] and/or exhibit greater movement and fragmentation during sleep.[15,16] Given the sleep characteristics of elite athletes, it is important to establish how well activity monitors detect sleep and wake in this population.

2. Methods

Sixteen endurance-trained cyclists (age: 19.3 ± 1.5 y; BMI: 21.6 ± 1.9 kg/m^2; mean ± SD) volunteered to participate in the study. Cyclists were competitive at a national level and took part in the study at the start of the domestic cycling season. Cyclists were not screened for sleep disorders prior to participation, but none of them reported a clinical diagnosis of a sleep disorder at the time of the study. Written informed consent was obtained from participants or their guardians. Ethics approval was granted by the Australian Institute of Sport Human Research Ethics Committee (approval number 20111202).

Data were collected as part of a larger study that examined the impact of cold water immersion on recovery.[17] Participants in the study completed a 6-wk training camp, which consisted of daily road rides and ergometer tests. On each day of the training camp, all participants completed the same type and duration of training as each other. Participants lived in shared apartments at the training facility and slept in their own private bedroom. Participants' sleep was assessed simultaneously using PSG and activity monitors on nine occasions: one night during the first week of the camp, one night at the end of the second week of the camp, two nights at the end of weeks 3 and 4, and one to two nights at the end of weeks 5 and 6 of the camp (i.e., 144 potential recordings). To determine the minimum number of observations required for comparison between PSG and activity monitors, the approach outlined by Bland and Altman was adopted.[18] Bland and Altman (1986) recommend a minimum of 100 observations to ensure that the confidence intervals for the 95% limits of agreement are narrow (i.e., $\pm 0.34 \times$ SD). Based on previous estimates of the difference between PSG and activity monitors,[10] 144 observations would give confidence intervals for the 95% limits of agreement for total sleep time of 14.5 min.

On sleep monitoring nights, participants were provided 9.5 hr in bed. Sleep recordings began at ~22:00 h when the bedroom lights were switched off and ended at ~07:30 h when the bedroom doors were opened. In the 30 min prior to bedtime, electrodes were applied to participants' heads and faces. Once all monitoring equipment was in place, the lights were switched off. The signals from each PSG device were transmitted to a laptop computer located in an adjacent room to be monitored overnight by a technician. Participants were instructed to remain in bed until woken at ~07:30 h, but were permitted to use the bathroom during the night.

Sleep was recorded using PSG equipment (Compumedics, Melbourne, Australia) with a standard montage of electrodes. The montage included two electroencephalograms (C4-M1, C3-M2), two electrooculograms (left and right outer canthus), and a submental electromyogram. Sleep records were blinded and analysed in 30-s epochs by a trained technician. Sleep stages were manually scored in accordance with established criteria.[19] The following variables were calculated from each record: sleep onset latency, sleep duration, wake after sleep onset, and sleep efficiency. In general, there is good agreement between attended PSG in the field and attended PSG in the laboratory with respect to signal quality and derived variables.[20]

Participants wore an activity monitor (Philips Respironics, Bend, USA) on their non-dominant wrist. The activity monitors were configured to sum and store data in 1-min epochs based on activity counts from a piezoelectric accelerometer with a sensitivity of 0.01 g and a sampling rate of 32 Hz. The estimation of sleep/wake duration from the activity monitors was a four-step process: (1) downloading the data from the activity monitors; (2) identifying time in bed using the 'lights out' and 'lights on' clock times recorded in the PSG record; (3) applying an algorithm to identify sleep onset and sleep offset; and (4) applying an algorithm to assign epochs as 'sleep' or 'wake'. This process was conducted using Actiware™-Sleep v3.1 software in conjunction with the Actiware™-Sleep scoring algorithm (Mini Mitter Co., Inc., Sunriver, USA). Sleep/wake estimates using the Actiware™-Sleep scoring algorithm are based on a comparison of activity counts with two activity thresholds. The first, called the immobility-mobility threshold, is used to identify sleep onset and sleep offset. The threshold assigns epochs within a given time in bed as either immobile (activity count <4) or mobile (activity count ≥4). The second threshold, called the sleep–wake threshold, assigns all epochs between sleep onset/offset as sleep or wake by comparing activity counts for the epoch in question to a user-defined sleep–wake threshold. Three sleep–wake thresholds were applied to this data set: low threshold—above 20 activity counts is scored as wake (low sensitivity to sleep); medium threshold—above 40 activity counts is scored as wake; and high threshold—above 80 activity counts is scored as wake (high sensitivity to sleep). The following variables were calculated from the data: sleep onset latency, sleep duration, wake after sleep onset, and sleep efficiency.

Epoch-for-epoch comparisons between activity monitors and PSG were performed to determine how well activity monitors detected sleep and wake. The percentage of epochs in agreement (i.e., percentage of all PSG sleep and wake epochs correctly detected by activity monitors), sensitivity (i.e., percentage of PSG sleep epochs correctly detected by activity monitors), and specificity (i.e., percentage of PSG wake epochs correctly identified by activity monitors) were calculated. In addition, Cohen's kappa (κ) was used to quantify better-than-chance agreement between activity monitors and PSG.[21] Agreement was interpreted using the guidelines established by Landis and Koch (1977).

To enable epoch-for-epoch comparisons, the 1-min epochs from activity monitors were split into two identical 30-s epochs and aligned with epochs from the corresponding PSG records.[6] Prior to the start of the study, clock time was manually synchronised on all devices (i.e., activity monitors, laptop computers, PSG devices). At the end of the study, there were some minor differences in clock time between devices. This was due to a small drift in clock time (~2–3 s per day) on the laptop computers. To minimise the effect of this difference, offset adjustments of three min in either direction were compared and the offset with the highest percentage agreement was selected for each sleep period.[6]

Agreement between PSG and activity monitors for sleep summary variables at each sleep–wake threshold were also compared using the limits of agreement method for repeated measurements.[23] For each sleep summary variable, the difference between each pair of measurements obtained from PSG and activity monitors were plotted against the mean of each pair of measurements. The mean difference between the two measurements (the bias) and the 95% limits of agreement (bias $\pm 1.96 \times$ SD) were also plotted. Plots were examined for heteroscedasticity and proportional bias using the Breusch-Pagan test and ordinary least squares regression, respectively. In cases where heteroscedasticity and/or

Please cite this article in press as: Sargent C, et al. The validity of activity monitors for measuring sleep in elite athletes. *J Sci Med Sport* (2015), http://dx.doi.org/10.1016/j.jsams.2015.12.007

Table 1
Epoch-for-epoch comparisons between activity monitors and polysomnography.

Sleep-wake threshold	Agreement (%)	Sensitivity (%)	Specificity (%)	Kappa (κ)
Low threshold (20)	80.9	80.6	82.3	0.38
Medium threshold (40)	86.5	87.5	77.1	0.46
High threshold (80)	89.5	91.9	67.0	0.50

proportional bias were present, the bias and 95% limits of agreement were adjusted accordingly.[24]

Paired *t*-tests were performed to examine differences in sleep summary variables between PSG and activity monitors at each sleep–wake threshold. Bonferroni corrections were applied to reduce the likelihood of committing type I errors. The critical *p* value for each *t*-test was 0.017 (i.e., .05/3). Prior to analyses, sleep summary variables were examined for normality using the Kolmogorov–Smirnov normality test.

3. Results

Originally, data collection was scheduled to occur on 10 occasions with each participant during the protocol (i.e., 160 potential recordings in total). However, the training camp was reduced by two days in the final week, resulting in 144 potential recordings in total. Of these 144 potential recordings, 122 were completed successfully and 22 were excluded from the analyses due to technical failure of the PSG recording device ($n = 4$); participant illness ($n = 1$); activity monitor not worn by the participant on the recording night ($n = 7$); and more than 10 epochs of unscorable PSG ($n = 10$). The numbers of recordings collected with participants were as follows: nine recordings ($n = 4$), eight recordings ($n = 8$), seven recordings ($n = 2$), six recordings ($n = 1$), and two recordings ($n = 1$).

The epoch-for-epoch agreement, sensitivity, specificity, and Cohen's kappa coefficients for the low, medium and high sleep–wake thresholds are presented in Table 1. The high sleep–wake threshold yielded the highest agreement, whereas the low sleep–wake threshold yielded the lowest agreement. The high sleep–wake threshold was the most sensitive to sleep, but the least specific in detecting wake. Conversely, the low sleep–wake threshold was the least sensitive to sleep, but yielded the highest specificity for wake. Cohen's kappa coefficients for the high and medium sleep–wake thresholds were similar and less than 0.6, which is considered "moderate" agreement.[22] Cohen's kappa coefficient was less than 0.4 for the low sleep–wake threshold, which is considered "fair" agreement.[22]

The level of agreement between PSG and activity monitors for each sleep summary variable is depicted in Fig. 1. The mean and standard deviation for each sleep summary variable is presented in Table 2. All sleep summary variables were normally distributed. Activity monitors significantly underestimated sleep duration and overestimated wake after sleep onset, regardless of the sleep–wake threshold that was applied. There was proportional bias at the low sleep–wake threshold ($R = .405$; $p < .0001$) and the medium sleep–wake threshold ($R = .251$; $p = .0053$) for sleep duration, as well as at the low sleep–wake threshold ($R = .486$; $p < .0001$) and the medium sleep–wake threshold ($R = .258$; $p = .0041$) for wake after sleep onset. That is, when these two thresholds were applied, activity monitors progressively underestimated sleep duration as average sleep duration decreased, and progressively overestimated wake after sleep onset as average wake after sleep onset increased. Activity monitors underestimated or overestimated sleep efficiency depending on the sleep–wake threshold that was applied. The low and medium sleep–wake thresholds significantly underestimated sleep efficiency, whereas the high sleep–wake threshold significantly overestimated sleep efficiency. Proportional bias was present for the low sleep–wake threshold ($R = .360$; $p < .0001$) and the high sleep–wake threshold ($R = .306$; $p = .0006$). Activity monitors progressively underestimated sleep efficiency as average sleep efficiency decreased when the low sleep–wake threshold was applied, whereas activity monitors progressively overestimated sleep efficiency as average sleep efficiency decreased when the high sleep–wake threshold was applied. Sleep latency did not differ between activity monitors and PSG. Since sleep latency is calculated using the immobility-mobility threshold, it remained unchanged, irrespective of the sleep–wake threshold applied.

4. Discussion

In the current study, the ability of activity monitors to measure the sleep of elite athletes was compared with PSG—the gold standard for monitoring sleep. The results indicate that overall agreement rates between PSG and activity monitors at each sleep–wake threshold were very high (81–90%) and were comparable to rates previously reported in validation studies with healthy adults (87–91%).[6,10,11] Based on this evidence, activity monitors are a valid alternative to PSG for measuring sleep in elite athletes. However, high rates of agreement between PSG and activity monitors are not entirely unexpected given that healthy adults spend most of the time in bed asleep.[13] In the present study, the average PSG sleep efficiency was 90%. Simply scoring all of the time in bed as sleep would result in an agreement rate of 90% between activity monitors and PSG. For this reason, additional measures such as sensitivity (i.e., ability to detect sleep), specificity (i.e., ability to detect wake), and Cohen's kappa coefficient (i.e., better than chance agreement) are required to provide a complete assessment of the validity of activity monitors for measuring the sleep of elite athletes.

In healthy populations, activity monitors are excellent at detecting sleep. Kosmadopoulos et al.[10] assessed the sensitivity of activity monitors using PSG as the gold standard for comparison. Data were collected with 22 participants over two nights in a sleep laboratory. On average, activity monitors had a sensitivity of 96%. This is not surprising given that the data was collected with healthy young adults under ideal sleeping conditions (i.e., cool, quiet, dark bedrooms). Slightly lower values for sensitivity have been reported for airline crew during a layover sleep (91%),[6] middle-aged women (88%),[25] and older adults (91%).[11] In the present study, sensitivity values obtained from athletes were high and ranged between 81 and 92% depending on the sleep–wake threshold that was applied. As expected, the high sleep–wake threshold was the most sensitive to sleep, whereas the low sleep–wake threshold was the least sensitive to sleep.

In contrast to the high values reported for sensitivity, activity monitors typically yield relatively low values for specificity. In line with previous validation studies, the ability of activity monitors to detect wake in this group of participants was much lower than the ability to detect sleep.[6,10,11] Interestingly, the values reported for specificity in the present study (67–82%) are higher than values previously reported for healthy adults (38–54%).[6,10,11] It is not clear from the present study why activity monitors are better at detecting wake in athletes compared with healthy adults. Put simply, when athletes were awake according to PSG, their level of movement was sufficiently high for it to be scored as wake by the activity monitor.

Please cite this article in press as: Sargent C, et al. The validity of activity monitors for measuring sleep in elite athletes. *J Sci Med Sport* (2015), http://dx.doi.org/10.1016/j.jsams.2015.12.007

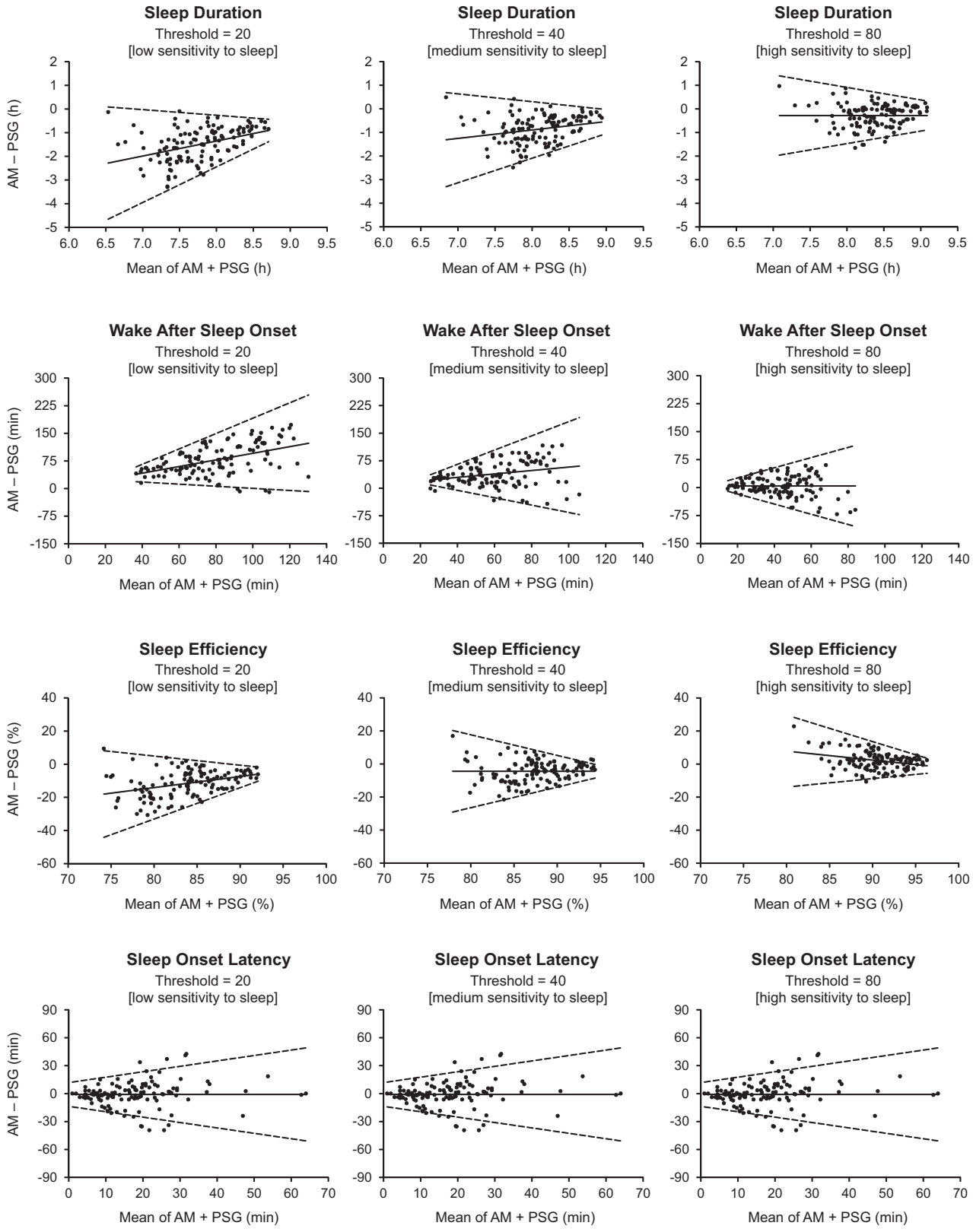

Fig. 1. Bland–Altman plots comparing PSG with activity monitors at the low (left column), medium (middle column), and high (right column) sleep–wake thresholds for sleep duration, wake after sleep onset, sleep efficiency, and sleep onset latency. The *x*-axes represent the mean of the values obtained from activity monitors and PSG and the *y*-axes represent the difference between the values, such that positive values indicate an overestimation by activity monitors relative to PSG, and negative values indicate an underestimation by activity monitors relative to PSG. Solid horizontal lines indicate the mean bias from PSG, and broken lines indicate the 95% limits of agreement (±1.96 SDs). AM—activity monitors, PSG—polysomnography.

Please cite this article in press as: Sargent C, et al. The validity of activity monitors for measuring sleep in elite athletes. *J Sci Med Sport* (2015), http://dx.doi.org/10.1016/j.jsams.2015.12.007

Table 2
Comparison of sleep parameters determined by polysomnography and activity monitors.

Sleep parameter	Value (mean ± SD)	Polysomnography vs. Activity Monitor		
		Difference (mean ± SD)	*t* (df 121)	*p*
Sleep duration (h)				
Polysomnography	8.5 ± 0.4	n/a	n/a	n/a
Activity Monitor Low Threshold (20)	7.0 ± 0.7	−1.5 ± 0.7	23.18	<0.0001
Activity Monitor Medium Threshold (40)	7.6 ± 0.6	−0.9 ± 0.6	16.17	<0.0001
Activity Monitor High Threshold (80)	8.2 ± 0.5	−0.3 ± 0.5	6.41	<0.0001
Wake after sleep onset (min)				
Polysomnography	41.0 ± 22.2	n/a	n/a	n/a
Activity Monitor Low Threshold (20)	118.5 ± 37.6	77.5 ± 42.0	−20.38	<0.0001
Activity Monitor Medium Threshold (40)	80.0 ± 28.8	39.0 ± 34.5	−12.49	<0.0001
Activity Monitor High Threshold (80)	45.1 ± 18.7	4.1 ± 27.2	−1.68	0.095
Sleep efficiency (%)				
Polysomnography	89.6 ± 4.7	n/a	n/a	n/a
Activity Monitor Low Threshold (20)	78.0 ± 6.9	−11.6 ± 8.0	15.94	<0.0001
Activity Monitor Medium Threshold (40)	85.2 ± 5.3	−4.4 ± 6.7	7.32	<0.0001
Activity Monitor High Threshold (80)	91.6 ± 3.4	2.0 ± 5.5	−4.11	<0.0001
Sleep latency (min)				
Polysomnography	18.3 ± 12.6	n/a	n/a	n/a
Activity Monitor Low Threshold (20)	17.3 ± 14.2	−0.9 ± 14.0	0.75	0.457
Activity Monitor Medium Threshold (40)	17.3 ± 14.2	−0.9 ± 14.0	0.75	0.457
Activity Monitor High Threshold (80)	17.3 ± 14.2	−0.9 ± 14.0	0.75	0.457

Specifically, if the athletes were awake, they were not lying still. It should be noted that the athletes in the present study were examined during a 6-wk period of intensified training. There is some evidence to suggest that during periods of intensified training, athletes exhibit less immobility during sleep[26] and more movement[16] compared with sleep during a phase of normal or reduced training. In the present study, athletes may have experienced some muscle soreness and as a result, were awake and moving during sleep periods in an attempt to get comfortable.

Cohen's kappa coefficients were 0.38, 0.46, and 0.50 for the low, medium, and high sleep–wake thresholds, respectively. The level of agreement between PSG and both the medium and high sleep–wake thresholds is considered moderate, whereas the level of agreement for the low sleep–wake threshold is considered fair. Similar kappa coefficients for PSG and activity monitors have been reported previously.[6,10] As noted by Signal et al.[6], such findings indicate that a large proportion of the agreement between PSG and activity monitors is due to chance alone. More specifically, although similar proportions of sleep periods are categorised as sleep and wake by PSG and activity monitors, the epochs scored as sleep and wake by the two methods are not necessarily the same.

In the present study, activity monitors underestimated sleep duration and overestimated wake after sleep onset in athletes, regardless of the sleep–wake threshold applied. The high sleep–wake threshold produced the smallest differences in sleep duration and wake after sleep onset compared with PSG, whereas the low sleep–wake threshold produced the largest differences compared with PSG. The results are in contrast to those reported by Kosmadopoulos et al.[10], who found that activity monitors overestimate sleep duration and underestimate wake after sleep onset in healthy young adults. Activity monitors underestimate sleep when the amount of movement in the light stage of sleep is so great that it is scored as wake.[13] This suggests that athletes exhibit different movement behaviours during sleep compared with healthy adults. Indeed, a number of studies have shown that athletes—particularly during heavy training phases—exhibit high levels movement and fragmentation during sleep.[15,16]

When considering the results of the current study, it must be recognised that it was conducted under controlled conditions with a sample of young elite male cyclists who were undertaking a program of intensified training. As such, caution should be shown in generalising the results beyond the conditions described in the present study. In particular, data were collected under ideal conditions, with a researcher present to record the start and end times of all sleep periods and to ensure that participants were wearing an activity monitor. In most field-based studies, participants must press an event marker on the activity monitor to signify 'lights out' and 'lights on' and/or record the information in a sleep diary. Collecting meaningful sleep/wake data with activity monitors therefore relies heavily on an individual's ability to accurately record time in bed and to wear the activity monitor.

5. Conclusion

The results from the present study indicate that activity monitors are a valid alternative to PSG for measuring the sleep of elite athletes. Overall, the high sleep–wake threshold provided the best combination of agreement, sensitivity, and specificity and provided the best estimates of sleep duration, wake duration and sleep efficiency. The default sleep–wake threshold used in most activity monitor scoring algorithms is the medium threshold (i.e., 40). This threshold has been applied in almost all of the studies that have examined the sleep of elite athletes[3,27,28]—although in some cases the threshold has not been reported.[15,26,29] The results of the present study suggest that at a medium sleep–wake threshold, sleep duration is underestimated by ~50 min and wake duration is overestimated by ~40 min. Thus, the choice of sleep–wake threshold should be considered, and at the very least reported, because it has important implications for the interpretation of findings related to the sleep/wake duration of elite athletes.

Practical implications

- Activity monitors are a valid alternative to polysomnography for measuring the sleep of elite athletes.
- Choosing the correct sleep–wake threshold is important when using activity monitors to measure the sleep of elite athletes. Thresholds that have a high sensitivity to sleep (i.e., 80) yield the best combination of agreement, sensitivity, and specificity.
- Thresholds that are moderately sensitive to sleep (i.e., 40) will underestimate sleep duration by ~50 min and overestimate wake duration by ~40 min.

Please cite this article in press as: Sargent C, et al. The validity of activity monitors for measuring sleep in elite athletes. *J Sci Med Sport* (2015), http://dx.doi.org/10.1016/j.jsams.2015.12.007

Acknowledgements

The authors would like to thank the Australian Institute of Sport and the Australian Research Council for their financial support; and the athletes for their cooperation.

References

1. Keenan SA, Hirshkowitz M. Monitoring and staging human sleep, in *Principles and Practice of Sleep Medicine*, 5th ed., Kryger MH, Roth T, Dement WC, editors, St. Louis, Elsevier Saunders, 2011, [Chapter 141].
2. Keenan SA. Polysomnographic technique: an overview, in *Sleep Disorders Medicine*, Chokroverty S, editor, Stoneham, Butterworth-Heinemann, 1994, [Chapter 6].
3. Robey E, Dawson B, Halson S et al. Sleep quantity and quality in elite youth soccer players: a pilot study. *Eur J Sport Sci* 2014; 14(5):410–417.
4. Roach GD, Schmidt WF, Aughey RJ et al. The sleep of elite athletes at sea level and high altitude: a comparison of sea level natives and high altitude natives (ISA3600). *Br J Sports Med* 2013; 47(Suppl. 1):i114–i120.
5. Sargent C, Lastella M, Halson SL et al. The impact of training schedules on the sleep and fatigue of elite athletes. *Chronobiol Int* 2014; 31(10):1160–1168.
6. Signal TL, Gale J, Gander PH. Sleep measurement in flight crew: comparing actigraphic and subjective estimates to polysomnography. *Aviat Space Environ Med* 2005; 76(11):1058–1063.
7. Kripke DF, Mullaney DJ, Messin S et al. Wrist actigraphic measures of sleep and rhythms. *Electroen Clin Neuro* 1978; 44(5):674–676.
8. de Souza L, Benedito-Silva AA, Nogueira ML et al. Further validation of actigraphy for sleep studies. *Sleep* 2003; 26(1):81–85.
9. Mullaney DJ, Kripke DF, Messin S. Wrist-actigraphic estimation of sleep time. *Sleep* 1980; 3(1):83–92.
10. Kosmadopoulos A, Sargent C, Darwent D et al. Validating alternatives to polysomnography (PSG): Wrist actigraphy and self-administered partial-PSG. *Behav Res Methods* 2014; 46(4):1032–1041.
11. Pacquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. *Sleep* 2007; 30(10):1362–1369.
12. 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(5):389–396.
13. Tryon WW. Issues of validity in actigraphic sleep assessment. *Sleep* 2004; 27(1):158–165.
14. Sargent C, Halson SL, Roach GD. Sleep or swim? Early morning training severely restricts the amount of sleep obtained by elite swimmers. *Eur J Sport Sci* 2014; 14(Suppl. 1):S310–S315.
15. Leeder J, Glaister M, Pizzoferro K et al. Sleep duration and quality in elite athletes measured using wristwatch actigraphy. *J Sports Sci* 2012; 30(6):541–545.
16. Taylor SR, Rogers GG, Driver HS. Effects of training volume on sleep, psychological, and selected physiological profiles of elite female swimmers. *Med Sci Sports Exerc* 1997; 29(5):688–693.
17. Halson SL, Bartram J, West N et al. Does hydrotherapy help or hinder adaptation to training in competitive cyclists? *Med Sci Sports Exerc* 2014; 46(8):1631–1639.
18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; i:307–310.
19. Iber C, Ancoli-Israel S, Chesson A et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, 1st ed. Westchester, American Academy of Sleep Medicine, 2007.
20. Mykytyn IJ, Sajkov D, Neill AM, McEvoy RD. Portable computerized polysomnography in attended and unattended settings. *Chest* 1999; 115:114–122.
21. Gale J, Signal TL, Gander PH. Statistical artefact in the validation of actigraphy. *Sleep* 2005; 28(8):1017–1018.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1):159–174.
23. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 2007; 17(4):571–582.
24. Ludbrook J. Confidence in Altman–Bland plots: a critical review of the method of differences. *Clin Exp Pharmacol Physiol* 2010; 37(2):143–149.
25. Jean-Louis G, Kripke DF, Cole RJ et al. Sleep detection with an accelerometer actigraph: comparisons with polysomnography. *Physiol Behav* 2001; 72(1–2):21–28.
26. Hausswirth C, Louis J, Aubry A et al. Evidence of disturbed sleep and increased illness in overreached endurance athletes. *Med Sci Sports Exerc* 2014; 46(5):1036–1045.
27. Lastella M, Roach GD, Halson SL et al. Sleep/wake behaviour of endurance cyclists before and during competition. *J Sports Sci* 2015; 33(3):293–299.
28. Lastella M, Roach GD, Halson SL et al. Sleep/wake behaviours of elite athletes from individual and team sports. *Eur J Sport Sci* 2014; 15(2):94–100.
29. Richmond L, Dawson B, Hillman DR et al. The effect of interstate travel on sleep patterns of elite Australian rules footballers. *J Sci Med Sport* 2004; 7(2):186–196.

Please cite this article in press as: Sargent C, et al. The validity of activity monitors for measuring sleep in elite athletes. *J Sci Med Sport* (2015), http://dx.doi.org/10.1016/j.jsams.2015.12.007