Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes

Fuller, K.L.; Juliff, L.E.; Gore, C.J.; et.al.

Fuller, K. L., Juliff, L. E., Gore, C. J., Peiffer, J. J., & Halson, S. L. (2017). Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes. Journal of Science and Medicine in Sport, 20(8), 756–760. https://doi.org/10.1016/j.jsams.2016.11.021
Document Version: Author's Version

Published Version: https://doi.org/10.1016/j.jsams.2016.11.021

Title: Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes

Authors: Kate L Fuller, Laura Juliff, Christopher J Gore, Jeremiah J Peiffer, Shona L Halson

PII: S1440-2440(17)30257-8
DOI: http://dx.doi.org/doi:10.1016/j.jsams.2016.11.021
Reference: JSAMS 1451

To appear in: *Journal of Science and Medicine in Sport*

Received date: 12-8-2016
Revised date: 26-10-2016
Accepted date: 28-11-2016

Please cite this article as: Fuller Kate L, Juliff Laura, Gore Christopher J, Peiffer Jeremiah J, Halson Shona L.Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes.*Journal of Science and Medicine in Sport* http://dx.doi.org/10.1016/j.jsams.2016.11.021

Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes

Kate L Fuller [a], Laura Juliff [b], Christopher J Gore [a], Jeremiah J Peiffer [b], Shona L Halson [a]

[a] Department of Physiology, Australian Institute of Sport, Australia

[b] School of Psychology and Exercise Science, Murdoch University, Western Australia

Corresponding author:

Kate Fuller

Kate.fuller@ausport.gov.au

Word count: 2816

Abstract word count: 245

Tables: 1

Figures: 1

Abstract

Objectives: Actical® actigraphy is commonly used to monitor athlete sleep. The proprietary software, called Actiware®, processes data with three different sleep-wake thresholds (Low, Medium or High), but there is no standardisation regarding their use. The purpose of this

study was to examine validity and bias of the sleep-wake thresholds for processing Actical® sleep data in team sport athletes.

Design: Validation study comparing actigraph against accepted gold standard polysomnography (PSG).

Methods: Sixty seven nights of sleep were recorded simultaneously with polysomnography and Actical® devices. Individual night data was compared across five sleep measures for each sleep-wake threshold using Actiware® software. Accuracy of each sleep-wake threshold compared with PSG was evaluated from mean bias with 95% confidence limits, Pearson moment-product correlation and associated standard error of estimate.

Results: The Medium threshold generated the smallest mean bias compared with polysomnography for total sleep time (8.5 min), sleep efficiency (1.8%) and wake after sleep onset (-4.1 min); whereas the Low threshold had the smallest bias (7.5 min) for wake bouts. Bias in sleep onset latency was the same across thresholds (-9.5 min). The standard error of the estimate was similar across all thresholds; total sleep time ~25 min, sleep efficiency ~4.5%, wake after sleep onset ~21 min, and wake bouts ~8 counts.

Conclusions: Sleep parameters measured by the Actical® device are greatly influenced by the sleep-wake threshold applied. In the present study the Medium threshold produced the smallest bias for most parameters compared with PSG. Given the magnitude of measurement variability, confidence limits should be employed when interpreting changes in sleep parameters.

Keywords

polysomnography, validity, Actical®, accelerometry.

Introduction

Sleep is widely accepted as a critical component of the recovery process for an elite athlete.[1,2] As such, monitoring an athlete's sleep has become commonplace as sport scientists look for ways to improve sleep, recovery, and optimise performance. Monitoring sleep using the accepted gold standard method of polysomnography (PSG) is impractical for most athletes since it requires specialist equipment and staff to collect and analyse the data. Also, because PSG monitoring often requires the subject to sleep in a laboratory or setting outside their home environment, long term monitoring of an individual's sleep, or monitoring multiple athletes simultaneously is problematic. For these reasons, actigraphy has become a popular low-cost, non-invasive alternative for collecting sleep data of athletes. Worn on the wrist, actigraph monitors contain a multidirectional accelerometer that detects movements and employs software algorithms to distinguish sleep from wakefulness based on the level of movement.[3] These small devices can store several days and nights of data before downloading to a computer, allowing users to monitor multiple athletes over consecutive nights in any environment; home or away at competition.

The Actical® (Philips Respironics) is an actigraph commonly employed by sport scientists to monitor sleep behaviour in elite athletes.[2,4] Data from the Actical® device can be converted into a format which allows for processing with the Actiware® analysis software (Philips Respironics). This software uses algorithms to process data based on one of three Actiware® sleep-wake threshold settings (Low, Medium and High). Although the sleep-wake threshold algorithms were originally developed and validated with sleep disordered patients, the algorithms and Actical® devices have been validated on a range of populations including sleep disordered and healthy adults.[5-8] There is however, currently no standardised protocol regarding the use of different threshold settings.

Previous research studies investigating the sleep behaviour of elite athletes have used the Medium sleep-wake threshold, based on the work of other industry researchers

using this threshold setting.[2,4] Recently, researchers compared the validity of wrist actigraphy across all three Actiware® threshold settings in elite endurance cyclists.[1] Whilst good agreement was observed between activity monitors and PSG for each of the three sleep-wake thresholds (81-90%), the devices underestimated sleep duration and overestimated wake duration depending on which threshold was applied. In contrast to studies using the Medium threshold, Sargent et al. (2015) recommended the High sleep-wake threshold be employed when using Actical® actigraphy with elite cyclists.

Considering the widespread use of actigraphy with elite athletes, we sought to expand to work of Sargent et al. (2015) to include elite team-sport athletes. Due to the lack of standardisation of the sleep-wake threshold settings used to analyse Actical® data, the aim of this study was to examine the validity and potential bias of the three software thresholds compared with polysomnography. Also, given the way the actigraphy and PSG data is used in a practical setting, only time matched, overall night data values were used for comparison rather than an epoch to epoch analysis which has been used by previous researchers.[3,7]

Methods

Participants were 21 male elite team-sport athletes (age: 22.5 ±2.7y) from the premier Australian Rules Football League (10) and Australian Rugby League (11). Participants completed a Pittsburgh Scale for Evaluation of Sleep Quality questionnaire to establish inclusion in the study.[9] Exclusion criteria included; shift workers, participants on medication which could impact study results, parents with newborns, presence of primary sleep disorders, and consumption of more than five caffeine beverages per day. Informed consent was obtained from each participant and the study was approved by the Ethics Committees of Murdoch University and the Australian Institute of Sport.

Participants' sleep was assessed using PSG and concurrent actigraphy on four occasions. All athletes were in pre-season training at the time of the study. Data was

collected as part of an intervention sleep study which was a randomised, parallel group, single blind experimental design comparing neurofeedback to a sham group. Sixty-seven successful observations were recorded, after some recordings were excluded due to technical issues (n = 16) or participant illness (n = 1). All participants slept in their own bedroom within an apartment. Bedtimes and awakening times were ad libitum; however, the time when bedroom lights were turned off (bedtime) and on (awakening time) was noted. Clocks on the PSG and Actical® devices were synchronised to align the two recording devices. For both devices, the following time-matched, summary measures were collected and calculated for each night: sleep onset latency (SOL), total sleep time (TST), sleep efficiency (SE), wake after sleep onset (WASO) and number of wake bouts. SOL was calculated as time from lights out until the onset of sleep. TST calculated as the total duration of epochs scored as sleep between lights off and on; SE was defined as the percentage of time asleep between lights off and on; WASO was calculated as the number of minutes spent awake between sleep onset and final awakening; wake bouts was defined as the number of discreet wake periods experienced after sleep onset and before final awakening.

Polysomnography (Compumedics Siesta 802 system; Compumedics, Texas, USA) was recorded following the technical specifications of the American Academy of Sleep Medicine manual for the scoring of sleep and associated events.[10] Polysomnograph montage included; four electroencephalogram (EEG) electrodes according to the international 10-20 electrode placement system (F4-A1, C4-A1, C3- A2, O2-A1); two electrooculogram electrodes (Left and Right eye); chin electromyogram (EMG1, EMG2) placed on the mentalis and submentalis; right and left anterior tibialis piezo EMG; thoracic and abdominal respiratory bands; pulse oximeter on the index finger of the non-dominant hand; oronasal airflow sensor; and a single modified lead 11 placement for electrocardiogram (ECG). Signals from each PSG system were stored in a data card within the system as well as transmitted to a laptop in an adjacent room where a researcher

monitored the signals throughout the night. All data was scored in 30 s epochs according to the American Academy of Sleep Medicine scoring criteria by a trained specialist, unaware of the participants' intervention condition.[10] The studies were reviewed according to the same criteria by a second sleep specialist blinded to the study design.

Actigraphy data were collected using Actical® Z series activity monitors (Actical® Z series part number 198-0200-03; Philips Respironics, Oregon, USA) worn on the non-dominant wrist. Each activity monitor contains a 3-axis piezoelectric accelerometer sampled at 32 Hz, which generates a voltage when it undergoes a change in acceleration. Sensitive to movements in the 0.5Hz to 3Hz range, the Actical® device records the mean of activity, or movement, sampled each second with the means summed to create activity counts for each 1 minute epoch. Actiware® 5.61 activity and sleep analysis software (Mini Mitter Philips/Respironics, Oregon, USA) was used to set up, download and process the data. An activity score was generated for each epoch as a weighted average of the activity count for the current epoch and that of the surrounding epochs (±2 min).[11] Data from the Actical® was assessed as sleep or wake based on whether or not the activity scores exceeded a set wake sensitivity threshold. For the purpose of this study, data from the actigraph devices was processed for all three wake sensitivity thresholds; Low (> 20 activity counts scored as wake), Medium (> 40 activity counts scored as wake), High (> 80 activity counts scored as wake). Time in bed was calculated using the 'lights off' and 'lights on' times recorded on the PSG system. SOL was calculated as the time from lights out until sleep onset and as such, the results for this sleep parameter do not change across the three sleep-wake thresholds.

In previous studies, agreement rates of epoch-by-epoch data have been used to compare PSG and actigraphy, however this technique is not considered fully appropriate as a measure of concordance.[3,12] For this reason, and due to the way PSG and actigraphy data are reported in a practical setting, time matched overall night data values (TST, SE, SOL, WASO and wake bouts) for PSG and Actical® threshold sensitivities (Low, Medium and High) were used for comparison.

Accuracy of each sleep-wake threshold compared with PSG was evaluated by determining mean bias and corresponding ninety-five percent confidence limits (95% CL), as well as the Pearson moment-product correlation and associated standard error of estimate (SEE). Magnitudes of the Pearson correlations were interpreted using the descriptors of Hopkins, low (0.1-0.3), moderate (0.3-0.5), large (0.5-0.7), very large (0.7-0.9).[13] Bland-Altman plots of absolute error in Actical® from the mean of the PSG and Actical® data across all sleep parameters were conducted.[14] The bias, correlation and Bland-Altman analyses were conducted with GraphPad Prism version 6.01 (GraphPad Software, La Jolla California USA), with magnitudes from specialised Excel spreadsheets.[15]

Results

Data comparing the three sleep-wake threshold settings on the Actical® devices to PSG are presented in Table 1. Bland-Altman plots comparing PSG to Actical® for each sleep-wake threshold are depicted in Figure 1. Compared to PSG, the Actical® devices underestimated total sleep time and sleep efficiency when the Low threshold was applied, but overestimated these measures on the Medium and High thresholds. For total sleep time, the SEE for all thresholds were similar and very large positive correlations to PSG were observed (r=0.78 to 0.85). Similarly, for sleep efficiency measures, the SEE for the three thresholds was almost identical, however when using the Low threshold, a low correlation to PSG was observed (r=0.27), and a moderate correlation was observed for the Medium and High thresholds (r=0.35 and 0.34).

The average amount of time athletes spent awake after sleep onset was overestimated by an average of 15.9 min using the Low threshold. Conversely, the Actical® devices underestimated wake time on the Medium and High thresholds by 4.1 and 19.3 min respectively. As with sleep efficiency, a low correlation to PSG was observed for the Low threshold (r=0.24) and a moderate correlation was observed for the Medium and High thresholds (r=0.33 and 0.37) and the SEE was similar across thresholds. Compared to PSG,

the Actical® monitor underestimated the number of wake bouts regardless of the sleep-wake threshold employed, however these results were not significantly different from that obtained from PSG. As with the other sleep parameters, the SEE for all thresholds was practically the same, however the Low threshold produced the smallest mean bias. A low correlation with PSG was observed with r values of 0.20 for Low, 0.12 for Medium and 0.16 for High thresholds. The Actical® devices underestimated sleep latency by an average of 9.5 min with a SEE of 15.2 min. The results for this sleep parameter were the same across the three thresholds and a low correlation to PSG was observed (r = 0.24).

Discussion

In a sport setting, Actical® devices are commonly used to identify athletes requiring further education or intervention about their sleep hygiene, and are also used in research settings as a measure of sleep quantity and quality.[2,4] The results of this study indicate that for elite team-sport athletes the interpretation of the Actical® data, and therefore feedback to athletes, can vary widely depending on the Actiware® software threshold used to process the data. In the present study, the Medium sleep-wake threshold of the Actiware® software produced the smallest mean bias compared with PSG for sleep duration, sleep efficiency and wake after sleep onset. These findings are in contrast to that of Sargent et al. (2015), who reported that the High threshold produced the smallest differences for the same sleep parameters compared to the PSG for elite endurance-trained cyclists. Our results are also in contrast to other validation studies using the Actical® device with non-athletic populations which have recommended using the Low or very low (activity count = 10) thresholds for better overall performance compared with PSG.[6,7]

The differences between the findings of the present study conducted on elite team-sport athletes and those of other validation studies noted above may relate to the different subject groups and conditions in which studies were conducted. Sargent (2015) used male endurance-trained cyclists who were measured during a six-week block of intensified

training. Sargent hypothesised that the heavy training load may have reduced the immobility of subjects during sleep, with the possibility that athletes moved more in their sleep due to muscle soreness induced by their training. The validation studies conducted by Kosmadopoulos and colleagues (2012 & 2014) did not use an elite athlete population, rather they used healthy young adults, male and female, sleeping in a laboratory. This setting may also influence results as factors such as noise, temperature and light are regulated in a sleep laboratory, thereby potentially influencing sleep behaviour.

Collectively the results of the current and previous validation studies highlight that, compared with PSG, actigraphy has limitations when applied to different athletic or non-athletic populations and should be interpreted with caution. One limitation of the present study was that only male athletes were used and as such, similar validation studies should be conducted on female athletes. However, given the varying results of the aggregated validation studies, the data on women may also be confounded by similar factors such as the type of athlete (endurance or team-sport), training phase, and if the data is collected in a private residence or a sleep laboratory.

The present study found that compared to PSG, the athletes' total sleep time (447 min) and sleep efficiency were underestimated on the Low threshold (~12 min) and overestimated with the Medium (~9 min) and High (~24 min) thresholds. In an applied sense, the latter error of about five percent is substantial. The practical relevance of the errors compared with PSG are larger for WASO (41 min), which was underestimated when the Medium (~37 min) and High (~22 min) thresholds were applied, and overestimated using the Low (~31 min) threshold. Indeed the value from the High threshold is approximately half that from PSG and would lead to a different interpretation of an athlete's sleep. Furthermore, regardless of the threshold applied, the Actical® devices significantly underestimated sleep onset latency by an average of 9.5 mins. Considering a normal sleep latency period is 10-20 mins, a bias of this magnitude is important when attempting to interpret sleep reports. In addition to these observed biases, it is important to note that in the present study the

correlations between the actigraphy and PSG for all sleep parameters, other than total sleep time, are only low to moderate. One of the limitations of actigraphy is that it uses movement or lack of movement as a surrogate to infer a state of wakefulness or sleep respectively[12], whereas PSG detects wakefulness and sleep using brain wave activity rather than subject mobility.

As well as investigating the potential systematic bias of each threshold on the different sleep parameters, another objective of the study was to understand the uncertainty of measures using the Actical® devices. These devices are often used with elite athletes to quantify sleep in research studies, and in routine servicing to identify athletes requiring sleep hygiene coaching. If the devices are used to longitudinally monitor an athlete's sleep, care must be taken when interpreting any change in measures. The standard error of the estimate (SEE) provides users with the typical 'noise' or variability of the measure. This estimate can be used to interpret changes in sleep reports for athletes. By quantifying the random uncertainty of each measure, normally distributed confidence limits (CL) can then be employed when interpreting changes in sleep parameters measured with the Actical® devices. If one wanted to be conservative, 95% CLs would be employed to help interpret a meaningful change in an individual's sleep measures, where the 95% CL for an individual change score is calculated as ± √2 x 1.96 x SEE. A less conservative approach might use a 52% CL for an individual change score, calculated as ± √2 x 0.71 x SEE.

This study found the SEE for each sleep parameter was similar regardless of the Low, Medium or High software threshold applied. For total sleep duration, the SEE across thresholds was approximately 25 min, sleep efficiency 4.5%, WASO 21 min and wake bouts approximately 8 counts. The SEE for sleep onset latency was 15.2 min. If one used 95% CLs, changes within an individual athlete would have to exceed ± 69 min for sleep duration, ± 12.5% for sleep efficiency, ± 58 min for WASO, ± 22 counts for Wake Bout, ± 42 min for sleep latency. These magnitudes highlight a limitation of using actigraphy on individuals to monitor changes longitudinally. Without an understanding of the 'noise' of each

measurement, changes in sleep measures may be interpreted as genuine, rather than being due to the measurement variability of the actigraphy device. In a practical sense, this may lead practitioners and researchers to believe a particular sleep intervention has been successful when in reality it has not.

Conclusions

Results from the present study suggest that whilst the 'noise' or imprecision of measures from the Actical® devices is similar for the different sleep-wake thresholds, there is less bias associated with the Medium threshold for sleep duration, sleep efficiency and WASO. Therefore, scientists using Actical® devices to monitor sleep in elite team-sport athletes should consider using thresholds that are moderately sensitive to sleep (Medium threshold) where activity counts are above 40. Additional validation studies of the Actical® devices with elite athlete populations, including female athletes, should be undertaken to understand the bias and imprecision of the different sleep-wake threshold settings on data analysis.

Practical Implications

- Sleep reports and research using the Actical® devices should indicate which sleep-wake threshold was used to process the data.
- A Medium sleep-wake threshold (activity counts above 40) should be used to process sleep data for team sport athletes.
- The imprecision of actigraphy highlights the importance of utilising confidence limits to assess the likelihood of a real change between sleep measures over time.

Acknowledgements

This study was funded by the Australian Institute of Sport through the Big Idea Grant. Special thanks to Jamie Plowman and Sarah Morris for their technical expertise and assistance throughout the project.

Conflict of interest

No competing agreements, professional relationships or financial interests existed where a third party may benefit from the presented results.

References

1. Sargent C, Lastella M, Halson S, Roach G. The validity of activity monitors for measuring sleep in elite athletes. *J Sci Med in Sport*. 2015 doi: 10.1016/j.jsams.2015.12.007.
2. Lastella M, Roach G, Halson S et al. (2015) Sleep/wake behaviours of elite athletes from individual and team sports. *Eur J Sport Sci* 2015; 15(2):94-100.
3. de Souza L, Benedito-Silva AA, Nogueira Pires ML et al. Further validation of actigraphy for sleep studies. *Sleep* 2003; 26(1):81-5.
4. Sargent C, Halson S, Roach G. Sleep or swim? Early morning training severely restricts the amount of sleep obtained by elite athletes. *Eur J Sport Sci* 2014; 14 (Supp. 1):S310-S315.
5. Robillard R, Lambert T, Rogers N. Measuring sleep-wake patterns with physical activity and energy expenditure monitors. *Biological Rhythm Res* 2012; 43:5, 555-562.
6. Kosmadopoulos A, Sargent C, Darwent D et al. Alternatives to polysomnography (PSG): A validation of wrist actigraphy and a partial-PSG system. *Behav Res Methods* 2014; 46(4):1032-41.
7. Kosmadopoulos A, Sargent C, Zhou X et al. Two for one? Validating an energy expenditure monitor against polysomnography for sleep/wake measurement. Chapter 4. *Sleep of Different Populations*. Zhou X, Sargent C (Eds). Adelaide, Australia Australasian Chronobiology Society 2012.
8. Kushida C, 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.

9. Buysse D, Reynolds III C, Monk T et al. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research* 1989; 28:193-213.
10. 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*: American Academy of Sleep Medicine 2007.
11. Weiss AR, Johnson NL, Berger NA et al. Validity of activity-based devices to estimate sleep. *J Clinical Sleep Med* 2010*;* 6(4): 336-342.
12. Tonetti L, Pasquini F, Fabbri M et al. Comparison of two different actigraphs with polysomnography in healthy young subjects. *Chronobiology International* 2008; 25(1): 145-53.
13. Hopkins W. A scale of magnitudes for effect statistics. A New View of Statistics. Available at: http://sportsci.org/resource/stats. Accessed 16 March 2016.
14. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; i: 307-310.
15. Hopkins W. Analysis of validity by linear regression. A New View of Statistics. Available at: http://sportsci.org/resource/stats.

Figures

Fig. 1 Bland-Altman plots comparing PSG with the Actical® devices at different sleep-wake thresholds (Low, Medium and High) for total sleep time, sleep efficiency, wake after sleep onset, wake bouts and sleep onset latency. The solid lines indicate the mean bias from PSG and the broken lines indicate 95% limits of agreement (±1.96 SDs). ACT - Actical®, PSG – polysomnography.

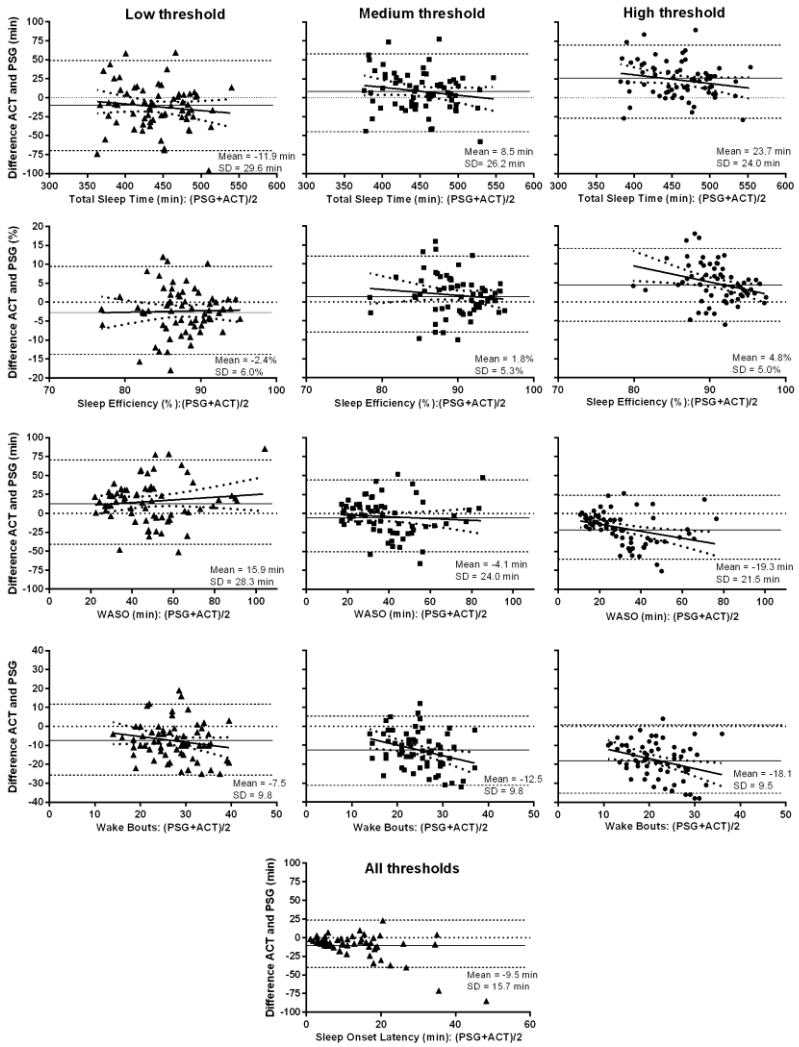

Table 1 Comparison of sleep parameters measured by polysomnography (PSG) and Actical® activity monitors. Positive values indicate an overestimation by activity monitors relative to PSG, and negative values indicate an underestimation by activity monitors relative to PSG.

Measure	n	Mean ± SD	Mean bias (95% CL)	SEE	*r*	p
Total Sleep Time (min)	67					
PSG		447.1 ± 46.1				
Actical Low (20)		435.2 ± 42.6	-11.9 (-19.1 to 4.6)	29.0	0.78	<0.0001*
Actical Medium (40)		455.6 ± 41.8	8.5 (2.1 to 14.9)	26.1	0.83	<0.0001*
Actical High (80)		470.8 ± 41.4	23.7 (17.8 to 29.6)	24.1	0.85	<0.0001*
Sleep Efficiency (%)	67					
PSG		88.7 ± 4.9				
Actical Low (20)		86.3 ± 5.0	-2.4 (-3.8 to 0.9)	4.7	0.27	0.0251*
Actical Medium (40)		90.5 ± 4.4	1.8 (0.5 to 3.7)	4.6	0.35	0.0036*
Actical High (80)		93.5 ± 3.7	4.8 (3.6 to 6.0)	4.6	0.34	0.0047*
Wake After Sleep Onset (min)	67					
PSG		41.0 ± 21.6				
Actical Low (20)		56.9 ± 24.0	15.9 (9.0 to 22.9)	21.1	0.24	0.0571
Actical Medium (40)		36.9 ± 19.9	-4.1 (-10.0 to 1.8)	20.5	0.33	0.0060*
Actical High (80)		21.7 ± 15.6	-19.3 (-24.6 to 14.0)	21.3	0.37	0.0024*
Wake Bouts	67					
PSG		31.0 ± 8.4				
Actical Low (20)		23.5 ± 7.0	-7.5 (-9.9 to 5.1)	8.3	0.20	0.1075
Actical Medium (40)		18.5 ± 6.1	-12.5 (-15.0 to 10.1)	8.4	0.12	0.3505
Actical High (80)		13.0 ± 6.1	-18.1 (-20.4 to 15.7)	8.3	0.16	0.1986
Sleep Onset Latency (min)	67					
PSG		16.0 ± 15.5				
Actical (Low, Medium and High thresholds)		6.3 ± 8.3	-9.5 (-13.4 to 5.7)	15.2	0.24	0.0496*

* Significant difference from PSG $p < 0.05$