Human Factors and Wearable Technologies, Vol. 29, 2022, 36–44
https://doi.org/10.54941/ahfe1001472

Actigraphy vs Polysomnography Measurements for Sleep Arousals

J P Little[1], S W K Ho[1], L Rayward[1], G Uhrle[2], D Green[3], and M Pearcy[1]

[1]Biomechanics and Spine Research Group, School of MMPE, Queensland University of Technology, Brisbane, Australia
[2]Mater Adult Sleep Service, Mater Misericordiae, Brisbane, Australia
[3]Sealy Australasia, Brisbane, Australia

ABSTRACT

This study sought to better understand how actigraphy may be practically applied to interpret sleep arousals when used in studies of home-based sleep. For this purpose, we analysed a small cohort of healthy adult sleep using polysomnographic (PSG) measurements and compared this with actigraphy measures for sleep quality parameters. Good agreement between PSG arousals related to arm movement and actigraphy awakenings highlighted the benefit of actigraphy in measuring sleep related arousals, although caution is needed when interpreting physiological 'body-related' awakenings vs limb-related motility associated with physiological awakenings.

Keywords: Sleep quality, Actigraphy, Polysomnography, Sleep biomechanics

INTRODUCTION

Polysomnography (PSG) refers to the measurement of multiple physiological biomarkers that define sleep quality, including brain activity (EEG), eye activity (EOG), muscle activity (EMG), heart activity (ECG), blood oxygenation, and respiratory effort. PSG is clinically considered the measurement modality of choice to achieve highly fidelic objective measures of physiological parameters defining sleep (Lotjonen et al. 2003, Roberts et al. 2020). However, actigraphy has been used for several decades as a portable means to monitor sleep wake rhythms and sleep movements, offering a more flexible alternative to PSG (Ferri et al. 2013).

Actigraphy offers several key advantages over PSG, in that it is portable, easily set-up, and involves a low burden of participation to the subject (Krystal and Edinger 2008) enabling sleep measurements to be recorded over extended periods of time. Conversely, the main shortcoming is that actigraphy is a surrogate measure of sleep physiology, providing only activity measurements and interpreting these as a proxy for sleep data (Krystal and Edinger 2008). The most reliable measurements of sleep obtained using actigraphy are total sleep time, and counts of the sleep-wake cycles, measured over multiple nights (Krystal and Edinger 2008). However, Paquet et al. (2007) note that actigraphy tends to overestimate total sleep time and sleep efficiency, as a result of its lower accuracy in detecting periods of wake and

high magnitude of false positives when detecting sleep. Overestimations of sleep are related to underestimation of limb movement counts (Kemlink et al 2008, Gschliesser et al. 2009) leading to more scored sleep time.

Studies exploring sleep behaviour in relation to the participant's experience of different sleeping surfaces/environments commonly utilize actigraphy to measure sleeping motility and sleep quality (eg. total sleep time (TST), sleep efficiency (SE), sleep onset latency (SOL) and wake after sleep onset (WASO)) (Monk et al. 1999, Estrella et al. 2021, Yu et al. 2020, Krauchi et al. 2018). Despite the afore-mentioned limitations in measurement of sleep parameters, for practical explorations of sleep behaviour, actigraphy still remains an attractive alternative to expensive, time-consuming, high-participant cost PSG studies. This is particularly the case when evaluating sleep quality over extended periods of time in the home.

For this reason, the current study sought to determine a reliable scoring algorithm for an Actiwatch Spectrum PRO (Phillips Respironics, Inc., Murrysville, USA) when using the device during a long-term home-based sleep study. The study compares the gold-standard PSG scoring of sleep quality with actigraph derived sleep parameters.

METHODS

Participants

Ten healthy adults between the ages of 22 - 32 years (mean age 25.9 [SD 3.2] years) participated in the study, including five females and five males. Participants were all physically active, healthy, and non-smokers, with a body mass index in the healthy range (18-25). Participants were excluded if they had a history of diagnosed sleep pathology, spinal conditions or recent orthopaedic surgery that may influence sleep comfort, or were using medication to aid sleep. The study was performed with approval from local organizational Ethics Committee, and all subjects gave informed consent for their participation in the study.

Study Design and Procedure

This comparative study of sleep metrics measured from actigraphy and PSG was part of a larger exploration of sleep quality relative to sleep surface. The overall study design consisted of a 10-week home-trial of two different sleeping surfaces (mattresses of 'soft' and 'firm' feel). Over the course of the trial, the participants were asked to wear an Actiwatch Spectrum PRO (AW) on their left arm on each night of the study. In addition to this, participants underwent two, multi-night, ambulatory PSG studies, where they slept in their home on the provided mattresses with both the AW and the ambulatory PSG setup attached to their body. The PSG measurements were collected on two consecutive nights in the third week (on mattress_1) and two consecutive nights in the eighth week (on mattress_2) of the home trial, providing four nights of data (per participant) over which actigraphy and PSG measurements were compared. Participants were requested to wear the AW from one hour prior to their typical bedtime, to ensure baseline activity scores

could be determined prior to sleep. Participants were asked to return after the first 5-week mattress home-trial, for a mid-study download of the stored data and for the watch to be recharged (2 hours).

Measurements

Actigraphy

The AW has an inbuilt micro electro-mechanical (MEMS) accelerometer with a sampling rate of 32 Hz. The MEMS detects movement as an electrical signal (Paquet et al., 2007) which is converted into an activity count and scored as 'sleep' or 'wake'. This method has been validated on patients with sleep disorders (Kushida et al., 2001). For each participant, the watch configuration at set-up defined an epoch (minimum recorded division of time) of 15 seconds.

To calculate the weighted activity score in each epoch, the Actiware software (version 6.1.1, Phillips Respironics, Inc., Murrysville, USA) employed a validated algorithm (Kushida et al. 2001) based on the Cole-Kripke method (Cole et al. 1992). The weighted activity score (*A*) is calculated by applying a weighting algorithm to the activity counts within ±2 minutes (E_{-8} to E_8) of the epoch of interest. This relationship is shown in Equation 1.

$$\begin{aligned} A \;=\; & 0.04(E_{-8} + E_{-7} + E_{-6} + E_{-5} + E_5 + E_6 + E_7 + E_8) \\ & + \; 0.2(E_{-4} + E_{-3} + E_{-2} + E_{-1} + E_1 + E_2 + E_3 + E_4) \;+\; 1E_0 \end{aligned} \qquad (1)$$

Equation 1. Activity count algorithm (Kushida et al. 2001) – E_0 = epoch of interest.

A was used to score the wake/sleep status in any given epoch on the basis of whether the weighted activity score exceeded the predefined activity threshold used in the Actiware software. These predefined thresholds were '*low*', '*medium*' and '*high*', corresponding to activity counts of 20, 40, and 80. All three thresholds were used when analysing the data for 10 participants (MATLAB, version R2019B, Mathworks, Portola Valley, USA), in order to establish the threshold that resulted in the best agreement with PSG-scored sleep metrics.

Polysomnography

An ambulatory PSG study was carried out using a Nox A1 (Nox Medical, Reykjavik, Iceland) system which is a fully portable polysomnography device permitting clinical grade sleep metrics to be measured in locations other than a clinical sleep laboratory. A level 2 ambulatory PSG electrode attachment set-up was used (Figure 1). Electrodes were placed on the participant by a researcher (Authors SH and LR) under the supervision of a trained sleep technician (Author GU) when the participants arrived the afternoon of the study night. Signals recorded were EEG, EOG, EMG (left arm, right arm, right leg - Figure 1 D, E, F), ECG, oxygen saturation, nasal airflow, and respiratory effort (Figure 1).

Sleep staging and respiratory events were scored based on the American Academy of Sleep Medicine (AASM) criteria by a qualified sleep scientist (Author GU) using Noxturnal software (version 6.1.0, Nox Medical, Reykjavik, Iceland). All studies were scored in 30 second epochs and later discretized

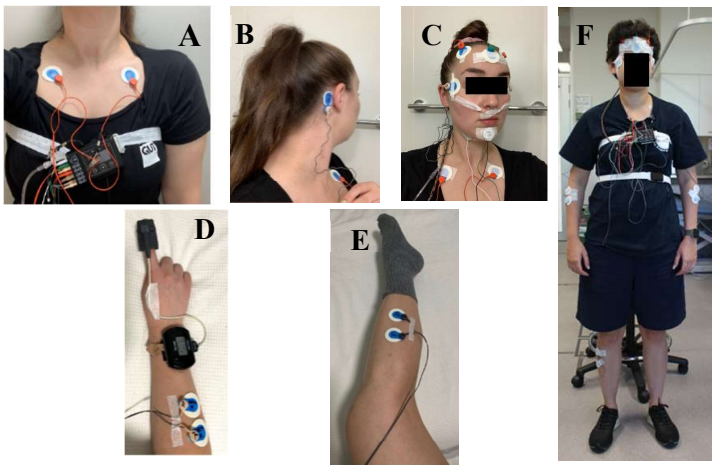

Figure 1: Ambulatory PSG setup for Nox 1, showing A) ECG connections with respiratory effort belt, B) EEG behind right (& left) ear, C) EEG/EOG over the head/chin, including a central ground (green), and respiratory measurement from nose, D) right arm EMG, spirometry and heart rate measurements, E) right leg EMG, F) Full set-up including abdominal effort belt and showing Actiwatch on left arm with EMG.

into 15 second epochs to match the AW. PSG awakenings were referred to as 'arousals' and were scored such that EEG and EMG signals indicating arousals and lasting <15 seconds were scored as 'sleep' and EEG and EMG signals indicating arousals but lasting ≥ 15 seconds were scored as 'wake/arousal'.

To better understand the relationship between PSG-scored and AW derived periods of wake (non-sleep), an additional PSG sleep scoring was carried out. Since the AW was worn on the left arm and measured motility of this arm, left arm movements (LM) were scored by the sleep scientist. These LM were defined as limb movements when the left arm EMG trace increased with an amplitude of >50% from the baseline and lasted between 3 – 15 seconds. Where a movement occurred and a body arousal was detected at the same time, it was considered a left arm arousal (LMA). With movement lasting more than 15 seconds, the stage was scored as a wake. Note that scoring was mediated manually to ensure only one arm movement within any 15 second period regardless of epoch. All arousals (AA) were recorded as arousals from all PSG signals (i.e., respiratory, periodic limb movement, spontaneous, and left arm related arousals). Parameters used in the study are summarized in Table 1.

Analysis

Ancoli-Israel et al. (2003) highlight that when comparing PSG and actigraphy, "time-locking" epochs for both devices is important to address drift in the actigraphy device. To ensure this, AW time was synchronised based on two reference points in the AW-PSG activity count data – characterized by short (1-3 epoch) bursts of LM EMG activity during the night.

A concordance comparison of the Actiware 'awakening' and 'motility count' (scored using *low*, *medium*, and *high* activity thresholds) with PSG

Table 1. PSG parameters for movement scoring.

Acronyms	Explanation
LM	Left arm movement, periods of left arm motility, EMG trace ↑ 50% compared to baseline & lasts 3-15 seconds
LMA	Left arm arousal, LM, where body arousal detected simultaneously
AA	All arousals, including all arousals measured by PSG due to any physiological signal
AAW	All arousals and wakes

scored LM, LMA, AAW, and AA was completed to determine the accuracy, sensitivity, and specificity of the AW when detecting arousals. Comparison of epoch-epoch results for these parameters was carried out using MATLAB. Statistical analysis was completed in IBM SPSS Statistics Version 27.0 (IBM, Armonk, NY). Spearman's rank-order correlation (Rho) was used due to the non-parametric nature of the data and two-tailed test of significance (p) was reported. The statistical significance of $p<0.05$ was used.

RESULTS AND DISCUSSION

Data Collection

Due to a problem with the underlying coding for the ambulatory Nox A1 devices, the second night of consecutive PSG sensor signals was not recorded for four participants. For this reason, these participants then repeated a second two-night PSG study and therefore, recorded a total of five nights of PSG data. Once the coding limitations were resolved, the remaining six participants recorded a total four nights PSG data. A total of 48 simultaneous AW and PSG study nights was completed. However, six nights had missing or corrupted PSG signals due to either participant error (four) or lack of discernable PSG signals which was likely due to poor connections at the electrode-body interface (two). One night of AW recording error occurred due to the participant not wearing AW. Therefore, 41 studies had a full data set of PSG and AW results that could be epoch-matched for comparison.

Comparison of AW and PSG

TST had a significantly strong positive correlation between PSG and AW (Rho 0.731 – 0.796, $p<0.001$). There was a weak-medium significant positive correlation seen between PSG and AW for SE (Rho 0.351-0.419, $p<0.05$), and non-significant poor correlations for WASO. Kushida et al. (2001) similarly showed good agreement between PSG and actigraphy predicted TST, and particularly for the highest wake sensitivity setting, which was also the case in the current results.

The mean difference between PSG-scored and AW-predicted TST, WASO and SE were calculated (ie. PSG – AW) and compared with the absolute PSG-score using the methods of Bland and Altmann (1986). Mean differences for *low*, *medium*, and *high* AW activity threshold were: −28min (SD 29min), −12min (SD 28min), and −3min (SD 27min) for TST; −15min (SD 17min),

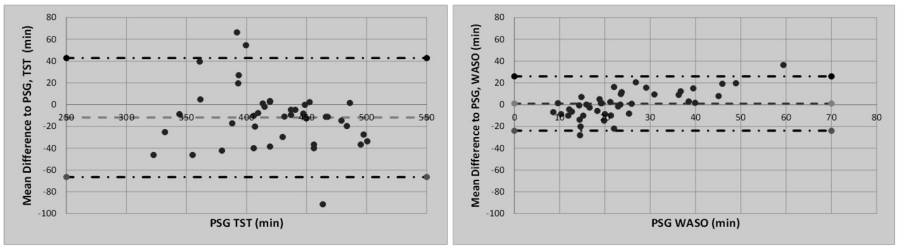

Figure 2: Bland and Altmann (1986) plots of differences between PSG and AW predicted TST (min), and WASO (min), with a medium AW activity threshold. Dark dashed lines indicate 95% confidence, Light dashed lines indicate mean difference.

1min (SD 13min) and 10min (SD 12min) for WASO; and −7% (SD 5%), −3% (SD 5%), and −2% (SD 5%) for SE, respectively. The *medium* activity threshold showed the best overall agreement with PSG-scored results and are shown graphically in Figure 2 for TST and WASO.

Roberts et al. (2020) found a tendency for other models of the Actiwatch Spectrum to over-classify sleep, resulting in overprediction of TST (mean difference from PSG 32min) and underprediction of WASO (mean difference from PSG −24min) for a *medium* AW threshold. However, for the Actiwatch Spectrum Pro used in the current study there was a closer agreement between TST and WASO, with no bias towards differing levels of agreement across the range of TST.

LM, LMA, and AA measured from PSG were compared with awakenings and motility count from AW (Table 2). By scoring the PSG to include LM and LMA, which is not standard for clinical sleep scoring, a significantly strong correlation was observed for both of these parameters compared to the AW awakenings scored at all activity thresholds. However, awakenings scored using the *medium* activity threshold demonstrated the stronger correlations, and of these LMA demonstrated the highest Rho when compared with AW awakenings. Weak, non-significant relationships were obtained when comparing AA.

Kripke et al. (2010) attempted to correlate PSG-scored arousals with leg movements as a potential correlate for actigraphy motility count, and found poor negative correlations, which they noted was because leg movements were not specifically scored via PSG. Results in the current study address this limitation by specifically scoring the PSG arousals related to arm movement (LMA) resulting in a positive strong correlation, which was significant.

Sensitivity, Specificity and Accuracy of AW Compared to PSG

For an epoch-epoch comparison, specificity was found to be lower than sensitivity in all events and scoring algorithms (Table 3). However, specificity was found to be lower (< 75%), with the AW incorrectly scoring wake/arousal events as sleep.

As in the current study, when comparing PSG and AW results, Paquet et al. (2007) and Roberts et al. (2020) found consistently higher sensitivity compared to specificity for the *medium* and *low* AW activity thresholds. Paquet

Table 2. PSG arousals, LM, LMA, AA and AAW vs AW awakenings and motility count. Spearman correlation (Rho) shown with *p* significance as 0.05. (n=41).

PSG	AW	AW Activity Threshold	Rho	*p*
LM	Awakening	*Low*	0.610	**<0.001**
		Medium	0.617	**<0.001**
		High	0.535	**<0.001**
LMA	Awakening	*Low*	0.625	**<0.001**
		Medium	0.660	**<0.001**
		High	0.650	**<0.001**
AA	Awakening	*Low*	0.292	0.064
		Medium	0.314	**0.045**
		High	0.390	**0.012**
AAW	Motility Count	-	0.348	**0.004**

Table 3. Sensitivity, specificity and accuracy from epoch-epoch comparisons.

PSG	AW	AW Threshold	Sensitivity (%)	Specificity (%)	Accuracy (%)
LM	Awakening	*Low*	93.9	74.7	93.1
		Medium	97.3	66.1	96.0
		High	98.9	54.3	97.1
LMA	Awakening	*Low*	93.5	68.4	92.5
		Medium	96.9	57.8	95.3
		High	98.5	46.3	96.4
AAW	Awakening	*Low*	96.2	49.0	90.9
		Medium	98.8	37.2	91.8
		High	99.7	26.7	91.4
AA	Awakening	*Low*	94.0	47.3	90.7
		Medium	97.2	37.4	92.9
		High	98.7	28.3	93.6

et al. (2007) also found improved accuracy in AW predictions for wake events using the *medium* threshold compared to low, which was also reflected in results from the current study.

Overall, the AW demonstrated high sensitivity and accuracy, indicating it was a good predictor of sleep events. But the AW showed a lower specificity, indicating its limited ability to accurately detect wake events related to left arm movement (LMA). For the latter, the AW incorrectly predicted 42% of the left arm arousals for a *medium* activity threshold. This is a reflection of the fact that left arm arousals detected via PSG are due to muscle activations in the forearm, which may not in all instances result in a movement of the arm, but are associated with physiological disturbance during sleep which will be scored as an arousal. Actigraphy relies on the successful interpretation of motility as either sleep or wake, and the current results indicate that in 97% of epochs, the AW will correctly score a sleep event for a healthy participant, which provides confidence in the use of this technology for this participant demographic. These results should be considered in light of

the study population, who were all healthy, young adults with no diagnosed sleep conditions and therefore, non-pathological levels of movement and sleep disturbances during the night.

CONCLUSION

Actigraphy is a valuable tool for ambulatory studies of sleep quality, offering particular utility when investigating sleep behaviour for participants in their own home, and when evaluating sleep over lengthy periods of time. Good agreement between PSG LM and AW awakenings highlighted the benefit of actigraphy in measuring sleep related arousals, although caution is needed when interpreting physiological 'body-related' awakenings vs limb-related motility associated with physiological awakenings. The low user cost associated with actigraphy for long-term home-based sleep studies, makes it a preferred alternative to PSG, and study results will assist to interpret actigraphy-based data in future studies of other populations.

REFERENCES

Ancoli-Israel, S., et al. 2003. "The role of actigraphy in the study of sleep and circadian rhythms: American Academy of Sleep Medicine Review Paper." *Sleep* 26 (3): 342-392.

Bland, J. M., and D. G. Altman. 1986. "Statistical methods for assessing agreement between two methods of clinical measurement." *Lancet* 1 (8476): 307-10.

Cole, R. J., et al. 1992. "Automatic sleep/wake identification from wrist activity." *Sleep* 15 (5): 461-9.

Ferri, R., et al. 2013. "Scoring of sleep related movements: Standard and advanced techniques." In *Sleep and movement disorders*, edited by S. Chokroverty, R. Allen, A. Walters and P. Montagna. UK: Oxford University Press.

Gschliesser, V., et al. 2009. "PLM detection by actigraphy compared to polysomnography: a validation and comparison of two actigraphs." *Sleep Med* 10 (3): 306-11.

Kemlink, D., et al. 2008. "A comparison of polysomnographic and actigraphic evaluation of periodic limb movements in sleep." *Neurol Res* 30 (3): 234-8.

Krauchi, K., et al. 2018. "Sleep on a high heat capacity mattress increases conductive body heat loss." *Physiol Behav* 185: 23-30.

Kripke, D. F., et al. 2010. "Wrist actigraphic scoring for sleep laboratory patients: algorithm development." *J Sleep Res* 19 (4): 612-9.

Krystal, A. D., and J. D. Edinger. 2008. "Measuring sleep quality." *Sleep Med* 9 Suppl 1: S10-7.

Kushida, C. A., et al. 2001. "Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients." *Sleep Med* 2 (5): 389-96.

Lotjonen, J., et al. 2003. "Automatic sleep-wake and nap analysis with a new wrist worn online activity monitoring device vivago WristCare." *Sleep* 26 (1): 86-90.

Monk, T. H., et al. 1999. "Wrist actigraphic measures of sleep in space." *Sleep* 22 (7): 948-54.

Moo Estrella, J., and G. Arankosky Sandoval. 2021. "Comparison and prediction of sleep quality in users of bed or hammock as sleeping device." *Sleep Health* 7 (1): 93-97.

Paquet, J., et al. 2007. "Wake detection capacity of actigraphy during sleep." *Sleep* 30 (10): 1362-9.
Roberts, D. M., et al. 2020. "Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography." *Sleep* 43 (7).
Yu, H., et al. 2020. "Effect of an Inflatable Air Mattress with Variable Rigidity on Sleep Quality." *Sensors (Basel)* 20 (18).