The validity of Actiwatch2 and SenseWear armband compared against polysomnography at different ambient temperature conditions

CrossMark

Mirim Shin[a,*], Paul Swan[b], Chin Moi Chow[a]

[a]Delta Sleep Research Unit, Exercise, Heath & Performance Research Group, Faculty of Health Sciences, The University of Sydney, Lidcombe, NSW, Australia
[b]Australian Wool Innovation Limited, The Woolmark Company, Sydney, NSW, Australia

ARTICLE INFO

ABSTRACT

There were no validation studies on portable sleep devices under different ambient temperature, thus this study evaluated the validity of wrist Actiwatch2 (AW2) or SenseWear armband (SWA) against polysomnography (PSG) in different ambient temperatures. Nine healthy young participants (6 males, aged 23.3 ± 4.1 y) underwent nine nights of study at ambient temperature of $17\,°C$, $22\,°C$ and $29\,°C$ in random order, after an adaptation night. They wore the AW2 and SWA while being monitored for PSG simultaneously. A linear mixed model indicated that AW2 is valid for sleep onset latency (SOL), total sleep time (TST) and sleep efficiency (SE) but significantly overestimated wake after sleep onset (WASO) at $17\,°C$ and $22\,°C$. SWA is valid for WASO, TST and SE at these temperatures, but severely underestimates SOL. However, at $29\,°C$, SWA significantly overestimated WASO and underestimated TST and SE. Bland–Altman plots showed small biases with acceptable limits of agreement (LoA) for AW2 whereas, small biases and relatively wider LoA for most sleep variables were observed in SWA. The kappa statistic showed a moderate sleep–wake epoch agreement, with a high sensitivity but poor specificity; wake detection remains suboptimal. AW2 showed small biases for most of sleep variables at all temperature conditions, except for WASO. SWA is reliable for measures of TST, WASO and SE at 17–$22\,°C$ but not at $29\,°C$, and SOL approximates that of PSG only at $29\,°C$, thus caution is needed when monitoring sleep at different temperatures, especially in home sleep studies, in which temperature conditions are more variable.

*Correspondence to: Delta Sleep Research Unit, Exercise, Health & Performance Research Group, Faculty of Health Sciences, The University of Sydney, 75 East Street, Discipline of Exercise and Sport Science, Lidcombe 2141, NSW, Australia. Tel.: +61 2 9351 9276; fax: +61 2 9351 9204.
E-mail address: mshi0568@uni.sydney.edu.au (M. Shin).

Peer review under responsibility of Brazilian Association of Sleep.

1. Introduction

In-home studies are becoming increasingly common for sleep monitoring as portable sleep devices are not only more affordable, but can be worn comfortably over extended periods of time without interfering with the participants' routine [1–3]. Two important factors govern the quality of the sleep data that are recorded in the home: the validity of the sleep monitoring instrument, and the sensitivity of the instrument to changes in temperature.

Various models of actigraphy have been validated against PSG under laboratory conditions [2,4,5] in healthy adolescents, young, and older adults [1,6], and in clinical populations of insomnia, major depressive disorder, dementia and sleep-disordered breathing [1,7], with an overall sleep and wake epoch agreement of 72.1–96.5% [4]. Similarly, sleep validation with the SWA (SenseWear armband) which is originally designed to measure energy expenditure and later adopted for sleep measures has been conducted in healthy subjects [8], children and adolescents [8,9] and patients with obstructive sleep apnea [10] with a sleep and wake epoch agreement of $79.9 \pm 1.6\%$ [10]. However, SWA differs from the numerous models of actigraphy in that in addition to employing an accelerometer for motion monitoring, it has a heat flux sensor, a skin temperature sensor and a galvanic skin response sensor that detects electrical conductivity. Despite an added feature of skin temperature measurement, previous validation studies found that SWA did not improve sleep onset detection compared to actigraphy [8,10].

Notably, the sleeping environment in which home studies are conducted is often not standardized especially with respect to ambient temperatures due to seasonal changes. According to the data collected from Summer 2012 to Summer 2014 in Australian houses by the Faculty of Architecture, The University of Sydney, the recorded overnight average temperatures (from 2200 to 0600 h) over the four seasons were: Summer, $24.3\,°C \pm 1.6\,°C$; Autumn, $21.6\,°C \pm 2.2\,°C$; Winter, $17.1\,°C \pm 2.6\,°C$; and Spring, $21.1\,°C \pm 1.9\,°C$ (Australian Research Council's Discovery Projects, DP 11010559). The observed large temperature variation, in spite of air-conditioning being used in the monitored rooms, would suggest an expected greater variation in homes that are not centrally heated. However, there have been no studies that examine the validity of actigraphy and SWA under different ambient temperature conditions. Given the differing principles of measurements between actigraphy and SWA, a validation study is warranted to evaluate the concordance rates for sleep variables simultaneously recorded from these devices and PSG. In this study, the Actiwatch 2 (AW2) (Phillips-Respironics) was used. Hence, agreement rates between AW2, SWA and PSG were examined under ambient temperatures of 17 °C, 22 °C and 29 °C. The objective of the study was to test the validity of AW2 and SWA for sleep assessment against PSG. We hypothesized that the performance of SWA may be more affected by ambient temperatures since it also measures skin temperature.

2. Method

2.1. *Participants*

Nine healthy participants (six males, aged 23.3 ± 4.1 y, BMI 22.6 ± 2.4 kg m^{-2}) were recruited. Participants with pre-existing medical conditions such as sleep disorders (insomnia, sleep apnea, periodic limb movement disorders and bruxism), cardio-respiratory conditions (hypertension, cardiovascular diseases, respiratory infections, chronic obstructive pulmonary diseases), and metabolic conditions (diabetes, metabolic syndrome) were excluded. Individuals on night shifts or medications/drugs, or who smoked or had travelled across trans-meridian borders in the last 2 weeks were also excluded. Participants avoided alcohol, caffeinated beverages, and vigorous exercise 8 h before their averaged bedtime on study days. The study was approved by the University of Sydney Human Research Ethics Committee.

2.2. *Procedures*

All participants completed a consent form and a questionnaire including demographics and medical history. They wore the AW2 for a week, prior to overnight sleep studies, to determine their average bedtime and rise time for study scheduling. They attended the sleep laboratory on ten occasions. After an adaptation night, all participants were randomized to nine different sleeping conditions with four nights at 17 °C, two nights at 22 °C and three nights at 29 °C. At both 17 °C and 22 °C, participants wore pyjamas (long sleeve top and pants) with bedding. At 29 °C, participants only wore shorts and a singlet. On each test night, participants ate a standardized mixed macronutrient meal 4 h before their average bedtime. Participants continuously wore SWA from the morning of each study session (removed only when taking a shower), whereas AW2 and PSG were applied 4 and 2 h before their bedtime, respectively. All measures of PSG, AW2 and SWA were collected simultaneously as participants slept at the University sleep laboratory. The data were collected as part of a study that investigated the effect of apparel and bedding type and ambient temperatures on sleep.

2.3. *Measures*

2.3.1. *Polysomnography*

Sleep parameters were measured using the Compumedics E-series or W-series Sleep system (Compumedics Australia Pty Ltd., Australia). EEG electrode placement (C3/A2, O2/A1 and F3/Cz for W-series or F3/A2 for E-series) was conducted in accordance with the International 10–20 system. EOG, submental EMG and ECG were continuously recorded. All electrode sites were referenced to the vertex (Cz), and a ground electrode was attached to the forehead (Fpz). On the adaptation night, left and right leg EMG, oxygen saturation, thoracic and abdominal breathing movements and airflow were also recorded to exclude sleep disorders. PSG data were scored by

two experienced scorers according to the American Academy of Sleep Medicine (AASM) guidelines [11].

2.3.2. Actigraphy

The AW2 (Phillips-Respironics, Murryville, PA) was placed on the non-dominant wrist. The data were collected in 30 s epochs. Rest intervals were manually marked in the actigraphy software according to the PSG timing of lights-out (bedtime) and lights-on (wake time). The standard factory-default algorithm was used for the estimation of sleep parameters using the Respironics Actiware v5.59.0015.

2.3.3. SenseWear armband

The SWA Pro3 (BodyMedia Inc., Pittsburgh, PA) was placed on the upper non-dominant arm over the triceps. Although the manufacturer has recommended placing the armband on the right arm, the measurements taken between the right and left arm were not significantly different [12]. Thus participants wore both the SWA and AW2 on the same, non-dominant arm for comparison purposes. The estimates of energy expenditure, physical activity duration, and sleep and wake parameters were extracted using proprietary algorithms (SenseWear professional 7.0 software).

2.4. Statistical analysis

Since all nine participants had completed nine study nights, the PSG and SWA collected 81 data points for each sleep variable except for AW2, which collected 79 data points due to technical issues. All data obtained from the PSG, AW2 and SWA were aligned to the timing of PSG lights-out and lights-on. A linear mixed model was applied to analyze differences in sleep variables between the devices at each temperature condition, with temperature conditions and sleep devices (PSG, AW2 and SWA) as fixed factors and participants as random factors. Further post-hoc pairwise comparisons were performed using the Fisher's Least Significant Difference procedure. The Bland–Altman (B–A) plots (MedCalc software, Belgium) displayed the mean bias (the average of the differences between two methods) and 95% limits of agreement (the mean bias plus or minus 1.96 times its SD) [13]. Additionally, B–A plot with multiple measurements per subject with the true value varies model was also performed due to the repeated measure design. Since PSG is considered the gold standard measurement for sleep, plots of the differences between AW2/ SWA and PSG against PSG rather than the mean of the two methods were displayed in this manuscript [14]. The linear regression was used to evaluate the associations between sleep parameters (SOL, WASO, TST and SE) collected from AW2/SWA and PSG (SPSS v20, Chicago, IL). Wake and sleep epoch (one epoch, 30 s) agreements were analyzed for AW2/SWA against PSG using the kappa statistic, which determines the amount of agreement that can be expected by chance [15]. The kappa statistic ranges from 1 which demonstrates perfect agreement, to 0 which demonstrates agreement based on chance alone, and to −1 which demonstrates complete disagreement [1]. As the SWA is limited to estimating sleep and wake in 1-min epochs, each 1 min output was divided to provide an equivalent measure in two 30 s epochs as reported previously [10]. For sleep and

wake epoch analysis, data were coded as 0=wake and 1=sleep. Overall agreement rates (percentage agreement), sensitivity, specificity and kappa statistic were calculated using SPSS v20. Sensitivity is a measure of the ability of the AW2 or SWA to detect sleep when the PSG has also scored sleep, and calculated as the number of true sleep epoch/ (number of true sleep+number of false wake epoch). Specificity is a measure of the ability of the AW2 or SWA to detect wake when the PSG indicated the same, and calculated as the number of true wake epoch/(number of true wake+number of false sleep epoch) [2].

3. Results

Table 1 shows the mean and standard deviations for SOL, WASO, TST and SE for PSG, AW2 and SWA at ambient temperature conditions of 17 °C, 22 °C, 29 °C. The sleep measures of SOL, TST and SE from AW2, at all temperature conditions, were not significantly different from those recorded during PSG. However, WASO was significantly overestimated when compared to PSG at 17 °C and 22 °C (Table 1). The sleep measures recorded from SWA show a significant underestimation of SOL (at 17 °C and 22 °C), and TST and SE (at 29 °C), but an overestimation of WASO (at 29 °C) compared to PSG. Fig. 1, B–A plots for single measurement, has been specifically chosen to display data sets from different ambient temperatures and the spread of data especially those that lie outside of the limits of agreement, i.e., outliers of WASO, TST and SE at 29 °C (Fig. 1F–H). Table 2 presents the differences between AW2/SWA and PSG (mean bias), and LoA from B–A plots with multiple measurements per subject. Consistent with the data presented in Table 1 and Fig. 1, WASO, TST and SE for SWA show larger mean bias and LoA at 29 °C than at 17 °C and 22 °C.

Although the correlation coefficients between AW2, SWA and PSG were generally acceptable with values well above .49,

Table 1 – Sleep indices (mean±SD) recorded from PSG (n=81), AW2 (n=79) and SWA (n=81).			
	PSG	AW2	SWA
SOL (min)			
17 °C	21.6±17.0	15.5±15.7	10.3±12.0*
22 °C	22.9±18.0	26.5±29.6	9.2±9.5*
29 °C	24.9±19.2	18.4±18.2	17.4±16.4
WASO (min)			
17 °C	20.1±16.1	33.2±23.7*	22.9±23.7
22 °C	21.5±17.3	33.0±21.8*	16.8±14.2
29 °C	27.1±21.0	34.3±17.1	48.1±40.2*
TST (min)			
17 °C	451.2±54.3	442.2±55.9	451.3±57.1
22 °C	433.4±47.7	415.8±58.2	438.8±57.9
29 °C	436.4±69.6	429.3±68.2	403.4±76.0*
SE (%)			
17 °C	91.3±5.1	89.4±5.8	91.9±5.2
22 °C	90.2±4.9	86.6±8.3	92.3±4.3
29 °C	88.5±8.1	87.2±7.3	83.7±10.5*

PSG, polysomnography; AW2, Actiwatch 2; SWA, SenseWear armband.
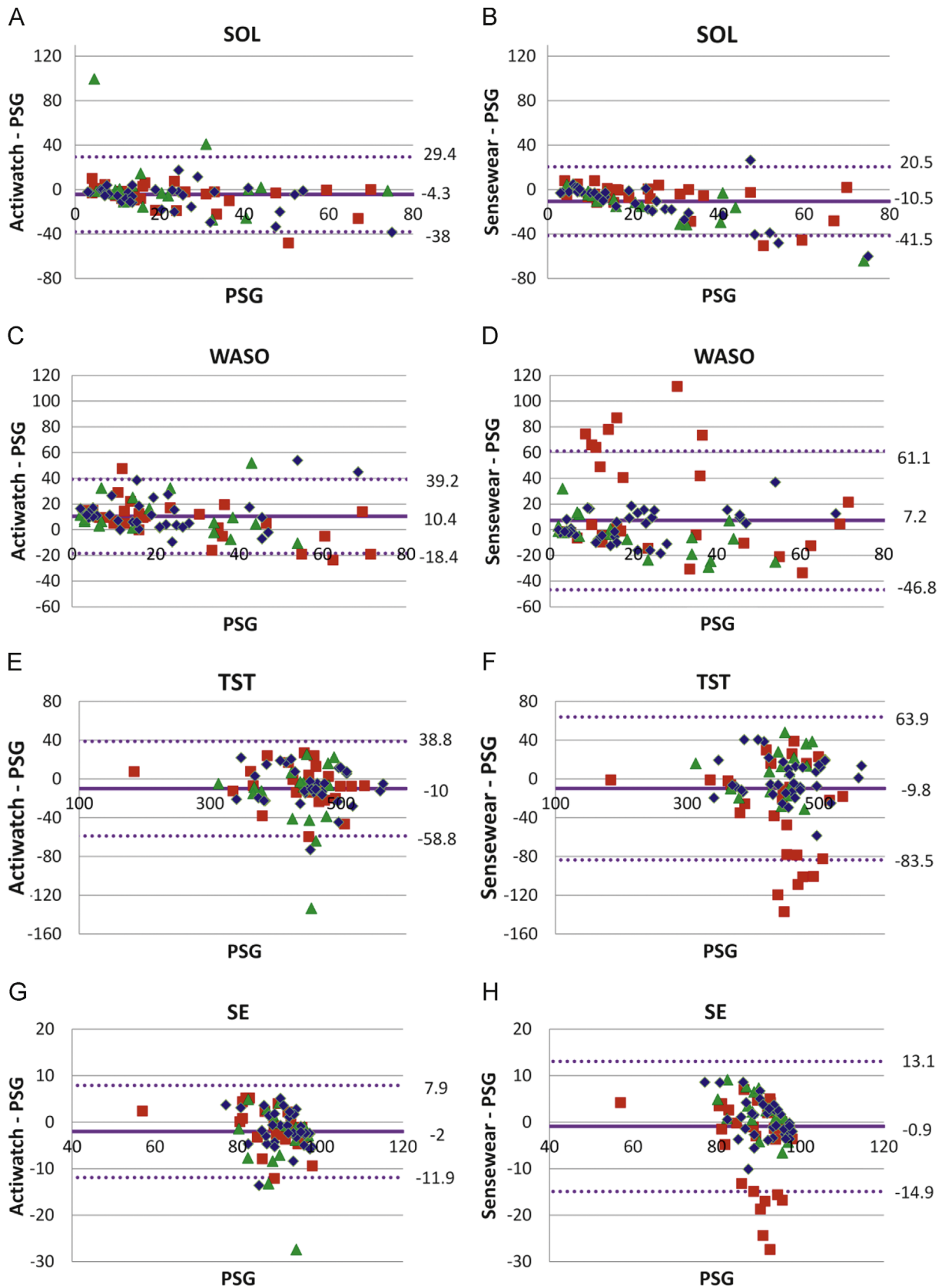* Significant difference between AW2, SWA and PSG, p<0.05.

Fig. 1 – The Bland–Altman plots for SOL, WASO, TST and SE, Left, B–A plots for AW2 and PSG; Right, B–A plots for SWA and PSG, Mean bias, middle horizontal line showing shift from zero; Limits of agreement (2 standard deviations of bias) as indicated by the dotted lines on either side of the mean bias line, SOL, WASO, TST are in minutes, SE in percentages, ◆ 17 °C, ▲ 22 °C and ■ 29 °C—temperature of the sleeping environment.

Table 2 – Mean bias and limits of agreement for sleep indices recorded from AW2 and SWA at each ambient temperature.

°C	SOL			WASO			TST			SE		
	29	22	17	29	22	17	29	22	17	29	22	17
AW2–PSG												
Bias	−6.5	3.6	−6.7	7.1	11.5	12.4	−7.1	−17.6	−8.3	−1.3	−3.6	−1.6
LoA	23.6	55.4	22.6	30.5	30.3	27.5	40.3	73.3	38.5	8.3	14.8	7.7
SWA–PSG												
Bias	−7.4	−13.7	−11.3	20.9	−4.7	2.8	−33.1	5.4	0	−4.9	2.1	0.6
LoA	28.7	32.5	32.1	80.6	31.1	24.1	100.2	46.9	41.6	19.5	8.3	8.1

Table 3 – Linear regression analysis for sleep indices between AW2 and PSG, SWA and PSG.

	SOL		WASO		TST		SE	
	R	SEE	R	SEE	R	SEE	R	SEE
AW2	.61	16.44	.73	14.49	.91	24.86	.72	4.92
SWA	.52	11.64	.49	27.57	.83	37.71	.54	6.95

R, Pearson's correlation coefficient; SEE, standard error of estimate.
AW2, Actiwatch 2 ($n=79$); SWA, SenseWear armband ($n=81$); PSG, polysomnography ($n=81$). ($p<.001$ for all results).

SWA fared poorer compared to AW2 (Table 3). The sleep and wake epoch agreement rates with PSG were high for both AW2 (87.6%) and SWA (89.6%). Although the sensitivity (ability to detect sleep) was high for AW2 (95%) and SWA (93%), the specificity (ability to detect wake) was low for both AW2 (45%) and SWA (57%). The Cohen's kappa coefficients showed moderate agreement for AW2 (.41) and SWA (.46), $p<.001$.

4. Discussion

The present study evaluated the concordance between AW2, SWA and PSG under ambient temperatures of 17 °C, 22 °C and 29 °C. AW2 shows good average agreements with PSG for SOL, TST and SE at all three temperatures, although it significantly overestimates WASO at 17 °C and 22 °C (Table 1). The mean bias for SOL were considered small (range from −6.7 to 3.6 min) and clinically acceptable, consistent with the finding of a previous study except that their study showed a relatively wider range of LoA [16]. The larger variability may be the result of the proprietary software default setting. In contrast, in the present study, lower variability may be explained by the use of a standardized method for manual detection of the rest interval (equivalent to time-in-bed) in which sleep variables are calculated. Although a small bias was observed for WASO (range from 7.1 to 12.4 min), AW2 significantly overestimated WASO at 17 °C and 22 °C but not at 29 °C (Tables 1 and 2 and Fig. 1B). This finding contradicted previous studies, which found that AW2 underestimated WASO [2,16]. These differences may be explained by the different model of actigraphy and algorithms used [5], or different level of activity threshold set [3]. Interestingly, good agreements for SOL and WASO were observed when sleep onset was short and wake bouts were low during the sleep period. Such agreements faded as sleep onset became delayed and wake bouts increased (Fig. 1A and B) contributing to the widening of LoA. The present study found good sleep detection (95%) but poor wake detection (45%) consistent with that reported in a review study [4]. Good sleep detection may reflect the good agreement rates in TST and SE at all temperatures (Fig. 1C and D). Poor wake detection may reflect the large variability in WASO (Fig. 1B and Table 2). The high correlations observed for the various sleep variables were consistent with previous reports [1,2,17]. The correlation coefficient is the highest for TST (Table 3). However, correlations may reflect how well data points from two measurements lie along any straight line or the line of equality [13].

For SWA, good agreements were observed for TST, WASO and SE at 17 °C and 22 °C. WASO was overestimated, and TST and SE were underestimated at 29 °C. SOL was significantly underestimated at 17 °C and 22°, although it was not statistically significantly different from PSG at 29 °C. Similar to a previous study reporting a bias of −8.7 min [9], the present study reported mean biases for SOL from −7.4 to −13.7 min (Table 2). Although SWA employed a skin temperature sensor and a heat flux sensor to improve the accuracy of estimating SOL, significant underestimations were observed at 17 °C and 22 °C ($p<.05$) (Table 1). The mechanism for this underestimation is unclear; presumably, at the lower temperatures (17 °C and 22 °C), skin warming over time, due to heat trapped under bedding, may cause vasodilation and may have led to earlier sleep onset detection. SWA showed a significantly higher WASO than PSG as well as larger LoA at 29 °C (Tables 1 and 2), in line with groups of outliers observed in the B–A plot (Fig. 1F). This discordance may reflect the principles of measurements and associated algorithms, in that SWA detects sleep based on skin temperature, heat flux, skin

conductance and movement. A study revealed that decreases in skin conductance provided a sensitive marker for autonomic arousal during sleep [18]. It may be speculated that an increase in WASO at high ambient temperatures was associated with a decrease in skin conductance. Accordingly, additional analysis was performed to see whether galvanic skin response values during wake bouts were different at the three ambient temperatures. ANOVA revealed non-significant findings suggesting that the outliers displayed by WASO at 29 °C could not be explained by a drop in galvanic skin conductance. Should the manufacturer's algorithm be known, then discordance would be explained and accuracy could be improved. Interestingly, we recorded a higher EEG arousal rate at 22 °C and 29 °C than at 17 °C, although the difference was not statistically significant. Notably, hot exposure increased wake time during the sleep period [19]. However, we cannot be ascertained of this explanation, since the SWA's sleep–wake detection algorithm is not accessible. This higher WASO at 29 °C may reflect a significantly lower TST and SE at the same temperature (Tables 1 and 2, Fig. 1G and H). Similar to AW2, SWA also showed good sleep detection (93%) but poor wake detection (57%), in agreement with a previous validation study [10].

The good sleep epoch but poorer wake epoch detection may be explained by the greater amount of time spent immobile, so that both AW2 or SWA detect sleep with greater ease resulting in high sensitivity [2,10,20]. The kappa coefficient adjusts the amount of agreement that can be expected by chance. Since a high proportion of sleep epochs occur during the sleep period, this correction for chance may have led to a relatively lower kappa statistic, compared to percentage agreement [16].

An analytic limitation was that 30-s epochs were recorded for PSG and AW2, whereas 1-min epochs were recorded for SWA. Hence, in the sleep–wake epoch analysis, we divided the SWA outputs to match each 30-s epoch [10] set for PSG and AW2. This methodology was likely biased to show poor ability of the SWA to detect sleep or wake epochs. In addition, SWA was placed on the non-dominant arm for comparison purposes. However, placing the SWA on the left arm may not yield equivalent results, given that the manufacturer-driven study compared seven subjects that wore the SWA on the right arm with two subjects that wore it on the left arm. This issue may cause discordance with PSG.

In summary, AW2 showed minimal bias for the measurements of SOL, TST and SE at all three temperatures, but significantly overestimated WASO at 17 °C and 22 °C. SWA also showed minimal bias for WASO, TST and SE but severely underestimated SOL at 17 °C and 22 °C. In addition, SWA significantly overestimated WASO and underestimated TST and SE at 29 °C. Wake detection cannot be ascertained under all temperature conditions for both devices. In conclusion, the results of this study show similar validity for sleep detection of SWA and AW2, except that SWA is dependent on ambient temperature. Unlike sleep studies conducted in sleep clinics, home sleep studies are conducted under more variable temperature conditions. Hence, given the current findings, monitoring of bedroom temperature is considered crucial in home sleep studies. Future studies are instigated to evaluate the concordance rates in sleep assessment for patients with delayed sleep onset, higher level of WASO and lower SE (for example, insomnia patients who have difficulty initiating/maintaining sleep, or obstructive sleep apnea patients who have fragmented sleep).

REFERENCES

[1] Kanady JC, Drummond SPA, Mednick SC. Actigraphic assessment of a polysomnographic-recorded nap: a validation study. J Sleep Res 2011;20:214–22.

[2] de Souza L, et al. Further validation of actigraphy for sleep studies. Sleep 2003;26(1):81–5.

[3] Ward TM, et al. Polysomnography and actigraphy concordance in juvenile idiopathic ahritis, arthma and healthy children. J Sleep Res 2012;21:113–21.

[4] Van De Water ATM, 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.

[5] Weiss AR, et al. Validity of activity-based devices to estimate sleep. J Clin Sleep Med 2010;6(4):336–42.

[6] Sadeh A, Sharkey K, Carskadon M. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep 1994;17:201–7.

[7] Cole R, et al. Automatic sleep/wake identification from wrist activity. Sleep 1992;15:461–9.

[8] Peterson BT, et al. Comparison of actigraphy and polysomnography to assess effects of zolpidem in a clinical research unit. Sleep Med 2012;13:419–24.

[9] Soric M, et al. Validation of a multi-sensor activity monitor for assessing sleep in children and adolescents. Sleep Med 2012;14(2):201–5.

[10] O'Driscoll DM, et al. Energy expenditure in obstructive sleep apnea: validation of a multiple physiological sensor for determination of sleep and wake. Sleep Breath 2013;17:139–46.

[11] Iber C, et al. The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specification, 1st ed.. Illinois: American Academy of Sleep Medicine; 2007.

[12] Sunseri M, et al. The SenseWear armband as a sleep detection device. BodyMedia; 2009.

[13] Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 1995;346:1085–7.

[14] Krouwer JS. Why Bland–Altman plots should use X, not (Y+X)/2 when X is a refernece method. Stat Med 2008;27:778–80.

[15] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37–46.

[16] Wang D, et al. The validity of wrist actimetry assessment of sleep with and without sleep apnea. J Clin Sleep Med 2008;4 (5):450–5.

[17] Sharif MM, BaHammam AS. Sleep estimation using BodyMedia's SenseWear™ armband in patients with obstructive sleep apnea. Ann Thorac Med 2013;8(1):53–7.

[18] Catcheside PG, et al. Noninvasive cardiovascular markers of acoustically induced arousal from non-rapid-eye-movement sleep. Sleep 2002;25(7):797–804.

[19] Okamoto-Mizuno K, Tsuzuki K, Mizuno K. Effects of mild heat exposure on sleep stages and body temperature in older men. Int J Biometeorol 2004;49:32–6.

[20] Paquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. Sleep 2007;30(10):1362–9.