Physiology & Behavior 118 (2013) 185–188

Contents lists available at SciVerse ScienceDirect

Physiology & Behavior

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

Brief communication

Polysomnographic validation of a wireless dry headband technology for sleep monitoring in healthy young adults☆,☆☆

Lorenzo Tonetti [a,*], Nicola Cellini [b], Massimiliano de Zambotti [b,c], Marco Fabbri [d], Monica Martoni [e], Stephan E. Fábregas [f], Luciano Stegagno [b], Vincenzo Natale [a]

[a] Department of Psychology, University of Bologna, Bologna, Italy
[b] Department of General Psychology, University of Padua, Padua, Italy
[c] Center for Health Sciences, SRI International, Menlo Park, CA, USA
[d] Department of Psychology, Second University of Naples, Caserta, Italy
[e] Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy
[f] Zeo Sleep Research Center, Zeo, Inc., Newton, MA, USA

HIGHLIGHTS

- A wireless headband technology (WS) was compared with polysomnography (PSG).
- Results showed an agreement ranging from moderate to high.
- The WS underestimated wake after sleep onset and deep sleep compared to PSG.
- The WS overestimated the length of REM sleep compared to PSG.

ARTICLE INFO

Article history:
Received 25 June 2012
Received in revised form 27 March 2013
Accepted 17 May 2013
Available online 26 May 2013

Keywords:
Polysomnography
Sleep
Wireless dry headband technology
Ambulatory
Automated
Reliability

ABSTRACT

The present study aimed to explore the validity and reliability of a wireless dry headband technology for sleep monitoring (WS), through a comparison with concurrent polysomnographic (PSG) recording in healthy young adults. Eleven volunteers (7 females; mean age ± SD: 24.75 ± 3.62 years) took part in the study, wearing the WS for two overnight PSG recordings in the sleep laboratory. The WS was compared to PSG in the identification of wake, light, deep and REM sleep. The WS sensitivity and specificity were 97.6% and 56.1%, respectively. The WS agreement with PSG, measured by Cohen's kappa, was 0.56 for light sleep, 0.70 for deep sleep, and 0.67 for REM sleep. The present results showed that the agreement ranged from moderate to high between PSG and the WS, while wakefulness detection was observed to be a limitation of the WS.

1. Introduction

Polysomnography (PSG) is the gold standard in differentiating sleep from wake, and in identifying the following sleep stages: stage 1, stage 2, stage 3, stage 4, and REM [1]. However, recording PSG is labour-intensive and costly, making it difficult to use for long term and home sleep monitoring. A recent review [2] has been put forward on the objective alternatives to PSG for assessing sleep in non-laboratory settings. Although several objective methods have been developed to assess sleep patterns in a home setting (e.g., actigraphy, bed/mattress sensors, non-contact biomotion sensors), the review of Van de Water et al. [2] suggested a need to develop inexpensive and non-invasive methods for sleep assessments in non-laboratory settings, because previous methods (except for actigraphy) are still at a prototype level. One method that may meet the requirements suggested by Van de Water et al. [2] is a recently developed wireless dry headband technology (WS) for automatic sleep monitoring (Zeo, Inc., Newton, MA) [3]. The WS utilizes dry fabric sensors that are integrated into a headband that wirelessly transmits sleep data to a base station for processing in real time. Sleep stages (i.e., light sleep, deep sleep, and REM) are scored automatically by a neural network algorithm.

☆ Financial support disclosure: The authors declare that they received no financial grants to carry out the study.
☆☆ Conflicts of interest disclosure: Lorenzo Tonetti received the wireless systems to carry out the study from Zeo, Inc. Stephan E. Fábregas has been a salaried employee of Zeo, Inc. Nicola Cellini, Massimiliano de Zambotti, Marco Fabbri, Monica Martoni, Luciano Stegano and Vincenzo Natale declare the absence of any conflicts of interest.

* Corresponding author at: University of Bologna, Department of Psychology, Viale Berti Pichat 5, 40127 Bologna, Italy. Tel.: +39 051 2091877; fax: +39 051 243086.
E-mail address: lorenzo.tonetti2@unibo.it (L. Tonetti).

0031-9384/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.physbeh.2013.05.036

The WS performance has been compared with that of actigraphy to assess reliability in long-term home sleep monitoring [4]. To date, only one study has compared the WS sleep scoring to that of PSG made by human sleep experts, showing that the WS appears to yield results similar to PSG in healthy adults [3]. Specifically, as regards different sleep stages, the epoch-by-epoch agreement (assessed through the positive predictive values) between PSG scorers and the WS for light (between 80 and 82%) and REM sleep (between 79 and 85%) was higher than that for deep sleep (between 60 and 67%) [3]. Moreover in the same study [3] the WS detected a shorter REM sleep latency and more time in REM sleep, as compared to PSG. The aim of the present study was to explore the validity of the WS, comparing its performance with PSG in the laboratory, in a sample of healthy young adults.

2. Material and Methods

2.1. Participants

Participants were 11 healthy volunteers (7 females) (mean age ± SD, 24.75 ± 3.62 years). Participants were undergraduate students at the University of Padua and they were recruited through advertisements throughout the university facilities. Exclusion criteria included the presence of clinically significant insomnia, depression, and anxiety; moreover it was verified that participants had no debilitating physical conditions and that they were not smokers and/or pregnant. Eligibility was assessed by an anamnestic interview and completion of three screening questionnaires: the Insomnia Severity Index (ISI) [5], the Beck Depression Inventory-II (BDI-II) [6], and the State-Trait Anxiety Inventory (Form Y) (STAI Form Y) [7]. Aiming to select healthy university students, we used the following cut-off scores: as regards the ISI, scores less than 11 with the highest scores equal to 28; with reference to the BDI-II, scores less than 12 for males and 15 for females with maximum scores of 63; regarding the STAI Form Y, scores less than 49 in males and 54 in females, with the highest scores of 80. Selected participants reported the following mean scores for the three screening tools: 5.6 on the ISI, 7.1 on the BDI-II, and 37.6 on the STAI Form Y. The study protocol was approved by the local ethics committee and complied with the tenets of the Declaration of Helsinki. Each participant involved in the research project provided informed consent and received 100 Euros as compensation for his/her participation in the research.

2.2. Polysomnographic recording

The polysomnographic recording was done with the Siesta 802 PSG acquisition system (Compumedics, Abbotsford, Australia). Unipolar electroencephalograms were recorded according to the international 10–20 system [8]: F4-A1, F3-A2, C4-A1, C3-A2, O2-A1 and O1-A2. Submental bi-polar electromyogram and two unipolar electrooculograms were also recorded. All signals were sampled with a frequency of 512 Hz. We used Ag/AgCl electrodes embedded in an elasticized cap montage (Easy-Cap, Germany), with an impedance lower than 5kΩ. Sleep records were visually scored in 30-sec epochs according to Rechtschaffen and Kales [1] standard scoring criteria by two trained scorers who were blinded to the WS data. The trained scorers were two authors of the present study (NC and MDZ), who work in the same laboratory. Discrepancies in sleep scoring were discussed by the assessors and a single reconciled consensus score was obtained analyzing together the epochs with discrepant scores, using the Rechtschaffen and Kales [1] manual. The single consensus score was used for data analyses. Sleep onset latency (SOL) was defined as the interval between lights off and the first epoch of sleep stage 2.

2.3. Wireless dry sensor system recording

The WS (Zeo, Inc., Newton, MA) [3] utilized a headband with a single bi-polar dry fabric sensor which wirelessly transmitted data to a base station, located on the night table, near the bed. The WS headband used 3 silverized, conductive, fabric leads (including a single ground lead) to collect a single electrical signal composed of brain activity, muscle tone and eye movements. The electrical signal collected is very small (typically 5–100 microvolts), with a sampling frequency of 128 Hz. Specific features (e.g. low frequency, high amplitude delta waves that are indicative of deep sleep) are extracted using signal processing techniques such as a Fast Fourier Transform (FFT). These individual features, often associated with different stages of sleep, are transmitted to a base station trough an ultra-low-power proprietary wireless protocol at 2.4 GHz. A microprocessor within the base station determinates the sleep stage from the signal in real time through an artificial neural network algorithm. The neural network algorithm uses a combination of specific features (i.e., both time- and frequency-domain) derived from the signal to determine the best estimation of sleep stage intended to match those described by Rechtschaffen and Kales [1].

The algorithm was previously trained by the developers of the WS [3] by iteratively improving the correlation between the algorithmic sleep staging and the corresponding sleep stage scoring made by a human scorer on the basis of the Rechtschaffen and Kales criteria [1].

The algorithm allocates wake or one of three sleep stages (i.e., light – stages 1 and 2; deep – stages 3 and 4; and REM) to every 2 s of data. These data are then summarized into 30 s epochs.

The WS algorithm defined SOL as the interval from headband on (lights off) to the first epoch of sleep.

2.4. Procedure

All participants completed two consecutive overnight recordings in the sleep laboratory wearing the WS, after one acclimation night. Synchronization of PSG and the WS was achieved by having set both devices with the same time. The separate recording systems were synchronized to begin at the point of lights off clock time (e.g., 00:00 h) and to end at lights on clock time (e.g., 08:00 h).

2.5. Data analysis

Aiming to evaluate the criterion validity, we determined the sensitivity (i.e., ability of the WS to detect sleep when the PSG also scores sleep) and specificity (i.e., ability of the WS to detect wakefulness when the PSG also scores wakefulness). To this end, PSG and WS records were analyzed with binary scores (0 = sleep, 1 = wakefulness).

In order to assess the inter-rater reliability, Cohen's kappa [9] was calculated as the primary statistic, which indicated the percentage of the scoring agreement not due to chance, providing a correction for effects of varying biases upon scoring agreement. A kappa value of 0–0.2 indicates the lack of agreement, 0.2–0.4 low agreement, 0.4–0.6 moderate agreement, 0.6–0.8 high agreement and 0.8–1.0 almost perfect agreement [10]. Moreover, as extra information, the epoch-by-epoch agreement was performed for light, deep and REM sleep between the PSG and the WS, computing the positive predictive values (i.e., the probability that the scoring for a given epoch by the WS was in agreement to the corresponding scoring of the PSG made by trained researchers) for the whole night. Aiming to compare the sleep stages, PSG data scored as stage 1 and stage 2 were combined into a single category for comparison with the WS light sleep. Additionally, PSG data scored as stage 3 and stage 4 were assembled into a unique category for comparison with the WS deep sleep.

A Bland-Altman plot [11] was generated to display the absolute agreement of the WS with PSG, with reference to TST. The difference between PSG- and WS-TST was calculated. A positive value indicates

a WS underestimation of TST compared with PSG, while a negative value indicates a WS overestimation of TST.

Finally, aiming to compare the overall night analyses as measures of performance, we performed a set of Kolmogorov–Smirnov tests to verify the normal distribution of the sleep variables. Afterwards, we carried out dependent sample *t* tests, focusing on these sleep measures: SOL, REM sleep latency (interval between the sleep onset and the appearance of the first epoch of two continuous epochs of REM sleep; REML), wake after sleep onset (sum, in minutes, of all wake epochs between sleep onset and lights on; WASO), number of awakenings lasting at least 2 min (NA ≥ 2), total sleep time (sum, in minutes, of all sleep epochs between sleep onset and rise time; TST), sleep efficiency (ratio of the total sleep time to time in bed multiplied by 100; SE), time in light sleep (sum of all light sleep epochs; t_{light}), time in deep sleep (sum of all the deep sleep epochs; t_{deep}) and time in REM sleep (sum of all the REM sleep epochs; t_{REM}).

3. Results

We lost 1 out of 22 PSG recordings (4.5%) and 7 out of 22 WS records (31.8%) due to technical problems. In total, 14,943 epochs were analyzed.

The WS sensitivity and specificity were 97.6% and 56.1%, respectively.

For light sleep, the agreement between PSG and the WS measured through Cohen's k was .56 (i.e., moderate agreement). The Cohen's k value for agreement between the two devices in the identification of deep sleep was .70 (i.e., high agreement), while the Cohen's k for the agreement in the REM sleep detection was .67 (i.e., high agreement). The positive predictive values for sleep stage between PSG and the WS are shown in Table 1.

The absolute agreement in TST between PSG and the WS is shown in Fig. 1. The mean difference between PSG_TST and WS_TST was −23.29, which indicated a WS overestimation compared to PSG.

Carrying out the Kolmogorov–Smirnov tests, the frequency distributions of all the sleep variables were normal. Regarding SOL, NA ≥ 2 and t_{light}, the dependent sample *t* tests did not show significant differences between PSG and the WS (Table 2).

On one hand, the WS significantly underestimated REML ($t_{13} = 4.14$; $p < .01$), WASO ($t_{13} = 2.51$; $p < .05$) and t_{deep} ($t_{13} = 5.17$; $p < .001$) in comparison to PSG (Table 2). On the other, the WS significantly overestimated TST ($t_{13} = -2.37$; $p < .05$), SE ($t_{13} = -2.37$; $p < .05$) and t_{REM} ($t_{13} = -4.92$; $p < .001$) compared to PSG (Table 2).

4. Discussion

The WS sensitivity was high (i.e., 97.6%), indicating a good ability of the device to detect sleep when PSG also scores sleep, and was similar to that of actigraphy, the most used and validated alternative method to PSG [2] (97% sensitivity reported by de Souza et al. [12] and 94.8% by Jean-Louis et al. [13]). However the WS specificity (i.e., 56.1%) was low but at the same time higher than that of actigraphy (44% of specificity in the work by de Souza et al. [12] and 40.6% reported by Jean-Louis et al. [13]) suggesting that the WS is in an intermediate position between the gold standard (PSG) and actigraphy.

Cohen's kappa showed an agreement between PSG and the WS that was not due to chance ranging from moderate (light sleep) to high (deep and REM sleep) [10].

The positive predictive values were acceptable (i.e., 78% for light sleep and 90% for REM sleep), except for deep sleep detection (i.e., 66.1%). Our data are in line with those reported by Shambroom et al. [3] who showed higher agreement between PSG scorers and WS regarding light (between 80 and 82%) and REM sleep (between 79 and 85%) than deep sleep (between 60 and 67%).

The WS weakness in the assessment of wakefulness and deep sleep has also been observed when the nightly sleep measures were compared between the two devices. The WS underestimated WASO and deep sleep. Moreover, the WS underestimated the REML and overestimated the length of REM sleep compared to PSG. The discrepancy in the evaluation of REM sleep onset and REM sleep length could be due to the absence of independent electrooculogram channels in the WS. The discrepancy could also be caused by sensor placement (WS is around Fp1-Fp2), where there is not much alpha activity being recorded (which we would expect in O3-O4). This may also account for the WS's under-reporting of wake, which is heavily defined by alpha activity. Other than the significant level of the differences detected between PSG and WS in the sleep measures shown in Table 2, it is interesting to note that the widest difference (i.e., 81.61 min) has been observed in relation to the REML. These data could have relevant clinical implications because REML is known to be shorter in narcoleptic patients and thus the use of the WS should be avoided to screen for such disorder. On the other side, bearing in mind that SE is commonly used to assess the overall sleep quality,

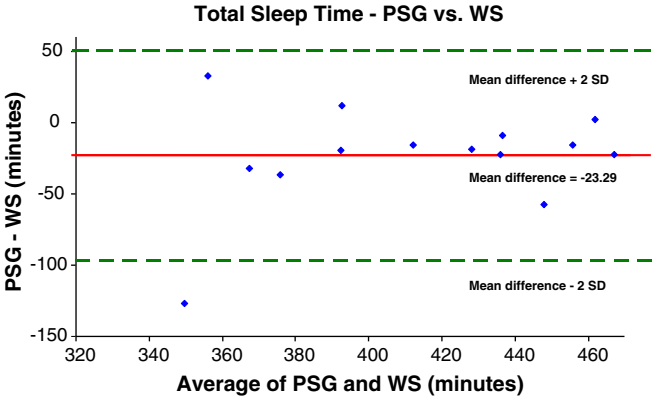

Fig. 1. Bland-Altman Plot for PSG_total sleep time (TST) vs. WS_TST. PSG refers to polysomnography and WS refers to the wireless dry headband technology for automatic sleep monitoring.

Table 1
Contingency tables for sleep stage between PSG and the WS. Row number of epochs and corresponding percentages are shown. The percentages in bold represent the positive predictive values.

	PSG			
		REM sleep	Light sleep	Deep sleep
WS	REM sleep	2.391	250	8
		90%	9.4%	0.3%
	Light sleep	969	5.328	241
		14.2%	**78%**	3.5%
	Deep Sleep	45	1.087	2.221
		1.3%	32.4%	**66.1%**

Table 2
Sleep measures (mean ± SD) for polysomnography (PSG) and the wireless system (WS). SOL refers to sleep onset latency (min); REML, REM sleep latency (min); WASO, wake after sleep onset (min); NA ≥ 2, number of awakenings lasting at least 2 min; TST, total sleep time (min); SE, sleep efficiency (%); t_{light}, time in light sleep (min); t_{deep}, time in deep sleep (min); t_{REM}, time in REM sleep (min.).

Sleep measure	PSG	WS
SOL	22.68 ± 13.74	15.86 ± 13.05
REML**	114.32 ± 60.12	32.71 ± 40.83
WASO*	51.57 ± 39.81	35.11 ± 33.22
NA ≥ 2	5.43 ± 4.94	4.79 ± 4.98
TST*	401.18 ± 47.96	424.46 ± 40.45
SE*	83.58 ± 9.99	88.43 ± 8.43
t_{light}	211.5 ± 29.80	223.04 ± 32.26
t_{deep}***	106.18 ± 32.86	80.46 ± 30.52
t_{REM}***	83.5 ± 19.01	120.96 ± 20.77

Significance: * $p < 05$; ** $p < .01$; *** $p < .001$.

it is comforting to note that this parameter showed the lowest difference (i.e., 4.85%) between PSG and WS.

Our data are in line with those reported by Shambroom et al. [3] with reference to SOL, REML, NA ≥ 2, t_{light} and t_{deep}, but disagree with their data as regards WASO, TST, SE and $t_{REM.}$ It is possible that this partial discrepancy could be due to the small sample size of both studies. Furthermore, the different mean age of our sample (i.e., 24.75 ± 3.62 years; age ranging between 19 and 30 years) in comparison to that of Shambroom et al. [3] (i.e., 38 ± 13 years; age ranging between 19 and 60 years), probably led to different polysomnographic sleep parameters between these two studies (see Table 2 of the present study and Table 3 reported in Shambroom et al. [3]).

The WS_TST overestimation compared to PSG, shown performing a set of dependent sample *t*-tests (Table 2), has been confirmed through the Bland-Altman Plot (Fig. 1). The WS overestimated TST compared to PSG regardless of short or long sleep duration and only one outlier was observed (Fig. 1). This overestimation is probably due to a reduced ability of the WS to detect wake during sleep, as shown by the significant WASO underestimation of the WS compared to PSG as well as by the low WS specificity; this low detection of wake could be related to the inability of the WS to detect alpha activity from the front of the brain.

We have to underline that seven WS recordings have been lost due to technical problems, which mainly occurred when the headband fell off participants' heads during the night. These data are in line with those reported by Griessenberger and colleagues [14] that have compared the staging accuracy of the WS with that of PSG recordings made by the semiautomatic scoring of Somnolyzer 24x7 [15], that served as their study standard, in a sample of 10 adults (mixed sample of controls and insomniac patients). They lost half of WS recoding nights mostly because participants lost the headband during the night. It is possible that this relatively high failure rate in both studies can be due to the presence of the electrodes for the EEG recording, that could facilitate the loss of the headband. The results of the study that compared WS with actigraphy [4] are in line with this explanation, because the failure rate, without the EEG electrodes, was much lower (i.e., 7.14%).

In order to improve performance of the WS in the identification of wakefulness, deep and REM sleep, it may be necessary to modify the WS algorithm in these areas. However, it is possible that due to the intrinsic features of the WS (i.e., the base station, the dry headband, the location of the sensor at Fp1-Fp2 that is not a good place to pick up alpha activity), it could prove challenging to obtain better algorithm performance.

5. Conclusions

To sum up, in comparison to the available data about actigraphy, the WS showed similar sensitivity and higher specificity, suggesting that it could be potentially used as an alternative to actigraphy when sleep architecture should be detected. The reliability testing (Cohen's kappa) showed an agreement ranging from moderate to high between PSG and the WS. Nevertheless the differences observed regarding sleep measures shed light on the necessity to use the WS with an understanding of its limitations, in particular the weaknesses in the identification of wake and deep sleep.

References

[1] Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Washington: National Institute of Health; 1968 .
[2] Van de Water ATM, Holems 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.
[3] Shambroom JR, Fábregas SE, Johnstone J. Validation of an automated wireless system to monitor sleep in healthy adults. J Sleep Res 2012;21:221–30.
[4] Tonetti L, Fábregas SE, Fabbri M, Occhionero M, Erbacci A, Martoni M, et al. Comparison of a wireless dry headband technology for sleep monitoring with actigraphy in healthy adults. Biol Rhythm Res 2013;44:333–8.
[5] Morin C. Insomnia: psychological assessment and management. New York, NY: Guilford Press; 1993 .
[6] Beck AT, Steer RA, Brown GK. Beck depression inventory—second edition: manual. San Antonio, TX: The Psychological Corporation; 1996 .
[7] Spielberger CD, Gorush RL, Lushene RE, Vagg PR, Jacobs GA. Manual for the State-Trait Anxiety Inventory (Form Y): self-evaluation questionnaire. Palo Alto, CA: Consulting Psychologists Press, Inc.; 1983 .
[8] Jasper HH. The ten–twenty electrode system of the International Federation. Electroencephalogr Clin Neurophysiol 1958;10:367–80.
[9] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
[10] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[11] Bland JM, Altman DG. Statistical method for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
[12] de Souza L, Benedito-Silva AA, Nogueira Pires ML, Poyares D, Tufik S, Calil HM. Further validation of actigraphy for sleep studies. Sleep 2003;26:81–5.
[13] Jean-Louis G, Kripke DF, Cole RJ, Assmus JD, Langer RD. Sleep detection with an accelerometer actigraph: comparisons with polysomnography. Physiol Behav 2001;72:21–8.
[14] Griessenberger H, Heib DPJ, Kunz AB, Hoedlmoser K, Schabus M. Assessment of a wireless headband for automatic sleep scoring. Sleep Breath 2013;17:747–52.
[15] Anderer P, Moreau A, Woertz M, Ross M, Gruber G, Parapatics S, et al. Computer-assisted sleep classification according to the standard of the American Academy of Sleep Medicine: validation study of the AASM version of the Somnolyzer 24 × 7. Neuropsychobiology 2010;62:250–64.