ORIGINAL ARTICLE

Validation of Photoplethysmography-Based Sleep Staging Compared With Polysomnography in Healthy Middle-Aged Adults

Pedro Fonseca, MSc[1,2]; Tim Weysen, MSc[1]; Maaike S. Goelema, MSc[1,3]; Els I.S. Møst, PhD[1]; Mustafa Radha, MSc[1,2]; Charlotte Lunsingh Scheurleer, MSc[1,3]; Leonie van den Heuvel, MSc[1]; Ronald M. Aarts, PhD[1,2]

[1]*Philips Group Innovation Research, Eindhoven, The Netherlands;* [2]*Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands;* [3]*Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands*

Study Objectives: To compare the accuracy of automatic sleep staging based on heart rate variability measured from photoplethysmography (PPG) combined with body movements measured with an accelerometer, with polysomnography (PSG) and actigraphy.

Methods: Using wrist-worn PPG to analyze heart rate variability and an accelerometer to measure body movements, sleep stages and sleep statistics were automatically computed from overnight recordings. Sleep–wake, 4-class (wake/N1 + N2/N3/REM) and 3-class (wake/NREM/REM) classifiers were trained on 135 simultaneously recorded PSG and PPG recordings of 101 healthy participants and validated on 80 recordings of 51 healthy middle-aged adults. Epoch-by-epoch agreement and sleep statistics were compared with actigraphy for a subset of the validation set.

Results: The sleep–wake classifier obtained an epoch-by-epoch Cohen's κ between PPG and PSG sleep stages of 0.55 ± 0.14, sensitivity to wake of 58.2 ± 17.3%, and accuracy of 91.5 ± 5.1%. κ and sensitivity were significantly higher than with actigraphy (0.40 ± 0.15 and 45.5 ± 19.3%, respectively). The 3-class classifier achieved a κ of 0.46 ± 0.15 and accuracy of 72.9 ± 8.3%, and the 4-class classifier, a κ of 0.42 ± 0.12 and accuracy of 59.3 ± 8.5%.

Conclusions: The moderate epoch-by-epoch agreement and, in particular, the good agreement in terms of sleep statistics suggest that this technique is promising for long-term sleep monitoring, although more evidence is needed to understand whether it can complement PSG in clinical practice. It also offers an improvement in sleep/wake detection over actigraphy for healthy individuals, although this must be confirmed on a larger, clinical population.

Keywords: Photoplethysmography, sleep tracker, actigraphy, computerized analysis, heart rate variability, scoring, statistics.

Statement of Significance

Polysomnographic studies are invaluable in the assessment and diagnosis of sleep disorders. However, this specialized and labor-intensive procedure is less suitable for long-term monitoring. Recent years have shown an explosion in the availability of wearable devices which claim to be able to track sleep and depict sleep architecture. Although practically well suited for long-term home sleep monitoring, the performance of most of these devices remains unpublished. We evaluated the sleep annotation performance of a wrist-worn photoplethysmography device in a population of healthy middle-aged adults. Further validation is needed for other age groups and sleep pathologies.

INTRODUCTION

Sleep polysomnography (PSG) is considered the gold standard for objectively evaluating sleep and is the preferred methodology to diagnose sleep disorders in clinical practice and often used in research trials. However, traditional in-lab PSG setups which comprise electroencephalography (EEG) have several limitations, such as their high cost, they are labor intensive, and they produce an impact to the patient negatively affecting sleep.[1] Yet, the major disadvantage of PSG is that the methodology is not well suited for long-term monitoring beyond one or two nights.

Actigraphy is a methodology more suited for long-term monitoring of sleep and is highly appropriate for examining sleep variability.[2] The American Academy of Sleep Medicine (AASM) indicated actigraphy as a suitable method to assist in the evaluation of patients with circadian disorders and sleep–wake disturbances and also to assess response to therapy of circadian disorders and insomnia.[3] The relative simplicity of actigraphy, together with the maturity and low-cost of accelerometer sensors, have fueled a growing trend in the area of health self-tracking, with many consumer devices in this area offering sleep monitoring as a feature of their products.[4–6] The first group of consumer sleep tracking (CST) devices, such as Jawbone Up (Jawbone, San Francisco, CA, USA), or Fitbit Ultra (Fitbit Inc., San Francisco, CA, USA), uses actigraphy to estimate periods of sleep and wake. However, because it relies exclusively on the measurement of gross body movements, actigraphy lacks the ability to describe sleep architecture. Furthermore, studies which validated actigraphy against PSG have shown that it tends to overestimate sleep as it is not well equipped to detect wakefulness when lying quietly.[7–9] The few CST devices that have been validated against PSG, such as the Jawbone Up[10] and the Fitbit Ultra (Fitbit Inc., San Francisco, CA, USA),[11,12] were found to overestimate total sleep time (TST) and sleep efficiency (SE) as compared to PSG,[10,12] confirming the general drawback of actigraphy-based sleep measurements.

The second group of more modern CST devices include, besides accelerometers, heart rate monitors based on reflective photoplethysmography (PPG). Devices such as the Jawbone Up3 (Jawbone, San Francisco CA, USA), Fitbit Alta HR (Fitbit Inc., San Francisco, CA, USA) or the now discontinued Basis Peak (Intel, San Francisco, CA, USA) claim, besides sleep and wake tracking, the ability to depict sleep architecture. However, nearly no validation information for these devices is available, and questions with respect to the accuracy of modern CST devices remain unanswered.[4–6]

Despite the lack of evidence for CST devices, the physiological link between cardiac activity, for example, indirectly measured with PPG, and sleep is relatively well understood. PPG is a technique whereby reflected or transmitted light shone on the skin is measured by a photosensor. Light is absorbed by the skin, by venous blood, and also by arterial blood and besides a slow-varying DC component, the measured signal has a pulsatile component related to the arterial blood volume

changes due to cardiac activity. Transmissive PPG, typically mounted on the finger tip, has been extensively used in various clinical applications and has been commercially available in pulse oximeters, vascular diagnostic tools, and beat-to-beat blood pressure devices for the past four decades.[13] When mounted on the wrist, reflective PPG (simply referred to as PPG in the remainder of the manuscript) measures blood volume changes in the microvascular bed of tissue of that part of the body. These sensors have been shown to accurately measure average heart rate, typically calculated over windows of a few seconds.[14] They also offer, in theory, the possibility of measuring heart rate variability (HRV) throughout the night based on the analysis of the distance between consecutive heart beats detected in the PPG signal. The relation between HRV, expressing characteristics of the autonomic nervous system (ANS) and its sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) divisions, and sleep stages has been extensively described in literature. For example, as nonrapid eye movement (NREM) sleep progresses from N1 to N3, there is an accompanying increase in cardiovagal drive and PNS activity[15,16] and a reduction in cardiac and SNS activity, translating to a decrease in heart rate and an increase in the respiratory mediation of HRV, visible in the high-frequency band (HF, 0.15 to 0.4 Hz), as compared to wake.[16] In contrast, rapid eye movement (REM) sleep is a state of autonomic instability where PNS and SNS activity fluctuate, producing abrupt changes in heart rate. The average heart rate and the power in the low-frequency band (LF, 0.04 to 0.15 Hz) of HRV is higher during REM than during NREM sleep, and there is a shift of the LF/HF ratio toward sympathetic dominance.[16,17] It has also been shown that these relations extend beyond just a different autonomic control for different sleep stages but that they are in fact continuously associated with different EEG characteristics during sleep. The power in the LF band and the LF/HF ratio, for instance, has been found to be correlated with delta EEG power,[18] and a high coherency was found between the power in the HF band and the power in the delta, theta, alpha, sigma, and beta EEG bands,[19] with nonlinear interactions between delta, theta, alpha, and HF power.[20] Machine learning techniques have successfully exploited the relations between HRV and sleep stages. Several algorithms were shown to be able to automatically score sleep stages based on HRV, typically measured with electrocardiogram (ECG), often in combination with respiratory effort[21–24] achieving moderate performance. Heart rate monitors embedded in CST devices could, in theory, enable similar PPG-based HRV characteristics to be measured, and, in combination with body movements measured with the accelerometers typically available in those devices, augment sleep–wake detection—already feasible with actigraphy—with more detailed sleep architecture information.

The current study has two objectives:

- To compare an automatic sleep staging method based on PPG-based HRV and body movements, with PSG.
- To evaluate whether the addition of PPG-based HRV to body movement information improves the estimation of sleep–wake statistics as compared with traditional wrist-worn actigraphy in the same task.

METHODS

The current study was based on three separately collected data sets, used to train the sleep staging algorithm ("Training set") and to validate it ("Validation set 1," "Validation set 2)."

Training Data Set

The training data set consists of a subset of the data collected during the SIESTA project in the period from 1997 to 2000[25] and which included 165 healthy participants with a mean ± standard deviation (SD) age of 46.7 ± 7.7 years. The SIESTA study was carried out in five different countries in seven different sleep laboratories. The study was approved by the local ethical committee of each research group. None of the participants were using or had a history of drug and/or alcohol use, were working at night, or had been diagnosed with a medical (or mental) disorder interfering with the aim of the study. All participants had a Mini-Mental State Examination score[26] >25, a Pittsburgh Sleep Quality Index (PSQI) score[27] of <6 and a bed time between 10:00 pm and 12:00 am. The total duration of the data collection per participant in the SIESTA data set was 15 days. At day 7 and day 8 participants were invited to sleep in the sleep laboratory to collect overnight PSG. Two hours before their habitual bed time, the participants were asked to perform a battery of psychometric tests and afterward the PSG was applied. The next morning, the PSG setup was removed and after washing, getting dressed, and breakfast, the same psychometric test battery was completed by the participants. All participants gave informed consent before participation. The PSG recordings of all participants were scored by two trained somnologists from different sleep centers and revised by a third expert who took the final decision in case of disagreement. More details regarding participants and study design were described by Klosh et al.[25]

For the purpose of our study, a total of 135 recordings of 101 healthy sleepers (57 females) were included, with an age range between 20 and 83 years. Additional demographics can be found in Table 1.

Validation Data Sets

The sleep staging method was validated with two hold-out data sets. None of the participants in the validation sets were part of the training set.

Participants

Validation set 1 was collected in 2014 and consisted of 16 healthy participants (eight females) with a mean age of 51.2 ± 8.4 years. The study lasted five nights, three measured at home which only included actigraphy and two in a hotel which included overnight PSG and PPG.

Validation set 2 was collected in 2015 and consisted of 35 healthy participants (20 females) with a mean age of 52.0 ± 6.9 years. The study lasted for 16 days, including 2 weeks of home monitoring including actigraphy and PPG and two nights at the end of the study in a hotel which included overnight PSG, PPG, and actigraphy.

Both studies included participants with no primary history of neurological, cardiovascular, psychiatric, pulmonary,

Table 1—Participant Demographics for the Set Used to Train the Algorithm (Training Set) and for the Two Hold-Out Sets Used to Validate the Algorithm (Validation Set 1 and Validation Set 2).

Parameter	Training set		Validation set 1		Validation set 2	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
N	101 participants, 135 recordings		16 participants, 26 recordings		35 participants, 54 recordings	
Sex	57 female participants (56.4%), 76 female recordings (56.3%)		8 female participants (50.0%), 13 female recordings (50.0%)		20 female participants (57.1%), 34 female recordings (63.0%)	
Age (year)	44.2 (17.2)	[20, 83]	51.2 (8.4)	[41, 66]	52.0 (6.9)	[41, 64]
BMI (kg/m^2)	23.9 (3.1)	[17.2, 31.3]	25.9 (2.8)	[20.90, 29.86]	26.2 (4.0)	[19.27, 36.23]

Abbreviations: BMI, body mass index; SD, standard deviation.

endocrinological, or sleep disorders. In addition, none of the participants were using sleep, antidepressant or cardiovascular medication, recreational drugs, excessive amounts of alcohol, nor were they pregnant or working in shifts, nor crossing more than two time zones in the 2 months prior to the investigation. All participants of both studies had a body mass index lower than 40 and a PSQI lower than 6.

The two studies were approved by the Internal Committee of Biomedical Experiments of Philips Research and were conducted in accordance with the Declaration of Helsinki. All participants gave informed consent before participation. The hotel gave approval to conduct the experiment on their premises.

Only the hotel night recordings (the last two nights in each study) were used because they comprised a complete PSG, needed to validate the sleep staging algorithm.

Procedure

Participants arrived at the hotel at 08:00 pm and were informed about the procedure. The recording devices were applied and participants were free to choose how to spend the time until their habitual bed time. Bed and lights off times were registered by the researcher, as well as wake up time and lights on time. Participants were free to choose how to spend the time between the two recording nights. The researcher only monitored the beginning and the end of the recording but not the entire night. The researcher was available at the hotel for the duration of both recording nights, and the participants could call the researcher during the night using the hotel room phone.

Recording Devices

Bed and wake times were logged by participants using the consensus sleep diary.[28] PSG was recorded with the Philips Respironics Alice PDx system (Philips Respironics Inc., Murrysville, PA, USA) using the standard 10–20 system electrode placement with a referential system montage, assisted by BraiNet (Jordan Neuroscience Inc., Redlands, CA, USA). The sleep scoring montage included the minimum set of three EEG channels recommended by AASM for offline scoring[29] (F4-A1, C4-A1, O2-A1) plus a backup electrode (C3-A2), left electrooculogram (LOC-A2), right electrooculogram (ROC-A1), a bipolar submental electromyogram (EMG1–EMG2), an ECG (modified lead II), and two respiratory inductance plethysmography (RIP) belts mounted around the thorax and the abdomen (RESP1 and RESP2).

In addition to PSG, a CE-marked logging device containing a PPG and three-axial accelerometer sensors (Royal Philips, Amsterdam, the Netherlands) was used during the same recording periods. The logging device was mounted on the nondominant wrist of the participant, with the sensor facing the skin on the dorsal side of the hand, above the ulnar styloid process.

Finally, the recordings of validation set 2 included actigraphy measured with Actiwatch Spectrum (Philips Respironics Inc., Murrysville, PA, USA) which uses a piezoelectric accelerometer to detect and log limb movements. It was worn on the dominant wrist of the participant, configured to measure activity counts in epochs of 30 seconds.

Analysis

The PSG data were analyzed by an external experienced somnologist (SleepVision, Nijmegen, the Netherlands) blind to the health condition of the participants and to the purpose of the study, and the sleep stages were manually scored in 30-second epochs according to the AASM guidelines.[29] No scoring of respiratory events was performed and the PSG did not highlight any sign of sleep disorders with any of the 51 participants. Table 2 indicates the PSG sleep statistics from the 80 recordings of the 51 participants in the two validation sets.

PPG-Based Sleep Stage Classification

In our earlier work,[24] we presented a machine learning approach to sleep staging based on HRV measured from ECG and respiratory effort measured from RIP. The system validated in the present paper uses a similar approach and similar set of HRV features, albeit computed from interbeat intervals detected from PPG instead of ECG. In short, the HRV feature set consists of a combination of time-domain features such as sample statistics of heart-beat interval durations and consecutive differences,[30] arousal likelihood ratios based on consecutive heart beats,[31] multiscale sample entropy[32] and detrended, progressive, and windowed fluctuation analysis of heart-beat intervals,[33] and measures of cardiorespiratory interaction based on visibility graphs.[34] The feature set also comprises frequency-domain features such as the spectral powers in the very low, low, and HF bands, with[35] and without[30] adapted spectral boundaries. HRV features were combined with features related to body movements, calculated based on the three-axial accelerometer signal.

Table 2—PSG Sleep Statistics, and Bias, 95% Limits of Agreement, and Root Mean Squared Error Between PPG and PSG in 80 Recordings of 51 Middle-Aged Adults.

Statistic	PSG		PPG–PSG		
	Mean (SD)	Range	Mean error (SD)	95% LoA	RMSE
SOL (minutes)	15.53 (8.23)	[3.50, 39.50]	−6.80 (7.69)	[−21.88, 8.28]	10.23
WASO (minutes)	33.76 (29.63)	[2.00, 164.50]	−3.19 (25.28)	[−52.74, 46.36]	25.32
TWT (minutes)	55.32 (36.04)	[14.00, 218.00]	−13.40 (31.74)	[−75.60, 48.80]	7.37
TIB (minutes)	462.42 (43.15)	[329.00, 531.50]	n/a	n/a	n/a
TST (minutes)	407.10 (50.13)	[232.00, 504.50]	13.40 (31.74)	[−48.80, 75.60]	34.27
SE (%)	88.10 (7.67)	[51.56, 96.93]	2.90 (6.82)	[−10.46, 16.26]	7.37
Time in N1 + N2 (minutes)	208.91 (44.58)	[107.00, 310.50]	−28.33 (42.22)	[−111.09, 54.42]	50.63
Time in N3 (minutes)	107.22 (34.48)	[39.50, 188.00]	0.26 (38.32)	[−74.85, 75.37]	38.08
Time in REM (minutes)	90.97 (28.87)	[28.00, 150.50]	41.47 (33.62)	[−24.43, 107.37]	53.25

Abbreviations: LoA, limits of agreement; PPG, photoplethysmography; PSG, polysomnography; REM, rapid eye movement; RMSE, root mean squared error; SD, standard deviation; SE, sleep efficiency; SOL, sleep onset latency; TIB, time in bed; TST, total sleep time; TWT, total wake time; WASO, wake after sleep onset.

Because no respiratory information is readily available from the wrist-worn sensors used, respiratory features were left out.

The sleep staging algorithm was trained with the same machine learning techniques described in[24], using the data of the training set (Table 1). Because no PPG data were available in the training set, the HRV features used to train the algorithm were estimated from ECG. Because the features are based on the distance between consecutive R-R intervals, they are essentially equivalent (in the absence of cardiovascular conditions) when computed from consecutive pulses measured from PPG. Using PPG-derived HRV, and body movement features, the algorithm was validated with a hold-out validation procedure on both validation sets (Table 1).

The sleep staging algorithm provides an estimation of the sleep stage (wake, combined N1 and N2, N3 and REM) of each 30 seconds epoch. Based on this automatic scoring, overall sleep–wake statistics such as sleep onset latency (SOL), wake after sleep onset (WASO), total wake time (TWT), TST, and SE as well as the percentage of each sleep stage were computed for each overnight recording.

Comparison With PSG

In order to validate the output of the sleep staging algorithm, the clocks of the PPG sensor and of the PSG needed to be synchronized. This was achieved by compensating the clock offset and drift based on an interpolated (at 4 Hz) time series comprised of the interbeat intervals detected from the PPG and another series from the interbeat intervals calculated from R-peaks detected from the ECG signal from the PSG. This guaranteed that the ground-truth reference and the output of the sleep staging algorithm were perfectly aligned to the start time and that there was no clock drift between the two recorders.

After synchronizing both signals, the hypnogram derived from PSG annotations (henceforth referred to as "reference hypnogram") and the hypnogram estimated by the PPG algorithm (henceforth referred to as "estimated hypnogram") were restricted to the period between lights off in the evening and lights on in the morning. The estimated hypnogram was compared with the reference hypnogram both in terms of epoch-by-epoch agreement, as well as in terms of sleep statistics calculated for each overnight recording.

Regarding the epoch-by-epoch agreement, two quality metrics were used: accuracy and Cohen's kappa coefficient of agreement[36] (or κ, in short). Accuracy indicates the percentage of 30-second epochs with correctly classified sleep stages. κ indicates the agreement between the two hypnograms corrected for agreement by chance. Though harder to interpret, the latter is more informative because of the imbalance in the occurrence of different sleep stages. κ is usually interpreted with the following terms: below 0.20 "slight agreement," 0.21 to 0.4 "fair agreement", 0.41 to 0.60 "moderate agreement," 0.61 to 0.8 "substantial agreement," and above 0.81 "almost perfect agreement."

Regarding the sleep statistics, each measure was evaluated by calculating the mean and SD of the error and the root mean squared error (RMSE) between the estimation obtained with the estimated hypnogram and the statistics calculated from the reference hypnogram. In addition, a Bland–Altman analysis of each statistic was performed, by calculating and plotting the mean and SD of the differences and the corresponding 95% limits of agreement (LoA) as the mean difference ± 1.96 times the SD. A positive mean difference value indicates that the algorithm tends to overestimate a particular statistic in comparison with reference PSG, while a negative value means that statistic is underestimated.

Similar to related work,[10,37,38] we defined a priori a satisfactory agreement if differences between the estimated and the PSG reference for WASO, TWT, and TST were smaller than 30 minutes and if the differences for SE were below 5%. Because the differences between PPG and PSG do not follow a normal distribution, the 99.9% confidence intervals (CIs) for these

errors were calculated with a nonparametric method based on Wilcoxon's signed rank test.[39] Statistically significant overall satisfactory agreement was considered (at $p < .001$) if both boundaries of the 99.9% CI were (in absolute terms) smaller than 30 minutes (for WASO, TWT, and TST) and 5% (for SE).[40]

It should be noted that the sleep statistic time in bed was not computed because the analysis was restricted to the lights off period according to protocol, corresponding not exactly to the time spent in bed but rather to the period of time when the participants had the intention to sleep.

Comparison With Actigraphy

Actigraphy-based sleep–wake statistics (SOL, WASO, TWT, TST, and SE) and epoch-by-epoch sleep–wake classification were estimated using the Actiware software (Philips Respironics Inc., Murrysville, PA, USA) with the default sensitivity settings (medium, 40).

The sleep–wake statistics were evaluated against PSG on the 53 recordings of validation set 2 for which actigraphy was available. The results were compared with the sleep–wake statistics obtained with PPG for the same recordings. A Wilcoxon signed-rank test was used to compare the estimation errors obtained with actigraphy and with PPG.

In order to evaluate the performance of epoch-by-epoch classification, the clocks of the actigraphy device and the PSG were synchronized. This was achieved by computing a surrogate measure of actigraphy from the artifacts in the respiratory effort signal recorded with PSG[41] and then finding the clock offset that maximized the correlation between the actigraphy and the surrogate actigraphy signals. All comparisons were manually reviewed to guarantee that the clocks were precisely synchronized. Because this method depended on the availability of a valid respiratory effort signal, it could only be applied on 49 recordings of validation set 2 and the comparison between PPG and actigraphy was restricted to that subset. The same metrics of κ, accuracy, specificity, and sensitivity (to wake) were used to evaluate the epoch-by-epoch classification agreement with ground-truth annotations from PSG. A Wilcoxon signed-rank test was used to compare the performance obtained with actigraphy and with PPG.

RESULTS

PPG-Based Sleep-Wake Statistics

SOL (minutes), WASO (minutes), TWT (minutes), TST (minutes), and SE (%) were computed from both the human annotation of the recordings as well as from the automated scoring by the PPG-based algorithm. Table 2 summarizes the mean differences, the 95% LoA, and the RMSE for these statistics, Figure 1 and Figure 2 illustrate the Bland–Altman analyses of the differences for SOL, WASO, and TWT and of TST and SE, respectively. The algorithm underestimates, in average, SOL and WASO by less than 10 minutes and TWT by less than 15 minutes and overestimates TST by less than 15 minutes and SE by less than 5% in comparison with PSG. The SD of the error for SOL is below 10 minutes and for SE, below 10%, but for the remaining sleep/wake statistics, WASO, TWT, and TST, it is close to 30 minutes. Closer inspection of the Bland–Altman plots of SOL and WASO reveals that despite the relatively low bias, and for SOL, the SD of the error, it tends to underestimate wake time when SOL and WASO increase beyond 15 minutes.

The algorithm achieves a satisfactory agreement according to the a priori defined margin of error of 30 minutes for 77.5% of the recordings for TST (62 recordings), 77.5% for TWT (62 recordings), 81.2% for WASO (65 recordings), and according to the a priori margin of 5% for 71.2% of the recordings for SE (57 recordings). Both boundaries of the 99.9% CI for TST ([−0.75, 21.00] minutes), TWT ([−21.00, 0.75] minutes), and WASO ([−10.00, 8.00] minutes) were contained within the priori interval of (−30, 30) minutes and the a priori defined satisfactory agreement can be statistically established ($p < .001$) for all statistics. Both boundaries of the 99.9% CI for SE ([−0.13, 4.61] %) were also contained within the a priori interval (−5, 5) % and the a priori defined satisfactory agreement can also be statistically established ($p < .001$).

PPG-Based Sleep Staging Performance

The mean differences, LoA, and RMSE for the duration of sleep stages N1 + N2, N3, and REM are given in Table 2. Figure 3 illustrates the Bland–Altman analyses of the differences for these sleep stages. The best agreement was obtained for the estimation of time in N3, time in N1 + N2 was underestimated,

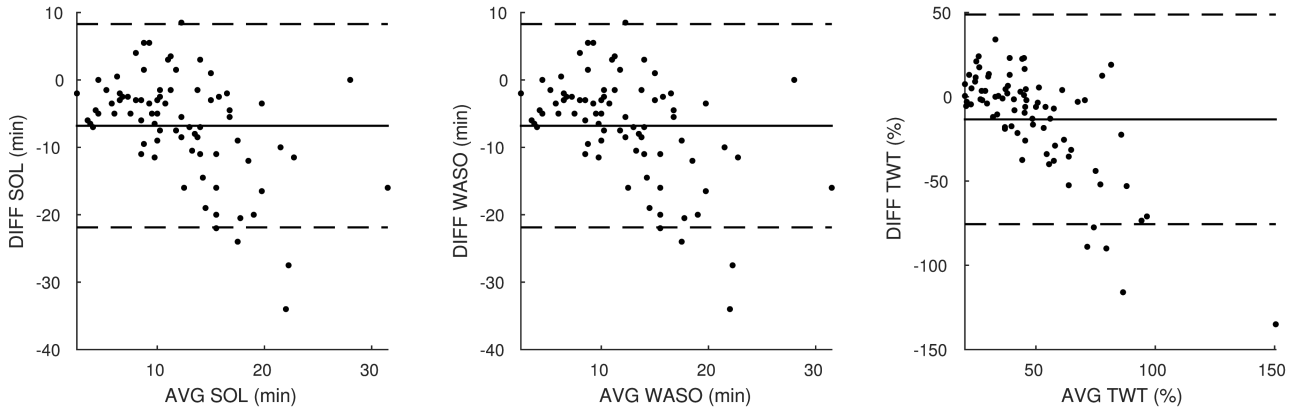

Figure 1—Bland–Altman plots for wake statistics. DIFF and AVG correspond to the difference and average between PPG and PSG, respectively. From left to right: sleep onset latency (SOL), wake after sleep onset (WASO), total wake time (TWT). PPG, photoplethysmography; PSG, polysomnography.

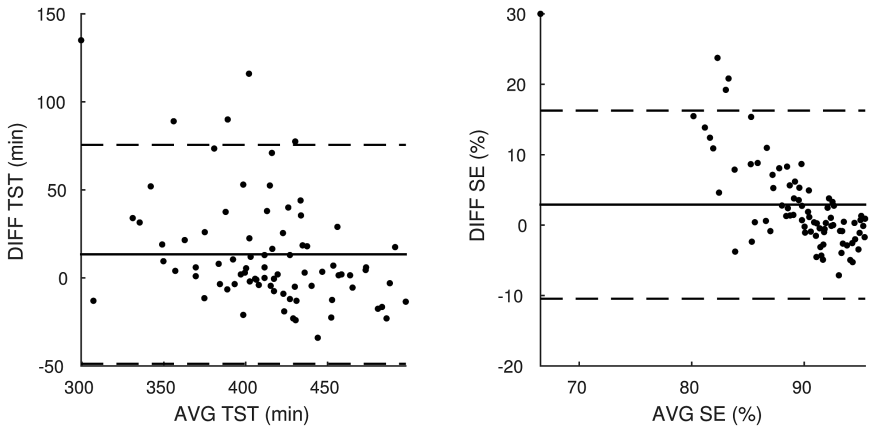

Figure 2—Bland–Altman plots for sleep-wake statistics. DIFF and AVG correspond to the difference and average between PPG and PSG, respectively. From left to right: total sleep time (TST), sleep efficiency (SE). PPG, photoplethysmography; PSG, polysomnography.

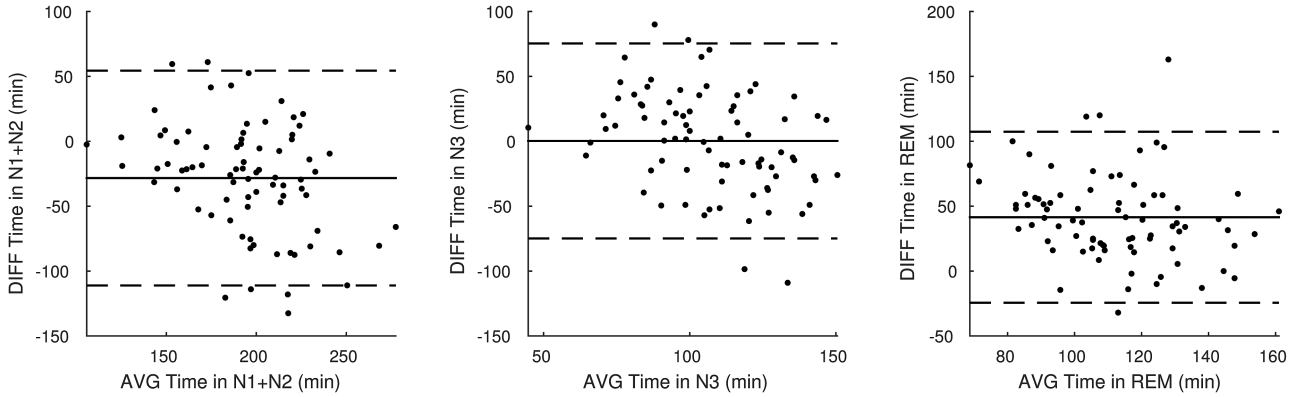

Figure 3—Bland–Altman plots for sleep stage statistics. DIFF and AVG correspond to the difference and average between PPG and PSG, respectively. From left to right: time in N1 + N2, time in N3, time in REM. PPG, photoplethysmography; PSG, polysomnography; REM, rapid eye movement.

in average, by nearly 30 minutes, and time in REM was overestimated by slightly more than 40 minutes.

The agreement between the automated system and the human annotation was also considered in terms of 30-second epoch-by-epoch agreement. Multiple cases were considered: first, the system was used for 4-class classification to distinguish between wake, REM, combined N1 + N2, and N3. In this case, the algorithm achieved a moderate agreement with a κ of 0.42 and accuracy of 59.3%. In the second case, the system was used for 3-class classification to distinguish between wake, REM, and NREM (combining N1, N2, and N3) and the κ increased to 0.46 and accuracy to 72.9%. Finally, the system was used for 2-class classification, distinguishing only between wake and sleep (combining all sleep classes), achieving a κ of 0.55 and accuracy of 91.5%.

Table 3 shows the κ and accuracy for these three cases. In addition, the table also shows κ, accuracy, sensitivity, and specificity in the detection of separate sleep stages. The worst single-class detection performances were obtained for N1 + N2 followed by REM. This is in line with the Bland–Altman analysis of these two statistics, illustrated in Figure 3.

Actigraphy-Based Sleep–Wake Classification Performance And Comparison With PPG

Regarding the error between the sleep–wake statistics obtained with actigraphy and with PSG, indicated in Table 4, SOL and WASO were underestimated by less than 10 minutes, TWT by less than 20 minutes, TST was overestimated by slightly more than 10 minutes, and SE by less than 5%. Table 4 also indicates the error calculated between the sleep–wake statistics estimated with PPG and PSG for the same subset of validation set 2. Although there is a decrease in the bias for SOL, WASO, and TWT when estimating the sleep–wake statistics with PPG as compared with actigraphy, only the decrease in the SOL estimation error is significant ($p < .05$).

Regarding the epoch-by-epoch sleep–wake classification performance, Table 5 indicates the results obtained between actigraphy and PSG, and between PPG and PSG for a subset of validation set 2. The κ obtained with PPG (0.55) is significantly ($p < .001$) higher than obtained with actigraphy (0.40). Additionally, there is a significant ($p < .001$) increase in sensitivity with PPG (57.4%) when compared with actigraphy (45.5%). There were no significant differences for accuracy and specificity, both above 90% for actigraphy and for PPG.

Table 3—Epoch-by-Epoch Agreement Between PSG and PPG: Mean (Standard Deviation) of Overall and per Class κ and Accuracy, and per Class Sensitivity and Specificity.

Classes	κ	Accuracy	Sensitivity	Specificity
Multiple classes[a]				
Wake/N1 + N2/N3/REM	0.42 (0.12)	59.3 (8.5)	n/a	n/a
Wake/NREM/REM	0.46 (0.15)	72.9 (8.3)	n/a	n/a
Single class detection[b]				
Wake	0.55 (0.14)	91.5 (5.1)	58.2 (17.3)	96.9 (2.0)
N1 + N2	0.30 (0.13)	65.7 (7.1)	55.6 (9.2)	74.2 (8.2)
N3	0.50 (0.19)	82.5 (6.3)	63.9 (18.9)	89.0 (4.8)
NREM	0.46 (0.16)	75.3 (7.7)	77.5 (7.8)	71.1 (12.0)
REM	0.43 (0.19)	78.9 (7.2)	70.7 (18.2)	81.4 (6.9)

Abbreviations: PPG, photoplethysmography; PSG, polysomnography; REM, rapid eye movement; NREM, non-REM.
[a]Agreement for multiple classes computed based on the comparison between PPG epoch-based classification and PSG annotations; for the 4-class task PSG N1 and N2 classes were merged into a single N1 + N2 class; for the 3-class task, PPG N1 + N2 and N3 were merged into a single NREM class and PSG N1, N2, and N3 were merged into a single NREM class.
[b]For each class detection task, the remaining classes were merged in a class, both for PPG and for PSG, and considered the negative class for purposes of sensitivity and specificity calculation.

Table 4—Bias, Standard Deviation, and Root Mean Squared Error in Sleep-wake Statistics Between Actigraphy and PSG, and Between PPG and PSG in 53 Recordings (Subset of Validation Set 2).

Statistic	Actigraphy versus PSG		PPG versus PSG		Comparison between PPG and actigraphy[c]
	Mean error (SD)[a]	RMSE	Mean error (SD)[b]	RMSE	
SOL (minutes)	−8.59 (9.05)	12.42	−7.48 (6.64)	9.96	$p < .05$
WASO (minutes)	−8.32 (29.92)	30.79	−5.25 (27.89)	28.12	$p = .18$
TWT (minutes)	−17.94 (28.09)	33.10	−17.22 (34.19)	37.99	$p = .18$
TST (minutes)	10.60 (30.29)	31.83	15.66 (34.95)	38.00	$p = .22$
SE (%)	3.66 (5.85)	6.85	3.71 (7.34)	8.16	$p = .51$

Abbreviation: PPG, photoplethysmography; PSG, polysomnography; RMSE, root mean squared error; SD, standard deviation; SE, sleep efficiency; SOL, sleep onset latency; TST, total sleep time; TWT, total wake time; WASO, wake after sleep onset.
[a]Actigraphy minus PSG.
[b]PPG minus PSG.
[c]Wilcoxon signed-rank comparison of estimation error between actigraphy versus PSG and PPG versus PSG.

Table 5—Epoch-by-epoch Sleep–Wake Classification Agreement Between Actigraphy and PSG and Between PPG and PSG in 49 Recordings (Subset of Validation Set 2).

Agreement metric	Actigraphy versus PSG, mean performance (SD)	PPG versus PSG, mean performance (SD)	Comparison between PPG and actigraphy[a]
κ (−)	0.40 (0.15)	0.55 (0.15)	$p < .001$
Accuracy (%)	91.8 (5.1)	91.3 (5.9)	$p = .58$
Sensitivity (%)	45.5 (19.3)	57.4 (17.7)	$p < .001$
Specificity (%)	97.1 (1.6)	97.4 (1.9)	$p = .21$

Abbreviations: PPG, photoplethysmography; PSG, polysomnography; SD, standard deviation.
[a]Wilcoxon signed-rank comparison of estimation error between actigraphy versus PSG and PPG versus PSG.

DISCUSSION

This manuscript describes the validation procedure and the results obtained with a PPG-based sleep staging algorithm when compared with human scored PSG and with wrist-worn actigraphy. Regarding the estimation of sleep–wake statistics, the algorithm achieved a satisfactory agreement for TST, TWT, WASO, and SE for more than 70% of the recordings and was overall, statistically significant ($p < .001$). Although it achieved a relatively small bias for SOL and WASO, inspection of the Bland–Altman plots revealed that it tends to underestimate wake time when SOL and WASO increase beyond 15 minutes. This is likely related to the behavior of participants who took longer to fall asleep or who woke up during the night and had difficulties falling asleep again. Besides lying still, as verified by actigraphy, an analysis of the characteristics of the respiratory effort signal of the PSG recordings suggested that some of these participants might have paced their breathing during these periods of wakefulness. This lowered the heart rate as a consequence and probably also led to an increase in the HF and a decrease in the LF component of HRV when compared with irregular, uncontrolled breathing[42,43] yielding HRV characteristics closer to those typically observed during N2 or even N3[44]. Although the SD of the error for SOL and SE was found to be relatively low (below 10 minutes and 10%, respectively), it reached nearly 30 minutes for the parameters WASO, TWT, and TST. This highlighted the between-participant variability in the accuracy of the algorithm.

Furthermore, in comparison with the sleep–wake statistic estimations obtained with actigraphy, we observed a significant improvement in the estimation of SOL, but no significant improvement (nor deterioration) for any other statistic. Interestingly, the error in the actigraphy estimation of statistics such as SE is lower than some of the values reported in literature.[7,9] This likely reflects the healthy nature of the individuals in our validation set, and it would be interesting to investigate whether the results hold when estimating wake in clinical populations or in participants with fragmented sleep and whether we would see a significant improvement in performance when using PPG to analyze sleep and wake in those cases.

In terms of epoch-by-epoch agreement with human annotated PSG, it is clear that the simpler the classification task, the better the overall performance is. This suggests that although the HRV features used in this work allow, to a certain extent, a reasonable partition of NREM classes into N1 + N2 and N3, these are not completely separable in the space of the HRV features used. When N1, N2, and N3 are merged into a single NREM class, the problem of lack of separability between these disappears, and the performance naturally increases.

It is also interesting to remark that the performance of the algorithm for 4- and 3-class tasks is lower than that reported in our earlier work (κ of 0.49 and accuracy of 69% for the 4-class task, and κ of 0.56 and accuracy of 80% for the 3-class task). Furthermore, while in that work the best single-class results were obtained for REM classification,[24] the worst single-class detection performances are now obtained for N1 + N2 followed by REM. One of the possible reasons for the deterioration in overall performance might be related to the highly susceptible nature of reflective PPG to participant and skin tissue motion, which are known to distort the PPG signal[45] and consequently decrease the precision of detected heart beats, interbeat intervals, and HRV features. Additionally, the absence of respiratory features in this study in comparison with our earlier work is likely to have a negative impact on performance. Although this alone is probably not sufficient to fully explain the decrease, other studies have shown the advantage of using features (besides body movements) in addition to HRV which help further express changes in sympathetic tone, important in the classification of REM. Herscovici et al.,[46] for instance, combined comparable time- and frequency-domain HRV features (referred to by the authors as interpulse periods, or IIP) with features describing variations in the amplitude of peripheral arterial tone (PAT) measured at the fingertip. The PAT signal provides a sensitive surrogate measure of changes in sympathetic tone and exhibits unique characteristics during REM, providing strong discriminative power for the detection of this sleep stage[46] and in general, overall automatic sleep staging based on autonomic features.[47] Although PAT is not directly available with the current sensor, the PPG signal actually contains respiratory-induced intensity variations (RIIV) which are related to ventilatory pressure and have been used, successfully, to estimate respiratory rate.[48] Although not explicitly related to RIIV, a preliminary study on ten sleep apnea patients by Uçar et al.[49] has shown that adding morphological features of PPG to a standard set of HRV features increases the performance of sleep stage classification. It remains to be confirmed whether features extracted from RIIV, together with the addition of dedicated HRV features which may better reflect changes in sympathetic tone characteristic of REM help alleviate or overcome this degradation.

Finally, the best single-class performance is obtained for sleep–wake classification, with a κ of 0.55, accuracy of 91.5%, sensitivity to wake of 58.2%, and specificity to wake of 96.9%. Comparing the performance with actigraphy-based epoch-by-epoch sleep–wake classification on a subset of the validation set, we found that PPG led to a substantial and significant increase in κ (from 0.40 with actigraphy to 0.55 with PPG on the same recordings, $p < .001$) and in sensitivity to wake (from 45.5% with actigraphy to 57.4% with PPG, $p < .001$). The results obtained with actigraphy in our work are slightly better than those reported in literature. Marino et al.[50] reported an accuracy of 86.3%, a sensitivity to sleep of 96.5%, and a specificity to sleep of 32.9%. We calculated a κ of 0.37 based on their published data. Once more, the differences can probably be explained by the exclusively healthy characteristics of the participants in our validation sets.

Notwithstanding, the increased performance in sleep–wake classification using PPG-HRV and body movements when compared with actigraphy is likely related to the presence of additional information captured by HRV characteristics which is better able to capture wake states, especially when the participants do not move. The reason why these improvements were less prominent in the computation of sleep–wake statistics is likely related to the high imbalance between sleep and wake classes (the participants in the validation sets had a mean SE of 88.1%) and especially to the relative lack of variability in the sleep–wake statistics to start with (the SD of SOL—computed

from PSG—was lower than 10 minutes). It is important to remark that the comparison with actigraphy is necessarily limited by the scope of our study. While actigraphy has been evaluated for many different clinical populations over the last decades and its performance and limitations are well understood for different clinical settings, our results only allow tentative conclusions regarding a healthy group of middle-aged participants and cannot be automatically extended to other groups. Furthermore, this technology has the disadvantage that it assumes an intact autonomic function, and its performance will be probably impacted by disorders of the ANS which affect the regulation of blood pressure, heart rate, and HRV. Besides, the accuracy of HRV features is intrinsically linked to the quality and availability of the PPG signal. Disorders such as peripheral artery disease may cause a restriction in the blood flow at peripheral arteries, limiting the quality of PPG or even rendering it unavailable. By relying simply on the measurement of gross body movements as an indicator of wakefulness, actigraphy might be more robust to the severity of such conditions.

Nevertheless, its relative low cost, ease of use, and comfort for the participant, and its moderate epoch-by-epoch agreement, and good agreement in terms of sleep statistics suggests the adequacy of this technique for long-term sleep monitoring, although more evidence is needed to understand whether it can complement PSG in clinical practice. These results also suggest that besides being more sensitivity to wake, the agreement obtained for four sleep stages with this technology (with a κ of 0.42) is in fact slightly higher than the overall agreement for sleep–wake classification with traditional actigraphy (κ of 0.40) in our data set. Provided that the level of evidence for PPG-based sleep staging increases in the upcoming years and that these findings are confirmed in relevant clinical populations, this technology may have the potential to eventually replace actigraphy in clinical practice with the benefit of offering additional insights into the sleep architecture of the participants under investigation besides only their sleep–wake patterns. Regardless of its potential, however, it will only be usable when it has been properly validated on relevant clinical populations. To our knowledge, it has not yet been validated in individuals with sleep disorders and therefore merits more research.

Long-term sleep monitoring was for a long time limited by the unavailability of small, comfortable sensors which could be used at home without assistance. A notable exception, accelerometers, enabled the development of actigraphy devices which albeit valuable in the assessment of some sleep disorders were limited to the analysis of sleep and wake patterns. Recent sensor developments have shown the feasibility of accurately monitoring other physiological parameters, with PPG, skin temperature, and skin conductance sensors offering the promise, in the long run, to be able to complement PSG with enough accuracy to enable unobtrusive sleep monitoring in clinical practice.

REFERENCES

1. Bruyneel M, Sanida C, Art G, et al. Sleep efficiency during sleep studies: results of a prospective study comparing home-based and in-hospital polysomnography. J Sleep Res. 2011; 20(1 Pt 2): 201–206.
2. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep. 2003; 26(3): 342–392.
3. Morgenthaler T, Alessi C, Friedman L, et al. Standards of Practice Committee; American Academy of Sleep Medicine. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. Sleep. 2007; 30(4): 519–529.
4. Lee J, Finkelstein J. Consumer sleep tracking devices: a critical review. Stud Health Technol Inform. 2015; 210: 458–460.
5. Russo K, Goparaju B, Bianchi MT. Consumer sleep monitors: is there a baby in the bathwater? Nat Sci Sleep. 2015; 7: 147–157.
6. Ko PR, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. J Clin Sleep Med. 2015; 11(12): 1455–1461.
7. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med. 2001; 2(5): 389–396.
8. Lichstein KL, Stone KC, Donaldson J, et al. Actigraphy validation with insomnia. Sleep. 2006; 29(2): 232–239.
9. Paquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. Sleep. 2007; 30(10): 1362–1369.
10. de Zambotti M, Baker FC, Colrain IM. Validation of sleep-tracking technology compared with polysomnography in adolescents. Sleep. 2015; 38(9): 1461–1468.
11. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. Sleep. 2015; 38(8): 1323–1330.
12. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. Sleep Breath. 2012; 16(3): 913–917.
13. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas. 2007; 28(3): R1–39.
14. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol. 2015; 39(5): 264–271.
15. Trinder J, Kleiman J, Carrington M, et al. Autonomic activity during human sleep as a function of time and sleep stage. J Sleep Res. 2001; 10(4): 253–264.
16. Lanfranchi PA, Pépin J-L, Somers VK. Cardiovascular physiology: autonomic control in health and in sleep disorders. In: Kryger M, Roth T, Dement WC, eds. Principles and Practice of Sleep Medicine. 6th ed. Philadelphia: Elsevier; 2016: 142–54.
17. Bonnet MH, Arand DL. Heart rate variability: sleep stage, time of night, and arousal influences. Electroencephalogr Clin Neurophysiol. 1997; 102(5): 390–396.
18. Ako M, Kawara T, Uchida S, et al. Correlation between electroencephalography and heart rate variability during sleep. Psychiatry Clin Neurosci. 2003; 57(1): 59–65.
19. Jurysta F, van de Borne P, Migeotte PF, et al. A study of the dynamic interactions between sleep EEG and heart rate variability in healthy young men. Clin Neurophysiol. 2003; 114(11): 2146–2155.
20. Dumont M, Jurysta F, Lanquart JP, Migeotte PF, van de Borne P, Linkowski P. Interdependency between heart rate variability and sleep EEG: linear/non-linear? Clin Neurophysiol. 2004; 115(9): 2031–2040.
21. Redmond SJ, de Chazal P, O'Brien C, Ryan S, McNicholas WT, Heneghan C. Sleep staging using cardiorespiratory signals. Somnologie-Schlafforschung und Schlafmedizin. 2007; 11(4): 245–256.
22. Domingues A, Paiva T, Sanches JM. Hypnogram and sleep parameter computation from activity and cardiovascular data. IEEE Trans Biomed Eng. 2014; 61(6): 1711–1719.
23. Willemen T, Van Deun D, Verhaert V, et al. An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification. IEEE J Biomed Health Inform. 2014; 18(2): 661–669.
24. Fonseca P, Long X, Radha M, Haakma R, Aarts RM, Rolink J. Sleep stage classification with ECG and respiratory effort. Physiol Meas. 2015; 36(10): 2027–2040.
25. Klösch G, Kemp B, Penzel T, et al. The SIESTA project polygraphic and clinical database. IEEE Eng Med Biol Mag. 2001; 20(3): 51–57.
26. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975; 12(3): 189–198.

27. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989; 28(2): 193–213.
28. Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep. 2012; 35(2): 287–302.
29. Iber C, Ancoli-Israel S, Chesson A, Quan SF. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. 1st ed. Westchester, IL: American Academy of Sleep Medicine; 2007.
30. Heart rate variability: standards of measurement, physiologic interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Eur Heart J. 1996; 17(3): 354–381.
31. Basner M, Griefahn B, Müller U, Plath G, Samel A. An ECG-based algorithm for the automatic identification of autonomic activations associated with cortical arousal. Sleep. 2007; 30(10): 1349–1361.
32. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Phys Rev Lett. 2002; 89(6): 068102.
33. Penzel T, Kantelhardt JW, Grote L, Peter JH, Bunde A. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. IEEE Trans Biomed Eng. 2003; 50(10): 1143–1151.
34. Long X, Fonseca P, Aarts RM, Haakma R, Foussier J. Modeling cardiorespiratory interaction during human sleep with complex networks. Appl Phys Lett. 2014; 105(20): 203701.
35. Long X, Fonseca P, Haakma R, Aarts RM, Foussier J. Spectral boundary adaptation on heart rate variability for sleep and wake classification. Int J Artif Intell Tools. 2014; 23(3): 1460002.
36. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20(1): 37–46.
37. Werner H, Molinari L, Guyer C, Jenni OG. Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns. Arch Pediatr Adolesc Med. 2008; 162(4): 350–358.
38. Meltzer LJ, Walsh CM, Traylor J, Westin AM. Direct comparison of two new actigraphs and polysomnography in children and adolescents. Sleep. 2012; 35(1): 159–166.
39. Hollander M, Wolfe DA, Chicken E. Nonparametric Statistical Methods. 3rd ed. Wiley; 2013.
40. Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. J Gen Intern Med. 2011; 26(2): 192–196.
41. Fonseca P, Aarts RM, Long X, Rolink J, Leonhardt S. Estimating actigraphy from motion artifacts in ECG and respiratory effort signals. Physiol Meas. 2016; 37(1): 67–82.
42. Novak V, Novak P, de Champlain J, Le Blanc AR, Martin R, Nadeau R. Influence of respiration on heart rate and blood pressure fluctuations. J Appl Physiol (1985). 1993; 74(2): 617–626.
43. Pitzalis MV, Mastropasqua F, Massari F, et al. Effect of respiratory rate on the relationships between RR interval and systolic blood pressure fluctuations: a frequency-dependent phenomenon. Cardiovasc Res. 1998; 38(2): 332–339.
44. Elsenbruch S, Harnish MJ, Orr WC. Heart rate variability during waking and sleep in healthy males and females. Sleep. 1999; 22(8): 1067–1071.
45. Wijshoff RW, Mischi M, Veen J, van der Lee AM, Aarts RM. Reducing motion artifacts in photoplethysmograms by using relative sensor motion: phantom study. J Biomed Opt. 2012; 17(11): 117007.
46. Herscovici S, Pe'er A, Papyan S, Lavie P. Detecting REM sleep from the finger: an automatic REM sleep algorithm based on peripheral arterial tone (PAT) and actigraphy. Physiol Meas. 2007; 28(2): 129–140.
47. Hedner J, White DP, Malhotra A, et al. Sleep staging based on autonomic signals: a multi-center validation study. J Clin Sleep Med. 2011; 7(3): 301–306.
48. Nilsson LM. Respiration signals from photoplethysmography. Anesth Analg. 2013; 117(4): 859–865.
49. Uçar MK, Bozkurt MR, Bilgin C, Polat K. Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques. Neural Comput Appl. 2016: 1–16. doi:10.1007/s00521-016-2365-x.
50. Marino M, Li Y, Rueschman MN, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep. 2013; 36(11): 1747–1755.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Tim Leufkens for the critical review of this manuscript and for suggestions for improvement.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication December, 2016
Submitted in final revised form April, 2017
Accepted for publication June, 2017
Address correspondence to: P. Fonseca, MSc, Philips Group Innovation Research, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands. Telephone: +31 63 1926848; Email: pedro.fonseca@philips.com

The work was conducted at Philips Group Innovation Research, Eindhoven, The Netherlands.

DISCLOSURE STATEMENT

At the time of writing all authors were affiliated and/or employed by Philips Group Innovation, part of Philips Electronics Nederland B.V. Philips is a manufacturer of consumer and medical electronic devices, and commercializes products in the area of sleep monitoring. The study reported in this manuscript was entirely funded by Philips Group Innovation. This trial was not a clinical trial; therefore, no registration number is available.