Sleep Lab validation of a wellness ring in detecting sleep patterns based on photoplethysmogram, actigraphy and body temperature

Hannu Kinnunen (Ouraring, Oulu, Finland)

Feb 9, 2016

INTRODUCTION

Wearable consumer devices and actigraphy provide meaningful long term data on sleep patterns in the home environment. However, medical sleep specialists have raised questions regarding their accuracy, especially their limited ability to identify wake during bedtime (Montgomery-Downs et al. 2012). A review paper by Water et al. (2011) provides reference data on the performance level of actigraphy in detecting sleep and wake, but very little has been published about their ability to detect sleep stages (Ancoli-Israel et al. 2015). Especially, the ability to identify deep sleep would be valuable, as it is an important component of restorative sleep. One example of a relatively good level of accuracy can be found from an FDA approved clinical device that sits on the wrist and has additional sensors on fingers (Hedner et al 2011). Their device does sleep staging based on peripheral arterial tone, actigraphy, pulse rate and pulse oximetry (Itamar Medical http://www.itamar-medical.com/watchpat-main/).

The gold standard to scientifically evaluate a person's sleep is a lab test, called polysomnography (PSG). PSG utilizes brain wave signals (EEG), eye movement signals (EOG), cardiac signals (ECG), muscle activity (EMG), and optionally finger photo-plethysmography (PPG). A simplified version of the lab test is performed based on EOG only.

The ŌURA ring is a lightweight device that measures the subjects' blood volume pulse waveform, and the time between heartbeats using an IR LED and sensors. It also measures movements using an accelerometer, and variations in body temperature. In line with its minimalistic design, the ring does not require any user actions - gestures or button presses - in order to initiate bedtime or sleep detection. Additionally, the ring evaluates daily activity and uploads its measurements to a smartphone application. The aim of this study was to validate the ŌURA ring in its ability to detect the amount of sleep and sleep quality against the gold standard sleep laboratory test.

METHODS

Fourteen subjects volunteered to participate in this study. Eight full polysomnograph recordings (four female and four male) and six EOG only recordings (two female and four male) were made. The recording and subsequent analysis was carried out by the sleep laboratory of the Finnish Occupational Health Institute, Helsinki, Finland[1)], an independent research institute. The research was funded by Ouraring.

A portable clinical instrumentation setup was used for the tests (Figure 1), and the subjects slept at their own or friend's home, or in a hotel room. The subjects recorded their lights off and on times on paper to help the sleep scientist in determining the period to be analyzed. An experienced sleep specialist made manual scoring of both the PSG and EOG recordings using the American Academy of Sleep Medicine (AASM) standard.

[1)] http://www.ttl.fi/en/Pages/default.aspx, Contact person Medical Physicist, Somnologist, PhD Jussi Virkkala

Figure 1. Portable PSG measurement instrumentation used in this project.

The subjects' age ranged between 9 and 48 years, with twelve of them lying between 35 and 48 years. None had a diagnosed sleep disorder or medication known to affect sleep. Each of the 14 subjects wore the ŌURA ring on one of the fingers of their non-dominant hand (left hand in 13 cases). In seven cases, to determine the effect of the dominant vs. non-dominant hand, the subject wore a second ŌURA ring on one of the fingers of their dominant hand. For epoch by epoch comparison, PSG or EOG based sleep stages N1 and N2 were combined and compared to the Light sleep phase determined by the ŌURA ring. PSG based stage N3 was compared to the Deep sleep stage determined by the ring (Figures 2 & 3).

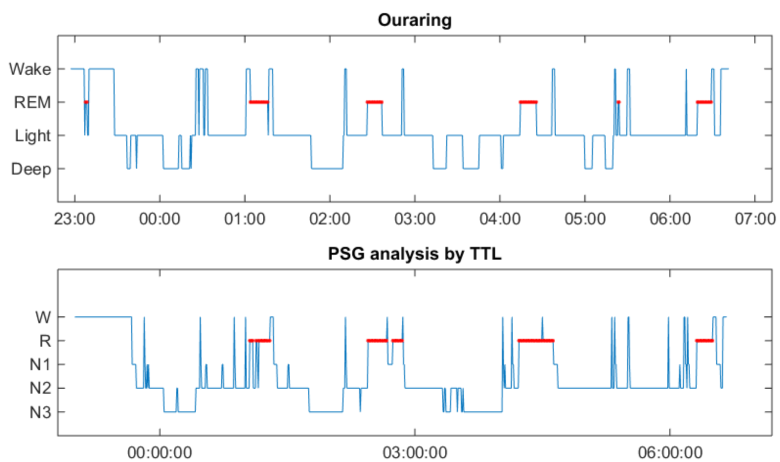

Figure 2. Example sleep patterns from one night, simultaneously measured by the ŌURA ring (top), and manually scored Polysomnography (PSG) made by the Finnish Occupational Health Institute (TTL) (bottom).

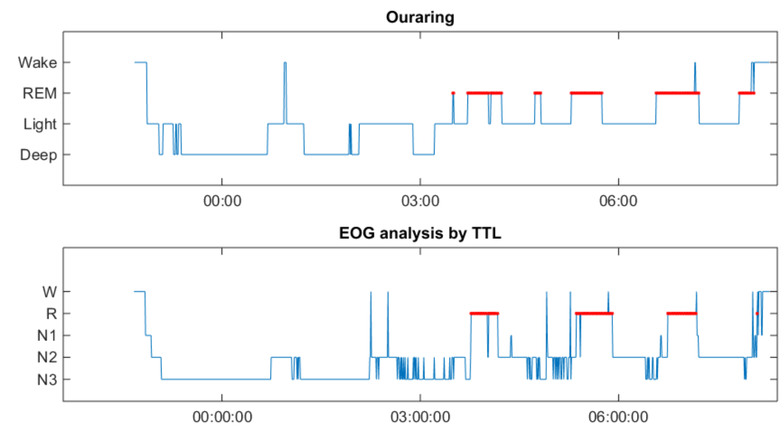

Figure 3. Example sleep patterns from one night, simultaneously measured by the ŌURA ring (top), and a home monitored (EOG) manually scored by the Finnish Occupational Health Institute (TTL) (Bottom).

In order to illustrate the current performance and agreement limits of the ŌURA ring when compared to the gold standard method, Scatter plots and Bland-Altman plots of Total Sleep Time (TST), Sleep Onset Latency (SOL), Wake After Sleep Onset (WASO), REM sleep, NREM sleep, Light sleep, and Deep sleep were drawn. The sleep onset was triggered only if followed by minimum of 5 minutes of persistent sleep.

To evaluate the accuracy of sleep staging, epoch by epoch agreement between the ŌURA ring and the PSG, and Cohen's kappa, were calculated across Wake, REM, Light (N1, N2) and Deep sleep (N3). Additionally, in order to evaluate the potential effect of the non-dominant vs. dominant hand, the sleep staging comparison was repeated with measured data from the dominant hand ring for those subjects that wore a ring on both hands.

RESULTS

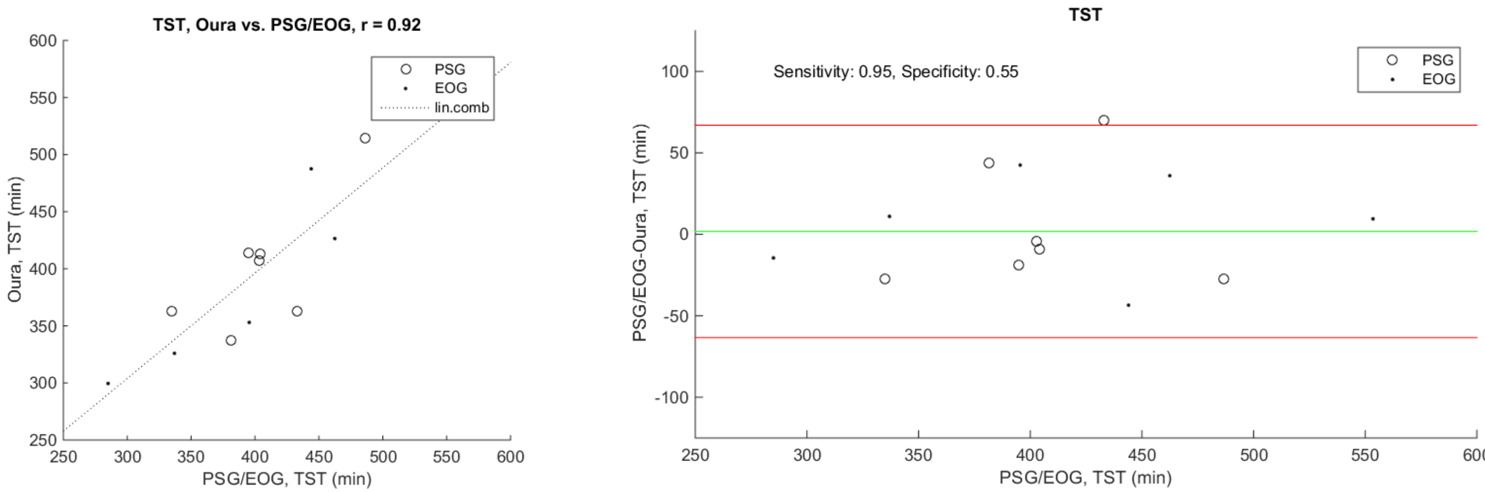

Figure 4. Scatter plot (left), and Bland-Altman plot (right) of Total Sleep Time (TST) comparing the ŌURA ring and the Sleep Lab (Polysomnography (PSG) or Electrooculogram (EOG)). Correlation coefficient is 0.92. Red lines indicate 95% limits of agreement (-64 - 66 min), and green line is the mean difference (+1.0 min) between the PSG/EOG and the ŌURA ring.

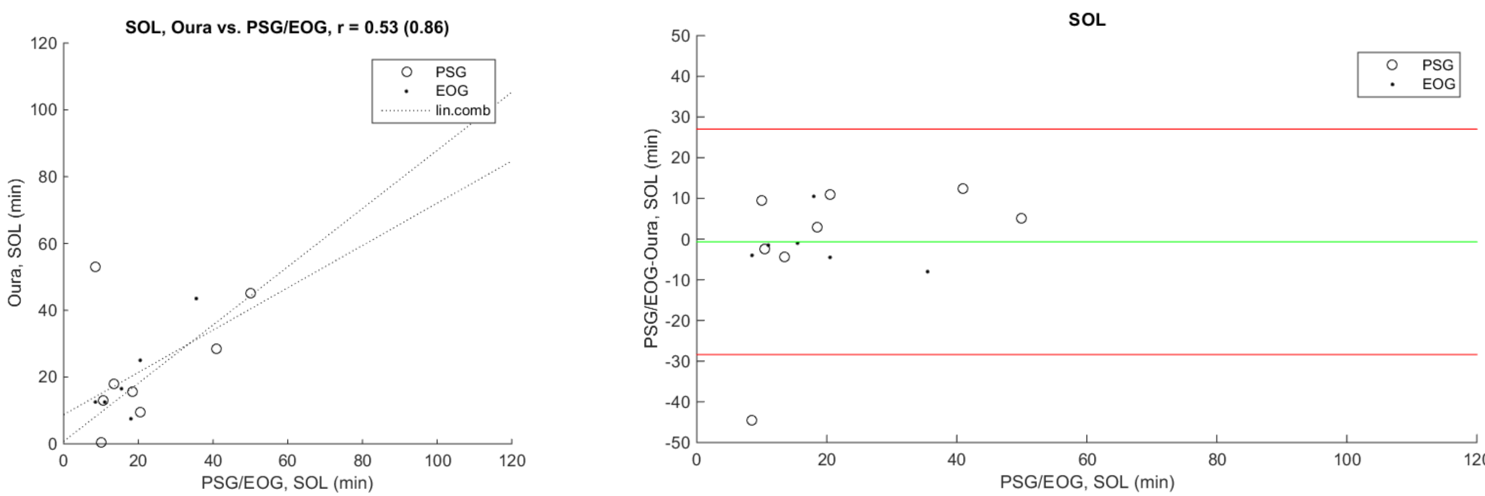

Figure 5. Scatter plot (left) and Bland-Altman plot (right) of Sleep Onset Latency (SOL) comparing the ŌURA ring and the Sleep Lab (Polysomnography (PSG) or Electrooculogram (EOG)). The coefficient of correlation was 0.53, or 0.86 if one outlier is removed. Red lines indicate 95% limits of agreement (-28 - 27 min), and green line indicates the mean difference (-0.7 min) between the PSG/EOG and the ŌURA ring. Except for one subject, SOL determined by the ŌURA ring and the Sleep Lab were within 13 minutes of each other.

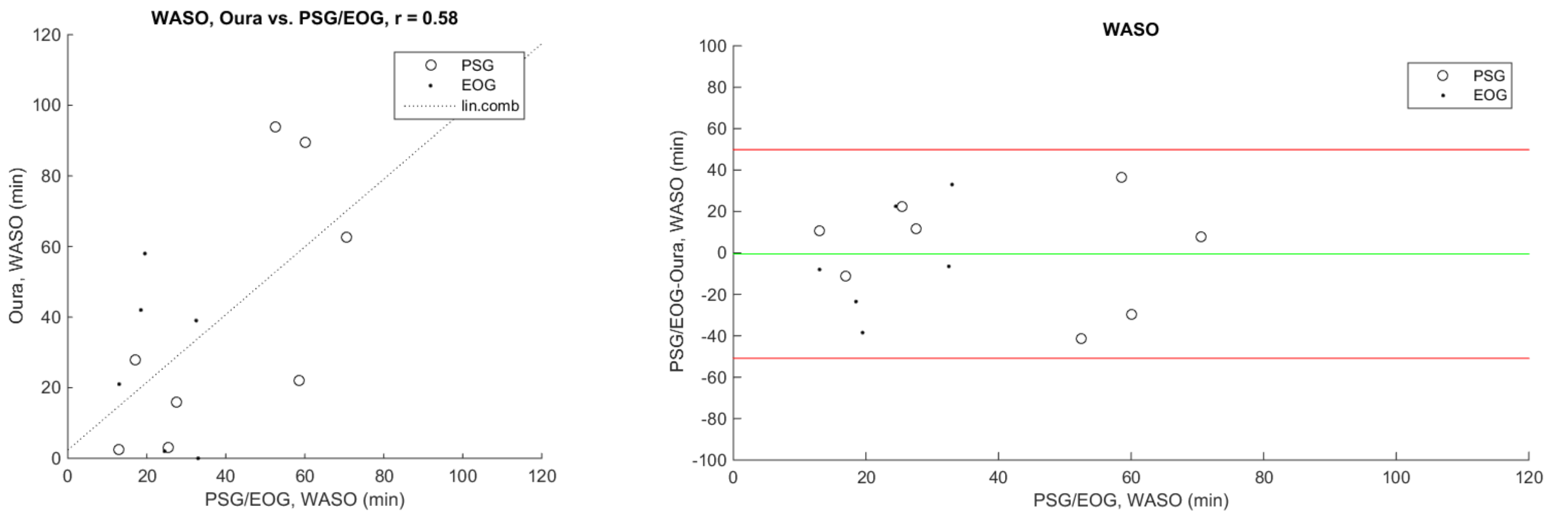

Figure 6. Scatter plot (left) and Bland-Altman plot (right) of Wake After Sleep Onset (WASO) comparing the ŌURA ring and the Sleep Lab (Polysomnography (PSG) or Electrooculogram (EOG)). The coefficient of correlation was 0.58. Red lines indicate 95% limits of agreement (-50 - 50 min), and green line indicates the mean difference (0.3 min) between the PSG/EOG and the ŌURA ring.

Table 1. Confusion matrix from epoch by epoch comparison of the sleep stages determined by the ŌURA ring and manually scored based on Polysomnography. Sleep stages were classified into Wake, REM, Light and Deep by both methods. The ŌURA ring was worn on the non-dominant hand. The ŌURA ring displays 65.3% agreement, Cohen's kappa 0.449.

		PSG scoring			
		Deep	Light	REM	WAKE
ŌURA	Deep	933	635	30	56
	Light	736	5036	627	394
	REM	78	932	1425	135
	WAKE	15	524	145	844

Table 2. Confusion matrix from epoch by epoch comparison of the sleep stages determined by the ŌURA ring and manually scored based on Polysomnography. Sleep stages were classified into Wake, REM, and Non-REM by both methods (combining light and deep into NREM). The ŌURA ring was worn on the non-dominant hand. The ŌURA ring displays 76.6% agreement, Cohen's kappa 0.505.

		PSG scoring		
		NREM	REM	WAKE
ŌURA	NREM	7340	657	450
	REM	1010	1425	135
	WAKE	539	145	844

Table 3. Potential effect of hand dominance: The confusion matrix from epoch by epoch comparison of the sleep stages determined by the ŌURA ring and manually scored based on Polysomnography. Sleep stages were classified into Wake, REM, Light and Deep by both methods. Compared to Table 1, 7 subjects' data has been taken from the dominant hand ring instead of the non-dominant hand ring. The ŌURA ring displays 66.0% agreement, Cohen's kappa 0.447.

		PSG scoring			
		Deep	Light	REM	WAKE
ŌURA	Deep	981	601	33	58
	Light	754	5263	734	433
	REM	70	847	1350	144
	WAKE	13	504	128	807

ANALYSIS OF RESULTS

The ŌURA ring determined bedtime start within ±5min when compared to the lights off time marked by the subjects. Our sample consisted of 14 subjects, which means that the critical value of correlation ($p<0.05$) is 0.53. Hence, a significant correlation between the ŌURA ring and the Sleep Lab was found in the following sleep parameters: TST, SOL, WASO, NREM, and Deep/N3 sleep.

None of the sleep amount or quality parameters evaluated - TST, SOL, WASO, NREM, REM, Deep, Light - showed a systematic bias when compared to the gold standard PSG. With the exception of one case, the ŌURA ring was able to detect SOL within 13 minutes of the Sleep Lab determined time. Yet the results revealed variability in individual accuracy.

If a measurement device can measure TST and WASO within 30 min of PSG, then it is considered meaningful for clinical sleep medicine. In this study, the ŌURA ring determined TST and WASO were within 45 minutes of the PSG determined values in all cases except for one subject's TST.

The ŌURA ring's sleep staging classification accuracy compared to the PSG classification was 65.3 % when the ring was worn on the non-dominant hand, Cohen's kappa 0.45.

Comparing the results in Table 1 and Table 3 provides suggestive evidence that use of the ring on the dominant hand instead of the non-dominant hand increases the amount of Deep and Light sleep and decreases the amount of REM and Wake determined by the ŌURA ring.

DISCUSSION

The main finding of this study was that the ŌURA ring was shown to be usable for sleep analysis in the home environment without need for user actions to initiate sleep measurements.

In studies where alternative methods, mainly actigraphy, have been compared to the Sleep Lab PST in healthy subjects, an overall sensitivity range of 72-97% and specificity range of 28-67%, and Pearson's correlation coefficients for TST (0.43-0.97), SOL (0.64-0.82) and WASO (0.36-0.39) have been reported (Water et al. 2011). The results of this study fell either at the highest tertile of the presented ranges, or exceeded their performance levels.

The current sleep evaluation accuracy of the ŌURA ring was shown to be on par with the performance levels reported for the FDA approved wrist/finger clinical device by Hedner et al. 2011, who reported 66.0% agreement in *Wake-REM-Light-Deep* classification with Cohen's kappa 0.475. Taking into account that the current algorithms of the ŌURA ring were developed based on fingertip PPG before the ring measured signals were available, it can be speculated that further improvements can be made when data measured by the rings simultaneously with the

Sleep Lab data is used to update the classification algorithms using machine learning techniques.

The comparison between the non-dominant hand and dominant hand suggested that use of the ring on the dominant hand may favor classification as NREM over REM sleep – this observation deserves to be studied further to analyze if the effect of wearing hand could be balanced in the sleep staging.

Even though no systematic error was observed in any sleep parameter, the variability in individual accuracy in some parameters informed the ŌURA team of clear opportunities for improvements. For example, the fact that the ŌURA ring missed very short awakenings implies the need adjust the ring's algorithms to be more sensitive to rapid changes. It should be noted that due to the discomfort of wearing the instrumentation, the clinical sleep measurement process negatively affects the subject's sleep quality and adds extra awakenings compared to normal sleep conditions.

CONCLUSION

In a sample of fourteen healthy subjects, the ŌURA ring provided unbiased and relevant data on sleep patterns, reaching comparable performance levels to the best scientifically validated alternative methods to Sleep Lab measurements.

References

Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. Sleep Breath. 2012 Sep;16(3):913-7. DOI: 10.1007/s11325-011-0585-y. Epub 2011 Oct 6

Hedner J, White DP, Malhotra A, Herscovici S, Pittman SD, Zou D, Grote L, and Pillar G. Sleep staging based on autonomic signals: a multi-center validation study. J. Clin. Sleep Med. 2011, 7(3):301–306.

Ancoli-Israel S, Martin JL, Blackwell T, Buenaver L, Liu L, Meltzer LJ, Sadeh A, Spira AP & Taylor DJ. The SBSM Guide to Actigraphy Monitoring: Clinical and Research Applications, Behavioral Sleep Medicine 2015, 13:sup1, S4-S38. DOI: 10.1080/15402002.2015.1046356and Research Applications.

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