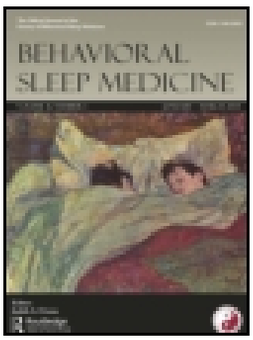

Behavioral Sleep Medicine

ISSN: 1540-2002 (Print) 1540-2010 (Online) Journal homepage: http://www.tandfonline.com/loi/hbsm20

The Sleep of the Ring: Comparison of the ŌURA Sleep Tracker Against Polysomnography

Massimiliano de Zambotti, Leonardo Rosas, Ian M. Colrain & Fiona C. Baker

To cite this article: Massimiliano de Zambotti, Leonardo Rosas, Ian M. Colrain & Fiona C. Baker (2017): The Sleep of the Ring: Comparison of the ŌURA Sleep Tracker Against Polysomnography, Behavioral Sleep Medicine, DOI: 10.1080/15402002.2017.1300587

To link to this article: http://dx.doi.org/10.1080/15402002.2017.1300587

Published online: 21 Mar 2017.

Behavioral Sleep Medicine, 00:1–15, 2017

ISSN: 1540-2002 print/1540-2010 online
DOI: 10.1080/15402002.2017.1300587

Taylor & Francis
Taylor & Francis Group

The Sleep of the Ring: Comparison of the ŌURA Sleep Tracker Against Polysomnography

Massimiliano de Zambotti and Leonardo Rosas
Center for Health Sciences, SRI International, Menlo Park, California

Ian M. Colrain
Center for Health Sciences, SRI International, Menlo Park, California
Melbourne School of Psychological Sciences, University of Melbourne, Parkville, Victoria, Australia

Fiona C. Baker
Center for Health Sciences, SRI International, Menlo Park, California
Brain Function Research Group, School of Physiology, University of the Witwatersrand, Johannesburg, South Africa

Objective/Background: To evaluate the performance of a multisensor sleep-tracker (ŌURA ring) against polysomnography (PSG) in measuring sleep and sleep stages. *Participants*: Forty-one healthy adolescents and young adults (13 females; Age: 17.2 ± 2.4 years). *Methods*: Sleep data were recorded using the ŌURA ring and standard PSG on a single laboratory overnight. Metrics were compared using Bland-Altman plots and epoch-by-epoch (EBE) analysis. *Results*: Summary variables for sleep onset latency (SOL), total sleep time (TST), and wake after sleep onset (WASO) were not different between ŌURA ring and PSG. PSG-ŌURA discrepancies for WASO were greater in participants with more PSG-defined WASO ($p < .001$). Compared with PSG, ŌURA ring underestimated PSG N3 (~20 min) and overestimated PSG REM (~17 min; $p < .05$). PSG-ŌURA differences for TST and WASO lay within the ≤ 30 min a-priori-set clinically satisfactory ranges for 87.8% and 85.4% of the sample, respectively. From EBE analysis, ŌURA ring had a 96% sensitivity to detect sleep, and agreement of 65%, 51%, and 61%, in detecting "light sleep" (N1), "deep sleep" (N2 + N3), and REM sleep, respectively. Specificity in detecting wake was 48%. Similarly to PSG-N3 ($p < .001$), "deep sleep" detected with the ŌURA ring was negatively correlated with advancing age ($p = .001$). ŌURA ring correctly categorized 90.9%, 81.3%, and

Correspondence should be addressed to Massimiliano de Zambotti, SRI International, 333 Ravenswood Avenue, Menlo Park, CA, 94025. E-mail: massimiliano.dezambotti@sri.com; maxdeze@gmail.com

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hbsm.

92.9% into PSG-defined TST ranges of < 6 hr, 6–7 hr, > 7 hr, respectively. *Conclusions*: Multisensor sleep trackers, such as the ŌURA ring have the potential for detecting outcomes beyond binary sleep–wake using sources of information in addition to motion. While these first results could be viewed as promising, future development and validation are needed.

The new wave of fitness trackers is booming. Distinct from the first accelerometer-based wearables, these new multisensory devices are able to collect a broad range of users' biosignals. The greater availability of more sophisticated devices that go beyond simple, user-friendly consumer products may provide the opportunity for sleep researchers to obtain a more detailed overview of sleep and physiological changes during sleep. However, validation of these commercial devices both in and outside of the laboratory is first required.

Standard actigraphy is a well-established measure of an individual's sleep-wake patterns (Sadeh, 2011). Although not measuring brain sleep states, actigraphy has the advantage of being relatively low-cost, nonintrusive, and easy to use (Ancoli-Israel et al., 2003), which allows for the tracking of individuals' sleep patterns over prolonged periods of time in nonlaboratory settings. Compared to PSG, actigraphy has high sensitivity (ability to detect sleep) although specificity (ability to detect wakefulness) is lower (Marino et al., 2013; Sadeh, 2011), with a wide range of accuracy, depending on the amount of nighttime wakefulness (Paquet, Kawinska, & Carrier, 2007), the algorithms used, and the particular population studied (Van de Water, Holmes, & Hurley, 2011). Most importantly, actigraphy relies on a single sensor, an accelerometer, and thus it provides a measure of motion from which it predicts sleep and wake states. However, information about sleep stage composition, fundamental in studying sleep and sleep disorders, is not provided.

Several consumer-grade sleep tracking devices based primarily on motion detection are available and have been compared with PSG in recent validation studies, with mixed results. Our group evaluated the validity of fitness-trackers Jawbone UP™ (de Zambotti, Baker, & Colrain, 2015; de Zambotti, Claudatos, Inkelis, Colrain, & Baker, 2015) and FitbitChargeHR™ (de Zambotti, Baker, et al., 2016) against PSG in adolescents and adults. Both devices had high sensitivity in detecting sleep, although specificity in detecting wake was lower, and accuracy for detecting sleep-wake states decreased as a function of more PSG-WASO. Roane and colleagues (2015) evaluated the accuracy of a multisensory armband (SenseWear® Pro3 Armband) in measuring sleep in 20 adolescents against PSG. In the study, authors also used a common standard actigraphic device (AMI Motionlogger®). SenseWear® Pro3 Armband sleep measures did not significantly differ from those obtained by PSG whereas, AMI Motionlogger® significantly overestimated sleep and underestimated wake. Similarly, a recent investigation in children and adolescents demonstrated that a commercial wristband (Jawbone UP™) performed similarly to a standard actigraphy (Actiwatch2) in detecting sleep and wake states compared to PSG (Toon et al., 2015). However, another study found that a consumer product (Fitbit Ultra) performed poorly in the assessment of sleep–wake in a group of children and adolescents (Meltzer, Hiruma, Avis, Montgomery-Downs, & Valentin, 2015). Factors such as frequent update of the device models and firmware, nonstandard definition of sleep parameters, and lack of access to proprietary algorithms make it difficult to compare results across studies and devices (de Zambotti, Godino, et al., 2016; Kolla, Mansukhani, & Mansukhani, 2016). Other limitations have recently emerged, with some devices claiming to assess sleep stages, which are defined using gold-standard PSG assessment, with multiple sources of information derived from

electroencephalogram, electrooculogram, and electromyogram. For example, Jawbone UP™, which uses motion sensors and proprietary algorithms to track daily sleep–wake activity, claims to be able to derive "sound" and "light" sleep. However, we found that Jawbone UP™ "sound sleep" was unrelated to PSG N3 sleep, rather being predicted by a combination of PSG N2 and REM sleep; similarly, Jawbone UP™ "light sleep" was unrelated to N1 sleep, being predicted by a combination of PSG N2 and N3 sleep (de Zambotti, Baker, et al., 2015). In a comparison of several actigraphy-based commercial devices to PSG, estimates of TST correlated highly with PSG measures for most devices; however, estimates of "deep" and "light" sleep were poor relative to PSG equivalents (Mantua, Gravel, & Spencer, 2016).

A novel, multisensory device that claims to be able to distinguish sleep stages, including REM sleep, has recently come on the market. The ŌURA ring (https://ouraring.com/) detects pulse rate, variation in interbeat intervals (IBIs) and pulse amplitude from the finger optical pulse waveform. The ring also measures motion and body temperature. Ōuraring (Oulu, Finland) claims to use these physiological signals (a combination of motion, heart rate, heart rate variability, and pulse wave variability amplitude) in combination with sophisticated machine learning-based methods to calculate deep (PSG N3), light (PSG N1 + N2) and rapid-eye-movement (REM) sleep in addition to sleep–wake states.

In the current study, we aimed to assess the accuracy of the ŌURA ring in assessing sleep–wake states as well as "light," "deep," and REM sleep compared to PSG during a laboratory night in a sample of 41 healthy adolescents and young adults (age range: 14–22 years). Adolescence is a period characterized by profound developmental changes, including dramatic changes in sleep stage composition (Colrain & Baker, 2011) and sleep-related behaviors (Gradisar, Gardner, & Dohnt, 2011). Insufficient sleep in adolescents has been recognized as a serious public health issue by the American Medical Association and American Academy of Sleep Medicine (American Medical Association & American Academy of Sleep Medicine, 2010) and "Sleep Health" has been recently added as a new target in the Healthy People initiative (https://www.healthypeople.gov/). The sleep wearable industry may offer an opportunity to monitor developmental trajectories of sleep in adolescents on a large scale, but the accuracy and limitations of these products still need to be determined.

METHODS

Participants

Forty-one healthy adolescents and young adults (14–22 years; 13 females; 35 Caucasian) with an average body mass index (BMI) of 21.6 ± 3.5 $kg.m^{-2}$ constituted the final sample. Participants were recruited from the San Francisco Bay Area as part of a longitudinal multisite study (the National Consortium on Alcohol and NeuroDevelopment in Adolescence, NCANDA). Participants had two overnight PSG assessments in the laboratory during each year of follow-up: a regular PSG recording and an evoked-potential recording. Data for the current study were collected from the regular PSG recording in Years 2 or 3 of the follow-up visits.

Details about recruitment and screening of the NCANDA sample are published elsewhere (Brown et al., 2015). All participants had an in-lab clinical interview and neuropsychological assessment, including a detailed medical history. None of the participants had severe medical

conditions (e.g., hypertension, diabetes) or current major DSM-IV (American Psychiatric Association, 2000) Axis I disorders (e.g., generalized anxiety disorder, major depressive disorder), and none of them currently used medications known to affect brain function and/or cardiovascular system (e.g., antidepressants, stimulants). An overnight clinical sleep evaluation reviewed for the presence of sleep pathology according to the guidelines of the American Academy of Sleep Medicine (AASM; Iber, 2007) confirmed that none of the participants had a sleep disorder (e.g., obstructive sleep apnea, periodic limb movement disorder).

The study was approved by the SRI International Institutional Review Board committee. Adult participants consented to participate and minors provided written assent along with consent from a parent or legal guardian. Participants were compensated for participation.

In-lab procedure

During one of their regular PSG follow-up laboratory overnight recordings, participants wore the ŌURA ring on a finger of the nondominant hand. All recordings were performed in sound-attenuated and temperature-controlled bedrooms at the human sleep research laboratory at SRI International. Participants self-selected lights-out and lights-on times.

Polysomnographic assessment

A six-lead electroencephalographic (EEG: F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1), submental electromyographic, and bilateral electromyographic recording was performed according to the AASM guidelines (Iber, 2007). The EEG signal was 256Hz sampled and 0.3–35Hz filtered. Sleep (wake, N1, N2, N3, and REM) was scored in 30-s epochs according to AASM rules (Iber, 2007) by an experienced scorer blinded to the ŌURA ring results. Time in bed (TIB, min) was determined as the period between lights-off and lights-on. Total sleep time was calculated as the time spent sleeping minus the time spent falling asleep and the amount of wakefulness after the sleep onset (WASO, min). The sleep onset latency (SOL, min) was determined as the time from the lights-off to the first epoch of any sleep stage. The time spent in N1, N2, N3 and REM sleep (min) was also calculated. Arousal (number of arousals per hour of sleep) and awakening (number of awakenings per hour of sleep) indexes were calculated as indices of sleep fragmentation.

The ŌURA ring

The ŌURA ring is a commercially available "sleep tracker" measuring and processing information from several user biosignals. Rings are waterproof, made in ceramic, and come with a dedicated mobile App. They come in different sizes (U.S. standard ring sizes 6–13) and weigh about 15 g with a battery life of about three days. The ring automatically connects via Bluetooth and transfers data to a mobile platform via the dedicated App.

In the current study, we used the first version of Ōuraring algorithm, which was not changed or updated during the course of the validation. We purchased two ring sizes (U.S. 7 and 11). For each participant, the finger demonstrating the best, snug fit for the ring was chosen. Twenty-one participants had the ŌURA ring on the index, 2 on the middle, 2 on the pinky, 11 on the ring, and 5 on the thumb.

Sleep lab technicians assured that the PSG recording was synchronized with the ŌURA mobile App time and that there was a connection between the ŌURA ring and the ŌURA mobile App. All data from the ŌURA ring and the PSG were anonymized using ad-hoc created codes. The app allows access to the summary night data but not the EBE data. Therefore, we requested the raw data from the Ōuraring company, which agreed to provide 30-s EBE data for each recording as well as technical information and support on the ŌURA ring and associated mobile App, allowing us to accurately perform EBE analysis. Each morning, the ŌURA ring data were sent to ŌURA tech staff, who subsequently provided 30-s-by-30-s data. Ōuraring was not involved in any other aspects of the study; Ōuraring did not have access to participant information nor access to the PSG staging.

Participants wore the ŌURA ring from the time they arrived at the lab until to the next morning and no action was required by them. The ŌURA ring collected data from the participants' fingers continuously and a proprietary algorithm determined sleep stages (wake, "light", "deep" and REM sleep). For each night, we calculated the following parameters, which were all aligned with PSG lights-off and lights-on time to match the PSG sleep staging): sleep onset latency (ŌURA-SOL, min), time spent in "deep sleep" (ŌURA-N3, min; equivalent of PSG N3 sleep), time spent in REM sleep (ŌURA-REM, min), time spent in "light sleep" (ŌURA-N1 + N2, min; equivalent of PSG N1 + N2 sleep), total time spent asleep (ŌURA-TST, min; equivalent of PSG TST), and periods of wakefulness after sleep onset (ŌURA-WASO, min; equivalent of PSG WASO). An example of a typical participant's PSG and ŌURA hypnogram (stages of sleep plotted as a function of time of the night) is provided in Figure 1.

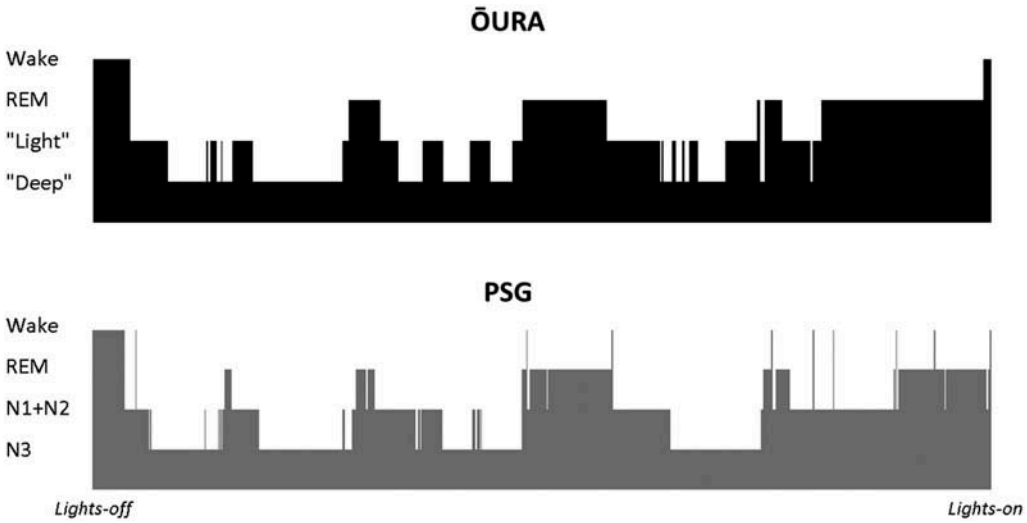

FIGURE 1 Hypnogram (sleep stages plotted as a function of time of the night) from the ŌURA ring and polysomnography (PSG) obtained from a participant's recording showing typical PSG-ŌURA discrepancies. REM, rapid eye movement.

Statistical Analyses

Summary all-night PSG and equivalent ŌURA sleep measures were compared using paired *t*-tests. The level of agreement between PSG and equivalent ŌURA sleep measures was assessed by the Bland-Altman plots (Bland & Altman, 1986). Mean difference (or Bias), standard deviation and ± 95% CI of the Bias, and lower and upper agreement limits (mean difference ± 1.96**SD*) between ŌURA and PSG sleep measures were calculated. Biases were tested against zero for significance. A positive Bias indicates that the ŌURA ring underestimates PSG sleep measures and a negative Bias indicates that the ŌURA ring overestimates them. The number of participants falling outside a-priori-set clinically satisfactory ranges for PSG outcomes, i.e., a difference between PSG and ŌURA ≤ 30 min for TST and WASO, was determined to allow more insight into the potential clinical relevance of the measurement and comparison with previous studies (de Zambotti, Baker, et al., 2015; de Zambotti, Baker, et al., 2016; Meltzer et al., 2015; Meltzer, Walsh, Traylor, & Westin, 2012; Montgomery-Downs, Insana, & Bond, 2012; Werner, Molinari, Guyer, & Jenni, 2008).

EBE analysis (30-s epochs based) was performed in order to obtain measures of sensitivity (proportion of PSG epochs identified correctly as "Sleep" by ŌURA), specificity (proportion of PSG epochs identified correctly as "Wake" by ŌURA), agreement with PSG in detecting "light sleep" (proportion of PSG N1+N2 epochs identified correctly as "light sleep" by ŌURA), "deep sleep" (proportion of PSG N3 epochs identified correctly as "deep sleep" by ŌURA), and REM sleep (proportion of PSG REM epochs identified correctly as "REM sleep" by ŌURA).

Additional analyses were also performed: (a) Multiple regression models were used to investigate the relationship between the PSG-ŌURA discrepancies in summary sleep measures (Dependent Variables: PSG-ŌURA discrepancies in WASO, "light sleep," "deep sleep," and REM sleep) and PSG metrics indicating sleep disruption (Independent Variables in each model: PSG WASO and arousal index). One participant was excluded from the WASO regression model (but kept in all other analyses) because their WASO was more than 3 *SD* greater than the mean. Additional models also tested age, BMI, and sex as potential confounders for PSG-ŌURA discrepancies. (b) When the ŌURA ring misclassified PSG REM sleep, we calculated the proportions of other sleep stages assigned in its place by the ŌURA algorithm (percentage of ŌURA wake, "light sleep" or "deep sleep"). (c) To explore the effect of "ring position" on PSG-ŌURA discrepancies, ANCOVA models were run with "ring position" as a three-level categorical factor ("index," "ring," and "other" fingers), using PSG WASO and arousal index as covariates.

Finally, we took advantage of the age range of our sample to assess if the ŌURA ring was able to detect effects of age well established in the literature and evident using PSG. Thus, Pearson's correlations were used to assess if the ŌURA ring was able to capture the well-established drop in the amount of N3 sleep with advancing age across adolescence (Baker et al., 2016; Colrain & Baker, 2011; Feinberg & Campbell, 2010). Considering the alarming evidence of insufficient sleep in this age group together with the detrimental health consequences of sleep loss (Hagenauer, Perryman, Lee, & Carskadon, 2009), we also investigated the percentage of participants ŌURA ring correctly categorized into three PSG-defined and commonly used TST ranges of < 6 hr (N = 11), between 6 and 7 hr (N = 16) and more than 7 hr (N = 14) at night. In all models, $p < .05$ was considered significant.

RESULTS

Comparisons Between Polysomnographic (PSG) and Equivalent ŌURA Sleep Measures

PSG and ŌURA sleep measures are provided in Table 1. Summary measures of TST, SOL, "light sleep," and WASO derived from the PSG and the ŌURA ring did not differ from each other. When compared to PSG, the ŌURA ring significantly underestimated time spent in N3 (or "deep sleep"; $p = .004$) and significantly overestimated time spent in REM sleep ($p = .034$).

Bland-Altman plots

Bland-Altman plots for TST, SOL, WASO, REM sleep, time in N1 + N2 ("light sleep"), and Time in N3 ("deep sleep") are provided in Figure 2. Biases, *SD* and ± 95% CI of the Biases, Bland-Altman plots upper and lower agreement limits (mean difference ± 1.96**SD*), and a-priori-set clinically satisfactory limits for TST and WASO (discrepancies ≥ 30 min) are provided in Table 2.

The ŌURA ring significantly underestimated PSG N3 and overestimated PSG REM ($p < .05$). None of the other ŌURA metrics significantly underestimated or overestimated the PSG parameters. Five participants (12% of the sample) exceeded the a-priori-set clinically satisfactory ranges for TST and six participants (15% of the sample) exceeded these ranges for WASO.

Epoch-by-epoch (EBE) analysis

Overall, ŌURA had 96% sensitivity (ability to detect sleep), 48% specificity (ability to detect wake), 65% agreement in detecting "light sleep," 51% agreement in detecting "deep sleep," and 61% agreement in detecting REM sleep, relative to PSG (see Table 3 for details).

Understanding PSG-ŌURA Discrepancies

The multiple regression model for PSG-ŌURA discrepancy in WAS0 was significant ($R^2 = .33$, $p < .001$), with the amount of PSG WASO as a significant factor ($\beta = .57$, $p < .001$). Having a greater amount of WASO was associated with a greater WASO discrepancy. Arousal Index was not a significant factor. None of the other models was significant. Age, BMI, and sex were not significant factors in any of the models.

When the ŌURA ring misclassified PSG REM sleep, the ŌURA algorithm assigned "light sleep" for 76 ± 23% (± 95% CI: 68–83%) of the time, "awake" for 16 ± 21% (± 95% CI: 9–23%) of the time, and "deep sleep" for 8 ± 13% (± 95% CI: 4–13%) of the time.

We also explored the potential effect of "ring position" on PSG-ŌURA discrepancies. ANCOVA models were significant for PSG-ŌURA discrepancies in "light sleep" ($F_{2,36} = 5.91$, $p = .006$) and REM sleep ($F_{2,36} = 10.10$, $p < .001$), with "ring position" being a significant factor. Bonferroni post-hoc tests indicated that the PSG-ŌURA discrepancies in "light sleep" ($p = .010$) and in "REM sleep" ($p = .034$) were greater in those participants having the ring on the ring finger compared to both the index and the other fingers (see Figure 3).

We also investigated if the ŌURA ring was able to detect a well-established literature finding, that is, the decline in N3 (slow wave) sleep with advancing age in adolescence. Both PSG N3

TABLE 1
Polysomnographic (PSG) and ŌURA Sleep Measures From an Overnight Laboratory Recording in a Sample of Forty-One Adolescents and Young Adults

	PSG			*ŌURA*				
	Mean ± SD	*± 95% CI*	*Min–Max*	*Mean ± SD*	*± 95% CI*	*Min–Max*	*t*	*p*
Lights-off (hh:mm)	24:04 ± 00:56	23:46–24:20	22:04–01:58	–	–	–	–	–
Lights-on (hh:mm)	07:14 ± 00:42	07:01–07:27	05:37–08:59	–	–	–	–	–
TIB (min)	429 ± 66	409–450	292–595	–	–	–	–	–
TST (min)	392 ± 59	373–410	282–563	393 ± 61	374–413	276–544	–.39	.700
SOL (min)	12 ± 11	8–15	0–59	12 ± 12	8–16	0–47	–.22	.825
WASO (min)	26 ± 21	19–32	4–80	24 ± 26	16–32	0–143	.47	.639
Awakening Index (N awakenings per hour of sleep)	3.0 ± 1.1	2.7–3.3	1.2–5.3	–	–	–	–	–
Arousal Index (N arousals per hour of sleep)	9.0 ± 4.2	7.7–10.3	4.0–24.9	–	–	–	–	–
Time in N1 (min)	20 ± 10	17–23	6–43	–	–	–	–	–
Time in N2 (min)	183 ± 52	167–199	92–285	–	–	–	–	–
Time in N1+N2 (“light sleep”) (min)	203 ± 58	185–221	110–310	206 ± 53	190–223	109–338	–.36	.722
Time in N3 (“deep sleep”) (min)	97 ± 34	87–108	27–171	78 ± 39	65–90	1–137	3.04	.004
Time in REM (min)	92 ± 26	83–100	43–147	109 ± 62	89–128	23–301	–2.20	.034

REM, rapid eye movement; SOL, sleep onset latency; TIB, time in bed; TST, total sleep time; WASO, wake after sleep onset.

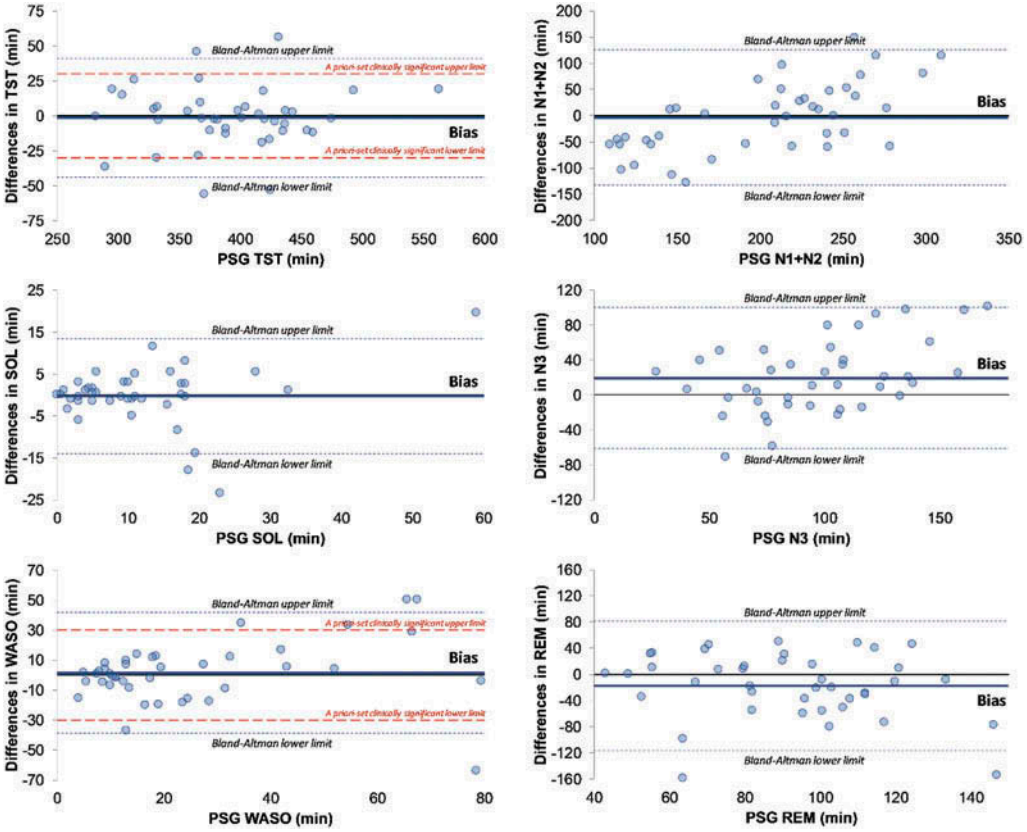

FIGURE 2 Bland-Altman plots for total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), time in N1 + N2 ("light sleep") and time in N3 ("deep sleep"). Individuals' PSG-ŌURA discrepancies on sleep metrics (y axis) are plotted as a function of the PSG metrics (x axis). Zero line and Biases are marked. The dotted lines represent the upper and lower Bland-Altman agreement limits (mean difference ± 1.96 * *SD*). The dashed lines represent the upper and lower a-priori-set clinically satisfactory limits for TST and WASO (± 30 min from the zero line).

sleep (R^2 = .46, p < .001) and equivalent ŌURA "deep sleep" (R^2 = .27, p = .001) were negatively related to participants' age (Figure 4). Finally, the percentage of participants that the ŌURA ring correctly categorized into the three PSG-defined TST ranges were, respectively, 90.9% for PSG TST < 6 hr, 81.3% for PSG TST 6–7 hr, and 92.9% for PSG TST > 7 hr.

DISCUSSION

The ŌURA ring showed good agreement with PSG in the whole night estimation of TST, SOL, WASO, and N1 + N2 ("light") sleep in this group of healthy adolescents and young adults, with 87.8% and 85.4% of the participants in the group lying within the a-priori-set clinically

TABLE 2
Biases, *SD* and ± 95% CI of the Biases, Upper and Lower Agreement Limits of Bland-Altman Plots for Polysomnographic (PSG), and Equivalent ŌURA Sleep Measures

	Bias ± SD	*± 95% CI of the Bias*	*Lower Agreement Limit*	*Upper Agreement Limit*
TST (min)	–1.3 ± 21.7	–7.8 – 5.3	–43.9	41.3
SOL (min)	–0.2 ± 7.0	–2.4 – 1.9	–14.0	13.5
WASO (min)	1.5 ± 20.7	–4.8 – 7.9	–39.0	42.0
Time in N1+N2 (min)	–3.7 ± 66.2	–23.9 – 16.5	–133.4	126.0
Time in N3 (min)	19.6 ± 41.2	7.0 – 32.2	–61.2	100.4
Time in REM (min)	–17.2 ± 50.2	–32.6 – –1.9	–115.5	81.1

REM, rapid eye movement; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset.

TABLE 3
Mean, *SD*, and ± 95% CI for Indices Derived From Epoch-By-Epoch (EBE) Analysis

	Mean ± SD	*± 95% CI of the Bias*
Sensitivity (in detecting sleep)	95.5 ± 4.5	94.1–96.9
Specificity (in detecting wake)	48.1 ± 19.1	42.0–54.1
PSG-ŌURA agreement for N1+N2—"light sleep"	64.6 ± 13.9	60.3–69.0
PSG-ŌURA agreement for N3—"deep sleep"	50.9 ± 24.5	43.2–58.6
PSG-ŌURA agreement for REM sleep	61.4 ± 22.8	54.2–68.6

REM, rapid eye movement

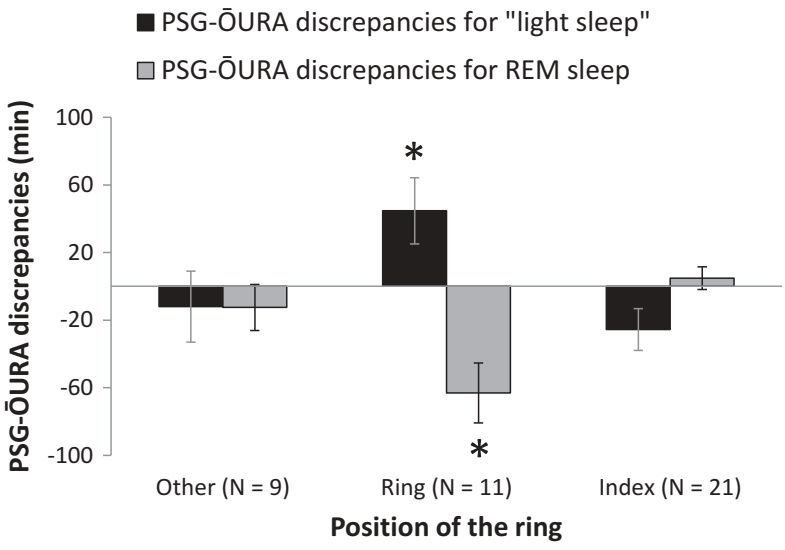

FIGURE 3 Polysomnographic (PSG)-ŌURA discrepancies in "light sleep" and in rapid-eye-movement (REM) sleep as a function of ring position. Asterisks indicate significant ($p < 0.05$) differences from both "other" and "index" fingers.

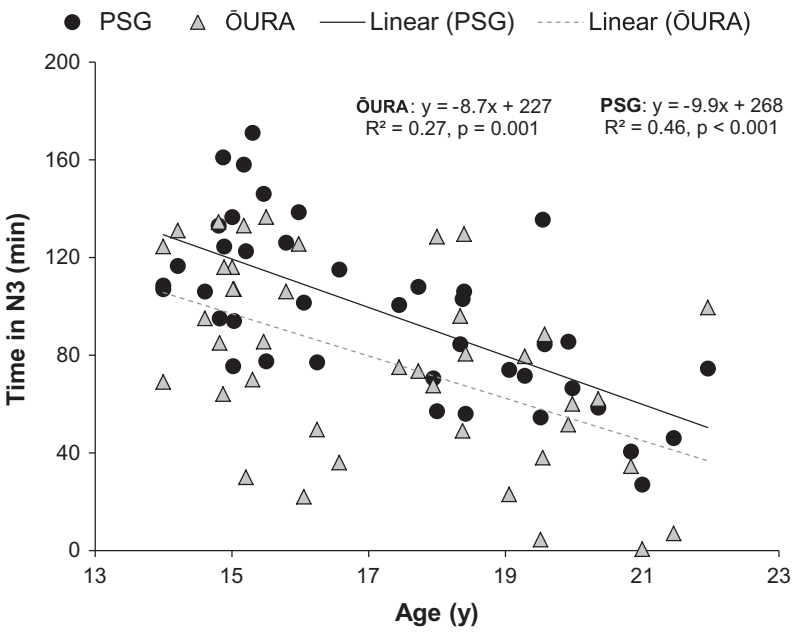

FIGURE 4 Relationships between polysomnographic (PSG) N3 sleep (circles) and ŌURA "deep sleep" (triangles) with participants' age.

satisfactory ranges for TST and WASO (≤ 30 min difference), respectively. As with other actigraphy devices and consumer wearable products, ŌURA is limited in specificity, that is, its ability to detect wake measured on an epoch-by-epoch basis. While the ŌURA ring significantly underestimated PSG N3 sleep (by about 20 min), which remained consistent across the age range, it was able to capture a significant relationship between "deep" sleep and age, with older participants having less "deep" sleep than younger participants, a well-known finding in the literature based on PSG data (Baker et al., 2016; Colrain & Baker, 2011; Feinberg & Campbell, 2010). In addition, the proportion of participants ŌURA correctly categorized into the three PSG-defined TST ranges of < 6 hr, 6–7 hr, > 7 hr were 90.9%, 81.3%, and 92.9% respectively. The ability of a device to accurately classify "short sleepers" in adolescents is important, considering the growing concern for the lack of sleep in this age group (Hagenauer et al., 2009). These results suggest that ŌURA ring is sensitive enough to capture overall differences in sleep patterns with limitations in detecting wake, as detected by PSG.

The Bland-Altman plot limits of agreement for SOL, TST, and WASO of the current study were narrow or comparable to that of previous investigations about the validity of other commercially available fitness-trackers in adolescent samples (de Zambotti, Baker, et al., 2015; de Zambotti, Baker, et al., 2016; Roane et al., 2015; Toon et al., 2015). PSG-ŌURA discrepancies and agreement limits were also comparable to those provided by a publicly available internal sleep lab validation from Ōuraring in a group of 14 participants (38.0 ± 10.2 years old, ranging in age between 9 and 48 years; Kinnunen, 2016). In that study, TST, SOL, and WASO as derived by standard PSG (in 8 participants) or EOG recordings (in 6 participants) did

not significantly differ from measures derived from the ŌURA ring. Compared to the ŌURA internal validation (Kinnunen, 2016), the agreement limits in our study were narrower for the sleep measures (for example, TST [our study: –44 min minus 41 min; ŌURA internal validation: –64 min minus 66 min], WASO [our study: –39 min minus 42 min; ŌURA internal validation: –50 min minus 50 min]). To our knowledge these are the only comparable and currently available data comparing PSG and the ŌURA ring.

The ŌURA ring did not show systematic PSG TST overestimation and PSG WASO underestimation. In contrast, some previous studies of fitness trackers found significant bias for sleep and wake assessment in adolescent samples (de Zambotti, Baker, et al., 2015; de Zambotti, Baker, et al., 2016) while others did not (Roane et al., 2015; Toon et al., 2015). However, there were greater PSG-ŌURA discrepancies in overnight total WASO for participants with more PSG WASO (in the current study, the PSG-ŌURA discrepancy was minimal when the amount of PSG WASO was about 18 min). Further, EBE analysis showed that while the ŌURA ring had a high sensitivity in detecting sleep (95.5%), it had lower specificity in detecting wake (48%), similar to findings by us and others for other sleep-trackers (de Zambotti, Baker, et al., 2015; de Zambotti, Baker, et al., 2016; de Zambotti, Claudatos, et al., 2015; Kolla et al., 2016; Montgomery-Downs et al., 2012), and standard actigraphy (Paquet et al., 2007; Sadeh, 2011). It is still unclear how the ŌURA ring integrates information from other biosensors, in addition to motion, to estimate wake. As speculated by others, part of the reason for a low specificity of actigraphy-based devices is the underestimation of wake time that may be due to their limited ability to identify periods of immobility as wake time (Marino et al., 2013). For multisensor devices, such as the ŌURA ring, the use of other signals including heart rate and its variability should theoretically increase the ability of the device to discriminate sleep and wake in immobile situations. In fact, several lines of evidence indicate that heart rate variability metrics show extensive changes from wake to sleep conditions, as well as between NREM and REM sleep stages (Trinder, 2007). Further development in the detection algorithm from the Ōuraring company or the introduction of other new multisensory devices able to discriminate sleep stages may ultimately reveal if the overall issue with specificity can be addressed or not by a multisensory approach combined with sophisticated analytic methods.

In this study, PSG-ŌURA discrepancies were independent from age, BMI, or sex, which is similar to findings for another sleep tracker in a group of children and adolescents (Soric et al., 2013). In contrast, we previously found a strong age-dependent effect in the accuracy of Jawbone UP™ in determining PSG outcomes in a different, younger sample of the NCANDA cohort (de Zambotti, Baker, et al., 2015). Similarly, Meltzer et al. (2012) tested for the validity of standard actigraphy against PSG and found a shift from underestimation of TST in children (3–12 years) to overestimation of TST in adolescents (13–18 years), and an inverted pattern for WASO, suggesting an age-dependent relationship for the discrepancies between actigraphy and PSG in children and adolescents. The reason for different findings between studies is unclear; however, we can speculate that an increase in motionless wakefulness (that would be misclassified as sleep) with age, may affect entirely motion-based detection of sleep–wake patterns, thus affecting actigraphy-based devices. On the other hand, multisensory devices like ŌURA, which use other biosignals in combination with motion to obtain information about wake and sleep states, may be less biased by changes in motion relationships to wakefulness and sleep.

EBE analysis showed that ŌURA accurately detected “light” and “deep” sleep in 65% and 51% of the epochs, respectively. It also accurately detected REM sleep epochs 61% of the time,

with an overall overestimation of PSG REM sleep (by about 17 min). When the ŌURA ring misclassified PSG REM sleep, the algorithm classified the epoch as "light sleep" (76%) for the majority of the time. Distinguishing sleep stages such as REM and N3 with non-EEG based systems has been challenging and is a goal of several commercial sleep-trackers, with mixed success. We previously reported that Jawbone UP's "sound sleep" was positively associated with PSG time in N2 and time in REM, but not with N3 sleep. "Light sleep" was positively associated with the PSG arousal index, awakening index, and N2 and N3 sleep (de Zambotti, Baker, et al., 2015). Other devices have classified "deep" sleep as a combination of N3 and REM, which they have tended to overestimate, or had varying results depending on the amount of deep sleep (Mantua et al., 2016).

The potential for devices to be able to detect sleep parameters beyond binary sleep–wake is attractive since it would allow estimates of sleep architecture to be determined in larger populations for longer periods of time than is currently possible with PSG. Algorithms that use information derived from heart rate analysis in addition to motion could potentially improve differentiation between sleep stages because of the established changes that are evident in heart rate variability indices in response to PSG sleep stages and phasic sleep events (Trinder, 2007) together with evidence of strong dynamic interplay between central and autonomic nervous systems during sleep. In particular, CNS measures of cortical electroencephalographic activity reflecting synchronization seem to be dynamically related to autonomic nervous system measures of heart rate variability of low sympathovagal balance (see Brandenberger, Ehrhart, & Buchheit, 2005; Brandenberger, Ehrhart, Piquard, & Simon, 2001; Otzenberger, Simon, Gronfier, & Brandenberger, 1997; Thomas et al., 2014). Clearly, further work is needed to determine what combination of sensors might be used to optimally develop an algorithm that differentiates sleep stages sufficiently well to detect real differences or changes in healthy and clinical populations.

Interestingly, we found that PSG-ŌURA discrepancies for "light sleep" and REM were greater on the ring finger compared to the other fingers, a result that was independent from the amount of PSG sleep fragmentation. Assuming that the main parameters that ŌURA uses to determine sleep stages are motion and optical sensor outputs, it is possible that the different blood supply among fingers may partially explain these results. For example, it has been shown that SpO_2 values differ between fingers as well as hands, suggesting a finger-dependent difference in accuracy of the pulse oximetry signal (Basaranoglu et al., 2015). Further studies should confirm and better characterize the dependency of the PSG-ŌURA discrepancies on the ring position by having the same participants simultaneously wear different rings on different fingers. It should also be noted that we had only two ring sizes available and chose the best-fitting finger for each participant. Possibly, if participants personally choose the ring that fits the finger of their choice, as suggested by Ōuraring, results may differ.

The current study is based on a single in-lab night used for the comparison and does not address the issue of reliability over time. Another consumer-based wearable device was reported to be unreliable over longitudinal assessments in a nonclinical population, with a large percentage of missing data (up to 70%) which was ascribed to device failure (Baroni, Bruzzese, Di Bartolo, & Shatkin, 2016). While we did not record any failure or malfunctioning of the ŌURA ring in this study, we have no data addressing reliability of the ŌURA ring over multiple nights or in nonlaboratory settings. Also, these data are from healthy adolescents and young adults, and we cannot generalize our results to different populations.

Despite these limitations and the fact that the ŌURA ring uses a proprietary algorithm, unknown to us, first results of the ability of the ŌURA ring to distinguish sleep stages could be viewed as promising; however, future development and validation are needed.

ACKNOWLEDGMENTS

We would like to thank Lena Kardos, Stephanie Claudatos, and Devika Nair for their effort in the data collection process. We would like to emphasize that this was an independent investigation, however, we would like to thank Ōuraring, who agreed to provide technical details of their product and the epoch-by-epoch data.

FUNDING

This study was supported by the National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA); grant: AA021696 (IMC+FCB).

REFERENCES

American Medical Association & American Academy of Sleep Medicine. (2010). *Resolution 503: Insufficient sleep in adolescents*. Retrieved from https://www.avma.org/Advocacy/National/Other/Documents/resolution_519.pdf

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, *26*(3), 342–392.

Baker, F., Willoughby, A., de Zambotti, M., Franzen, P., Prouty, D., Javitz, H.,… Colrain, I. (2016). Age-related differences in sleep architecture and electroencephalogram in adolescents in the National Consortium on Alcohol and NeuroDevelopment in Adolescence Sample. *Sleep*, *39*(7), 1429–1439.

Baroni, A., Bruzzese, J., Di Bartolo, C., & Shatkin, J. (2016). Fitbit Flex: An unreliable device for longitudinal sleep measures in a non-clinical population. *Sleep and Breathing*, *20*(2), 853–854.

Basaranoglu, G., Bakan, M., Umutoglu, T., Zengin, S. U., Idin, K., & Salihoglu, Z. (2015). Comparison of SpO2 values from different fingers of the hands. *SpringerPlus*, *4*, 561.

Bland, J., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*(8476), 307–310.

Brandenberger, G., Ehrhart, J., & Buchheit, M. (2005). Sleep stage 2: An electroencephalographic, autonomic, and hormonal duality. *Sleep*, *28*(12), 1535–1540.

Brandenberger, G., Ehrhart, J., Piquard, F., & Simon, C. (2001). Inverse coupling between ultradian oscillations in delta wave activity and heart rate variability during sleep. *Clinical Neurophysiology*, *112*(6), 992–996.

Brown, S. A., Brumback, T., Tomlinson, K., Cummins, K., Thompson, W. K., Nagel, B. J.,… Chung, T. (2015). The National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA): A multisite study of adolescent development and substance use. *Journal of Studies on Alcohol and Drugs*, *76*(6), 895–908.

Colrain, I., & Baker, F. (2011). Changes in sleep as a function of adolescent development. *Neuropsychology Review*, *21*(1), 5–21.

de Zambotti, M., Baker, F., & Colrain, I. (2015). Validation of sleep-tracking technology compared with polysomnography in adolescents. *Sleep*, *38*(9), 1461–1468.

de Zambotti, M., Baker, F., Willoughby, A., Godino, J., Wing, D., Patrick, K., & Colrain, I. (2016). Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiology & Behavior*, *158*, 143–149.

de Zambotti, M., Claudatos, S., Inkelis, S., Colrain, I., & Baker, F. (2015). Evaluation of a consumer fitness-tracking device to assess sleep in adults. *Chronobiology International*, *37*(2), 1024–1028.

de Zambotti, M., Godino, J., Baker, F., Cheung, J., Patrick, K., & Colrain, I. (2016). The boom in wearable technology: Cause for alarm or just what is needed to better understand sleep? *Sleep, 39*(9), 1761–1762.

Feinberg, I., & Campbell, I. (2010). Sleep EEG changes during adolescence: An index of a fundamental brain reorganization. *Brain and Cognition*, *72*(1), 56–65.

Gradisar, M., Gardner, G., & Dohnt, H. (2011). Recent worldwide sleep patterns and problems during adolescence: A review and meta-analysis of age, region, and sleep. *Sleep Medicine*, *12*(2), 110–118.

Hagenauer, M., Perryman, J., Lee, T., & Carskadon, M. (2009). Adolescent changes in the homeostatic and circadian regulation of sleep. *Developmental Neuroscience*, *31*(4), 276–284.

Iber, C. (2007). *The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications*. Darien, IL: American Academy of Sleep Medicine.

Kinnunen, H. (2016). *Validity of the ŌURA ring in determining sleep quantity and quality. Sleep lab validation of a wellness ring in detecting sleep patterns based on photoplethysmogram, actigraphy and body temperature*. Retrieved from https://support.ouraring.com/support/solutions/6000114605

Kolla, B. P., Mansukhani, S., & Mansukhani, M. P. (2016). Consumer sleep tracking devices: A review of mechanisms, validity and utility. *Expert Review of Medical Devices*, *13*(5), 497–506.

Mantua, J., Gravel, N., & Spencer, R. M. (2016). Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. *Sensors*, *16*(5), 646.

Marino, M., Li, Y., Rueschman, M. N., Winkelman, J., Ellenbogen, J., Solet, J.,… Buxton, O. M. (2013). Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*, *36*(11), 1747–1755.

Meltzer, L., Hiruma, L., Avis, K., Montgomery-Downs, H., & Valentin, J. (2015). Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep*, *38*(8), 1323–1330.

Meltzer, L., Walsh, C., Traylor, J., & Westin, A. (2012). Direct comparison of two new actigraphs and polysomnography in children and adolescents. *Sleep*, *35*(1), 159–166.

Montgomery-Downs, H., Insana, S., & Bond, J. (2012). Movement toward a novel activity monitoring device. *Sleep and Breathing*, *16*(3), 913–917.

Otzenberger, H., Simon, C., Gronfier, C., & Brandenberger, G. (1997). Temporal relationship between dynamic heart rate variability and electroencephalographic activity during sleep in man. *Neuroscience Letters*, *229*(3), 173–176.

Paquet, J., Kawinska, A., & Carrier, J. (2007). Wake detection capacity of actigraphy during sleep. *Sleep*, *30*(10), 1362–1369.

Roane, B., Van Reen, E., Hart, C., Wing, R., & Carskadon, M. (2015). Estimating sleep from multisensory armband measurements: Validity and reliability in teens. *Journal of Sleep Research*, *24*(6), 714–721.

Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: An update. *Sleep Medicine Reviews*, *15*(4), 259–267.

Soric, M., Turkalj, M., Kucic, D., Marusic, I., Plavec, D., & Misigoj-Durakovic, M. (2013). Validation of a multi-sensor activity monitor for assessing sleep in children and adolescents. *Sleep Medicine Reviews*, *14*(2), 201–205.

Thomas, R. J., Mietus, J. E., Peng, C.-K., Guo, D., Gozal, D., Montgomery-Downs, H.,… Goldberger, A. L. (2014). Relationship between delta power and the electrocardiogram-derived cardiopulmonary spectrogram: Possible implications for assessing the effectiveness of sleep. *Sleep Medicine*, *15*(1), 125–131.

Toon, E., Davey, M., Hollis, S., Nixon, G., Horne, R., & Biggs, S. (2015). Comparison of commercial wrist-based and smartphone accelerometers, actigraphy, and PSG in a clinical cohort of children and adolescents. *Journal of Clinical Sleep Medicine*, *12*(3), 343–350.

Trinder, J. (2007). Cardiac activity and sympathovagal balance during sleep. *Sleep Medicine Clinics*, *2*(2), 199–208.

Van de Water, A., Holmes, A., & Hurley, D. (2011). Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography: A systematic review. *Journal of Sleep Research*, *20*(1 Pt 2), 183–200.

Werner, H., Molinari, L., Guyer, C., & Jenni, O. (2008). Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns. *JAMA Pediatrics*, *162*(4), 350–358.