Sleep Breath (2018) 22:139–147
DOI 10.1007/s11325-017-1547-9

SLEEP BREATHING PHYSIOLOGY AND DISORDERS • ORIGINAL ARTICLE

Comparison of two home sleep testing devices with different strategies for diagnosis of OSA

Tyler Gumb[1] · Akosua Twumasi[2] · Shahnaz Alimokhtari[3] · Alan Perez[3] · Kathleen Black[3] · David M. Rapoport[1,2] · Jag Sunderram[4] · Indu Ayappa[1,2]

Received: 30 December 2016 / Revised: 24 July 2017 / Accepted: 31 July 2017 / Published online: 19 August 2017

Abstract
Purpose Home sleep testing devices are being widely used in diagnosis/screening for obstructive sleep apnea (OSA). We examined differences in OSA metrics obtained from two devices with divergent home monitoring strategies, the Apnea Risk Evaluation System (ARES™, multiple signals plus forehead reflectance oximetry) and the Nonin WristOx$_2$™ (single channel finger transmission pulse oximeter), compared to differences from night-night variability of OSA.
Methods One hundred fifty-two male/26 female subjects (BMI = 30.3 ± 5.6 kg/m^2, age = 52.5 ± 8.9 years) were recruited without regard to OSA symptoms and simultaneously wore both ARES™ and Nonin WristOx$_2$™ for two nights (n = 351 nights). Automated analysis of the WristOx$_2$ yielded oxygen desaturation index (ODI$_{Ox2}$, ≥4% O2 dips/h), and automated analysis with manual editing of ARES™ yielded AHI4$_{ARES}$ (apneas + hypopneas with ≥4% O2 dips/h) and RDI$_{ARES}$ (apneas + hypopneas with ≥4% O2 dips/h or arousal surrogates). Baseline awake oxygen saturation, percent time < 90% O2 saturation (%time < 90%O2Sat), and O2 signal loss were compared between the two methods.
Results Correlation between AHI4$_{ARES}$ and ODI$_{Ox2}$ was high (ICC = 0.9, 95% CI = 0.87–0.92, p < 0.001, bias ± SD = 0.7 ± 6.1 events/h). Agreement values for OSA diagnosis (77–85%) between devices were similar to those seen from night-to-night variability of OSA using a single device. Awake baseline O2 saturation was significantly higher in the ARES™ (96.2 ± 1.6%) than WristOx$_2$™ (92.2 ± 2.1%, p < 0.01). There was a significantly lower %time < 90%O2Sat reported by the ARES™ compared to WristOx$_2$ (median (IQR) 0.5 (0.0, 2.6) vs. 2.1 (0.3, 9.7), p < 0.001), and the correlation was low (ICC = 0.2).
Conclusions OSA severity metrics predominantly dependent on change in oxygen saturation and metrics used in diagnosis of OSA (AHI4 and ODI) correlated well across devices tested. However, differences in cumulative oxygen desaturation measures (i.e., %time < 90%O2Sat) between the devices suggest that caution is needed when interpreting this metric particularly in populations likely to have significant hypoxia.

Keywords Sleep disordered breathing · Oximetry · Portable monitoring · Obstructive sleep apnea · Limited channel monitoring

Work performed at NYU Sleep Disorders Center and Rutgers Robert Wood Johnson Medical School

✉ Indu Ayappa
indu.ayappa@mssm.edu

[1] Division of Pulmonary, Critical Care and Sleep Medicine, New York University, New York, NY, USA

[2] Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1232, New York, NY 10029, USA

[3] Environmental and Occupational Health Sciences Institute, Rutgers Biomedical and Health Sciences, Piscataway, NJ, USA

[4] Department of Medicine, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ 08903, USA

Introduction

Diagnosis of obstructive sleep apnea (OSA) is increasingly being made with limited channel monitoring (level III and IV devices) in both clinical and epidemiological settings. However, the devices being used vary from single channel oximeters to devices that monitor multiple signals such as airflow, breathing efforts, oxygen saturation, and surrogates for arousal. When disease is moderate to severe, both level

IV and level III devices have been shown to have value in screening as well as in establishing the diagnosis of OSA [1, 2]. The American Academy of Sleep Medicine (AASM) has suggested that only level III devices should be used to obtain OSA severity indices comparable to that obtained with full polysomnography (PSG) [3–7]. However, as noted by the recent AASM guidelines document for diagnostic testing for OSA [8], there are limited data in the literature assessing the impact of the number of parameters or technologies being used, particularly in different clinical settings. We were interested in assessing the impact of using maximally different diagnostic devices on derivation of indices of OSA severity and the agreement between devices for OSA diagnosis.

One potentially important difference between devices is the technology used for oximetry as well as the measurement site which may lead to true physiologic differences in oxygen saturation. Pulse oximetry can be measured using either transmission or reflectance techniques. Transmission oximetry detects light after it passes through relatively translucent placement sites, such as the fingertips or earlobes, whereas reflectance oximetry detects reflected light from tissue that is opaque (e.g., forehead sensors). There are known differences in the measurement of oxygen saturation (O2Sat) using these two methods [9], but, e*ven by a single technique*, we and others have shown a significant effect of pulse oximeter brand and sample averaging duration on OSA diagnosis [10–12]. No study has specifically addressed the impact on OSA diagnosis of using reflectance vs. transmission oximetry technique with currently used devices.

Additionally, identifying sleep-related respiratory events (apneas/hypopneas) using only their consequences (e.g., O2 desaturation) may differ from detecting events by reduction in airflow and assessing their significance from the oximetry signal. Finally, some devices also use surrogates for electroencephalographic (EEG) arousal and use these to maximize similarity to polysomnography detection of hypopnea. Given all these differences, the purpose of the present study was to compare the effect on diagnosis of OSA using two devices that are maximally different in their approach to measure OSA severity. The present study compares data from a widely used stand-alone level IV device, the Nonin WristOx$_2$™ (Model 3150, Plymouth, MN), and a level III device, the Apnea Risk Evaluation System (ARES™, Watermark Medical, Boca Raton, FL) that uses reflectance oximetry combined with other signals (Fig. 1). We have previously shown that the ARES™ has good agreement and adequate sensitivity and specificity with in-laboratory polysomnography for diagnosis of OSA [4]. For the present study, we compared these devices in a non-sleep clinic population which, in contrast to sleep clinic populations, does not have a high pre-test probability of OSA. Specifically, we compared (i) the amount of data loss with each device and type of oximetry, (ii) baseline awake O2 saturation and %time below 90% O2 saturation, and (iii) agreement between the values of OSA severity, and agreement for diagnosis of OSA from these devices, compared to the physiologic night-to-night variability of these metrics. We hypothesized that there would be significant differences in sleep disordered breathing (SDB) indices and diagnostic agreement for OSA when using finger pulse oximetry alone vs. a device that measures multiple parameters including airflow, forehead oxygen saturation, and surrogates for the EEG arousal.

Methods

Data from the first 178 subjects who enrolled in an ongoing study in the World Trade Center Responder population (Clinical Trials No. NCT01753999) were analyzed. The parent study collects exclusively ambulatory overnight home studies to evaluate the relationship between new onset OSA following the WTC disaster on September 11, 2001 and nasal pathology. Subjects were recruited without regard to OSA symptoms, but were not eligible if prior to September 11, 2001 they had documented evidence of OSA or significant snoring or if they were currently on treatment for OSA. For the present substudy, subjects were instructed to wear the WristOx$_2$™ and ARES™ simultaneously for two consecutive nights. Both devices were initialized on the same computer to synchronize the internal clocks. Devices were given to the patients during an in-person visit. A research coordinator provided verbal instructions that took < 5 min, and a one-page pamphlet with written instructions was also provided to the subjects. No additional interaction with the patient was required. Both devices are easy to self-apply, and we have used the ARES™ successfully with just the written instructions in previous studies [4]. The ARES™ measures oximetry on the forehead, and the subjects wore the WristOx$_2$™ on whichever hand they were more comfortable. Data were excluded from analysis if both devices were not used on the same night or if the duration of recording on either device was less than 2 h (tabulated as data loss).

WristOx$_2$™: Oximetry by transmission was sampled at 1 Hz with an averaging time of four beats. Automated analysis of the WristOx$_2$ data, using Nonin nVision data management software version 6.3, provided an index of OSA severity, the oxygen desaturation index (ODI$_{ox2}$), and % time below 90% O2 saturation (%time < 90%O2Sat) for each night. ODI$_{Ox2}$ was defined as the number of drops in saturation by at least 4% lasting a minimum of 10 s per hour of valid recording time as defined by the Nonin software. We inferred the awake baseline O2 saturation from a period at the beginning of the study within the first 10 min from the start of the recording. Poor signal quality was identified and excluded in the oximetry data using the automated algorithm on the WristOx$_2$. The

signal was manually inspected to ensure validity of the algorithm, and no further editing was required.

ARES™: Oximetry by reflectance was sampled at 100 Hz with an averaging time of three to five beats depending on signal quality and displayed at 1 Hz. Automated analysis of SDB events was followed by manual inspection and editing of events by investigator as per device use instructions. ODI is not provided by the ARES™ automated algorithm, so a direct comparison of transmittance and reflectance saturation-derived SDB was not possible in our study. The closest measure to ODI provided by the ARES™ is the AHI4 which counts apneas and those hypopneas with a ≥ 4% desaturation. The apnea hypopnea index 4% from ARES™ ($AHI4_{ARES}$) is calculated as the sum of apneas and hypopneas 4% divided by total sleep time (TST); apneas were defined as a reduction in flow amplitude of > 90% of baseline; hypopneas 4% were defined as a reduction in flow amplitude > 30% followed by ≥ 4% oxygen desaturation, or a visible reduction in flow amplitude along with a change in shape suggesting inspiratory flow limitation (IFL) followed by ≥4% oxygen desaturation. The ARES™ provides an estimate of total sleep time obtained from a combination of actigraphy and automated analysis of single channel forehead EEG recording [13].

The ARES™ also provides a respiratory disturbance index (RDI_{ARES}): sum of apneas, hypopneas 4% and hypopneaArousal divided by TST; hypopneaArousal was defined by a visible reduction (usually > 30%) in flow amplitude along with a change in shape suggesting inspiratory flow limitation and followed by arousal surrogates that included an abrupt change in head position or an increase in flow amplitude to > 2 times the amplitude during the event along with a normalization of shape.

For ARES™ oximetry data, automated algorithms were first applied to exclude areas with poor quality oximetry. This was followed by additional manual editing to exclude areas with sustained drops in O2 saturation following body position changes but not associated with SDB events.

For normally distributed variables, data are presented as mean ± SD and groups compared using paired *t* test. Data are presented as median (IQR), and Wilcoxon rank sum tests were used to compare groups when data was not normally distributed. Concordance of SDB indices was assessed by performing intraclass correlations (ICC) between ODI_{Ox2} and $AHI4_{ARES}$, and both devices' baseline O2 saturation level and %time < 90%O2Sat [14]. Bland-Altman plots were used to measure bias and differences between these measures, and Pearson correlation coefficients (*r*) were used to assess relationships between other related variables [15]. Agreement, sensitivity, and specificity for diagnosis of OSA were examined using standard cutoffs for OSA (i.e., ≥ 5/h, > 15/h). Analyses were performed between devices using all nights in all subjects and between pairs of data for comparisons across nights using the same device.

Results

Three hundred fifty-one nights of data were analyzed from 178 subjects (85.3% male, 14.6% female, BMI = 30.3 ± 5.6 kg/m^2, age = 52.5 ± 8.9 years, Epworth Sleepiness Scale (ESS) = 8.5 ± 5, 37% with ESS > 10). The reported prevalence of congestive heart failure = 2.3%, hypertension = 25%, stroke = 0.6%, MI = 2%, diabetes = 10%, gastro-esophageal reflux disease = 40%, chronic rhinosinusitis = 42%, and mental health conditions (depression, post-traumatic stress disorder, panic disorder) = 15%. Eight percent were current smokers. The prevalence of OSA in this dataset was 28% using a cutoff AHI4 ≥ 15/h (data from multiple nights collected for each subject combined) and 62% for AHI4 ≥ 5/h. The prevalence values were 10 and 22% when the coexistence of excessive daytime somnolence (ESS > 10) was included in the definition of OSA.

Majority of the subjects (73%) used both devices simultaneously for two nights, 4% for three nights, and the remaining 23% of subjects used both devices for only one night. Of data, 11.7% (*n* = 41/351 nights) were excluded for having < 2 h of data on either device (10.8% $WristOx_2$; 2.3% ARES™) leaving 310 nights with simultaneous ARES™ and $WristOx_2$ data, and 130 studies with 2 nights of data with each of the devices.

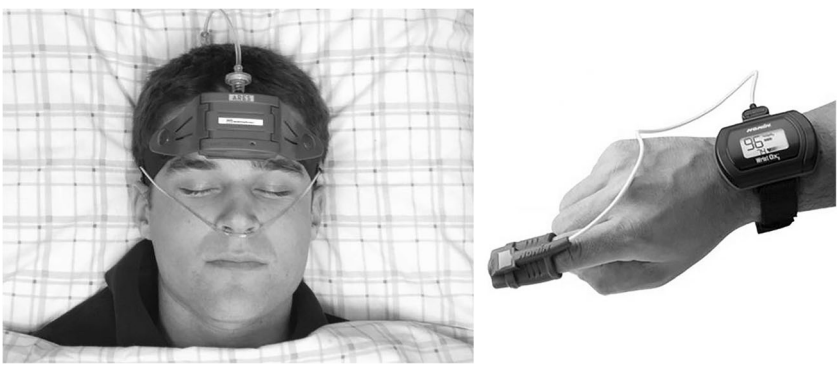

Fig. 1 ARES™ and Nonin $WristOx_2$ devices

Table 1 Average and standard deviation and median (IQR) for pertinent variables from ARES and $WristOx_2$

	All subjects, all nights (n = 310) Mean ± SD Median (IQR)	
	Ares	$WristOx_2$
Duration (h)	5.8 ± 1.5 5.6 (4.4, 6.2)	6.2 ± 1.8** 6.4 (4.9, 7.4)**
$AHI4_{ARES}/ODI_{Ox2}$	12.8 ± 14.1 8.0 (3.0, 18.0)	12.1 ± 13.3* 7.4 (3.4, 15.8)
RDI_{ARES}	26.0 ± 16.9 21.5 (13.8, 35.0)	n/a
% Time below 90% O2Sat	2.6 ± 5.0 0.5 (0.0, 2.6)	9.6 ± 17.1** 2.1 (0.3, 9.7)***
Baseline O2Sat (%) (n = 295)	96.2 ± 1.6 96.6 (95.6, 97.2)	92.2 ± 2.1** 92.0 (91.0, 94.0)***
% Artifact O2Sat	14.6 ± 15.9 8.9 (3.6, 20.3)	2.1 ± 7.9** 0.7 (0.2, 1.4)**

*$p < 0.05$; **$p < 0.01$ for comparison between ARES™ and $WristOx_2$; ***$p < 0.001$ Wilcoxon signed rank test

Comparison between devices The average duration of data recorded was 5.8 ± 1.5 h/night on the ARES™ and 6.2 ± 1.8 h/night on the $WristOx_2$™. Over the 310 nights analyzed, there was a significantly greater percentage of time with poor quality O2 saturation on the ARES™ device compared to the $WristOx_2$™ (median (IQR) 8.9% (3.6, 20.3) vs. 0.7% (0.2, 1.4), $p < 0.01$; see Table 1). Figure 2 shows the histogram of % time with poor quality O2 saturation for each device. There was no correlation between the amount of data loss that occurred on the two devices ($r = 0.1$, p = NS), suggesting that this was a device-determined, and not patient-specific finding. In addition, when we identified the individual subjects with the most data loss (top 5%) for each device, there was no overlap between individuals who showed large data loss with the ARES™ or $WristOx_2$™.

Table 1 shows significant differences in reported %time < 90%O2Sat between the ARES™ and $WristOx_2$ devices. Figure 3 shows the poor correlation in this metric between the devices, and that the ARES™ reported less %time < 90%O2Sat, especially when higher levels of percent time below 90% were measured by the $WristOx_2$. Only 213 nights (60.7%) had < 5% difference in the %time < 90%O2Sat recorded by the two devices. Baseline awake O2Sat was significantly higher with the ARES™ than the $WristOx_2$™ (Table 1). The correlation between the baseline awake O2Sat values from the two devices was poor and did not reach statistical significance ($r = 0.1$, $p = 0.08$).

BMI was correlated with the %time < 90%O2Sat in both the ARES™ ($r = 0.3$, $p < 0.01$) and the $WristOx_2$™ ($r = 0.4$, $p < 0.01$). BMI also correlated with the difference in %time < 90%O2Sat between the two devices ($r = 0.3$, $p < 0.01$).

Table 1 shows the SDB indices obtained using the ARES ($AHI4_{ARES}$, RDI_{ARES}) and $WristOx_2$ (ODI_{Ox2}). Correlation between $AHI4_{ARES}$ and ODI_{Ox2} was high (ICC = 0.9, bias ± SD = 0.7 ± 6.1, see Fig. 4). A small bias was observed with $AHI4_{ARES}$ systematically slightly higher than ODI_{Ox2} (but with wide limits of agreement). Table 2 shows the % agreement for diagnosing OSA, the associated kappa, and sensitivity, specificity, and false negative rates using cutoffs of ≥5/h and ≥ 15 events/h with ARES™ as the gold standard. If sleepiness was required for the definition of OSA, the agreement rate was unchanged when a cutoff of ≥5/h was used and slightly improved at 93% when a cutoff of ≥15/h was used. If the more sensitive metric, RDI_{ARES} (with a cutoff of ≥15/h), was used for diagnosis of OSA 20 (6.4%), additional

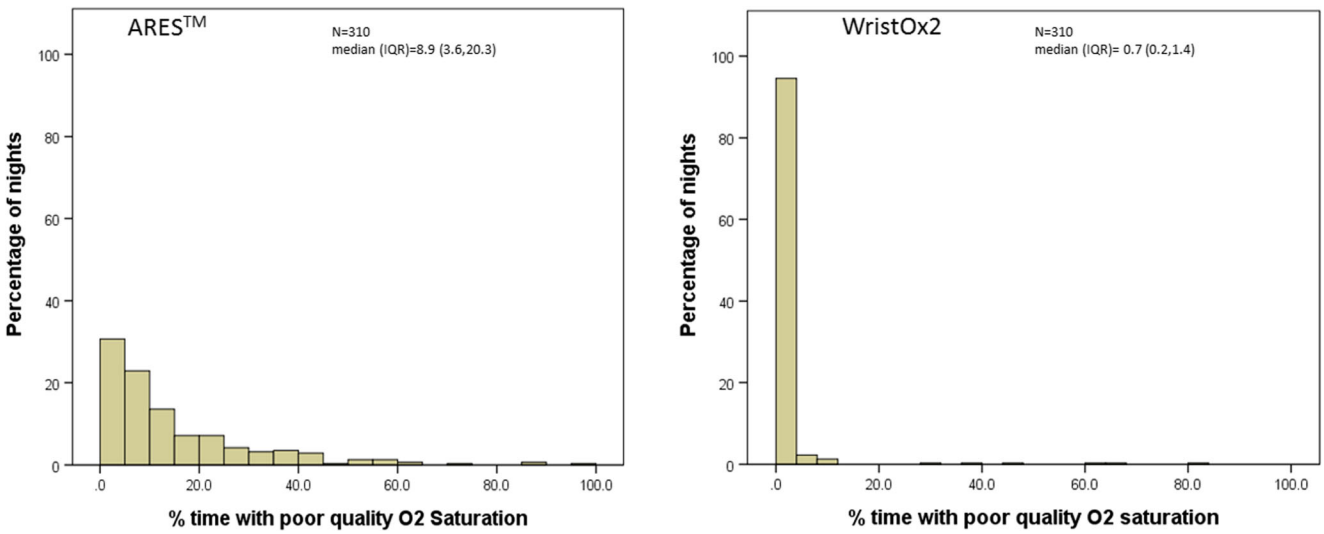

Fig. 2 Histograms showing the percentage of time with poor quality O2 saturation in both devices on the x-axis and the percentage of nights on the y-axis

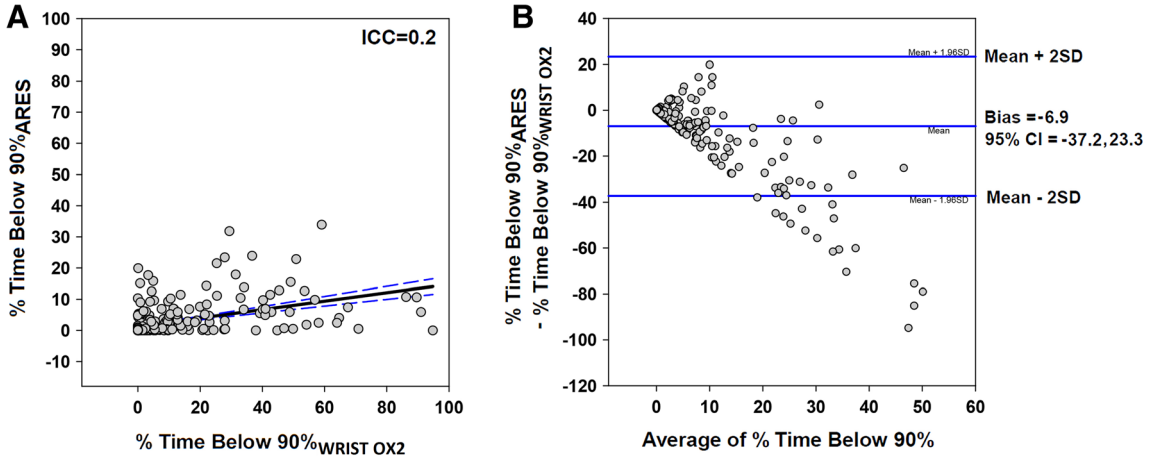

Fig. 3 **A** Scatterplot of %time < 90%O2Sat forWristOx$_2$™ and ARES™ for n = 310 nights showing significant differences in cumulative O2 desaturation time reported by the devices (ICC = 0.2, 95% CI = 0.14–0.35, $p < 0.001$). **B** Bland-Altman plot shows bias and limits of agreement

diagnoses of OSA are identified compared to AHI4 ≥ 5/h and 22 (7%) additional diagnoses are identified compared to ODI ≥ 5/h. Table 3 shows agreement for OSA diagnosis between N1 and N2.

Comparison of differences in OSA metrics between *devices* vs. between *nights using the same device* is shown in Table 4. The magnitude of the difference (bias) in the indices (AHI/ODI) between night 1 (N1) and night 2 (N2) was not statistically different and consistent with data in the literature with no significant first-night effect [16–19]. This difference in AHI4/ODI seen when the same device is used on two separate nights was of similar magnitude as the (absolute) difference in the same index between the two different devices when used simultaneously. There were small differences in %time < 90%O2Sat and baseline awake O2Sat between N1 and N2 consistent with the stated accuracy of the devices. The difference in the %time < 90% and awake O2Sat between nights (using the same device) was of lower magnitude than differences *between the two devices* used simultaneously on the same night.

Discussion

This is the first study to our knowledge that examines the differences in SDB metrics obtained from divergent home monitoring strategies using different types of oximetry and contrasting a level III forehead device and a level IV finger device. As both the ARES™ and WristOx$_2$™ devices are widely used, our results have implications for comparing clinical and epidemiological datasets. In the simultaneously recorded home data of this large dataset of non-sleep clinic subjects, there were significant differences in the baseline awake O2Sat and %time < 90%O2Sat obtained from the two devices. Despite this, the correlation between ODI_{Ox2} and $AHI4_{ARES}$ was high. Agreement and rate of false

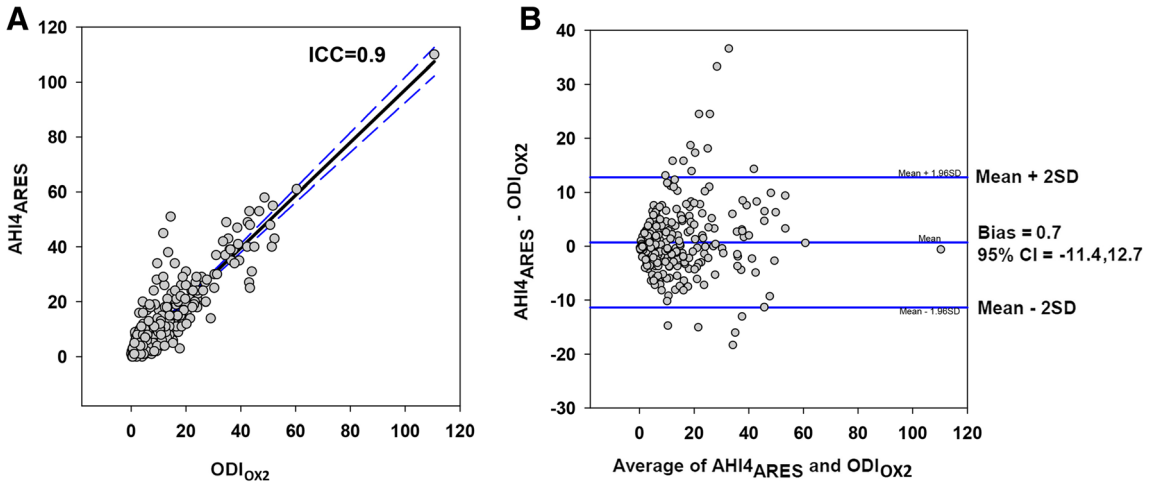

Fig. 4 **a** Scatterplot of ODI_{Ox2} and $AHI4_{ARES}$ for n = 310 nights, showing a good correlation (ICC = 0.9, 95% CI = 0.87–0.92, $p < 0.001$). **b** Bland-Altman plot showing bias and limits of agreement

Table 2 Agreement, sensitivity, and specificity when defining disease using cutoffs: (i) $AHI4_{ARES} \geq 5$/h and $ODI_{Ox2} \geq 5$, (ii) $RDI_{ARES} \geq 15$/h and $AHI4_{ARES} \geq 5$/h, and (iii) $RDI_{ARES} \geq 15$/h and $ODI_{Ox2} \geq 5$

Definition for OSA diagnosis	Agreement, kappa	Sensitivity	Specificity	False negative rate
		Gold standard ARES		
$ODI_{Ox2} \geq 5$ vs. $AHI4_{ARES} \geq 5$	82.5%, 0.62	85.9%	76.5%	14%
$AHI4_{ARES} \geq 5$ vs. $RDI_{ARES} \geq 15$	84.5%, 0.65	84.5%	84.6%	15.5%
$ODI_{Ox2} \geq 5$ vs. $RDI_{ARES} \geq 15$	77.4%, 0.49	79.0%	73.6%	21%

negatives identified for diagnosis of OSA were similar to those seen due to night-to-night variability in SDB.

We and others have compared limited channel monitoring devices (levels III and IV) and shown good sensitivity (80–89%) and specificity (86–94%) for diagnosis of OSA compared to full laboratory PSG [1, 2, 20–23]. A recent study compared a level III device (ApneaLink Plus, which uses transmittance finger pulse oximetry) and a stand-alone finger pulse oximeter (Pulsox 300i) used at home against an in-lab NPSG in a sleep clinic population with suspected OSA [24]. This report showed good agreement between indices of OSA that required ≥4% oxygen desaturation, calculated from either a single pulse oximetry or multichannel PSG recording. We have also previously found that significant differences occur in AHI4 obtained when different brands of finger pulse oximeters were used in the same patient [10]: the biases found in AHI4 in that study ranged from 0.3 ± 1.7 to 7.1 ± 9.6/h and were in the same order of magnitude for biases in AHI4/ODI (0.7 ± 6.1) seen in the present study between finger transmission and forehead reflectance oximetry. These data suggest that differences in measurement may be device dependent in addition to varying with the method of oximetry used [10].

In contrast to the results for the SDB index, we found significant and systematic device-dependent differences in the %time < 90%O2Sat. As neither device is the gold standard for oxygen saturation and we did not obtain a blood gas, it is impossible to know which of the two measures better reflected actual O2 saturation levels. We believe that a significant contributor to the systematic difference in oximetry observed in our study is the oximeter type. However, differences in averaging or sampling rates may have contributed [10]. It is also likely that calibration of the oximetry signal will account for differences in percent time below 90%; in support of this, our data showed differences in the simultaneously measured baseline awake saturation across the two devices. In addition, physiologic O2 saturation may be different at the forehead and finger as temperature and blood flow have been shown to influence O2 saturation and may differ [25]. The WristOx$_2$™ O2 saturation measurement at the beginning of the night was lower than the ARES device, consistent with the overall greater percent time < 90%O2Sat reported by the WristOx$_2$. In addition, the manual editing of the ARES data may have contributed by being more aggressive during periods of low O2 saturation/poor signal quality. This likely will result in removal of more periods of low O2 saturation ("poor signal") with the ARES algorithm than with the fully automated WristOx$_2$ algorithm. Irrespective of the cause, the significant observed differences in reported %time < 90%O2Sat suggest that care must be taken whenever comparing these data across studies using portable monitors with different oximeter types or algorithms.

Although data loss from oximetry was < 10% for both devices, the ARES™ device had significantly greater data loss than the WristOx$_2$. This may be due to a lower signal to noise ratio in forehead reflectance oximetry compared to transmission. Review of oximetry tracings suggests that the forehead signal is particularly susceptible to data loss during movement (as during position changes) possibly due to positional changes in venous blood flow to the forehead. On the other hand, our data showed that in the WristOx$_2$ studies, we had more nights with insufficient duration of data recorded for analysis (nights with < 2 h) than in the ARES™ studies. Possible reasons include the more precarious placement of the WristOx$_2$ on the fingertip, allowing for the device to fall off the patient more easily than

Table 3 Agreement, sensitivity, and specificity when defining disease using cutoffs: $AHI4_{ARES} \geq 5$/h, $RDI_{ARES} \geq 15$/h, and $ODI_{Ox2} \geq 5$ N1 vs. N2, using N1 as the gold standard

Definition for OSA diagnosis	Agreement, kappa	Sensitivity	Specificity	False negative rate
		Gold Standard N1		
$AHI4_{ARES} \geq 5$ N1 vs. N2	73.4%, 0.44	83%	60%	17%
$ODI_{Ox2} \geq 5$ N1 vs. N2	83.0%, 0.64	88.0%	74.5%	11.4%
$RDI_{ARES} \geq 15$ N1 vs. N2	81.5%, 0.57	92.8%	61%	7%

Table 4 Comparison of differences between devices and between nights using the same device

Measurement	Comparison	Number of comparisons	Bias	Abs diff mean ± SD or median (IQR)
$AHI4_{ARES}/ODI_{Ox2}$	ARES vs. $WristOx_2$	310	0.7 ± 6.1	2.5 (1.2, 5.2)
	N1 vs. N2 ARES	130	0.6 ± 9.9	3.0 (1.0, 6.0)
	N1 vs. N2 $WristOx_2$	130	0.7 ± 8.8	2.5 (1.1, 5.6)
%time < 90%O2Sat	ARES vs. $WristOx_2$	310	− 7.0 ± 15.4	1.9 (0.4, 7.2)
	N1 vs. N2 ARES	130	0.9 ± 5.1	0.6 (0.2, 2.9)
	N1 vs. N2 $WristOx_2$	130	0.7 ± 15.2	1.4 (0.3, 6.8)
Awake baseline 02Sat[a]	ARES vs. $WristOx_2$	295	4.0 ± 2.6	4.1 ± 2.4
	N1 vs. N2 ARES	130	− 0.1 ± 2.3	1.7 ± 1.6
	N1 vs. N2 $WristOx_2$	125	− 0.9 ± 8.8	1.5 ± 1.6

N1 night 1, *N2* night 2

[a] O2 saturation during a period in the first 10 min of data recording. Stable O2 saturation level could not be reliably obtained from the $WristOx_2$ at the beginning of the study on some nights

with the ARES™, which is secured around the forehead by an elastic strap and a nasal cannula.

The impact of the ability to score any SDB events *without* 4% O2 desaturation (e.g., those with arousal surrogates) using the ARES™ is reflected in the additional 20 (6.4%) diagnoses that were made when using RDI data from this device. This suggests that mild OSA captured by the RDI may be missed by the oximeter alone. Based on our prior work, we used a cutoff of ≥15/h for OSA diagnosis when events with O2 desaturation *and/or arousal* were included in the index (i.e., RDI) as opposed to the AASM guideline cutoff ≥ 5/h for diagnosis of OSA when using an SDB index that includes hypopneas defined by O2 desaturation alone [26]. We and others have demonstrated the validity of this higher cutoff for OSA whenever using the more inclusive definition of RDI [4, 23, 27]. If one is interested in more than just the overall SDB and oximetry indices, the ARES™ device provides additional information, such as sleep position and indirect confirmation of obstruction from snoring and from the shape of the airflow signal (inspiratory flow limitation), but this comes at a cost of additional time required for review and manual editing of data. A recent study by Chai-Coetzer et al. showed excellent agreement of indices for diagnosis for moderate to severe OSA of both level III and level IV devices compared to PSG. However, this study also showed slightly worse functional outcomes when level IV (oximetry alone) was used for OSA management, but similar outcomes between PSG and level III analysis [2]. The authors suggested that this was due to reduced physician confidence when using only a single channel, which is consistent with our previously published work [28].

Although our population composition was closer to an epidemiologic population (subjects were recruited without regard to symptoms of OSA and do not have significant comorbid cardiovascular and metabolic conditions), the majority of our subjects were overweight, middle-aged men. This likely contributed to the high prevalence of OSA. It remains to be tested whether our results can be generalized to a population with comorbid conditions showing significant hypoxemia.

Limitations of our study include the lack of a definitive reference measure of oximetry and not having comparison to in-laboratory PSG. Also, AHI4 by design includes apneas irrespective of any desaturation, and thus differs from the ODI, which only counts ≥4% desaturations. It has been estimated that at least 20% of apneas may not have accompanying desaturation [29]. This may explain why the $AHI4_{ARES}$ was slightly higher than the ODI_{Ox2} (i.e., due to inclusion of apnea events without desaturation).

Conclusion

In our study population, the level IV device ($WristOx_2$) showed good agreement for diagnosis of OSA compared to the level III device (ARES™) despite different approaches to signals monitored. The disagreement between the devices for the same index and between OSA diagnoses based on AHI4 ≥ 5/h and RDI ≥ 15/h with a single device was comparable to night-night variability in our study and reported in previous publications. This suggests that using oximetry alone with the ODI ≥ 5/h as a criterion to diagnose OSA may detect OSA with a clinically acceptable success in populations composed of middle-aged overweight males with few comorbidities. Whether the number of patients missed by ODI vs. AHI4 or RDI will increase above 10% in other populations, such as pediatrics or patients with minimally desaturating events, will need to be tested separately in future studies.

Acknowledgements This study was supported by NIOSH/CDC U01OH010415 and NIH K24 grant HL109156.

Funding This study was funded by NIOSH/CDC U01OH010415 and NIH K24 grant HL109156.

Compliance with ethical standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Conflict of interest TG, AT, SA, AP, and KB have no conflicts of interest to disclose.

DMR has received support for research from the industry in the past 24 months: grants from Fisher & Paykel Healthcare and speaking and consulting engagements for Fisher & Paykel Healthcare. DMR holds multiple US and foreign patents covering techniques and analysis algorithms for the diagnosis of OSA and techniques for administering CPAP. Several of these have been licensed to Biologics, Fisher & Paykel Healthcare, Advanced Brain Monitoring, and Sefam Medical.

JS has received support for speaker training from Merck Pharmaceuticals.

IA has received support for research from the industry in the past 24 months: grants from Fisher & Paykel Healthcare. IA holds multiple US and foreign patents covering techniques and analysis algorithms for the diagnosis of OSA and techniques for administering CPAP. Several of these have been licensed to Fisher & Paykel Healthcare and Advanced Brain Monitoring.

References

1. Hang LW, Wang HL, Chen JH, Hsu JC, Lin HH, Chung WS et al (2015) Validation of overnight oximetry to diagnose patients with moderate to severe obstructive sleep apnea. BMC Pulm Med 15:24
2. Chai-Coetzer CL, Antic NA, Hamilton GS, McArdle N, Wong K, Yee BJ et al (2017) Physician decision making and clinical outcomes with laboratory polysomnography or limited-channel sleep studies for obstructive sleep apnea: a randomized trial. Ann Intern Med 166(5):332–340
3. Collop NA, Anderson WM, Boehlecke B, Claman D, Goldberg R, Gottlieb DJ et al (2007) Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. J Clin Sleep Med 3(7):737–747
4. Ayappa I, Norman RG, Seelall V, Rapoport DM (2008) Validation of a self-applied unattended monitor for sleep disordered breathing. J Clin Sleep Med 4(1):26–37
5. Vazquez JC, Tsai WH, Flemons WW, Masuda A, Brant R, Hajduk E et al (2000) Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea. Thorax 55(4):302–307
6. Zou D, Grote L, Peker Y, Lindblad U, Hedner J (2006) Validation a portable monitoring device for sleep apnea diagnosis in a population based cohort using synchronized home polysomnography. Sleep 29(3):367–374
7. Westbrook PR, Levendowski DJ, Cvetinovic M, Zavora T, Velimirovic V, Henninger D et al (2005) Description and validation of the apnea risk evaluation system: a novel method to diagnose sleep apnea-hypopnea in the home. Chest 128(4): 2166–2175
8. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K et al (2017) Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine Clinical Practice Guideline. J Clin Sleep Med 13(3):479–504
9. Kisch-Wedel H, Bernreuter P, Kemming G, Albert M, Zwissler B (2009) Does the estimation of light attenuation in tissue increase the accuracy of reflectance pulse oximetry at low oxygen saturations in vivo? IEEE Trans Biomed Eng 56(9):2271–2279
10. Zafar S, Ayappa I, Norman RG, Krieger AC, Walsleben JA, Rapoport DM (2005) Choice of oximeter affects apnea-hypopnea index. Chest 127(1):80–88
11. Davila DG, Richards KC, Marshall BL, O'Sullivan PS, Osbahr LA, Huddleston RB et al (2003) Oximeter's acquisition parameter influences the profile of respiratory disturbances. Sleep 26(1):91–95
12. Cross TJ, Keller-Ross M, Issa A, Wentz R, Taylor B, Johnson B (2015) The impact of averaging window length on the "desaturation indexes during overnight pulse oximetry at high-altitude". Sleep 38(8):1331–1334
13. Stepnowsky C, Levendowski D, Popovic D, Ayappa I, Rapoport DM (2013) Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters. Sleep Med 14(11):1199–1207
14. Flemons WW, Littner MR (2003) Measuring agreement between diagnostic devices. Chest 124(4):1535–1542
15. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1(8476):307–310
16. Levendowski D, Steward D, Woodson BT, Olmstead R, Popovic D, Westbrook P (2009) The impact of obstructive sleep apnea variability measured in-lab versus in-home on sample size calculations. Int Arch Med 2(1):2
17. Newell J, Mairesse O, Verbanck P, Neu D (2012) Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples. Psychiatry Res 200(2–3):795–801
18. Le Bon O, Hoffmann G, Tecco J, Staner L, Noseda A, Pelc I et al (2000) Mild to moderate sleep respiratory events: one negative night may not be enough. Chest 118(2):353–359
19. Prasad B, Usmani S, Steffen AD, Van Dongen HP, Pack FM, Strakovsky I et al (2016) Short-term variability in apnea-hypopnea index during extended home portable monitoring. J Clin Sleep Med 12(6):855–863
20. Nigro CA, Aimaretti S, Gonzalez S, Rhodius E (2009) Validation of the WristOx 3100 oximeter for the diagnosis of sleep apnea/hypopnea syndrome. Sleep Breath 13(2):127–136
21. Romem A, Romem A, Koldobskiy D, Scharf SM (2014) Diagnosis of obstructive sleep apnea using pulse oximeter derived photoplethysmographic signals. J Clin Sleep Med 10(3):285–290
22. Chiner E, Signes-Costa J, Arriero JM, Marco J, Fuentes I, Sergado A (1999) Nocturnal oximetry for the diagnosis of the sleep apnoea hypopnoea syndrome: a method to reduce the number of polysomnographies? Thorax 54(11):968–971
23. Collop NA, Tracy SL, Kapur V, Mehra R, Kuhlmann D, Fleishman SA et al (2011) Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation. J Clin Sleep Med 7(5):531–548
24. Dawson A, Loving RT, Gordon RM, Abel SL, Loewy D, Kripke DF et al (2015) Type III home sleep testing versus pulse oximetry: is the respiratory disturbance index better than the oxygen

desaturation index to predict the apnoea-hypopnoea index measured during laboratory polysomnography? BMJ Open 5(6): e007956

25. Yamaura K, Nanishi N, Higashi M, Hoka S (2014) Effects of thermoregulatory vasoconstriction on pulse hemoglobin measurements using a co-oximeter in patients undergoing surgery. J Clin Anesth 26(8):643–647
26. The Report of an American Academy of Sleep Medicine Task Force (1999) Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. Sleep 22(5):667–689
27. Hosselet J, Ayappa I, Norman RG, Krieger AC, Rapoport DM (2001) Classification of sleep-disordered breathing. Am J Respir Crit Care Med 163(2):398–405
28. Masdeu MJ, Ayappa I, Hwang D, Mooney AM, Rapoport DM (2010) Impact of clinical assessment on use of data from unattended limited monitoring as opposed to full-in lab PSG in sleep disordered breathing. J Clin Sleep Med 6(1):51–58
29. Ayappa I, Rapaport BS, Norman RG, Rapoport DM (2005) Immediate consequences of respiratory events in sleep disordered breathing. Sleep Med 6(2):123–130