Abstract

Background: Obstructive sleep apnea (OSA) is the most prevalent respiratory sleep disorder occurring in 9% to 38% of the general population. About 90% of patients with suspected OSA remain undiagnosed due to the lack of sleep laboratories or specialists and the high cost of gold-standard in-lab polysomnography diagnosis, leading to a decreased quality of life and increased health care burden in cardio- and cerebrovascular diseases. Wearable sleep trackers like smartwatches and armbands are booming, creating a hope for cost-efficient at-home OSA diagnosis and assessment of treatment (eg, continuous positive airway pressure [CPAP] therapy) effectiveness. However, such wearables are currently still not available and cannot be used to detect sleep hypopnea. Sleep hypopnea is defined by ≥30% drop in breathing and an at least 3% drop in peripheral capillary oxygen saturation (SpO2) measured at the fingertip. Whether the conventional measures of oxygen desaturation (OD) at the fingertip and at the arm or wrist are identical is essentially unknown.

Objective: We aimed to compare event-by-event arm OD (arm_OD) with fingertip OD (finger_OD) in sleep hypopneas during both naïve sleep and CPAP therapy.

Methods: Thirty patients with OSA underwent an incremental, stepwise CPAP titration protocol during all-night in-lab video-polysomnography monitoring (ie, 1-h baseline sleep without CPAP followed by stepwise increments of 1 cmH2O pressure per hour starting from 5 to 8 cmH2O depending on the individual). Arm_OD of the left biceps muscle and finger_OD of the left index fingertip in sleep hypopneas were simultaneously measured by frequency-domain near-infrared spectroscopy and video-polysomnography photoplethysmography, respectively. Bland-Altman plots were used to illustrate the agreements between arm_OD and finger_OD during baseline sleep and under CPAP. We used t tests to determine whether these measurements significantly differed.

Results: In total, 534 obstructive apneas and 2185 hypopneas were recorded. Of the 2185 hypopneas, 668 (30.57%) were collected during baseline sleep and 1517 (69.43%), during CPAP sleep. The mean difference between finger_OD and arm_OD was 2.86% (95% CI 2.67%-3.06%, t667=28.28; P<.001; 95% limits of agreement [LoA] –2.27%, 6.00%) during baseline sleep and 1.83% (95% CI 1.72%-1.94%, t1516=31.99; P<.001; 95% LoA –2.54%, 6.19%) during CPAP. Using the standard criterion of 3% saturation drop, arm_OD only recognized 16.32% (109/668) and 14.90% (226/1517) of hypopneas at baseline and during CPAP, respectively.

Conclusions: arm_OD is 2% to 3% lower than standard finger_OD in sleep hypopnea, probably because the measured arm_OD originates physiologically from arterioles, venules, and capillaries; thus, the venous blood adversely affects its value. Our findings demonstrate that the standard criterion of ≥3% OD drop at the arm or wrist is not suitable to define hypopnea because it could provide large false-negative results in diagnosing OSA and assessing CPAP treatment effectiveness.

(J Med Internet Res 2021;23(7):e24171) doi: 10.2196/24171
KEYWORDS
obstructive sleep apnea; wearable devices; smartwatch; oxygen saturation; near-infrared spectroscopy; continuous positive airway pressure therapy; photoplethysmography

Introduction
Monitoring health using wearables, such as smartwatches and armbands, is becoming a new lifestyle [1-4]. Hundreds of millions of smartwatches and armbands are being used daily, and the number is still sharply increasing. Sleep monitoring is one of the most popular functions of such wearables [2,5-7], because sleep is a critical determinant of an individual’s health and well-being. Obstructive sleep apnea (OSA) is the most prevalent respiratory sleep disorder occurring in 9% to 38% of the general population [8], and it is a high-risk factor for cardiovascular and cerebrovascular diseases [9,10]. Nevertheless, about 90% of suspected patients with OSA remain undiagnosed [11] due to the lack of sleep laboratories or specialists and the high cost associated with an in-lab polysomnography (PSG) diagnosis (ie, the gold-standard for sleep disorder diagnosis [12]), thus leading to decreased quality of life and increased health care burden in the aging society. Relatively simple and less-expensive diagnostic methods such as portable home respiratory polygraphy have been developed [13,14]. However, its signal quality is compromised, with failure rates ranging from 3% to 18% [14], mostly due to incorrect handling of the device or sensors by the people performing the test (ie, the patients themselves) [15]. Home respiratory polygraphy also has the risk of false diagnosis because it cannot measure sleep/wake, so patients may be awake during the night. The cost of such diagnostic methods is still relatively high in many low- and middle-income countries, limiting their broad application in the general population. Therefore, using low-cost and easy-to-use wearable devices, such as smart finger rings, smartwatches, or armbands, for at-home diagnosis of OSA and to assess treatment effectiveness would substantially contribute to public health worldwide [6,16].

However, using the aforementioned low-cost wearable devices to evaluate OSA is still not clinically viable because, currently, no product has been licensed or certified as a medical diagnostic device by the United States Food and Drug Administration (FDA) or CE marking. We hypothesize that one of the major limitations in measuring OSA with wearables is the detection of sleep hypopnea. Currently, most consumer-grade wearables can measure surrogate markers for breathing and heartbeats or heart rate variability [3,17-20]. Although sleep apneas consisting of a complete pause in breathing for ≥10 seconds are relatively easy to assess by analyzing breathing frequency [6,12,21-23], it is challenging to detect sleep hypopnea, which is defined as a ≥30% drop in airflow lasting ≥10 seconds accompanied by either an arousal or a ≥3% drop in peripheral capillary oxygen saturation (SpO2) measured at the fingertip [12]. Smart rings that can measure SpO2 at the fingertip (eg, Sleepon [24]) are likely to accurately quantify the drops in SpO2 because, essentially, they are similar to fingertip pulse oximetry. However, these devices have not only a very tiny market share compared to other popular wearables such as smartwatches and armbands [25]. Whether the measures of oxygen desaturation (OD) at the fingertip and at the arm or wrist are physiologically identical in sleep hypopneas is essentially unknown. Some recent smartwatches (eg, Fitbit [26], Garmin [27], Huami [28], and Huawei [29]), armbands (eg, Humon [30,31], Moxy [32,33], PortaMon [34], and Biofourmis [35]), and prototypes [36,37] claim to measure SpO2 or muscle tissue oxygen saturation (StO2) at the arm or wrist. However, to the best of our knowledge, these devices have not been clinically validated for use in patients with OSA. We were able to find only one registered clinical validation study (Trial Registration: ClinicalTrials.gov NCT03775291) measuring OSA using a smartwatch, which was initiated by Fitbit in December 2018 [38]. However, the latest update of the study’s recruitment status as of November 2019 was still “active, not recruiting.” The study aimed to only compare PSG-assessed Apnea–Hypopnea Index (AHI) with the AHI derived from the smartwatch, rather than performing an event-by-event comparison of the apnea or hypopnea events measured by these two devices. Thus, even if we assume that Fitbit may have completed their validation work recently, their study still cannot answer the key question as to whether the hypopnea diagnostic standard criterion of ≥3% OD at the fingertip can be equally applied to the OD at the arm or wrist.

Therefore, we, for the first time, aimed to compare event-by-event OD at the fingertip (finger_OD) measured by the gold-standard in-lab PSG transmission photoplethysmography (T-PPG) with OD at the arm (arm_OD) measured by frequency-domain multidistance (FDMD) near-infrared spectroscopy (NIRS) in sleep hypopneas during naïve sleep and continuous positive airway pressure (CPAP) therapy. FDMD-NIRS is a well-validated [39-41] advanced, noninvasive optical technique that can quantify hemodynamic changes, including OD, in the measured tissues for long-term recordings such as all-night sleep with high temporal resolution [42,43]. Our results can conclusively demonstrate whether physiologically arm_OD measures can directly replace finger_OD to define sleep hypopnea. Thus, this study will have a broad appeal to the general public, sleep clinicians and scientists, health care insurance providers, and wearable technology developers who are aiming to measure OSA at-home by using wearable devices.

Methods
Patients
All patients underwent video-PSG measurement for diagnosis in our sleep laboratory. The following day, those patients who were diagnosed with OSA and recommended to use CPAP therapy by clinicians were recruited and gave their written informed consent for participation in the study. Patients with unstable coronary or cerebral artery disease, severe arterial hyper- or hypotension, respiratory diseases, or a history of a sleep-related accident were excluded. Finally, 30 newly diagnosed patients with OSA (mean age 54.2, SD 13.8 years, IQR 42-65 years; male: n=27; mean BMI 35.9, SD 7.5 kg/m²,
IQR 31.8–42.0 kg/m²: mean AHI 53.4, SD 24.7 per hour, IQR 32–71 per hour) participated in this study. This study was approved by the local ethical commission of Northwest Switzerland, and it was in compliance with the Declaration of Helsinki.

Protocol

The patients underwent incremental stepwise CPAP (AirSense 10, ResMed) titration combined with video-PSG and FDMD-NIRS recordings in one nocturnal sleep episode. This sleep episode consisted of 1 hour of baseline sleep without CPAP, followed by incremental stepwise titration of 1 cmH₂O pressure per hour starting from 5-8 cmH₂O depending on the individuals. The 1-h baseline sleep allowed us to compare arm_OD with finger_OD during normal sleep in patients with OSA. We included the CPAP titration protocol because (1) the stepwise CPAP titration protocol can increase the number of hypopneas for data analysis in our patients, since low pressures cannot fully open the airways to restore the apneas but instead cause hypopneas and (2) auto-CPAP that can automatically adjust the pressures within a given range (ie, automatic titration), and CPAP with fixed pressures are the most efficient and popular therapy for OSA [44]. Thus, the comparison between arm_OD and finger_OD in hypopneas under various CPAP pressures could allow us to test the feasibility of monitoring treatment efficacy by measuring arm_OD.

Video-PSG

Video-PSG (Embla RemLogic, Natus Medical Incorporated) is a comprehensive recording of physiological signals during sleep, including electroencephalography at electrode locations of C3, C4, O1, O2, F3, and F4 according to 10 to 20 system, eye movements (electrooculogram), muscle activation (electromyogram), electrocardiogram, breathing functions, heart rate, fingertip SpO₂ (left index fingertip in this study), and movement during sleep. Two experienced sleep technologists independently scored the sleep stages, respiratory and limb movement events, and motion artifacts in 30-second epochs according to the 2017 American Academy of Sleep Medicine manual [12]. Sleep hypopneas were defined as an at least 30% drop in airflow lasting for at least 10 seconds with either an arousal or >3% drop in SpO₂. The discrepancy between these two technologists was resolved by discussion or recommendation by an experienced neurophysiologist. The hypopneas were excluded from analysis if their SpO₂ desaturations were larger than 15% (n=31) to exclude outliers and potentially unreliable measurements caused by instrument errors.

FDMD-NIRS Measurements

In this study, FDMD-NIRS (Imagent, ISS) measurements were conducted over the middle of the left biceps muscle. Imagent is currently the only commercial benchtop FDMD-NIRS device [40,42,45] and has been CE-approved for research. Its light emitters, four laser diodes at 690-nm wavelength and four laser diodes at 830-nm wavelength, are coupled into four light sources and are high frequency modulated at 110 MHz. The light can penetrate the measured tissues with a depth of several centimeters when the four light sources are aligned and placed at 2, 2.5, 3, and 3.5 cm away from an optical fiber bundle connected to the photomultiplier tube detector.

The most common commercially available NIRS devices, including wearable NIRS devices (eg, Humon [30,31], Moxy [32,33], and PortaMon [34]), are continuous-wave NIRS (CW-NIRS), which measure the hemodynamic changes in human tissues based on the modified Beer–Lambert law (MBLL) [46-48]. As illustrated in Figure 1A, in the original Beer–Lambert law, the light extinction is proportional to the concentration C multiplied by the constant extinction coefficient ε for the particular absorber and the length d of the absorbing medium when light passes through a nonscattering but absorbing media [49]. C × ε is also called the absorption coefficient µ of the absorbing media. However, biological tissues are highly scattering media, and scattering will increase the path-length of light, thus increasing the probability of both light absorption and loss of light. Scattering also makes it possible that some light can go out of the tissue from the same side of the light source (ie, backscattering light), as shown in Figure 1B. Thus, reflectance photoplethysmography (R-PPG) pulse oximeter [50-52] and the NIRS device in which the light source and detector are placed on the same side of the measured tissues were developed based on MBLL. In MBLL, the light propagation due to scattering is taken into account by introducing the differential path-length factor (DFP). The real path-length of light in the tissue is then calculated as DPF multiplied by the source-detector distance r. DPF varies between different biological tissues and different individuals, and it also depends on other factors such as light wavelength and age and gender of the individual [48,53]. CW-NIRS devices use fixed value of DPF in the range of 3 to 6 for different light wavelengths [53]. They can only estimate the relative changes in the main absorbing chromophores, that is, oxygenated hemoglobin (HbO₂) and deoxygenated hemoglobin (HHb) in the measured biological tissues. R-PPG and T-PPG pulse oximeters usually use two near-infrared wavelengths that are mainly absorbed by HbO₂ and HHb respectively. The influences of scattering on light attenuation are approximately assumed to be cancelled out in the calculation of SpO₂ that is derived from the ratio of the light intensities of the two wavelengths, ie, assuming the path-lengths of the two wavelengths are identical in the tissues [50-52]. However, this assumption is actually not valid because DPF varies between different wavelengths. Experimental calibrations against in vitro measurement of arterial oxygen saturation (SaO₂) in extracted arterial blood (ie, invasive co-oximetry) thus must be performed in commercial pulse oximeters during their research and development period to correct measurement errors [50-52]. Several calibration-free methods were also proposed, including calculating the real light path-length and absorption by using the frequency-domain NIRS algorithm [52,54-57].
Figure 1. Beer–Lambert law and frequency-domain multidistance near-infrared spectroscopy (FDMD-NIRS) measurement. (A) The original Beer–Lambert law describing the light propagation in a nonscattering absorbing media. The attenuation of light intensity in this absorbing media is proportional to the concentration C multiplied by the constant extinction coefficient ε for the particular absorber and the length d of the absorbing media. (B) The travelling of light in biological tissue (i.e., highly scattering media). Light travels a longer pathway in the tissue than the light source-detector distance r due to scattering. (C) Basic principle of FDMD-NIRS measurement. The blue sine wave represents the high-frequency modulated light source. I_{DC0} and I_{AC0} are its light intensity and modulation amplitude, respectively. The two black sine waves are the output light detected after passing the measured tissues. They are detected by detectors 1 and 2 placed at different distances away from the light source. The light intensities and modulation amplitudes of the two black sine waves are smaller than those of the light source, and their phases are delayed because of the absorption and scattering in the tissues. Detector 1 is closer to the light source than detector 2. Therefore, the light intensity I_{DC1} and modulation amplitude I_{AC1} detected at detector 1 are larger than the light intensity I_{DC2} and modulation amplitude I_{AC2} detected at detector 2. The phase delay φ1 at detector 1 is smaller than the phase delay φ2 at detector 2 because the light reaching detector 2 has travelled a longer distance in the tissue. Similarly, the light intensity and modulation amplitude will be further decreased, and the phase delay will be further increased, when the light reaches the other detectors placed farther away than detector 2.

As illustrated in Figure 1C, in FDMD-NIRS, the light emitted from light source can be detected by detectors placed at different distances away from the source. The light intensity (I_{DC}) and modulation amplitude (I_{AC}) of the detected light decrease, and a phase delay (φ) occurs between the detected light and the source light due to absorption and scattering. The detected I_{DC} and I_{AC} are smaller but the φ is larger at the detector further away from the light source. The I_{DC}, I_{AC} and φ from different light source-detector distances vary linearly [42,45]. Therefore, to submit the measured I_{DC}, I_{AC} and φ to linear regression, we can obtain the following equations derived from the photon diffusion equation in a semi-infinite geometry [40,45,58-60]:

\[ \ln (r^2 I_{AC}) = r S_{AC} + C_{AC} \]
\[ \ln (r^2_{DC}) = rS_{DC} + C_{DC} \]  
\[ \phi = rS_{\phi} + C_{\phi} \]

Where \( r \) is the known source-detector distance, \( S_{AC}, S_{DC} \) and \( S_{\phi} \) are the slopes and \( C_{AC}, C_{DC} \) and \( C_{\phi} \) are the intercepts. The linearity of the equation is monitored by the \( R^2 \) of the fitted linear regression. By combing any two of these three slopes (eg, we chose \( S_{AC} \) and \( S_{\phi} \) in the following equations), we can calculate the absorption coefficient \( \mu_a \) and the reduced scattering coefficient \( \mu_s' \) of the measured tissue [40,45,58-60]:

\[ \mu_a = \frac{\omega}{2\pi} \left( S_{AC} - S_{DC} / S_{\phi} \right) \]  
\[ \mu_s' = \frac{S_{AC} - S_{\phi}}{3\mu_a - \mu_s} \]

where \( \omega/2\pi \) is the modulation frequency and \( v \) is the velocity of light in the tissue. FDMD-NIRS uses two wavelengths. The \( \mu_a \) and \( \mu_s' \) of both these wavelengths can be calculated individually by using the same equations (4) and (5). Equation (4) gives us the absorption coefficient of the measured tissues calculated by taking the influence of scattering into account. It is equal to \( C \times \varepsilon \) as mentioned above. In NIRS, the main contributions to absorptions in tissues are HbO\(_2\) and HHb, so we have the following equation:

\[ \mu_a^\lambda = \varepsilon_{HbO_2}^\lambda C_{HbO_2} + \varepsilon_{HHb}^\lambda C_{HHb} \]

where \( \mu_a^\lambda \) is the absorption coefficient of the measured tissue at wavelength \( \lambda \). \( \varepsilon_{HbO_2}^\lambda \) and \( \varepsilon_{HHb}^\lambda \) are the known extinction coefficients at wavelength \( \lambda \) for HbO\(_2\) and HHb, respectively. \( C_{HbO_2} \) and \( C_{HHb} \) are the concentrations of HbO\(_2\) and HHb, respectively. Using two wavelengths \( \lambda_1 \) and \( \lambda_2 \), we can then calculate \( C_{HbO_2} \) and \( C_{HHb} \) with the following equations:

\[ C_{HbO_2} = \left( \mu_a^{\lambda_1} \varepsilon_{HbO_2}^{\lambda_1} - \mu_a^{\lambda_2} \varepsilon_{HbO_2}^{\lambda_2} \right) / \left( \varepsilon_{HbO_2}^{\lambda_1} \varepsilon_{HbO_2}^{\lambda_2} \right) \]
\[ C_{HHb} = \left( \mu_a^{\lambda_2} \varepsilon_{HHb}^{\lambda_2} - \mu_a^{\lambda_1} \varepsilon_{HHb}^{\lambda_1} \right) / \left( \varepsilon_{HHb}^{\lambda_1} \varepsilon_{HHb}^{\lambda_2} \right) \]

Therefore, \( StO_2 \) can be further derived as:

\[ StO_2 = 100 \times C_{HbO_2} / \left( C_{HbO_2} + C_{HHb} \right) \]

In this study, the sampling rate of FDMD-NIRS recording was 5.2 Hz. The reliability and accuracy of FDMD-NIRS measurements depend on the linearity of the measured optical signals on distances, because \( \mu_a \) and \( \mu_s' \) are derived from the slopes of equations (1-3). The linear dependence \( R^2 \) of the modulated light amplitude and phase shift over the measured distances should be highly close to 1 in each light wavelength. Thus, before the start of the recording, our Imagent system was calibrated on an optical phantom block with known \( \mu_a \) and \( \mu_s' \), that is, the light intensity of each light source was adjusted so that \( R^2 \) was equal to 1 and the measured \( \mu_a \) and \( \mu_s' \) of the optical phantom block were equal to their known values. This calibration step can exclude the uncertainty of our measurements due to machine errors such as light source and detector errors.

The raw optical data were discarded if the \( R^2 \) was smaller than 0.95 in either modulated light amplitude or phase shift in any wavelength to exclude poor-quality data arising from improper probe-skin contact and shunted light reaching the detector without travelling through the tissue [68,69]. The NIRS data were then subjected to a low-pass (<0.08 Hz), zero-phase filter designed using Hanning window to remove the physiological noises, including heart rate, respiratory noise, and spontaneous slow hemodynamic oscillations [70,71]. The filtered data were smoothed with moving average smooth method (robust locally weighted scatter plot smoothing [70,72]).

**Statistical Analysis**

Bland-Altman plots were used to illustrate the agreements between arm_OD and finger_OD during baseline sleep and under CPAP, respectively. We used \( t \) test to determine whether the differences between arm_OD and finger_OD were significantly different from 0. In order to check whether arm_OD could replace finger_OD in subgroups of hypopneas (ie, hypopneas with different degrees of desaturations), we used Spearman correlation to evaluate the relationship between arm_OD and finger_OD in different subgroups defined by finger_OD greater than or equal to specific cut-offs. The cut-offs ranged from 2% to 8%. \( P \) values <.05 indicated statistical significance for both analyses. Pre-processing of FDMD-NIRS signals was carried out in MATLAB (MathWorks, Inc.). All statistical analyses were performed using R statistical software (version 3.2.4; R Foundation for Statistical Computing).

**Results**

In total, 2185 hypopneas (median 67, IQR 41-82) were analyzed, including 668 (median 16, IQR 11-35) recorded during baseline sleep and 1517 (median 41, IQR 25-64) recorded during CPAP. Figure 2 illustrates the typical changes that occurred in arm_OD and finger_OD in hypopnea events. Indeed, sleep hypopneas cause OD in arm StO2. The absolute (mean) values of StO2 are 68.8% (SD 6.5%) at baseline before the start of hypopneas. The distributions of finger_OD and arm_OD (Figure 3) suggest larger OD in the finger than in the arm both at the baseline and during CPAP, as the peaks of the distributions of arm_OD are smaller than those of finger_OD.
Figure 2. Typical oxygen desaturation (OD) at fingertip (finger_OD) and at arm (arm_OD) during hypopneas. Arrows indicate the degree of OD. SpO$_2$ is measured at the fingertip by polysomnography transmission photoplethysmography, and StO$_2$ is measured at the biceps muscle by frequency-domain multidistance near-infrared spectroscopy. SpO$_2$: peripheral capillary oxygen saturation; StO$_2$: peripheral tissue oxygen saturation.
**Figure 3.** Distributions of oxygen desaturation at fingertip (finger\_OD) and at arm (arm\_OD) at (A) baseline (n=668) and (B) under continuous positive airway pressure (CPAP: n=1517).

Bland-Altman plots (Figure 4) show that the mean difference between finger\_OD and arm\_OD is 2.86% (95% CI 2.67%-3.06%, t\textsubscript{667}=28.28; \(P<.001\)) during baseline sleep and 1.83% (95% CI 1.72%-1.94%, t\textsubscript{1516}=31.99; \(P<.001\)) under continuous positive airway pressure (CPAP) sleep, with broad 95% limits of agreement (LoA) as [-2.27%, 8.00%] and [-2.54%, 6.19%], respectively. Using the criterion of arm\_OD \(\geq 3\%\), we can only define 16.32% (109/668) and 14.90% (226/1517) of hypopneas at baseline and during CPAP sleep, respectively.
Figure 4. Bland-Altman plots of oxygen desaturation at fingertip (finger_OD) and at arm (arm_OD) at (A) baseline (n=668) and (B) under continuous positive airway pressure (CPAP: n=1517). The x-axes show the mean between the two measures, whereas the y-axes represent the differences (ie, finger_OD – arm_OD). The horizontal dotted lines indicate the mean difference and the 95% limits of agreement between the measures, ie, mean difference ± 1.96 × SD. The distribution of the mean difference is shown at the right margin of the plot, which is a normal distribution.

To test whether arm_OD could replace finger_OD in subgroups of hypopneas such as those causing severe OD, we correlate arm_OD and finger_OD in different subgroups with finger_OD ≥ specific cut-offs (from 2% to 8%). Again, we could only observe weak correlations (ie, correlation coefficients < 0.4) between them at baseline and under CPAP (Figure 5).
Since the baseline values of StO$_2$ (mean 68.8% SD 6.5%) are obviously much smaller than those of SpO$_2$ (usually above 90%), it is possible that the relative change rather than the absolute (raw) value of arm_OD has better agreement with the finger_OD. We therefore normalize the arm_OD to its baseline. The mean difference between finger_OD and normalized arm_OD is 1.68% (95% CI 1.46-1.90%, $t_{667}=14.99$; $P<.001$) during baseline sleep and 0.88% (95% CI: 0.75-1.01%, $t_{1516}=12.91$; $P<.001$) during CPAP sleep, respectively. The mean differences are smaller than those between absolute arm_OD and finger_OD. However, the 95% LoA are still equally broad, which are [-3.99%, 7.34%] during baseline sleep and [-4.32%, 6.08%] during CPAP sleep respectively, suggesting poor agreement. Only 41.6% (278/668) of hypopneas at baseline and 31.7% (481/1517) during CPAP sleep can be defined using the criterion of normalized arm_OD $\geq$ 3%, suggesting that the normalization still has poor sensitivity in detecting hypopneas.

**Discussion**

**Principal Findings**

In this study, we compare, for the first time, event-by-event finger_OD to arm_OD in sleep hypopneas during naive sleep and CPAP therapy sleep. We choose the gold-standard reference methods for both arm_OD and finger_OD—advanced FDMD-NIRS for arm_OD and PSG fingertip transmission photoplethysmography for finger_OD. Arm_OD is 2% to 3% smaller than finger_OD, probably because finger_OD is caused only by arterioles, whereas arm_OD is physiologically determined by mixed sources of arterioles, venules, and capillaries [42,73,74]. The lower value of arm_OD is thus most likely due to the contribution of the venous blood pool. The significant difference between finger_OD and arm_OD and their broad LoA (Figure 4) suggest that arm_OD cannot directly replace finger_OD to define hypopneas. If the standard criterion of $\geq$ 3% drop is applied, it will cause a high rate of false-negative results in diagnosing OSA and in assessing the efficacy of CPAP treatment. The poor agreement between finger_OD and arm_OD and the low sensitivity of arm_OD in measuring hypopneas cannot essentially be improved even after normalization, in which raw arm_OD is normalized to its baseline before the start of hypopnea. The poor correlations (Figure 5) between finger_OD and arm_OD across hypopnea severity further suggest that the arm_OD cannot be used to define hypopneas, that is, correlations remain low even in severe hypopneas that are associated with much larger finger_ODs.

The armband and smartwatch are two of the most popular wearable technologies. Our results have direct implications for wearable armbands and portable oximeters using NIRS techniques to measure OSA, as we discuss below.

First, the FDMD-NIRS system used in our study is well recognized as the most robust and reliable reference NIRS technique [31,42,61]. Arm StO$_2$ measured by our device should be identical to that measured by consumer-grade wearable NIRS armbands (eg, Humon [30,31], Moxy [32,33], and PortaMon [34]). It should also be equal to that measured by FDA-cleared medical-grade portable NIRS oximeters, such as INVOS 5100C (Medtronic) [66], FORE-SIGHT (CAS Medical Systems) [67], SenSmart (Nonin Medical Inc) [75].
(Hutchinson Technology Inc) [76], and ViOptix ODISsey (ViOptix Inc) [77]. Our results suggest that the absolute or relative (ie, normalized to baseline) OD at arm muscle measured by these aforementioned devices cannot be directly used to define sleep hypopnea.

Second, the NIRS StO2 is the proportion of HbO2 in arterial, capillary, and venous compartments of the measured tissue. It can be expressed by its two major compounds as StO2 = \( a \times \text{SaO}_2 + (b \times \text{SvO}_2) \), where SaO2 and SvO2 are the arterial and venous oxygen saturation [68,73]. Fingertip SpO2 is the best noninvasive estimate of SaO2 [50,52]. The ratio of the coefficients \( a/b \) is called the arterio-venous ratio (AVR). The AVR of different tissues (eg, brain and muscles) and different populations (eg, healthy people and patients, adults, and children) has been determined by invasive measurements under different conditions but not during sleep [73,78-88]. Commercially available NIRS oximeters, including the aforementioned FDA-certificated medical devices take the fixed AVR value as either 0.25/0.75 or 0.30/0.70 but have not validated it in OSA [78-88]. Our results of poor correlations between arm_OD and finger_OD (Figure 5) indicate that such an uncritical AVR adoption is not justified. According to the mathematical model of fixed relationship between StO2 and SaO2, we expect a strong correlation between StO2 and SpO2 that we could not confirm using our data. These conflicting results suggest that the model is not valid for OSA. Therefore, developers of wearable or portable NIRS devices should first study the AVR in muscle tissues and validate it against invasive blood sample measurements before using their devices commercially for OSA diagnosis.

Our findings also indicate that appropriate arterial oxygen saturation measurement is not possible with armbands using the R-PPG method. As mentioned in the Methods section, R-PPG essentially has the same theoretical limitations as CW-NIRS, such that the scattering of light in the human tissues cannot be calculated. R-PPG and T-PPG pulse oximeters estimate SpO2 under the assumption that changes in blood volume only occur in arterial but not in the venous compartment. SpO2 is measured from the fingertip or the superficial forehead because these locations are well perfused by arteries [36,52,89]. It is not recommended to measure SpO2 by using R-PPG at the wrist or the arm, because of its low signal-to-noise ratio. Compared to the more precise fingertip T-PPG method, the low signal-to-noise ratio of R-PPG at the wrist or arm is about 10 times weaker due to various factors such as relative low blood perfusion and sensitivity to pressure and ambient light sources [51,89]. Additionally, the key assumption of constant venous blood volume is no longer valid at the wrist or arm [89,90]. Although recently, some smartwatches [26-29] and a few armbands [35-37] claimed that they can measure SpO2 at the arm or wrist by using R-PPG, they actually measure both arterial and venous blood [89,90] similar to NIRS. A main difference between NIRS and R-PPG is the measurement depth. Whether R-PPG measures the blood only in the skin or in both the skin and muscle depends on the distance between the light source and detector [91]. The detector can detect the light passing through deeper tissues at a larger separation distance. In vivo studies suggest that R-PPG can obtain its best signal-to-noise ratio at a separation distance of 3 to 6 mm [92]. Thus, a separation distance of several millimeters is used in the design of R-PPG pulse oximetry [36]. A recent study modeled the R-PPG light propagation in human skin [37]. The authors found that even at a separation distance of only 0.6 mm, many light rays reaching deeper into the muscle can still be received by the detector because of the random nature of light scattering [37]. Therefore, similar to NIRS armbands, armband devices using R-PPG capture the oxygen saturation in both skin and muscle. There are challenges in using both techniques to differentiate SaO2 and SvO2 desaturations from the measured StO2 desaturation to define sleep hypopneas.

We did not include a smartwatch in this study, although recently, leading smartwatch companies like Apple, Fitbit, Garmin, Huami, and Huawei have all added the function of measuring oxygen saturation in their products. This is because these commercially available products cannot or are unwilling to export their raw data for analyses, or their temporal resolutions are simply too low for an event-by-event comparison necessary for our study because usually these consumer-grade smartwatches upload their data to their cloud servers in minute resolution. Nevertheless, our results may have indirect implications for smartwatches. R-PPG smartwatch is different from the fingertip T-PPG in measuring SpO2. Lee et al [89] found that the raw light signals measured by wrist R-PPG and fingertip T-PPG change differently during breath-holding, indicating different SpO2 values are calculated by these two techniques. Abay et al [36] reported that wrist R-PPG results in lower SpO2 values than fingertip T-PPG at rest, and during venous occlusion, fingertip T-PPG SpO2 does not exhibit desaturation but wrist R-PPG SpO2 drops similarly as the simultaneously measured NIRS StO2 at the same arm. Their findings also indicate that although smartwatches measure the wrist and NIRS measures the arm muscle, the measured changes in oxygen saturation by these two techniques are likely to be the same. Thus, the venous blood influences in our NIRS StO2 measurements are also likely to be observed in smartwatch measurements.

Conclusions

Our study warns consumers, health care insurance companies, and sleep clinicians and scientists to interpret the AHI provided by smartwatches and armbands with caution until those products are clinically and experimentally validated. An AHI >5/hour suggests the diagnosis of OSA [12]. Our results suggest that AHI is likely to be underestimated if using the criterion of arm_OD ≥3% to define hypopneas. Wearable technology developers who are validating their products can learn from this study and take into account the mismatch between the ODs measured by their products and by the gold-standard technique fingertip pulse oximetry. Developing new parameters (eg, estimated-oxygen-variation provided by Fitbit smartwatches [26]) or combining smartwatches with external fingertip T-PPG sensors [93] may be a more promising strategy to measure OSA. Nevertheless, validations of these new approaches are necessary before releasing them for clinical use. In addition, our finding
of a weak correlation between finger_OD and arm_OD indicates that (1) prediction of finger_OD using arm_OD may be possible but will require development and implementation of sophisticated data-mining, such as machine learning algorithms [90], and (2) as previous studies have validated NIRS oximeters as medical devices, protocols that quantify the arterial and venous contributions to the arm_OD are needed. Arm_OD can then be calibrated to indicate the changes in the arterial ODs at the arm or wrist.

Acknowledgments

This work was supported by Clinic Barmelweid Scientific Foundation. The data acquisition work was supported by the Research Fund of the Swiss Lung Association No. 2014-22.

Conflicts of Interest

None declared.

References


24. GoZleep SE. URL: https://www.sleepon.us/go2sleepse/ [accessed 2020-08-07]


27. Top FAQs for the Pulse Ox Feature on Garmin Watches. URL: https://support.garmin.com/en-US/FAQ/aq=SK2Y99aBp5D0n4sXnFBG7?f [accessed 2020-08-07]


34. PortaMon. URL: https://www.artinis.com/portamon [accessed 2020-08-07]


75. REGIONAL OXIMETER. U.S. Food and Drug Administration (FDA) 510(k) Database. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpdmn/pmn.cfm?ID=K102715

76. INSPECTRA TISSUE SPECTROMETER SYSTEM (INSPECTRA), MODEL 325. U.S. Food and Drug Administration (FDA) 510(k) Database. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpdmn/pmn.cfm?ID=K1012759

77. VIOPTIX ODISSEY TISSUE OXIMETER TISSUE OXIMETER, MODEL OXY-2. U.S. Food and Drug Administration (FDA) 510(k) Database. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpdmn/pmn.cfm?ID=K042657


Abbreviations

- AHI: Apnea–Hypopnea Index
- arm_OD: arm oxygen desaturation
- AVR: arterio-venous ratio
- CPAP: continuous positive airway pressure
- CW-NIRS: continuous wave near-infrared spectroscopy
- DPF: differential path-length factor
- FDMD: frequency-domain multidistance
- finger_OD: fingertip oxygen desaturation
- HbO₂: oxygenated hemoglobin
- HHb: deoxygenated hemoglobin
- LoA: limit of agreement
- MBLL: modified Beer–Lambert law
- NIRS: near-infrared spectroscopy
- OD: oxygen desaturation
- OSA: obstructive sleep apnea
- PSG: polysomnography
- R-PPG: reflectance photoplethysmography
- SaO₂: arterial oxygen saturation
- SNR: signal-to-noise ratio
- SpO₂: peripheral capillary oxygen saturation
- StO₂: peripheral tissue oxygen saturation
- SvO₂: venous oxygen saturation
- T-PPG: transmission photoplethysmography

©Zhongxing Zhang, Ming Qi, Gordana Hügli, Ramin Khatami. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any
medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.