Sleep Medicine 60 (2019) 45–52

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

Pediatric pulse oximetry-based OSA screening at different thresholds of the apnea-hypopnea index with an expression of uncertainty for inconclusive classifications

Ainara Garde [a, b, *], Xenia Hoppenbrouwer [a], Parastoo Dehkordi [b], Guohai Zhou [c], Aryannah Umedaly Rollinson [e], David Wensley [d], Guy A. Dumont [b], J. Mark Ansermino [e]

[a] Biomedical Signals and Systems Group, Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente, Enschede, the Netherlands
[b] The Department of Electrical & Computer Engineering, The University of British Columbia, Vancouver, British Columbia, Canada
[c] Center for Outcomes Research & Evaluation, School of Medicine, Yale University, New Haven, United States
[d] Division of Critical Care, The University of British Columbia and BC Children's Hospital, Vancouver, British Columbia, Canada
[e] The Department of Anesthesiology, Pharmacology & Therapeutics, The University of British Columbia, Vancouver, British Columbia, Canada

ARTICLE INFO

Article history:
Received 29 June 2018
Received in revised form
27 August 2018
Accepted 29 August 2018
Available online 22 September 2018

Keywords:
Pediatric sleep apnea
Pulse oximetry
Signal analysis
Pulse rate variability
Oxygen saturation dynamics
Mobile health solutions

ABSTRACT

Background: Assessments of pediatric obstructive sleep apnea (OSA) are underutilized across Canada due to a lack of resources. Polysomnography (PSG) measures OSA severity through the average number of apnea/hypopnea events per hour (AHI), but is resource intensive and requires a specialized sleep laboratory, which results in long waitlists and delays in OSA detection. Prompt diagnosis and treatment of OSA are crucial for children, as untreated OSA is linked to behavioral deficits, growth failure, and negative cardiovascular consequences. We aim to assess the performance of a portable pediatric OSA screening tool at different AHI cut-offs using overnight smartphone-based pulse oximetry.

Material and methods: Following ethics approval and informed consent, children referred to British Columbia Children's Hospital for overnight PSG were recruited for two studies including 160 and 75 children, respectively. An additional smartphone-based pulse oximeter sensor was used in both studies to record overnight pulse oximetry [SpO_2 and photoplethysmogram (PPG)] alongside the PSG.

Features characterizing SpO_2 dynamics and heart rate variability from pulse peak intervals of the PPG signal were derived from pulse oximetry recordings. Three multivariate logistic regression screening models, targeted at three different levels of OSA severity (AHI $\geq$ 1, 5, and 10), were developed using stepwise-selection of features using the Bayesian information criterion (BIC). The "Gray Zone" approach was also implemented for different tolerance values to allow for more precise detection of children with inconclusive classification results.

Results: The optimal diagnostic tolerance values defining the "Gray Zone" borders (15, 10, and 5, respectively) were selected to develop the final models to screen for children at AHI cut-offs of 1, 5, and 10. The final models evaluated through cross-validation showed good accuracy (75%, 82% and 89%), sensitivity (80%, 85% and 82%) and specificity (65%, 79% and 91%) values for detecting children with AHI $\geq$ 1, AHI $\geq$ 5 and AHI $\geq$ 10. The percentage of children classified as inconclusive was 28%, 38% and 16% for models detecting AHI $\geq$ 1, AHI $\geq$ 5, and AHI $\geq$ 10, respectively.

Conclusions: The proposed pulse oximetry-based OSA screening tool at different AHI cut-offs may assist clinicians in identifying children at different OSA severity levels. Using this tool at home prior to PSG can help with optimizing the limited resources for PSG screening. Further validation with larger and more heterogeneous datasets is required before introducing in clinical practice.

* Corresponding author. Biomedical Signals and Systems Group, Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente, Enschede, the Netherlands.
E-mail address: a.gardemartinez@utwente.nl (A. Garde).

https://doi.org/10.1016/j.sleep.2018.08.027
1389-9457/

1. Introduction

Obstructive sleep apnea (OSA), the most common form of sleep-disordered breathing, affects 2–6% of children [1,2]. The primary cause for OSA in children is enlarged tonsils and adenoids, which contribute to airway obstruction. The highest OSA prevalence is in children aged 2–5 years when adenoids and tonsils are largest relative to the airway [3]. Untreated OSA in children has been linked to cognitive and behavioral deficits, growth disorders, metabolic disorders, systemic inflammation, and serious cardiovascular consequences [4,5]. Thus, prompt diagnosis and treatment of OSA are vital for the healthy growth and development of many children.

Polysomnography (PSG), the gold standard for OSA diagnosis, calculates the average number of apneas/hypopneas per hour (AHI), which provides information about OSA severity. However, PSG is highly resource-intensive involving an overnight stay in a specialized sleep laboratory [6]. Moreover, there is a lack of resources for pediatric OSA evaluation, with pronounced geographical disparities, because specialized sleep laboratories capable of performing pediatric PSGs are available only in large urban centers [7]. In Canada in particular, the demand for polysomnography is 10-fold the capacity [7]. For instance, in the province of British Columbia (BC), all pediatric PSGs are performed at BC Children's Hospital (BCCH), where less than 250 can be done annually. As a result, there are long waitlists and delays in the testing and diagnosis of children in need.

These limitations have led to an extensive search for simpler OSA diagnostic and management tools for children. Overnight pulse oximetry has been widely studied as a potential standalone method to screen for pediatric OSA [8–12]. Pulse oximetry, an essential component of the PSG, is a simple, non-invasive means of measuring blood oxygen saturation (SpO_2) and the photoplethysmogram (PPG), a waveform that indicates blood volume changes in tissue. In general, most of the studies have focused on SpO_2 analysis alone [8,9,11,12]; however, there are OSA events that occur in the absence of significant SpO_2 desaturation [13], which will be missed by this analysis. In our previous study, we demonstrated that combining SpO_2 with pulse rate variability (PRV) – a surrogate measure of heart rate variability (HRV) – improved OSA screening performance beyond the results achieved using SpO_2 analysis alone, considering AHI $\geq$ 5 as OSA positive [10]. Additionally, pulse oximetry can be interfaced with a smartphone, increasing the portability of the device [14,15]. For example, the Phone Oximeter is a mobile device that integrates a commercially available and Federal Drug Administration (FDA) approved microcontroller-based pulse oximeter (Masimo Set iSpO_2 Pulse Oximetry Cable) with a mobile smartphone. The Phone Oximeter facilitates the acquisition, monitoring, and analysis of pulse oximetry both at-home and in hospital.

The majority of studies suggest that the threshold which indicates a need for pediatric OSA treatment (eg, adenotonsillectomy), is a PSG-determined AHI $\geq$ 5 [5,16]. However, current guidelines suggest an AHI threshold of 1, 5 and 10 to define mild, moderate and severe OSA in children. Therefore, in this study our goal is to extensively assess the performance of the smartphone-based pulse oximetry to identify children with OSA at varying AHI cut-offs, combining the characterization of SpO_2 dynamics and PRV analysis. However, a single numerical point index should not be an appropriate division between levels of OSA severity [5]. Moreover, the binary constraint of the ROC curve approach may not fit the reality of clinical or screening practice. Thus, we propose to apply the “Gray Zone” statistical approach as a novel alternative to allow for uncertainty when screening for pediatric OSA. This technique identifies cases corresponding to predictions that are not precise enough for a diagnostic decision [17]. Thus, we aim to develop and assess the performance of three multivariate logistic regression models that allow uncertainty through the “Gray Zone” approach to identify children at three different OSA severity thresholds (AHI $\geq$ 1, $\geq$5, and $\geq$10).

2. Materials

2.1. Datasets

Two separate datasets referred to as the Sleep dataset and Screen My Sleep (SMS) dataset, were used in this study to develop and validate the logistic models for identifying children at different OSA severity levels. Both study protocols were approved by the University of British Columbia and Children's and Women's Health Centre of British Columbia Research Ethics Board (H11-01769 and H14-02241). Parental/guardian written informed consent was obtained for all subjects, and written assent was obtained for all subjects over the age of 11 years. Children with cardiac arrhythmia or abnormal hemoglobin were excluded.

The Sleep dataset consisted of 160 children, referred to the BCCH for a PSG sleep study, who were recruited between January 2012 and May 2014, under the study protocol (H11-01769). Eight patients withdrew from the study, and nine children were excluded from the study because the total duration of the sleep time or the collected signals from the PSG or the Phone Oximeter (PPG and SpO_2) were shorter than 3 h (see Ref. [10] for a detailed description of this dataset). SMS was a subsequent clinical study focused on validating the Phone Oximeter's OSA screening performance at home [18]. It consisted of 75 children, referred to the BCCH for a PSG sleep study, recruited between June 2015 and March 2016, under the ongoing study protocol (H14-02241). After an overnight stay in the hospital, these patients were asked to record two additional nights at home with the Phone Oximeter. One patient withdrew from the study, three patients did not start the recording with the Phone Oximeter and have no smartphone-based data, seven children were excluded from this analysis because the total duration of the sleep time or the collected signals from the PSG or the Phone Oximeter was shorter than 3 h [18].

2.2. Data acquisition

Data acquisition was carried out in a dedicated facility where standard PSGs are performed. A standard PSG study includes the overnight measurement of electrocardiography (ECG), electroencephalography (EEG), pulse oximetry, chest movement, and nasal airflow, as well as video recording using the Embla Sandman S4500, specifically designed to meet the American Academy of Sleep Medicine (AASM) accreditation requirements (see Table 1). A second pulse oximeter sensor was applied to the finger adjacent to the one used during the standard PSG. This sensor was attached to the smartphone-based pulse oximeter, and SpO_2 and PPG signals were recorded alongside PSG. The SpO_2 (0.1% resolution) and PPG signals, recorded by the Phone Oximeter, were sampled at 1 Hz and 62.5 Hz, respectively.

3. Methods

3.1. Pulse oximetry characterization

Overnight SpO_2 and PPG signals, recorded with the Phone Oximeter, were segmented into two-minute signal segments without overlap; these segments were characterized in time and frequency domains. To evaluate overnight SpO_2 and PRV dynamics, we studied the statistical distribution of each feature through its mean (M), median (Me), standard deviation (S), and interquartile range (I).

Table 1
Illustrates the demographic information and PSG-based and Phone Oximeter-based information (mean and standard deviation) of the population in both datasets (Sleep and SMS) at different AHI cut-offs (1, 5 and 10). PSG-based features include the apnea-hypopnea index (AHI), total bedtime (TBT), total sleep time (TST), the percentage of time spent in the different sleep stages, number of awakenings, number of respiratory arousals, and arousal index. Phone Oximeter-based information includes the oximetry signal quality percentage (SQP), and mean and minimal SpO_2 value of the overnight recording. Only subjects containing a total duration of PSG or Phone Oximeter recordings longer than 3 hours of reliable data are included (n = 207 children).

	AHI 1				AHI 5				AHI 10			
	Sleep		SMS		Sleep		SMS		Sleep		SMS	
	AHI ≥ 1	AHI < 1	AHI ≥ 1	AHI < 1	AHI ≥ 5	AHI < 5	AHI ≥ 5	AHI < 5	AHI ≥ 10	AHI < 10	AHI ≥ 10	AHI < 10
Clinical information												
N (F,M)	103 (64,39)	40 (22,18)	51 (25,26)	13 (9.4)	56 (38,18)	87 (48,39)	24 (10,14)	40 (24,16)	35 (25,10)	108 (61,47)	11 (4.7)	53 (30,23)
Age (y)	9.2 (4.3)	8.9 (4.2)	7.8 (4.1)	9.9 (2.5)	8.8 (4.6)	9.3 (4.1)	6.6 (4.0)	9.2 (3.5)	8.7 (4.6)	9.2 (4.2)	5.6 (4.8)	8.8 (3.5)
AHI	11.7** (16.8)	0.5 (0.3)	7.3** (7.5)	0.5 (0.3)	19.7** (19.5)	1.4 (1.1)	12.4** (8.4)	2 (1.3)	27.5** (21.2)	2.4 (2.4)	18.8** (8.7)	3.2 (2.6)
BMI	21.8* (8)	19.4 (5.2)	19.8 (6.2)	19.8 (4.8)	23.4* (8.3)	19.6 (6.4)	18.5 (3.9)	20.6 (6.8)	24.5 (9.5)	20.1 (6.3)	20.1 (4.4)	19.7 (6.2)
PSG information												
TST (min)	367 (79.8)	360.8 (73)	403.3 (46.1)	385.6 (77.6)	362.1 (82.6)	368 (74.9)	410.9 (36.5)	393 (61.3)	346.2 (90.4)	372 (72.5)	410.6 (33.8)	397.5 (57)
TBT (min)	480.5 (34.4)	481.8 (21.5)	482.4 (23.5)	481.3 (11.5)	479.9 (40)	481.4 (24.4)	482.6 (16.7)	481.9 (24.2)	476.8 (48.6)	482.1 (23.3)	476.9 (21.3)	483.3 (21.6)
Sleep efficiency (%)	76.3 (15.9)	75.1 (15.6)	83.6 (8.6)	80.2 (16.1)	75.1 (16.2)	76.5 (15.5)	85.2 (7.3)	81.5 (11.9)	72.2 (18)	77.2 (14.8)	86.1 (6.3)	82.2 (11.1)
NREM (%)	80.4 (7)	83.1 (6)	80.6 (10.2)	81.4 (4.0)	80 (7.7)	81.9 (6.1)	79.6 (14.2)	81.5 (4.3)	80.5 (8.4)	81.4 (6.2)	75 (20.8)	81.9 (4.2)
REM (%)	19.7 (7.1)	16.9 (6)	18.2 (5.4)	18.6 (4.0)	20.2 (8)	18.1 (6.1)	17.6 (6.4)	18.7 (4.3)	19.9 (8.7)	18.6 (6.3)	18.4 (8.5)	18.2 (4.3)
Awakenings (nr)	20.6* (11)	17.7 (5.7)	15.5 (14.5)	14.6 (5.0)	21.2 (10.6)	18.8 (9.3)	18.2 (20.1)	14 (5.7)	21.3 (11.3)	19.2 (9.4)	14.8 (10.1)	15.7 (13.7)
Respiratory arousals (nr)	44.4** (58.3)	2 (1.4)	22.2** (25.7)	1.8 (1.7)	73.3** (66.3)	6.3 (6)	38** (30.2)	6 (5.1)	98.5** (72.5)	11.1 (12.8)	51.9* (39.1)	11 (11.4)
Arousal Index	8.1** (11.8)	0.3 (0.2)	3.5** (4.5)	0.3 (0.3)	13.6** (13.9)	1 (0.9)	6** (5.5)	0.9 (0.8)	18.9** (15.3)	1.7 (1.9)	8.5 (7.3)	1.6 (1.7)
Phone Oximeter information												
SQP	96 (10.6)	95.9 (11.5)	96.6 (12)	98 (4.4)	95.6 (12.5)	95.9 (10.6)	95.2 (16.2)	98.4 (2.3)	94.5 (14.3)	96.5 (9.6)	94.3 (19.1)	98.1 (4.3)
Mean SpO_2	97.3 (1.2)	97.6 (1)	97.8 (1.1)	97.9 (0.7)	97.3 (1.1)	97.5 (1.1)	97.5 (1.3)	97.9 (0.8)	97.2 (1.1)	97.5 (1.2)	97.4 (1.5)	97.9 (0.7)
Min SpO_2	85.5* (6.1)	88.4 (5.1)	87.4 (5.2)	89.2 (4.2)	84.7* (6.2)	87.5 (5.5)	87** (5.4)	89.2 (4.2)	83.9* (6.6)	87.4 (5.2)	86.8** (5.4)	88.9 (4.3)

* p-value < 0.01, ** p-value < 0.001 comparing OSA and Non OSA.

This characterization was performed offline in Matlab (Mathworks Inc, Natick, USA).

1. SpO_2 Analysis

The SpO_2 samples below 70% or above 100% were considered artifacts, and SpO_2 signal segments containing more than 50% of the signal contaminated by artifacts were excluded from analysis.

SpO_2 signal segments were characterized in the time domain by descriptive statistics, variability measures and indices related to SpO_2 desaturation events, which were previously proposed to predict the presence of OSA [19]. These features included the mean, median, standard deviation and interquartile range of the SpO_2 (SpO_2mean, SpO_2median, SpO_2std, and SpO_2iqr, respectively); the Delta variability index (SpO_2delta) [20]; the number of desaturations from baseline below 3% and 4% (nb3 and nb4, respectively); and the cumulative time spent below an SpO_2 value of 96%, 94%, 92% (tb96, tb94 and tb92, respectively). The Delta index quantifies SpO_2 variability and was computed as the average of absolute differences of the mean oxygen saturation between successive 12-sec intervals.

To characterize the SpO_2 in the frequency domain, we applied a parametric power spectral density (PSD) analysis to the SpO_2 signal segments. The optimum model order was selected according to the minimum description length criterion from Rissanen [21]. Three spectral parameters were extracted from the PSD: (1) the normalized power (P) in the modulation frequency band, which consisted of a frequency interval (0.02 Hz) centered around the modulation frequency peak located between 0.005 and 0.1 Hz; (2) the ratio (R) between the power of the discriminant frequency band and total power; and (3) the Shannon entropy (SE) of the PSD.

2. PRV Analysis

The PPG signal quality was evaluated using an algorithm that assigns a signal quality index (SQI) between 0 and 100 (100 being the best quality). The method is based on pulse segmentation and cross-correlation of consecutive pulse cycles [22]. PRV analysis is very sensitive to artifacts. Thus PPG signal segments containing four or more cycles with low SQI (less than 50) were rejected from further analysis.

A peak detection algorithm based on zero-crossing was applied to locate the pulse peaks in the PPG signal segments, and obtain pulse-to-pulse interval time series (PPIs) [23]. Three time domain parameters were extracted from the PPIs time series, including the mean of the PPIs (meanPP), the standard deviation of the PPIs (SDPP) and the root mean square of the difference of the successive PPIs (RMSSD).

For the frequency analysis, each segment of PPIs was resampled into an equivalent, uniformly spaced time series (sampling rate of 4 Hz), and PSD was computed. The power in each frequency band was computed by integrating the area under the power spectral density curve bound by the band of interest: Total spectral power (TP; 0.01–0.04 Hz), Very Low Frequency (VLF; 0.01–0.04 Hz), Low Frequency (LF; 0.04–0.15 Hz) and High Frequency (HF; 0.15–0.4 Hz). LF and HF powers were normalized (LFn and HFn) by the total spectral power between 0.04 and 0.4 Hz. The ratio of the low-to-high frequency power (LF/HF ratio) was also computed.

3.2. Multivariate model development

Three multivariate logistic regression models were developed to classify children at the selected AHI cut-offs. A stepwise selection method was applied to select the most relevant features for each multivariate logistic regression model. The stepwise selection

method added/dropped one feature with each iteration and stopped when further inclusion/exclusion no longer improved the model, as determined by the Bayes Information Criterion (BIC). All features were included as candidate predictors and a minimum value of five events per variable was retained to avoid overfitting [24]. All data analysis was conducted using R v3.2.0 (R Foundation for Statistical Computing, Vienna, Austria).

3.3. Threshold and "Gray Zone" approach

The "Gray Zone" is an alternative to the binary decision of the ROC curve approach based on only one probability threshold. This technique proposes using two probability thresholds, defined by a diagnostic tolerance value, to form the borders of the "Gray Zone." The first threshold allows the exclusion of the diagnosis (in this study, OSA with specific severity, AHI $\geq$ 1, AHI $\geq$ 5, AHI $\geq$ 10) with near certainty (ie, high sensitivity). The second threshold is chosen to include the diagnosis with near certainty (ie, high specificity). Prediction values that fall within the "Gray Zone," between the two thresholds, correspond to an inconclusive diagnostic prediction. The performance of the "Gray Zone" approach was explored for different diagnosis tolerance values, ranging from 5 to 30%, using leave one out cross-validation. For a specific tolerance value of 5% for example, the "Gray Zone," or inconclusive responses, are defined for threshold values with a sensitivity lower than 95% or specificity values lower than 95% [25]. These thresholds were estimated separately for each model using the different tolerance values.

3.4. Validation

Two methods were used to validate the performance of models regarding accuracy, sensitivity, specificity, and inconclusive or "Gray Zone" percentage. First, we combined both datasets (Sleep + SMS) and applied leave one out cross-validation to develop and validate the models on n = 207 children. The second validation method consisted of using the Sleep dataset (n = 143) to train the models and SMS (n = 64) to validate their performance.

4. Results

4.1. "Gray Zone" tolerance

The effect of the diagnostic tolerance value varied for the different models (see Fig. 1). The optimal tolerance values were selected as the ones providing optimal classification performance while keeping the percentage of children within the "Gray Zone" below 40%. The optimal tolerance values selected were 15%, 10% and 5% for the models classifying children with AHI $\geq$ 1, AHI $\geq$ 5, and AHI $\geq$ 10, respectively.

4.2. Model development and validation

4.2.1. Cross-validation approach

The results of the logistic regression models using optimal diagnostic tolerance values and validated using 207 children showed that the performance of the pulse oximetry-based screening tool increases for higher AHI cut-offs (Table 2).

4.2.2. External validation approach (train with sleep, validate with SMS)

The final model to identify children with an AHI $\geq$ 1 selected two features (see Table 3) and provided an area under the ROC curve (AUC) of 0.80 with a 95% confidence interval (CI) of 0.73–0.88. The model equation was: logit(probability(AHI $\geq$ 1)) = 14.485 + 7.425*(I_SpO$_2$iqr) − 6.06*(M_SE).

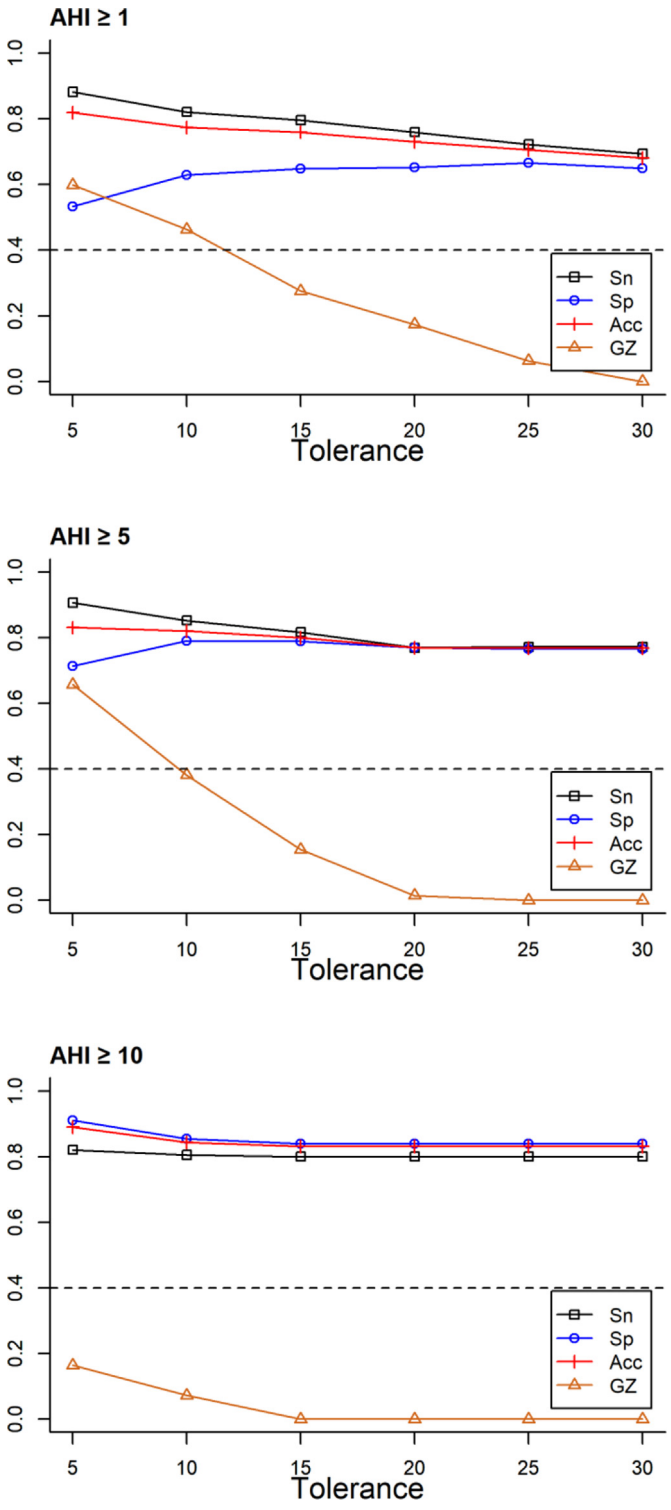

Fig. 1. The performance of the logistic regression models classifying children with AHI $\geq$ 1, AHI $\geq$ 5, AHI $\geq$ 10, for different diagnostic tolerance values validated through cross-validation (n = 207 children). The classification performance is represented as accuracy (Acc), Sensitivity (Sn), Specificity (Sp) and "Gray Zone" percentage (GZ).

The final model to identify children with an AHI $\geq$ 5 selected five features (see Table 3) and provided an AUC of 0.89 with a 95% confidence interval (CI) of 0.82–0.95.

The model equation was: logit(probability(AHI $\geq$ 5)) = 36.13 + 4.279*(I_SpO$_2$iqr) − 6.01*(Me_SE) + 7.64*(S_nb3) − 202.52*(M_P) − 10.87*(S_SpO$_2$delta).

Table 2
Shows the classification results, applying optimal diagnostic tolerance values, regarding accuracy (Acc), Sensitivity (Sn), Specificity (Sp) and inconclusive or "Gray Zone" percentage (GZ) when using cross-validation on both datasets (n = 207 children).

Cross-validation	Acc (%)	Sn (%)	Sp (%)	GZ (%)
AHI ≥ 1	75	80	65	28
AHI ≥ 5	82	85	79	38
AHI ≥ 10	89	82	91	16

The final model to identify children with an AHI ≥ 10 selected seven features (see Table 3) and provided an AUC of 0.92 with a 95% confidence interval (CI) of 0.94–1.

The model equation was: logit(probability(AHI ≥ 10)) = −19.15 − 5.45*(M_HF) − 0.07*(Me_tb96) + 19.37*(M_SpO$_2$delta) + 128.89*(M_SDPP) + 15.64*(I_SpO$_2$std) − 9.36*(M_MeanPP) + 19.76*(M_R).

These models provided good training accuracy, sensitivity and specificity results (see Fig. 2). However, this performance considerably decreased when validated with the SMS dataset (Table 4). In general, the most discriminating features selected by the logistic regression models to screen for OSA at different OSA severity levels, are related to overnight SpO$_2$ variability, overnight modulation of SpO$_2$ reflected in the spectral domain and PRV changes due to intermittent apnea/hypopnea events during sleep.

5. Discussion

We have developed and assessed the performance of a pediatric OSA screening tool at different AHI cut-offs (1, 5, 10) using smartphone-based pulse oximetry through the analysis of overnight SpO$_2$ and PRV dynamics. In addition, we have explored the use of the "Gray Zone" approach, applied with different diagnostic tolerance values to the developed logistic regression models, to allow for screening outputs near to the decision boundary. Including the "Gray Zone" improves the screening performance of the models and highlights inconclusive results when the predictions are not reliable enough for screening decision-making. The classification accuracy not allowing for uncertainty (68%, 77% and 83% for AHI thresholds of 1, 5, 10, respectively) increases to 75%, 82% and 89% when the "Gray Zone" approach is implemented. However, this comes with the cost of having a percentage of children that need to repeat the OSA screening test or be referred for a PSG due to the inconclusive output. The screening tool performed better at higher OSA severity levels providing sensitivity-specificity values of 80–65%, 85–79% and 82–90% for models classifying children with AHI ≥ 1, AHI ≥ 5, and AHI ≥ 10 when using cross-validation. Identifying children with mild (AHI ≥ 1) or moderate (AHI ≥ 5) OSA was more challenging, as illustrated by lower classification performance and the higher percentage of inconclusive children that fell within the "Gray Zone" area (28% and 38%, respectively).

The accuracy of the models in detecting mild and moderate cases decreased (from 75% to 71% and from 82% to 78%) when the models were developed using the Sleep dataset and tested on the SMS dataset. This performance reduction is due to the difference in the AHI distribution of the datasets, especially in the AHI ranges identifying children with mild and moderate OSA (see Fig. 3). The percentage of children with mild OSA (AHI ≥ 1) was 72% in the Sleep dataset compared to 80% in the SMS dataset. Also, the percentage of children with moderate OSA (AHI ≥ 5) was 37% in the Sleep dataset compared to 39% in the SMS dataset. To properly train and validate our screening tool the training and testing data should have similar AHI distributions. This can be guaranteed when using large datasets or by using cross-validation. Thus, in our opinion, the results obtained through cross-validation are more reliable.

This study provides an extensive assessment of the Phone Oximeter as a screening tool for children at different OSA severity levels. The main advantage of this smartphone-based pulse oximeter is that it provides an easy-to-use device to record overnight pulse oximetry. Thus, the proposed Phone Oximeter-based solution will allow anyone, including parents, to perform OSA screening, with the download of an easy-to-use, intuitive application on any existing smartphone. In addition, the results could be computed, stored and communicated to a physician using the connectivity features of the phone. The evaluation of OSA at home rather than in a sleep facility will result in more natural sleep. Furthermore, the at-home monitoring could also allow for performing the OSA screening on multiple nights, in case the results are inconclusive, a

Table 3
Distribution (quartiles) and univariate/multivariate logistic regression odds ratio (OR) with the 95% confidence interval [CI] analysis of the features selected for each final model trained with the Sleep dataset.

Model (AHI ≥ 1)					
Feature	Children with AHI < 5[a]	Children with AHI ≥ 1[a]	p-value	Univariate OR (95% CI)	Multivariate OR (95% CI)
I_SpO$_2$iqr	0.31 (0.27, 0.36)	0.48 (0.34, 0.60)	7E^{-08}	2.10 (1.49,2.96)[b]	2.10 (1.49,2.96)[b]
M_SE	6.96 (6.89, 7.08)	6.92 (6.77, 7.02)	0.02	0.75 (0.60,0.93)[b]	0.79 (0.62,0.96)[b]
Model (AHI ≥ 5)					
Feature	Children with AHI < 5[a]	Children with AHI ≥ 5[a]	p-value	Univariate OR (95% CI)	Multivariate OR (95% CI)
I_SpO$_2$iqr	0.33 (0.28, 0.43)	0.54 (0.45, 0.68)	2E^{-10}	1.88 (1.45,2.43)[b]	1.53 (1.03,2.29)[b]
Me_SE	7.02 (6.95, 7.11)	6.92 (6.78, 7.02)	2E^{-04}	0.69 (0.56,0.85)[b]	0.55 (0.41,0.74)[b]
S_nb3	0.20 (0.15, 0.30)	0.39 (0.27, 0.61)	4E^{-08}	1.91 (1.44,2.52)[b]	2.15 (1.37,3.37)[b]
M_P	0.003 (0.002, 0.005)	0.007 (0.008, 0.014)	2E^{-09}	1.23 (1.06,1.43)[c]	0.36 (0.21,0.62)[c]
S_SpO$_2$delta	0.16 (0.14, 0.20)	0.26 (0.21, 0.38)	3E^{-10}	1.74 (1.30, 2.33)[b]	2.96 (1.27,6.90)[b]
Model (AHI ≥ 10)					
Feature	Children with AHI < 10[a]	Children with AHI ≥ 10[a]	p-value	Univariate OR (95% CI)	Multivariate OR (95% CI)
M_HF	0.29 (0.13,0.85)	0.30 (0.16,0.60)	0.4	0.94 (0.87,1.01)[b]	0.58 (0.43,0.78)[b]
Me_tb96	0 (0,1)	3.5 (0,26.75)	0.001	1.01 (0.99,1.02)	0.93 (0.90,0.97)
M_SpO$_2$delta	0.23 (0.20, 0.31)	0.41 (0.34,0.53)	3E^{-11}	5.78 (2.90,11.53)[b]	6.9 (2.10,22.94)[b]
M_SDPP	0.05 (0.04,0.08)	0.05 (0.04,0.07)	0.6	0.99 (0.98,1.01)[d]	1.14 (1.05,1.23)[d]
I_SpO$_2$std	0.29 (0.24,0.36)	0.47 (0.40, 0.69)	6E^{-11}	3.01 (1.91,4.92)[b]	4.78 (1.78,12.81)[b]
M_MeanPP	0.79 (0.69,0.88)	0.69 (0.64,0.79)	0.004	0.70 (0.52,0.95)[b]	0.39 (0.18,0.87)[b]
M_R	0.047 (0.44,0.49)	0.49 (0.47,0.52)	0.002	3.62 (1.44,9.12)[b]	7.2 (1.39,37.44)[b]

[a] Presented as median (first quartile, third quartile) for each feature.
[b] Associated with an increase of 0.1 times the feature unit.
[c] Associated with an increase of 0.005 times the feature unit.
[d] Associated with an increase of 0.001 times the feature unit.

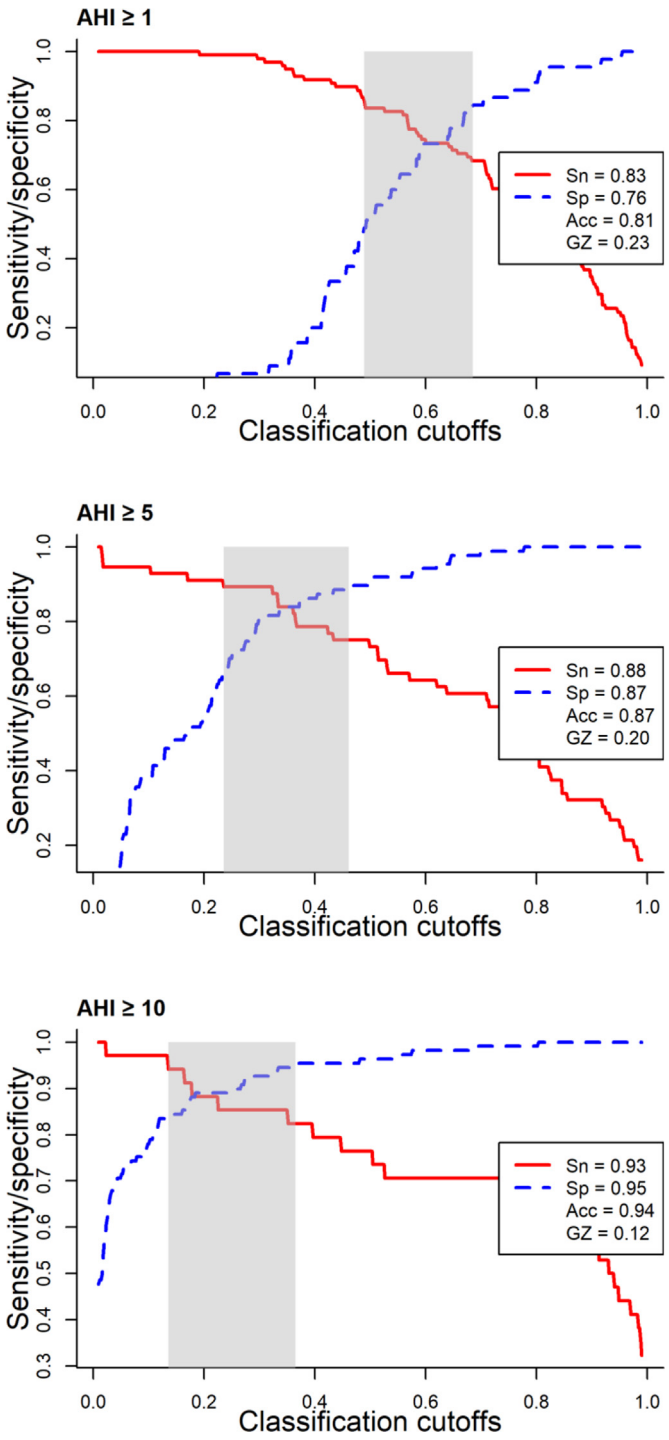

Fig. 2. Representation of sensitivity and specificity obtained with the training dataset (Sleep dataset) of the logistic regression models classifying children with AHI ≥ 1, AHI ≥ 5, AHI ≥ 10. The "Gray Zone" (inconclusive area), which was defined for the optimal diagnostic tolerance values (15%, 10%, and 5% respectively), is represented with a grey background for all models. The training performance of each model is illustrated by its accuracy (Acc), sensitivity (Sn), specificity (Sp), and inconclusive or "Gray Zone" percentage (GZ).

benefit over single night hospital tests. Thus, this at-home screening tool could facilitate the evaluation of OSA in children and optimize the use of PSG in Canada.

The objective information provided by the OSA screening tool, added to clinical history, will help clinicians to more efficiently

Table 4
Shows the classification results, applying optimal diagnostic tolerance values, in terms of accuracy (Acc), Sensitivity (Sn), Specificity (Sp) and "Gray Zone" percentage (GZ), inconclusive area, when using Sleep dataset (n = 143) to train the models and SMS dataset (n = 64) to test the model.

SMS validation	Acc (%)	Sn (%)	Sp (%)	GZ (%)
AHI ≥ 1	71	68	86	30
AHI ≥ 5	78	58	89	16
AHI ≥ 10	88	90	87	11

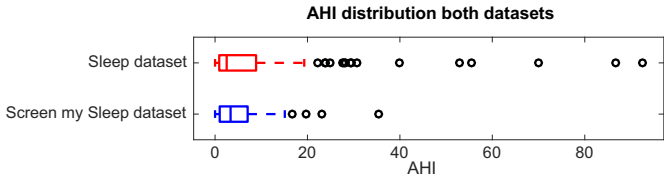

Fig. 3. Boxplot of the AHI for the two datasets: Sleep is represented in red and SMS in blue. Whiskers are used to represent the most extreme values within 1.5 times the interquartile range from the median. Outliers (data with values beyond the ends of the whiskers) are displayed as circles. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

identify and treat pediatric OSA. This screening tool could be set at different AHI cut-offs for different purposes. For example, by setting the AHI cut-off to 5, we could identify children with moderate OSA that should be referred for a PSG. From both datasets, we could predict that less than 40% of the children suspected of having OSA have an AHI ≥ 5. One could halve the number of children referred to the hospital for a PSG and thus increase OSA assessment coverage with at home screening. Based on our results, using our implemented algorithm including the "Gray Zone" approach, the Phone Oximeter would correctly identify 85% of the children with OSA that need a PSG. In addition, 21% of children referred for PSG would not have OSA, and 38% of children will require a re-evaluation at-home. Alternatively, we could set the AHI cut-off to 10 to identify children with severe OSA that need prompt treatment. Due to the limited availability and high cost of PSG, the majority of the children who are referred for adenotonsillectomy (the most common treatment for pediatric OSA) is based on clinical history alone, without a confirmed, objective evaluation of OSA [26]. For instance, at BCCH, only 5–10% of children referred for adenotonsillectomy have a preoperative PSG test [27]. Thus, objectively identifying severe cases, might help clinicians identify the children that require urgent surgery, optimize the selection of postoperative disposition and assist anaesthesiologists with optimization of opioid dosage during surgery. In addition, the use of the implemented "Gray zone" approach will allow us to identify inconclusive subjects. These children should first repeat the overnight pulse oximetry at home, and when the results are still inconclusive again, they should be referred for a PSG, unless the clinical history alone provides enough diagnostic information.

In our previous research, we showed that using advanced signal processing to characterize SpO_2 dynamics (micro-desaturations) can improve identification of children with OSA [25]. HRV has also been shown to improve the detection of OSA. We have previously established that PRV, derived from the PPG signals a valid a surrogate measure of HRV [28]. PRV provided accurate estimates of HRV in the time domain, while some differences were found in the frequency domain. The low sampling frequency of the PPG and pulse transit time variability might introduce uncertainty in the estimation of PRV in the spectral domain, mainly in the respiratory band (HFn) [29,30]. However, previous research has concluded that

these differences are sufficiently small to suggest the use of PRV as an alternative measure of HRV [28,29]. We conducted an additional investigation into the effects of OSA on PRV during different sleep stages and confirmed that the modulation of PRV could improve assessment of OSA in children [23,31]. There are some instances of airway obstruction that do not cause sufficient desaturation to be classified as a desaturation 'event', but are clinically significant as they result in arousal from sleep. We showed that PRV is sensitive to these obstructions [13] and that combining SpO_2 and PRV improved the accuracy (85% estimated using cross-validation) in detecting children with AHI ≥ 5[10]. In the present study, the accuracy of the logistic regression model at this AHI cut-off is slightly below (82%) and allows uncertain outputs, but we used a larger dataset (n = 207).

Recently, Crespo et al., evaluated the performance of different binary classifiers identifying pediatric OSA using overnight oximetry alone, at AHI cut-offs of 1, 3 and 5 [9]. They concluded that logistic regression provided the highest diagnostic accuracy (84%, 78%, and 83%) for identifying children with AHI $\geq$ 1, AHI $\geq$ 3, and AHI $\geq$ 5, respectively. Their dataset consisted of 176 children and the results were validated through bootstrapping. In addition, Hornero et al., performed an extensive validation of an automated neural network algorithm based on overnight oximetry features using a dataset of 4191 children [8]. They showed an accuracy of 75%, 82% and 90% identifying children with AHI $\geq$ 1, AHI $\geq$ 5, and AHI $\geq$ 10, respectively. Our results are in agreement with these findings, showing higher performance in detecting children with moderate and severe OSA compared to children with mild OSA. In our opinion, the screening performance is highly dependent on the particular dataset, device, and settings used and the validation method [32]. However, all these studies demonstrate that pulse oximetry can be an alternative to PSG, particularly considering the current lack of resources to meet the demand.

One of the limitations of the study is that our datasets (Sleep and SMS) showed different AHI distributions, which shows that using cross-validation is a more appropriate approach to validate the results reliably and that greater datasets are required to validate the screening tool externally. This is shown by Hornero et al., who successfully trained and externally validated the neural network with a dataset of 4191 children combining nocturnal single-channel oximetry from multiple hospitals [8]. However, all our studies were performed using a new device, the Phone Oximeter, which is a limitation for expanding our dataset. Nevertheless, when compared with the majority of the studies using pulse oximetry to detect pediatric OSA, our dataset (n = 207) is quite extensive. Similarly, the results of these studies were validated through cross-validation or bootstrapping.

Another limitation of this study is that overnight pulse oximetry was recorded at the BCCH under the supervision of a sleep technician. Yet, the real setting for the use of this tool is at home, where we expect artifacts caused by sensor displacement to be more severe. This could degrade pulse oximetry signal quality. Our future research will further validate the use of the Phone Oximeter and the proposed OSA assessment tool at home through the analysis of the SMS study, which also contains the pulse oximetry recordings of two additional nights at home using the Phone Oximeter [18].

In this study, we focused on assessing the performance of the Phone Oximeter as a screening tool. However, we also plan to develop an OSA severity assessment tool. We will further investigate the fusion of the developed multivariate regression models to provide not only OSA screening but also OSA severity information. The implementation of the "Gray Zone" approach could be also helpful for this purpose. Because some information about OSA severity could still be deduced from the children that fell in the "Gray Zone" for one of the models, though the results of the other models.

6. Conclusion

This study evaluates the performance of a portable pediatric OSA screening tool at different AHI cut-offs using overnight smartphone-based pulse oximetry through the characterization of SpO_2 dynamics and PRV analysis. The novel "Gray zone" approach allows providing inconclusive OSA-screening results and may increase its utility and performance. Implementing this method on the Phone Oximeter will provide a tool hosted on a smartphone that can be used at home by parents or caregivers. Including this objective OSA assessment information in the clinical history of the children suspected with OSA might assist clinicians in the management of pediatric OSA.

Acknowledgments

The authors would also like to especially thank the Pediatric Anesthesia Research Team at the British Columbia Children's Hospital for their efforts in data collection and continuous support. The authors would also like to thank the clinical staff of the sleep laboratory at the BC Children's Hospital and especially Heather Beals, for their assistance in conducting this study.

Conflicts of interest

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: https://doi.org/10.1016/j.sleep.2018.08.027.

References

[1] Rosen CL, Larkin EK, Kirchner HL, et al. Prevalence and risk factors for sleep-disordered breathing in 8- to 11-year-old children: association with race and prematurity. J Pediatr 2003;142(4):383–9. http://www.ncbi.nlm.nih.gov/pubmed/12712055.
[2] Wildhaber JH, Moeller A. Sleep and respiration in children: time to wake up! Swiss Med Wkly 2007;137(49–50):689–94. http://www.ncbi.nlm.nih.gov/pubmed/18197483.
[3] Arens R, Marcus CL. Pathophysiology of upper airway obstruction : a developmental perspective. Sleep 2004;27(5):997–1019. http://www.ncbi.nlm.nih.gov/pubmed/15453561. [Accessed 10 August 2016].
[4] The Section on Pediatric Pulmonology and Subcommittee on Obstructive Sleep Apnea Syndrome. Clinical practice guideline: diagnosis and management of childhood obstructive sleep apnea syndrome. Pediatrics 2002;109(4):704–12.
[5] Marcus CL, Brooks LJ, Draper KA, et al. Diagnosis and management of childhood obstructive sleep apnea syndrome. Pediatrics 2012;130(3):576–84. https://doi.org/10.1542/peds.2012-1671.
[6] Kushida C a, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 2005;28(4):499–521. http://www.ncbi.nlm.nih.gov/pubmed/16171294.
[7] Katz SL, Witmans M, Barrowman N, et al. Paediatric sleep resources in Canada: the scope of the problem. Paediatr Child Health 2014;19(7):367–72. http://www.ncbi.nlm.nih.gov/pubmed/25332676. [Accessed 31 May 2018].
[8] Hornero R, Kheirandish-Gozal L, Gutiérrez-Tobal GC, et al. Nocturnal oximetry–based evaluation of habitually snoring children. Am J Respir Crit Care Med 2017;196(12):1591–8. https://doi.org/10.1164/rccm.201705-0930OC.
[9] Crespo A, Álvarez D, Kheirandish-Gozal L, et al. Assessment of oximetry-based statistical classifiers as simplified screening tools in the management of childhood obstructive sleep apnea. Sleep Breath February 2018. https://doi.org/10.1007/s11325-018-1637-3.
[10] Garde A, Dehkordi P, Karlen W, et al. Development of a screening tool for sleep disordered breathing in children using the phone oximeter™. PLoS One 2014;9(11). https://doi.org/10.1371/journal.pone.0112959.
[11] Nixon GM, Kermack AS, Davis GM, et al. Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry. Pediatrics 2004;113(1):e19–25.
[12] Brouillette RT, Morielli A, Leimanis A, et al. Nocturnal pulse oximetry as an abbreviated testing modality for pediatric obstructive sleep apnea. Pediatrics 2000;105(2):405–12. https://doi.org/10.1542/peds.105.2.405.
[13] Garde A, Dehkordi P, Wensley D, et al. Pulse oximetry recorded from the phone oximeter for detection of obstructive sleep apnea events with and without oxygen desaturation in children. Proc IEEE Conf Eng Med Biol 2015:7692–5. https://doi.org/10.1109/EMBC.2015.7320174.

[14] Karlen W, Dumont GA, Petersen C, et al. Human-centered Phone oximeter interface design for the operating room. Pulse oximeter interfaced to a mobile device for anesthesia monitoring in the developing world. Proc Int Conf Heal Informatics 2011:433–8. https://doi.org/10.5220/0003335204330438 (August 2015).
[15] Petersen CL, Chen TP, Ansermino JM, et al. Design and evaluation of a low-cost smartphone pulse oximeter. Sensors (Basel) 2013;13(12):16882–93. https://doi.org/10.3390/s131216882.
[16] Kaditis AG, Luz M, Alvarez A, et al. Obstructive sleep disordered breathing in 2- to 18-year-old children: diagnosis and management. *Task force report*. Eur Respir Soc Statement 2016:69–94. https://doi.org/10.1183/13993003.00385-2015.
[17] Cannesson M, Le Manach Y, Hofer CK, et al. Assessing the diagnostic accuracy of pulse pressure variations for the prediction of fluid responsiveness. Anesthesiology 2011;115(2):231–41. https://doi.org/10.1097/ALN.0b013e318225b80a.
[18] Hoppenbrouwer XLR, Dehkordi P, Rollinson AU, et al. Night to night pulse oximetry variability in children with suspected sleep apnea. In Proc IEEE Conf Eng Med Biol 2018:179–82 (in press).
[19] Alvarez D, Hornero R, Marcos JV, et al. Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis. Int J Neural Syst 2013;23(5):1350020. https://doi.org/10.1142/S0129065713500202.
[20] Levy P, Pepin JL, Paramelle B, et al. Accuracy of disturbances oximetry for detection in sleep respiratory apnea syndrome. Chest 1996:395–9. https://doi.org/10.1378/chest.109.2.395.
[21] Rissanen J. Modeling by shortest data description. Automatica 1978;14:465–71.
[22] Karlen W, Kobayashi K, Ansermino JM, et al. Photoplethysmogram signal quality estimation using repeated Gaussian filters and cross-correlation. Physiol Meas 2012;33(10):1617–29. https://doi.org/10.1088/0967-3334/33/10/1617.
[23] Dehkordi P, Garde A, Karlen W, et al. Evaluation of cardiac modulation in children in response to apnea/hypopnea using the phone oximeter™. Physiol Meas 2016;37(2):187. https://doi.org/10.1088/0967-3334/37/2/187.
[24] Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49(12):1373–9. http://www.ncbi.nlm.nih.gov/pubmed/8970487. [Accessed 12 August 2016].
[25] Goarin JP, Lehot J, Vallet B, et al. Assessing the diagnostic accuracy of pulse pressure. Anesthesiology 2011;89(2):231–41. https://doi.org/10.1097/ALN.0b013e318225b80a.
[26] Kaditis A, Kheirandish-Gozal L, Gozal D, et al. Algorithm for the diagnosis and treatment of pediatric OSA: a proposal of two pediatric sleep centers. Sleep Med 2012;13(3):217–27. https://doi.org/10.1016/j.sleep.2011.09.009.
[27] Allison E, Garde A, Dunsmuir D, et al. Overnight oximetry at home before and after adenotonsillectomy. 2016.
[28] Dehkordi P, Garde A, Karlen W, et al. Pulse rate variability compared with heart rate variability in children with and without sleep disordered breathing. In Proc IEEE Conf Eng Med Biol 2013:6563–6. https://doi.org/10.1109/EMBC.2013.6611059.
[29] Gil E, Orini M, Bailón R, et al. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. Physiol Meas 2010;31(9):1271–90. https://doi.org/10.1088/0967-3334/31/9/015.
[30] García-González M a, Fernández-Chimeno M, Ramos-Castro J. Bias and uncertainty in heart rate variability spectral indices due to the finite ECG sampling frequency. Physiol Meas 2004;25(2):489–504. https://doi.org/10.1088/0967-3334/25/2/008.
[31] Dehkordi P, Garde A, Karlen W, et al. Pulse rate variability in children with disordered breathing during different sleep stages. In Comput Cardiol 2013:1015–8.
[32] Garde A, Karlen W, Dehkordi P, et al. Oxygen saturation resolution influences regularity measurements. In: Proc. IEEE Conf Eng Med Biol; 2014. p. 2257–60. https://doi.org/10.1109/EMBC.2014.6944069.