

Nocturnal Oximetry–based Evaluation of Habitually Snoring Children

Roberto Hornero¹, Leila Kheirandish-Gozal², Gonzalo C. Gutiérrez-Tobal¹, Mona F. Philby², María Luz Alonso-Álvarez³, Daniel Álvarez^{1,4}, Ehab A. Dayyat⁵, Zhifei Xu⁶, Yu-Shu Huang⁷, Maximiliano Tamae Kakazu⁸, Albert M. Li⁹, Annelies Van Eyck^{10,11}, Pablo E. Brockmann¹², Zarmina Ehsan^{13*}, Narong Simakajornboon¹³, Athanasios G. Kaditis¹⁴, Fernando Vaquerizo-Villar¹, Andrea Crespo Sedano⁴, Oscar Sans Capdevila¹⁵, Magnus von Lukowicz¹⁶, Joaquín Terán-Santos³, Félix Del Campo^{1,4}, Christian F. Poets¹⁶, Rosario Ferreira¹⁷, Katalina Bertran¹⁵, Yamei Zhang⁶, John Schuen⁸, Stijn Verhulst^{10,11}, and David Gozal²

¹Biomedical Engineering Group, University of Valladolid, Valladolid, Spain; ²Section of Sleep Medicine, Department of Pediatrics, Pritzker School of Medicine, Biological Sciences Division, University of Chicago, Chicago, Illinois; ³Unidad Multidisciplinar del Sueño, Centro de Investigación Biomédica en Red Respiratorio, Hospital Universitario de Burgos, Burgos, Spain; ⁴Sleep-Ventilation Unit, Pneumology Service, Río Hortega University Hospital, Valladolid, Spain; ⁵Division of Child Neurology, Department of Pediatrics, LeBonheur Children's Hospital, University of Tennessee Health Science Center, School of Medicine, Memphis, Tennessee; ⁶Sleep Unit, Beijing Children's Hospital, Capital Medical University, Beijing, People's Republic of China; ⁷Department of Child Psychiatry and Sleep Center, Chang Gung Memorial Hospital and University, Taoyuan, Taiwan; ⁸Spectrum Health, Michigan State University, Grand Rapids, Michigan; ⁹Department of Pediatrics, Prince of Wales Hospital, Chinese University of Hong Kong, Hong Kong, China; ¹⁰Laboratory of Experimental Medicine and Pediatrics and ¹¹Department of Pediatrics, University of Antwerp and Antwerp University Hospital, Antwerp, Belgium; ¹²Sleep Medicine Center, Department of Pediatric Cardiology and Pulmonology, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; ¹³Division of Pulmonary and Sleep Medicine, Cincinnati Children's Medical Center, Cincinnati, Ohio; ¹⁴Pediatric Pulmonology Unit, Sleep Disorders Laboratory, First Department of Pediatrics, National and Kapodistrian University of Athens School of Medicine and Aghia Sophia Children's Hospital, Athens, Greece; ¹⁵Sleep Unit, Department of Neurology, Sant Joan de Deu, Barcelona Children's Hospital, Barcelona, Spain; ¹⁶Department of Neonatology and Sleep Unit, University of Tübingen, Tübingen, Germany; and ¹⁷Pediatric Respiratory Unit, Department of Pediatrics, Hospital de Santa Maria, Academic Medical Center of Lisbon, Lisbon, Portugal

ORCID IDs: 0000-0003-3332-1057 (L.K.-G.); 0000-0001-8195-6036 (D.G.).

Abstract

Rationale: The vast majority of children around the world undergoing adenotonsillectomy for obstructive sleep apnea–hypopnea syndrome (OSA) are not objectively diagnosed by nocturnal polysomnography because of access availability and cost issues. Automated analysis of nocturnal oximetry (nSpO₂), which is readily and globally available, could potentially provide a reliable and convenient diagnostic approach for pediatric OSA.

Methods: Deidentified nSpO₂ recordings from a total of 4,191 children originating from 13 pediatric sleep laboratories around the world were prospectively evaluated after developing and validating an automated neural network algorithm using an initial set of single-channel nSpO₂ recordings from 589 patients referred for suspected OSA.

Measurements and Main Results: The automatically estimated apnea–hypopnea index (AHI) showed high agreement with AHI from conventional polysomnography

(intraclass correlation coefficient, 0.785) when tested in 3,602 additional subjects. Further assessment on the widely used AHI cutoff points of 1, 5, and 10 events/h revealed an incremental diagnostic ability (75.2, 81.7, and 90.2% accuracy; 0.788, 0.854, and 0.913 area under the receiver operating characteristic curve, respectively).

Conclusions: Neural network–based automated analyses of nSpO₂ recordings provide accurate identification of OSA severity among habitually snoring children with a high pretest probability of OSA. Thus, nocturnal oximetry may enable a simple and effective diagnostic alternative to nocturnal polysomnography, leading to more timely interventions and potentially improved outcomes.

Keywords: childhood obstructive sleep apnea–hypopnea syndrome; nocturnal oximetry; blood oxygen saturation; automated pattern recognition; neural network

(Received in original form May 11, 2017; accepted in final form July 28, 2017)

*Present address: Pulmonology and Sleep Medicine, Children's Mercy Kansas City, Kansas City, MO.

Supported in part by project VA037 U16 from the Consejería de Educación de la Junta de Castilla y León and the European Regional Development Fund (FEDER), project RTC-2015-3446-1 from the Ministerio de Economía y Competitividad and FEDER, and project 153/2015 of the Sociedad Española de Neumología y Cirugía Torácica (SEPAR). L.K.-G. is supported by NIH grant 1R01HL130984. M.F.P. was supported by a Fellowship Educational grant award from the Kingdom of Saudi Arabia. D.Á. was in receipt of a Juan de la Cierva grant from the Ministerio de Economía y Competitividad. The funders played no role in the study design, data collection, data analysis, interpretation, and writing of the manuscript.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 196, Iss 12, pp 1591–1598, Dec 15, 2017

Copyright © 2017 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201705-0930OC on July 31, 2017

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Obstructive sleep apnea (OSA) is a prevalent condition in children that currently requires an overnight sleep study for diagnosis. In view of the relative scarcity and cost of sleep studies for children, major delays occur in sleep apnea diagnosis and treatment, and a large proportion of children are treated without a formal diagnosis.

What This Study Adds to the

Field: An automated neural network algorithm based on overnight oximetry recordings provides accurate identification of OSA severity among habitually snoring children with a high pretest probability of OSA.

Since its initial description, pediatric obstructive sleep apnea–hypopnea syndrome (OSA) has become recognized not only as a prevalent condition in children, but has also been associated with increased risk for major morbidities affecting neurocognitive, behavioral, cardiovascular, and metabolic functioning, ultimately resulting in overall health and quality of life declines (1, 2). These adverse consequences, along with the increased health care use and associated costs (3, 4), have prompted several consensus statements and guidelines advocating for timely diagnosis of OSA using nocturnal polysomnography (NPSG) (1, 5, 6). However, the relative unavailability of pediatric sleep laboratories around the world, the overall high costs and labor-intensive nature associated with NPSG testing, and their inconvenience to both parents and children have resulted in only a minority of children being objectively evaluated (7). Such findings have prompted the search for simplified

approaches that would increase the accessibility and effectiveness of OSA diagnosis, and a large array of proposed methodologies ranging from questionnaires to biomarkers has emerged (1, 2, 5, 6, 8–12).

Nocturnal oximetry ($nSpO_2$) was initially proposed as a screening tool for OSA in symptomatic children (13–17), and although based on relatively small pediatric cohorts, this approach appears to provide high specificity but limited sensitivity (10). However, the relatively low accuracy and interscorer reliability of visually scored $nSpO_2$ and the inability of such approaches in providing an accurate estimate of OSA severity, particularly at the low end of its severity spectrum, reduced the enthusiasm for wider implementation of $nSpO_2$ recordings.

In a preliminary study, we developed a neural network–based signal-processing technique that appeared to improve remarkably the diagnostic ability of single-channel $nSpO_2$ recordings (15). In the present study, we expanded the derivation and validation of $nSpO_2$ recordings by prospectively assessing a large cohort of children, using polysomnography, across 13 pediatric sleep laboratories around the world. The cumulative findings lend support to the use of automated signal-processing algorithms of $nSpO_2$ recordings in children to diagnose OSA and estimate its severity.

Methods

Patients

A total of 4,191 habitually snoring children (2,517 boys and 1,674 girls), ranging in age from 2 to 18 years and who were referred for clinical suspicion of OSA and underwent NPSG, composed the population under study. Deidentified recordings of $nSpO_2$ were extracted from each NPSG along with pertinent demographic and clinical information including NPSG-derived measures. As

indicated, all patients underwent physician-directed in-laboratory NPSG because of habitual snoring and/or witnessed breathing pauses during sleep as reported by their parents or caregivers. The study was approved by the ethics review committee of each of the 13 participating centers.

Sleep Studies

All NPSGs were originally scored manually at each participating center, based on the 2012 American Academy of Sleep Medicine criteria (17). Included NPSG studies required at least 6 hours of recorded sleep. The apnea–hypopnea index (AHI) obtained from each individual NPSG was used as the “gold standard” for OSA diagnosis. In this study, the following common clinically used AHI cutoff points were assessed: 1, 5, and 10 events per hour (e/h). Table 1 shows the patient demographics and OSA prevalence, overall and for each participating center, according to the aforementioned AHI cutoffs.

Automatic $nSpO_2$ Signal Analysis

Four automatic signal analysis stages were implemented to obtain useful information from $nSpO_2$. A scheme of the automatic methodology applied to the $nSpO_2$ signals is shown in Figure E1 in the online supplement. The first phase was a preprocessing step to standardize the signals obtained from the 13 different pediatric sleep centers. These centers recorded $nSpO_2$ using one or several sampling rates ranging from 1 to 500 Hz, as well as different decimal resolution. Hence, all the $nSpO_2$ recordings acquired were resampled to 25 Hz, as recommended by the American Academy of Sleep Medicine (18). In addition, signals were all rounded to the second decimal place. Finally, artifacts were removed according to the automatic method suggested by Magalang and colleagues (19).

After preprocessing, a feature extraction phase was used to characterize

Author Contributions: R.H. and L.K.-G. conceptualized the study, analyzed data, and drafted components of the manuscript. G.C.G.-T. performed analyses and contributed to manuscript editing. M.F.P., M.L.A.-Á., D.Á., E.A.D., Z.X., Y.-S.H., M.T.K., A.M.L., A.V.E., P.E.B., Z.E., N.S., A.G.K., A.C.S., O.S.C., M.v.L., J.T.-S., F.D.C., C.F.P., R.F., K.B., Y.Z., J.S., and S.V. recruited subjects and contributed cases to the database. F.V.-V. contributed to data analyses. D.G. conceptualized the study, coordinated the database, performed data analyses, drafted components, and edited the manuscript.

Correspondence and requests for reprints should be addressed to David Gozal, M.D., M.B.A., Section of Pediatric Sleep Medicine, Department of Pediatrics, Pritzker School of Medicine, Biological Sciences Division, University of Chicago, KCBD, Room 4100, 900 E. 57th Street, Mailbox 4, Chicago, IL 60637. E-mail: dgozal@uchicago.edu

Table 1. Demographic and Clinical Data for Each Participating Site

Center	Participants (n)	Age (yr) (Mean ± SD)	Male (%)	BMI (kg/m ²) (Mean ± SD)	AHI (per Hour TST) (Mean ± SD)	OSA for AHI = 1 e/h (%)	OSA for AHI = 5 e/h (%)	OSA for AHI = 10 e/h (%)
UofC	981	6.1 ± 3.4	61.4	19.7 ± 7.3	9.3 ± 17.2	82.2	41.3	23.3
UofTn	611	7.2 ± 4.6	54.6	23.3 ± 10.1	5.8 ± 11.3	68.1	29.6	17.7
HUBU	578	4.1 ± 2.2	61.8	17.1 ± 4.2	5.9 ± 11.3	64.5	26.3	15.2
BCH	558	6.3 ± 5.3	66.3	17.8 ± 3.7	5.8 ± 11.7	65.1	27.4	17.0
MSU	499	6.5 ± 5.0	55.5	17.8 ± 11.2	6.2 ± 9.3	85.8	22.0	14.8
CGMH	283	9.9 ± 3.2	72.4	19.5 ± 4.6	4.3 ± 10.0	72.4	21.9	8.1
UofHK	202	10.0 ± 2.4	62.9	18.7 ± 4.6	4.9 ± 7.5	70.3	26.2	10.4
PUCC	183	5.4 ± 4.8	52.5	17.8 ± 4.2	3.7 ± 9.1	60.1	18.6	7.6
UofA	130	11.7 ± 3.1	37.7	30.3 ± 5.7*	3.2 ± 7.1	63.1	22.3	10.0
SJDCH	60	8.4 ± 4.8	58.3	19.5 ± 5.2	4.2 ± 6.1	76.7	25.0	11.6
ASCH	51	7.0 ± 3.4	66.7	20.6 ± 6.4	10.6 ± 13.8	90.2	54.9	37.2
UofTu	36	10.4 ± 3.5	61.1	21.0 ± 8.0	6.9 ± 12.9	72.2	27.8	16.7
HSM	19	6.5 ± 3.8	47.4	19.1 ± 6.8	11.0 ± 15.2	73.7	47.4	36.8
All	4,191	6.7 ± 4.4	60.0	20.0 ± 7.0	6.4 ± 12.5	72.9	29.6	16.8

Definition of abbreviations: AHI = apnea-hypopnea index; ASCH = Aghia Sophia Children's Hospital (Greece); BCH = Beijing Children's Hospital (China); BMI = body mass index; CGMH = Chang Gung Memorial Hospital (Taiwan); e/h = events per hour; HSM = Hospital de Santa Maria (Portugal); HUBU = Hospital Universitario de Burgos (Spain); MSU = Michigan State University (USA); OSA = obstructive sleep apnea-hypopnea syndrome; PUCC = Pontificia Universidad Católica de Chile (Chile); SJDCH = San Joan de Deu Children's Hospital (Spain); TST = total sleep time; UofA = University of Antwerp (Belgium); UofC = University of Chicago (USA); UofHK = University of Hong Kong (Hong Kong, China); UofTn = University of Tennessee (USA); UofTu = University of Tuebingen (Germany).

*This cohort was specifically aimed at verifying the accuracy of the neural network-based algorithm in an obese pediatric population.

pediatric OSA in each recording. Consequently, two complementary analytical approaches consisting of time-domain and frequency-domain analyses were used. The latter is justified because of the recurrence of apneic events, whereas the

former has already shown its ability to quantify the statistical and nonlinear information inherent to biomedical signals (20). Thus, up to 23 features were obtained from each recording, which are summarized in Table 2. All of them have

been successfully evaluated in previous studies involving patients with OSA (20–24).

Such comprehensive characterization of the nSpO₂ recordings may lead to redundant features (25). Hence, the third phase of our automatic signal-processing methodology was a feature selection stage. The fast correlation-based filter was applied to the extracted features to evaluate the relevance and redundancy of their OSA-related information. In a first step, the fast correlation-based filter algorithm conducts a relevancy analysis by comparing the information shared by each feature and a reference variable (26), that is, AHI. Then, the features are ranked higher as more information is shared with the reference. A second step consists of comparing the information shared by each feature with the other ones, that is, conducting a redundancy analysis. Features sharing more information with other, higher-ranked ones than with AHI are discarded because of redundancy (26).

At this point of the processing, subjects under study were characterized by a vector, whose components are their corresponding values of the nonredundant extracted features. Thus, the fourth stage consisted of the training of a multilayer perceptron (MLP) model with the ability to automatically estimate AHI from these patterns. MLP is an artificial neural network

Table 2. Time and Frequency Domain Features Extracted from the nSpO₂ Recordings

Feature	Description
Time domain	
Mt1–Mt4	First, second, third, and fourth statistical moment of a time series
CTM	Central tendency measure to quantify variability
LZC	Lempel–Ziv complexity
SampEn	Sample entropy to measure irregularity
ODI3	3% oxygen desaturation index
Frequency domain	
MA	Full-spectrum amplitude maximum
mA	Full-spectrum amplitude minimum
Mf1–Mf4	First, second, third, and fourth statistical moment of the full spectrum
MF	Median frequency of the full spectrum to estimate the distribution of the power of the spectrum
SpecEn	Spectral entropy to measure the full spectrum flatness
WD	Wootters' distance to estimate the statistical distance of the full spectrum and a uniform distribution
MA _{BOI} *	Spectrum amplitude maximum from the band of interest*
mA _{BOI}	Spectrum amplitude minimum from the band of interest
Mf1 _{BOI} –Mf4 _{BOI}	First, second, third, and fourth statistical moment of the spectral band of interest

Definition of abbreviation: nSpO₂ = nocturnal oximetry.

*The band of interest (BOI) was defined as the one showing the highest statistical differences among obstructive sleep apnea severity groups of the training set (21).

that is typically arranged in three layers of mathematical units called “neurons”: input, hidden, and output (27). The input has as many neurons as features used to train the network, that is, the number of nonredundant features. By contrast, the output is composed of only a single neuron, which provides an AHI value for each subject. Finally, the number of neurons of the hidden layer (N_H) is a tuning parameter experimentally determined in the training set (27). The value of a regularization parameter (α) is another common design choice used to minimize the chances of overfitting training data (27).

Statistical Analyses

SPSS Statistics version 20 software (SPSS/IBM, Chicago, IL) was used to perform statistical analyses. Normality and homoscedasticity analyses revealed that oximetric features derived from the population under study were not normally distributed and variances were unequal. Therefore, descriptive analysis of features was presented in terms of their median and interquartile range. In addition, the nonparametric Mann-Whitney U test was applied to search for statistical significant differences between OSA-negative and OSA-positive groups. A *P* value less than 0.05 was considered significant.

MATLAB R2016b (MathWorks, Natick, MA) was used to implement feature extraction, selection, and classification stages. A Bland-Altman plot and intraclass correlation coefficient (ICC) were used to directly assess the agreement between the NPSG-derived AHI and the neural network-derived AHI estimate. The agreement in the four-class classification between the NPSG AHI and our estimate was measured using Cohen’s κ . The four OSA groups were defined according to the aforementioned cutoffs ($AHI < 1$, $1 \leq AHI < 5$, $5 \leq AHI < 10$, $AHI \geq 10$). In addition, the diagnostic performance for each cutoff ($AHI = 1$, 5, and 10 e/h) was assessed by means of sensitivity, specificity, positive predictive value, negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR-), accuracy (Acc), and area under the receiver operating characteristic curve (AUC).

Study Validation

Sixty percent of subjects were randomly selected from the initial database at the University of Chicago and were included in

the training set ($n = 589$). The remaining 3,602 subjects from all 13 centers composed the test set. The training set was used for three purposes: (1) selection of relevant and nonredundant features, (2) optimization of MLP neural network parameters (N_H and α), and (3) training the specific MLP model to be tested. Each of them was differently implemented for the sake of results generalization. Feature selection was conducted along with a sampling with replacement procedure (bootstrap) repeated 1,000 times to obtain a robust optimum subset (25). MLP parameters were optimized by computing Cohen’s κ for a representative range of (N_H , α) pairs. This measurement of agreement for optimization was chosen to prioritize the correct group classification over the exact AHI estimation. Each κ value was obtained after a leave-one-out cross-validation procedure (28). Finally, the specific MLP model was derived from the entire training set by the use of the optimum (N_H , α) pair previously found. The test set was only used to estimate the diagnostic performance of our candidate algorithm in a large group of previously untested recordings.

Results

Optimum Feature Subset

The number of times that each extracted feature was selected over the 1,000 bootstrap repetitions exhibited substantial redundancy (Figure E2), such that only two

features were retained as the optimum set for the MLP training model: ODI3 (selected 995 times) and $Mf3_{BOI}$ (selected 621 times).

Derivation of the MLP Model

Optimization of N_H and α was conducted by training MLP models through a leave-one-out cross-validation procedure and was applied to the optimum selected features from the training set. As shown in Figure E3, the Cohen’s κ obtained for each (N_H , α) pair evaluated, representing an average of 10 repetitions, was conducted to minimize the effect of the MLP random initialization, with ultimately optimum parameters being set to $\alpha = 7$ and $N_H = 6$ for the sake of model complexity, because no increase in the third decimal of the κ value was found with increased iterations. A specific MLP model was thus derived by training a new neural network using these parameters and the entire training group, that is, without the leave-one-out cross-validation procedure.

Agreement with NPSG AHI

Figure 1 displays the Bland-Altman plot comparing the NPSG AHI of the subjects from the test group with our corresponding AHI estimation. It also shows the ICC between these two measurements. As can be observed, a low mean positive difference (slight AHI overestimation) is reached (0.230), with a 95% confidence interval of $[-13.80, 14.26]$. In addition, a high ICC is reached (0.785). A scatter plot comparing NPSG-derived AHI against our AHI

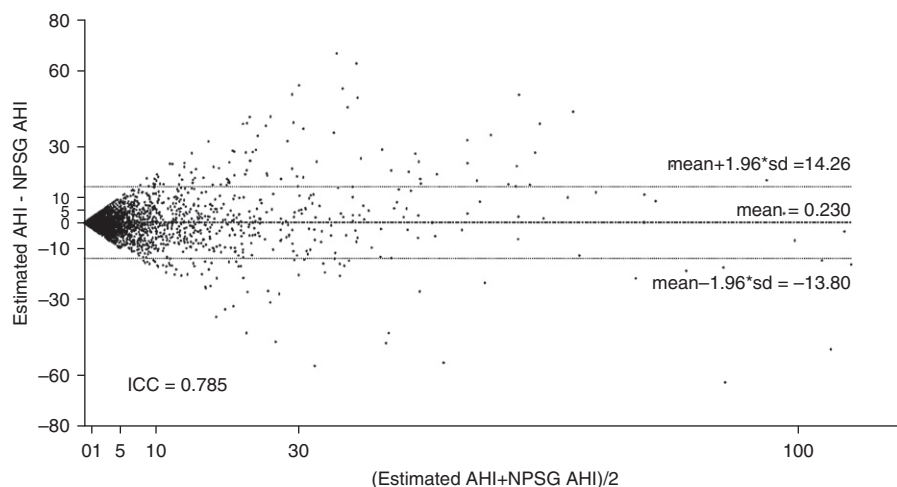


Figure 1. Bland-Altman plot comparing NPSG AHI with the estimated AHI, using the neural network-based algorithm. AHI = apnea-hypopnea index; ICC = intraclass correlation coefficient; NPSG = nocturnal polysomnography.

estimation can be found in Figure E4. In addition, the lack of any significant differences across either subject ages or body mass index are also provided in Tables E1 and E2.

Diagnostic Performance

Table 3 shows the four-class confusion matrix comparing the classification derived from the NPSG AHI with the classification of our AHI estimation in the test set. General accuracy over the four classes (sum of the main diagonal of the matrix) was 54.7%. Accordingly, Cohen's κ was 0.348. Table 4 displays sensitivity, specificity, Acc, positive predictive value, NPV, LR+, and LR- for the AHI = 1, 5, and 10 e/h cutoffs, derived from the confusion matrix. Our AHI estimation showed an increasing degree of diagnostic ability as the cutoff increased. The highest Acc (90.2%) was reached when AHI = 10 e/h was defined as the OSA threshold, and was accompanied by specificity, 94.1%; NPV, 94.3%; and LR+, 11.64. Figure 2 displays the receiver operating characteristic curves for the three AHI cutoffs for the diagnosis of OSA. The corresponding AUC for each diagnostic cutoff was high, with improving values as the AHI cutoff increased. Accordingly, the maximum AUC value (0.913) was reached for AHI = 10 e/h.

Discussion

This study shows that neural network-based analytic approaches of nSpO₂ recordings are capable of reliably identifying children with OSA, using any of three commonly used AHI clinical cutoff values, while also allowing for accurate estimates of NPSG-

derived AHI. Considering the extensive derivation and validation precautions undertaken in this work, the current findings should enable their expanded and widespread use in clinical settings, particularly when pediatric sleep laboratory facilities are not readily available.

Before we discuss some of the clinical implications of our findings, several methodological considerations should be mentioned. In our aim to maximize the diagnostic ability of nSpO₂, a careful automatic signal analysis was conducted, because sampling rates, resolution, and averaging time settings may impose significant influence on the collected data, which could affect time response and reproducibility of nSpO₂ (29). Thirteen medical centers were involved in this study, each of them using its own oximetry data acquisition settings. However, we *a posteriori* proceeded to standardize the nSpO₂ recordings to 25 Hz as the sampling rate recommended by the American Academy of Sleep Medicine (18). In addition, two decimal places were set *post hoc* for all SpO₂ measures, which avoided intercenter inequalities when conducting time domain analyses (14). These precautionary steps enabled expanded application of our analyses to all recordings, regardless of the center where they had been acquired. Other steps aimed at enhancing both the validation of our methodology and its generalizability were also implemented, such that 86% of samples were used only for testing purposes and included subjects from the 13 different centers involved in the study, but only an untested proportion of the database from which the neural network algorithm was originally derived. Moreover, the three

Table 4. Diagnostic Performance of Neural Network Estimated Apnea–Hypopnea Index for the Cutoffs 1, 5, and 10 Events per Hour

	AHI Cutoff		
	1 e/h	5 e/h	10 e/h
Se, %	84.0	68.2	68.7
Sp, %	53.2	87.2	94.1
PPV, %	81.6	68.6	67.7
NPV, %	57.3	87.0	94.3
LR+	1.79	5.32	11.64
LR-	0.30	0.36	0.33
Acc, %	75.2	81.7	90.2

Definition of abbreviations: Acc = accuracy; AHI = apnea–hypopnea index; e/h = events per hour; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; NPV = negative predictive value; PPV = positive predictive value; Se = sensitivity; Sp = specificity.

objectives pursued with the analysis of the training set (feature selection, MLP parameter optimization, and MLP model derivation) were reached with three different validation approaches (bootstrapping, leave-one-out, and hold-out, respectively), thereby enhancing the robustness of the approach. Of note, a high degree of redundancy in the nSpO₂ information provided by the features commonly used to evaluate OSA in adults is present when applied to children. This may be due to the more restrictive rules that are commonly used to diagnose pediatric OSA, which led to many of the adult-related features being suboptimal for pediatrics, and should prompt future exploration of nSpO₂ features that enable improved specific ability to analyze nSpO₂ recordings in children. Nevertheless, our MLP model showed high diagnostic ability even when using information from only two features (ODI3 and Mf3_{BOI}).

Overall, the accuracy of our proposed approach ranged between 75 and 90% for AHI estimates of 1 to 10 e/h, respectively, along with the anticipated shift from higher sensitivity at lower AHI to higher specificity at the higher AHI cutoff (Table 4). Thus, the maximum benefit of our automated methodology in terms of simplicity and screening capability can be achieved when using the more widely used and clinically relevant cutoff for OSA (i.e., ≥ 5 e/h), which corresponds to the point of upward inflection in morbidity risks associated with

Table 3. Four-Class Confusion Matrix Showing Classification Agreement of Neural Network–based AHI Estimate and Nocturnal Polysomnography–derived AHI

	MLP AHI			
	Estimated Severity			
	AHI < 1	1 \leq AHI < 5	5 \leq AHI < 10	10 \leq AHI
NPSG AHI				
AHI < 1	551	427	44	14
1 \leq AHI < 5	356	892	206	63
5 \leq AHI < 10	51	193	149	104
10 \leq AHI	3	87	83	379

Definition of abbreviations: AHI = apnea–hypopnea index; MLP = multilayer perceptron; NPSG = nocturnal polysomnography.

Bold indicates patients correctly classified into their severity degree category.

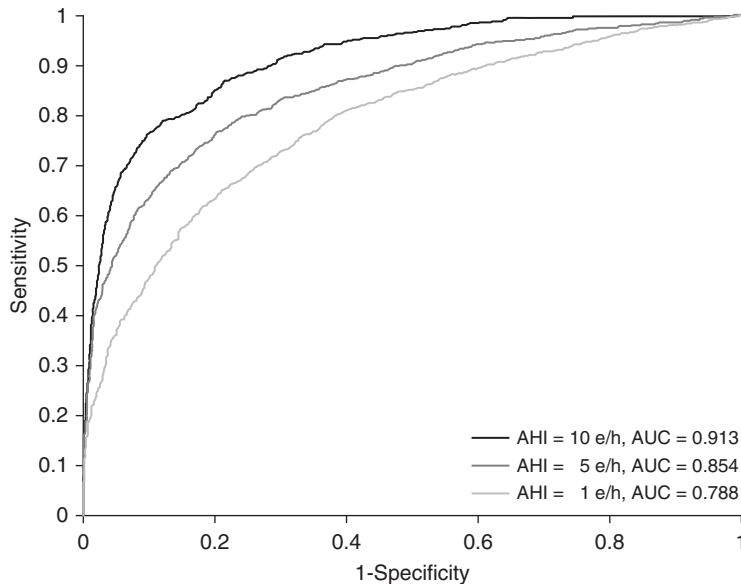


Figure 2. Receiver-operating characteristic curves of neural network–based AHI estimate in the test set for 1, 5, and 10 events per hour (e/h). AHI = apnea–hypopnea index; AUC = area under the curve.

sleep-disordered breathing in children (30–33). In other words, the accuracy of the surrogate diagnostic method, using the $n\text{SpO}_2$ -based approach described here, is increasingly robust as the severity of NPSG-based OSA increases, such that we can confidently both confirm and discard cases that would or would not fulfill OSA criteria. According to the confusion matrix shown in Table 3, 94.4% of children predicted to have an estimated AHI less than 1 e/h would have an actual NPSG AHI no greater than 5 e/h. In addition, 94.4% of children predicted to have an estimated AHI ≥ 5 e/h would actually have an NPSG AHI greater than 1 e/h. A diagnostic protocol could then be derived from these results as follows: (1) if our neural network model predicts an estimated AHI less than 1 e/h, OSA would be discarded because most patients (94.4%) probably will have an NPSG AHI value less than 5 e/h; (2) if our neural network model predicts an estimated AHI ≥ 5 e/h, consider referring for treatment because most patients (94.4%) probably will have an NPSG AHI value ≥ 1 e/h in the context of snoring children; and (3) if our neural network model predicts an estimated AHI ≥ 1 e/h but less than 5 e/h, we would recommend referral for NPSG, because doubts arise about the actual definitive diagnosis. Such a protocol would potentially reduce the need for 52.6% of NPSGs, while indicating treatment for only 5.7% of the snoring

children with an actual NPSG AHI < 1 e/h; in addition, this approach would not lead to treatment for only 5.5% of the children with an actual NPSG AHI ≥ 5 e/h. In the latter case, persistence of clinical symptoms for 2–3 months should then prompt referral for NPSG.

Notwithstanding such considerations, the relatively small proportion of children who would be potentially missed using $n\text{SpO}_2$ might be minimized by repeating the oximetry-based test within a short period of time (weeks to a few months) if the child's symptoms persist. Alternatively, highly symptomatic children identified through the use of existing alternative tools such as the sleep medical record (34), and whose $n\text{SpO}_2$ assessments were negative using the methodology proposed above, could then be referred for NPSG, thereby markedly reducing the overall need for more costly testing in resource-constrained settings. Because a large number of sleep laboratories appear to use the AHI cutoff of 1 e/h for interpretation of NPSG (1), the elevated sensitivity and lower specificity features that emerged in our $n\text{SpO}_2$ approaches to predict AHI ≥ 1 e/h would increase the number of positive diagnoses by the proposed methodology, while reducing the rate of false negative cases. Considering the broad variance in AHI cutoffs applied for therapeutic decision making in the field, such an approach may be preferred by those clinicians who are

more inclined to advocate treatment such as adenotonsillectomy at the low end of abnormal AHI levels. Thus, similar to the options offered by NPSG to use the AHI as one of the major parameters guiding clinical management decisions, the $n\text{SpO}_2$ neural network–derived AHI would offer similar options, that is, different AHI cutoffs at a fraction of the cost and effort involved in NPSG testing.

Multiple alternatives to NPSG have been examined over the years in an effort to improve the accessibility of habitually snoring symptomatic children to a timely diagnosis and treatment. To this end, a large number of approaches have been advocated ranging from questionnaires to unattended NPSG at home. The overall consensus from such efforts indicates that clinical history and physical examination or questionnaire-based instruments lack the required diagnostic accuracy, precluding their use as a routine diagnostic tool for OSA (35, 36). However, respiratory polygraphy is becoming increasingly accepted as a surrogate diagnostic approach in adults and children, even if its accuracy is reduced at the low end of OSA severity (9, 37). In the present study, our findings in a large and diverse clinical cohort indicate that automated analysis of single-channel $n\text{SpO}_2$ is at least as accurate as respiratory polygraphy in the diagnosis of OSA in children, further confirming the adequacy and potential limitations of such approaches as reported in other studies (9, 24).

The obvious simplicity of $n\text{SpO}_2$ recordings has prompted efforts to evaluate their potential diagnostic properties; to date, commonly used oximetric indices such as the number of desaturations, clusters of events within a particular timeframe, and percentage of time spent with a SpO_2 below a particular threshold have been examined (9, 38–43). However, such conventional $n\text{SpO}_2$ analytical approaches have not achieved the desirable diagnostic performance for detecting OSA in children. Conversely, as shown by Garde and colleagues (14), more sophisticated mathematical analyses of the oximeter signal achieved 88.4% sensitivity and 83.6% specificity for an AHI cutoff of 5 e/h, albeit in a small single-center cohort. Notwithstanding these considerations, the methodological approach presented herein aims to confirm or discard the presence of OSA, that is, to perform binary

classification. Although three commonly used AHI cutoff points in clinical practice were assessed to classify the absence or presence of OSA (1, 5, and 10 e/h), it may be useful to couple the current approach to a pattern recognition methodology that would aim at not only classifying high pretest symptomatic pediatric patients into the four common categories of severity (no disease and mild, moderate, and severe

disease) as performed here, but further enable estimates of the actual PSG-derived AHI of each patient.

Conclusions

This study provides extensive validation of the satisfactory diagnostic performance of automated analysis of nocturnal single-channel oximetry as a low-cost alternative to standard NPSG in the context of

childhood OSA. Therefore, the current findings indicate that automated processing of the nSpO₂ signal provides an accurate and widely implementable diagnostic tool for childhood OSA, particularly in resource-constrained environments. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

- Marcus CL, Brooks LJ, Draper KA, Gozal D, Halbower AC, Jones J, Schechter MS, Ward SD, Sheldon SH, Shiffman RN, *et al.*; American Academy of Pediatrics. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics* 2012;130:e714–e755.
- Gileles-Hillel A, Philby MF, Lapping-Carr G. Insights into selected aspects of pediatric sleep medicine. *Am J Respir Crit Care Med* 2015; 191:1459–1461.
- Tarasjuk A, Simon T, Tal A, Reuveni H. Adenotonsillectomy in children with obstructive sleep apnea syndrome reduces health care utilization. *Pediatrics* 2004;113:351–356.
- Mukherjee S, Patel SR, Kales SN, Ayas NT, Strohl KP, Gozal D, Malhotra A; American Thoracic Society *Ad Hoc* Committee on Healthy Sleep. An official American Thoracic Society statement: the importance of healthy sleep: recommendations and future priorities. *Am J Respir Crit Care Med* 2015;191:1450–1458.
- Alonso-Álvarez ML, Canet T, Cubell-Alarco M, Estivill E, Fernández-Julián E, Gozal D, Jurado-Luque MJ, Lluch-Roselló A, Martínez-Pérez F, Merino-Andreu M, *et al.*; Sociedad Española de Sueño; Área de Sueño de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR). Consensus document on sleep apnea-hypopnea syndrome in children (full version) [article in Spanish]. *Arch Bronconeumol* 2011; 47(Suppl 5):1–18.
- Kaditis AG, Alonso Álvarez ML, Boudewyns A, Alexopoulos EI, Ersu R, Joosten K, Larramona H, Miano S, Narang I, Trang H, *et al.* Obstructive sleep disordered breathing in 2- to 18-year-old children: diagnosis and management. *Eur Respir J* 2016;47:69–94.
- Weatherly RA, Mai EF, Ruzicka DL, Chervin RD. Identification and evaluation of obstructive sleep apnea prior to adenotonsillectomy in children: a survey of practice patterns. *Sleep Med* 2003;4:297–307.
- Kaditis A, Kheirandish-Gozal L, Gozal D. Algorithm for the diagnosis and treatment of pediatric OSA: a proposal of two pediatric sleep centers. *Sleep Med* 2012;13:217–227.
- Alonso-Álvarez ML, Terán-Santos J, Ordax Carbajo E, Cordero-Guevara JA, Navazo-Egúia AI, Kheirandish-Gozal L, Gozal D. Reliability of home respiratory polygraphy for the diagnosis of sleep apnea in children. *Chest* 2015;147:1020–1028.
- Kaditis A, Kheirandish-Gozal L, Gozal D. Pediatric OSAS: oximetry can provide answers when polysomnography is not available. *Sleep Med Rev* 2016;27:96–105.
- Brouillette RT, Morielli A, Leimanis A, Waters KA, Luciano R, Ducharme FM. Nocturnal pulse oximetry as an abbreviated testing modality for pediatric obstructive sleep apnea. *Pediatrics* 2000;105:405–412.
- Ayas NT, Drager LF, Morrell MJ, Polotsky VY. Update in sleep-disordered breathing 2016. *Am J Respir Crit Care Med* 2017;195: 1561–1566.
- Tsai CM, Kang CH, Su MC, Lin HC, Huang EY, Chen CC, Hung JC, Niu CK, Liao DL, Yu HR. Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children. *Int J Pediatr Otorhinolaryngol* 2013;77:1286–1290.
- Garde A, Dehkordi P, Karlen W, Wensley D, Ansermino JM, Dumont GA. Development of a screening tool for sleep disordered breathing in children using the phone Oximeter™. *PLoS One* 2014;9: e112959.
- Gutiérrez-Tobal GC, Kheirandish-Gozal L, Álvarez D, Crespo A, Philby MF, Mohammadi M, Del Campo F, Gozal D, Hornero R. Analysis and classification of oximetry recordings to predict obstructive sleep apnea severity in children. *Conf Proc IEEE Eng Med Biol Soc* 2015;2015: 4540–4543.
- Álvarez D, Kheirandish-Gozal L, Gutiérrez-Tobal GC, Crespo A, Philby MF, Mohammadi M, Del Campo F, Gozal D, Hornero R. Automated analysis of nocturnal oximetry as screening tool for childhood obstructive sleep apnea-hypopnea syndrome. *Conf Proc IEEE Eng Med Biol Soc* 2015;2015:2800–2803.
- Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, Marcus CL, Mehra R, Parthasarathy S, Quan SF, *et al.*; American Academy of Sleep Medicine; Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. *J Clin Sleep Med* 2012;8:597–619.
- Iber C, Ancoli-Israel S, Chesson A, Quan SF. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, Vol. 1. Westchester, IL: American Academy of Sleep Medicine; 2007.
- Magalang UJ, Dmochowski J, Veeramachaneni S, Draw A, Mador MJ, El-Solh A, Grant BJ. Prediction of the apnea-hypopnea index from overnight pulse oximetry. *Chest* 2003;124:1694–1701.
- Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of biological signals. *Phys Rev E Stat Nonlin Soft Matter Phys* 2005;71:021906.
- Álvarez D, Hornero R, Marcos JV, Wessel N, Penzel T, Del Campo F. Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis. *Int J Neural Syst* 2013;23:1350020.
- Álvarez D, Hornero R, García M, del Campo F, Zamarrón C. Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive sleep apnea detection by means of central tendency measure. *Artif Intell Med* 2007;41:13–24.
- Gutiérrez-Tobal GC, Hornero R, Álvarez D, Marcos JV, del Campo F. Linear and nonlinear analysis of airflow recordings to help in sleep apnea-hypopnea syndrome diagnosis. *Physiol Meas* 2012;33:1261–1275.
- Álvarez D, Alonso-Álvarez ML, Gutiérrez-Tobal GC, Crespo A, Kheirandish-Gozal L, Hornero R, Gozal D, Terán-Santos J, Del Campo F. Automated screening of children with obstructive sleep apnea using nocturnal oximetry: an alternative to respiratory polygraphy in unattended settings. *J Clin Sleep Med* 2017;13:693–702.
- Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003;3:1157–1182.
- Yu L, Liu H. Feature selection for high-dimensional data: a fast correlation-based filter solution. Presented at the Proceedings of the 20th International Conference on Machine Learning (ICML-2003). August 21–24, 2003, Washington, DC; pp. 856–863.
- Bishop CM. Neural networks for pattern recognition. New York: Oxford University Press; 1995.
- Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
- Böhning N, Schultheiss B, Eilers S, Penzel T, Böhning W, Schmittendorf E. Comparability of pulse oximeters used in sleep medicine for the screening of OSA. *Physiol Meas* 2010;31:875–888.
- Gozal D, Kheirandish-Gozal L. Obesity and excessive daytime sleepiness in prepubertal children with obstructive sleep apnea. *Pediatrics* 2009;123:13–18.

31. Bhattacharjee R, Kim J, Alotaibi WH, Kheirandish-Gozal L, Capdevila OS, Gozal D. Endothelial dysfunction in children without hypertension: potential contributions of obesity and obstructive sleep apnea. *Chest* 2012;141:682–691.
32. Gozal D, Kheirandish-Gozal L, Bhattacharjee R, Spruyt K. Neurocognitive and endothelial dysfunction in children with obstructive sleep apnea. *Pediatrics* 2010;126:e1161–e1167.
33. Hunter SJ, Gozal D, Smith DL, Philby MF, Kaylegian J, Kheirandish-Gozal L. Effect of sleep-disordered breathing severity on cognitive performance measures in a large community cohort of young school-aged children. *Am J Respir Crit Care Med* 2016;194:739–747.
34. Villa MP, Pietropaoli N, Supino MC, Vitelli O, Rabasco J, Evangelisti M, Del Pozzo M, Kaditis AG. Diagnosis of pediatric obstructive sleep apnea syndrome in settings with limited resources. *JAMA Otolaryngol Head Neck Surg* 2015;141:990–996.
35. Spruyt K, Gozal D. Pediatric sleep questionnaires as diagnostic or epidemiological tools: a review of currently available instruments. *Sleep Med Rev* 2011;15:19–32.
36. Spruyt K, Gozal D. Screening of pediatric sleep-disordered breathing: a proposed unbiased discriminative set of questions using clinical severity scales. *Chest* 2012;142:1508–1515.
37. Collop NA, Anderson WM, Boehlecke B, Claman D, Goldberg R, Gottlieb DJ, Hudgel D, Sateia M, Schwab R; Portable Monitoring Task Force of the American Academy of Sleep Medicine. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. *J Clin Sleep Med* 2007;3:737–747.
38. Kirk VG, Bohn SG, Flemons WW, Remmers JE. Comparison of home oximetry monitoring with laboratory polysomnography in children. *Chest* 2003;124:1702–1708.
39. Pavone M, Cutrera R, Verrillo E, Salerno T, Soldini S, Brouillette RT. Night-to-night consistency of at-home nocturnal pulse oximetry testing for obstructive sleep apnea in children. *Pediatr Pulmonol* 2013;48:754–760.
40. Sahadan DZ, Davey MJ, Horne RSC, Nixon GM. Improving detection of obstructive sleep apnoea by overnight oximetry in children using pulse rate parameters. *Sleep Breath* 2015;19:1409–1414.
41. Nixon GM, Davey MJ, Weichard AJ, Horne RS. Oximetry for suspected obstructive sleep apnea: does removal of awake data affect the result? *Pediatr Pulmonol* 2016;51:1409–1413.
42. Garde A, Dehkordi P, Wensley D, Ansermino JM, Dumont GA. Pulse oximetry recorded from the Phone Oximeter for detection of obstructive sleep apnea events with and without oxygen desaturation in children. *Conf Proc IEEE Eng Med Biol Soc* 2015; 2015:7692–7695.
43. Van Eyck A, Lambrechts C, Vanheeswijck L, Van Hoorenbeeck K, Haentjens D, Boudewyns A, De Winter BY, Van Gaal L, De Backer W, Verhulst SL. The role of nocturnal pulse oximetry in the screening for obstructive sleep apnea in obese children and adolescents. *Sleep Med* 2015;16:1409–1412.