Nocturnal Oximetry–based Evaluation of Habitually Snoring Children

Roberto Hormero1, Leila Kheirandish-Gozal2, Gonzalo C. Gutiérrez-Tobal1, Mona F. Philby2, María Luz Alonso-Álvarez3, Daniel Álvarez1,4, Ehab A. Dayyat5, Zhei Xi6, Yu-Shu Huang7, Maximiliano Tamae Kakazu8, Albert M. Li9, Annelies Van Eyck10,11, Pablo E. Brockmann12, Zarmina Ehsan13a, Narong Simakajoomboon13, Athanasios G. Kaditis14, Fernando Vaquerizo-Villar1, Andrea Crespo Sedano1, Oscar Sans Capdevila15, Magnus von Lukowicz16, Joaquín Terán-Santos8, Félix Del Campo1,4,14, Christian F. Poets16, Rosario Ferreira17, Katalina Bertran15, Yamei Zhang6, John Schuen6, Stijn Verhulst10,11, 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 Sueno, 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; 8Department of Pediatrics, Prince of Wales Hospital, Chinese University of Hong Kong, Hong Kong, China; 9Laboratory of Experimental Medicine and Pediatrics and 10Department of Pediatrics, University of Antwerp and Antwerp University Hospital, Antwerp, Belgium; 11Sleep Medicine Center, Department of Pediatric Cardiology and Pulmonology, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; 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 Tubingen, Tubingen, Germany; and 17Pediatric Respiratory Unit, 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

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 (nSpO2) 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 nSpO2 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 nSpO2 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 nSpO2 recordings (15). In the present study, we expanded the derivation and validation of nSpO2 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 nSpO2 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 nSpO2 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 nSpO2 Signal Analysis
Four automatic signal analysis stages were implemented to obtain useful information from nSpO2. A scheme of the automatic methodology applied to the nSpO2 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 nSpO2 using one or several sampling rates ranging from 1 to 500 Hz, as well as different decimal resolution. Hence, all the nSpO2 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...
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 1. Demographic and Clinical Data for Each Participating Site

<table>
<thead>
<tr>
<th>Center</th>
<th>Participants (n)</th>
<th>Age (yr) (Mean ± SD)</th>
<th>Male (%)</th>
<th>BMI (kg/m²) (Mean ± SD)</th>
<th>AHI (per Hour TST) (Mean ± SD)</th>
<th>OSA for AHI = 1 e/h (%)</th>
<th>OSA for AHI = 5 e/h (%)</th>
<th>OSA for AHI = 10 e/h (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UoC</td>
<td>981</td>
<td>6.1 ± 3.4</td>
<td>61.4</td>
<td>19.7 ± 7.3</td>
<td>9.3 ± 17.2</td>
<td>82.2</td>
<td>41.3</td>
<td>23.3</td>
</tr>
<tr>
<td>UoTn</td>
<td>611</td>
<td>7.2 ± 4.6</td>
<td>54.6</td>
<td>23.3 ± 10.1</td>
<td>5.8 ± 11.3</td>
<td>68.1</td>
<td>29.6</td>
<td>17.7</td>
</tr>
<tr>
<td>HUBU</td>
<td>578</td>
<td>4.1 ± 2.2</td>
<td>61.8</td>
<td>17.1 ± 4.2</td>
<td>5.9 ± 11.3</td>
<td>64.5</td>
<td>26.3</td>
<td>15.2</td>
</tr>
<tr>
<td>BCH</td>
<td>558</td>
<td>6.3 ± 5.3</td>
<td>66.3</td>
<td>17.8 ± 3.7</td>
<td>5.8 ± 11.7</td>
<td>65.1</td>
<td>27.4</td>
<td>17.0</td>
</tr>
<tr>
<td>MSU</td>
<td>499</td>
<td>6.5 ± 5.0</td>
<td>55.5</td>
<td>17.8 ± 11.2</td>
<td>6.2 ± 9.3</td>
<td>85.8</td>
<td>22.0</td>
<td>14.8</td>
</tr>
<tr>
<td>CGMH</td>
<td>283</td>
<td>9.9 ± 3.2</td>
<td>72.4</td>
<td>19.5 ± 4.6</td>
<td>4.3 ± 10.0</td>
<td>72.4</td>
<td>21.9</td>
<td>8.1</td>
</tr>
<tr>
<td>UoHK</td>
<td>202</td>
<td>10.0 ± 2.4</td>
<td>62.9</td>
<td>18.7 ± 4.6</td>
<td>4.9 ± 7.5</td>
<td>70.3</td>
<td>26.2</td>
<td>10.4</td>
</tr>
<tr>
<td>Pucc</td>
<td>183</td>
<td>5.4 ± 4.8</td>
<td>52.5</td>
<td>17.8 ± 4.2</td>
<td>3.7 ± 9.1</td>
<td>60.1</td>
<td>18.6</td>
<td>7.6</td>
</tr>
<tr>
<td>UoFa</td>
<td>130</td>
<td>11.7 ± 3.1</td>
<td>37.7</td>
<td>30.3 ± 5.7*</td>
<td>3.2 ± 7.1</td>
<td>63.1</td>
<td>22.3</td>
<td>10.0</td>
</tr>
<tr>
<td>SDCH</td>
<td>60</td>
<td>8.4 ± 4.8</td>
<td>58.3</td>
<td>19.5 ± 5.2</td>
<td>4.2 ± 6.1</td>
<td>76.7</td>
<td>25.0</td>
<td>11.6</td>
</tr>
<tr>
<td>ASCH</td>
<td>51</td>
<td>7.0 ± 3.4</td>
<td>66.7</td>
<td>20.6 ± 6.4</td>
<td>10.6 ± 13.8</td>
<td>90.2</td>
<td>54.9</td>
<td>37.2</td>
</tr>
<tr>
<td>UoTu</td>
<td>36</td>
<td>10.4 ± 3.5</td>
<td>61.1</td>
<td>21.0 ± 8.0</td>
<td>6.9 ± 12.9</td>
<td>72.2</td>
<td>27.8</td>
<td>16.7</td>
</tr>
<tr>
<td>HSM</td>
<td>19</td>
<td>6.5 ± 3.8</td>
<td>47.4</td>
<td>19.1 ± 6.8</td>
<td>11.0 ± 15.2</td>
<td>73.7</td>
<td>47.4</td>
<td>36.8</td>
</tr>
<tr>
<td>All</td>
<td>4,191</td>
<td>6.7 ± 4.4</td>
<td>60.0</td>
<td>20.0 ± 7.0</td>
<td>6.4 ± 12.5</td>
<td>72.9</td>
<td>29.6</td>
<td>16.8</td>
</tr>
</tbody>
</table>

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); SDCH = San Joan de Deu Children’s Hospital (Spain); TST = total sleep time; UoC = University of Chicago (USA); UoHK = University of Hong Kong (Hong Kong, China); UoTn = University of Tennessee (USA); UoTu = University of Tuebingen (Germany).

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

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

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time domain</td>
<td>First, second, third, and fourth statistical moment of a time series</td>
</tr>
<tr>
<td>Mf1–Mf4</td>
<td>CTM</td>
</tr>
<tr>
<td>MA</td>
<td>LzC</td>
</tr>
<tr>
<td>MA</td>
<td>SampEn</td>
</tr>
<tr>
<td>MA</td>
<td>Odi3</td>
</tr>
<tr>
<td>Frequency domain</td>
<td>Full-spectrum amplitude maximum</td>
</tr>
<tr>
<td>Mf1–Mf4</td>
<td>MA</td>
</tr>
<tr>
<td>MF</td>
<td>SpecEn</td>
</tr>
<tr>
<td>WD</td>
<td>Wooters’ distance to estimate the statistical distance of the full spectrum</td>
</tr>
<tr>
<td>MA_BOI</td>
<td>Spectrum amplitude maximum from the band of interest</td>
</tr>
<tr>
<td>MA_BOI</td>
<td>Mf1</td>
</tr>
<tr>
<td>Mf1–Mf4</td>
<td>Mf1–Mf4_BOI</td>
</tr>
</tbody>
</table>

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 \(\alpha\) 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 significance 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 \(\kappa\). 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 \(\alpha\)), 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 \(\kappa\) for a representative range of \((N_h, \alpha)\) pairs. This measurement of agreement for optimization was chosen to prioritize the correct group classification over the exact AHI estimation. Each \(\kappa\) 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 \((N_h, \alpha)\) 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 Mf3BOI (selected 621 times).

### Derivation of the MLP Model

Optimization of \(N_h\) and \(\alpha\) 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 \(\kappa\) obtained for each \((N_h, \alpha)\) 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 \(\kappa\) 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 \(-13.80\) to \(14.26\). 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.](image-url)
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 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 Mf3BOI).

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

<table>
<thead>
<tr>
<th>AHI Cutoff</th>
<th>1 e/h</th>
<th>5 e/h</th>
<th>10 e/h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se, %</td>
<td>84.0</td>
<td>68.2</td>
<td>68.7</td>
</tr>
<tr>
<td>Sp, %</td>
<td>53.2</td>
<td>87.2</td>
<td>94.1</td>
</tr>
<tr>
<td>PPV, %</td>
<td>81.6</td>
<td>68.6</td>
<td>67.7</td>
</tr>
<tr>
<td>NPV, %</td>
<td>57.3</td>
<td>87.0</td>
<td>94.3</td>
</tr>
<tr>
<td>LR+, %</td>
<td>1.79</td>
<td>5.32</td>
<td>11.64</td>
</tr>
<tr>
<td>LR−, %</td>
<td>0.30</td>
<td>0.36</td>
<td>0.33</td>
</tr>
<tr>
<td>Acc, %</td>
<td>75.2</td>
<td>81.7</td>
<td>90.2</td>
</tr>
</tbody>
</table>

*Definition of abbreviations: Se = sensitivity; Sp = specificity.*

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

<table>
<thead>
<tr>
<th>Estimated Severity</th>
<th>AHI &lt;1</th>
<th>1 ≤ AHI &lt; 5</th>
<th>5 ≤ AHI &lt; 10</th>
<th>10 ≤ AHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPSG AHI AHI &lt; 1</td>
<td>551</td>
<td>427</td>
<td>44</td>
<td>14</td>
</tr>
<tr>
<td>1 ≤ AHI &lt; 5</td>
<td>356</td>
<td>892</td>
<td>206</td>
<td>63</td>
</tr>
<tr>
<td>5 ≤ AHI &lt; 10</td>
<td>51</td>
<td>193</td>
<td>149</td>
<td>104</td>
</tr>
<tr>
<td>10 ≤ AHI</td>
<td>3</td>
<td>87</td>
<td>83</td>
<td>379</td>
</tr>
</tbody>
</table>

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

**Bold indicates patients correctly classified into their severity degree category.**
sleep-disordered breathing in children (30–33). In other words, the accuracy of the surrogate diagnostic method, using the nSpO2-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 nSpO2 might be minimized by repeating the oximeter-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 nSpO2 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 nSpO2 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 nSpO2 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 nSpO2 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 nSpO2 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 SpO2 below a particular threshold have been examined (9, 38–43). However, such conventional nSpO2 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
References


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


