AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

Nocturnal Oximetry-Based Evaluation of Habitually Snoring Children.

Roberto Hornero, PhD, [1] Leila Kheirandish-Gozal, MD, MSc,[2] Gonzalo C. Gutiérrez-Tobal, PhD,[1] Mona F. Philby, MD,[2] María Luz Alonso-Álvarez, MD,[3] Daniel Álvarez, PhD,[1,4] Ehab A. Dayyat, MD, [5] Zhifei Xu, MD,[6] Yu-Shu Huang, MD,[7] Maximiliano Tamae Kakazu, MD,[8] Albert M. Li, MD,[9] Annelies Van Eyck, PhD,[10] Pablo E. Brockmann, MD,[11] Zarmina Ehsan, MD*, [12] Narong Simakajornboon, MD,[12] Athanasios G. Kaditis, MD,[13] Fernando Vaquerizo-Villar, MSc,[1] Andrea Crespo Sedano, MD,[4] Oscar Sans Capdevila, MD,[14] Magnus von Lukowicz, MD, [15] Joaquín Terán-Santos, MD,PhD,[3] Félix Del Campo, MD, PhD,[1,4] Christian F. Poets, MD,[15] Rosario Ferreira, MD,[16] Katalina Bertran, MD, [14] Yamei Zhang, MD,[6] John Schuen, MD,[8] Stijn Verhulst, MD,PhD,[10] David Gozal, MD, MBA.[2]

[1] Biomedical Engineering Group, University of Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain.

[2]Section of Sleep Medicine, Dept. of Pediatrics, Pritzker School of Medicine, Biological Sciences Division, The University of Chicago, Chicago, IL, 60637, USA

[3]Unidad Multidisciplinar de Sueño, CIBER Respiratorio, Hospital Universitario de Burgos, Burgos, Spain.

[4] Sleep-Ventilation Unit, Pneumology Service, Río Hortega University Hospital, c/ Dulzaina 2, 47012 Valladolid, Spain.

[5] Division of Child Neurology, Dep of Pediatrics, LeBonheur Children's Hospital, University of Tennessee Health Science Center, School of Medicine, Memphis, TN 38015, USA

[6] Sleep Unit, Beijing Children's Hospital, Capital Medical University, Beijing, People's Republic of China

[7] Department of Child Psychiatry and Sleep Center, Chang Gung Memorial Hospital and University,Taoyuan, Taiwan

[8] Spectrum Health, Michigan State University, Grand Rapids, Michigan, USA

[9] Department of Pediatrics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, China

[10] Laboratory of Experimental Medicine and Pediatrics and Department of Pediatrics, University of Antwerp and Antwerp University Hospital, Antwerp, Belgium

[11] Sleep Medicine Center, Department of Pediatric Cardiology and Pulmonology, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.

[12] Division of Pulmonary and Sleep Medicine, Cincinnati Children's Medical Center

[13] Pediatric Pulmonology Unit, Sleep Disorders Laboratory, First Department of Pediatrics, National and Kap,odistrian University of Athens School of Medicine and Aghia Sophia Children's Hospital, Athens, Greece

[14] Sleep Unit, Department of Neurology; Sant Joan de Deu, Barcelona Children´s Hospital Barcelona, Spain

[15] Department of Neonatology and Sleep Unit, University of Tubingen, Tubingen, Germany

[16] Pediatric Respiratory Unit, Department of Pediatrics, Hospital de Santa Maria, Academic Medical Centre of Lisbon

*Current institutional affiliation: Pulmonology and Sleep Medicine, Children's Mercy Kansas City, Email: zehsan@cmh.edu

Running head: oximetry-based screening of childhood OSA

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

Funding. This research has been partially supported by project VA037U16 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. Kheirandish-Gozal is supported by NIH grant 1R01HL130984. MFP was supported by a Fellowship Educational grant award from the Kingdom of Saudi Arabia. D. Álvarez 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.

Conflicts of interest. No potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Corresponding author: David Gozal, MD, MBA Section of Pediatric Sleep Medicine, Department of Pediatrics, Pritzker School of Medicine, Biological Sciences Division, The University of Chicago, KCBD, Room 4100, 900 E. 57th Street, Mailbox 4, Chicago, IL, 60637. (773) 702-3360- TEL; (773) 926-0756- FAX; email: dgozal@uchicago.edu

Word Count: 3,128

Authors Contributions:

RH and LKG conceptualized the study, analyzed data, and drafted components of the manuscript.

GCGT performed analyses and contributed to manuscript editing.

MFP, MLAA, DA, EAD, ZX, YH, MTK, AML, AVE, PEB, ZE, NS, AK, AC, OSC, MVL, JT, FDC, CFP, RF, KB, YZ, JS, and SV recruited subjects and contributed cases to the database.

FVV contributed to data analyses,

DG conceptualized the study, coordinated the database, performed data analyses, drafted components and edited the manuscript.

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

What is the scientific knowledge: Obstructive sleep apnea 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 is treated without a formal diagnosis.

What this study adds: An automated neural network algorithm based on overnight oximetry recordings provides accurate identification of OSA severity among habitually snoring children with a high pre-test probability of OSA.

Abstract

Background: The vast majority of children around the world undergoing adenotonsillectomy for obstructive sleep apnea-hypopnea syndrome (OSA) are not objectively diagnosed with nocturnal polysomnography (NPSG) due to access availability and cost issues. Automated analysis of nocturnal oximetry ($nSpO_2$), which is readily and globally available, could potentially provide a reliable and convenient diagnostic approach for pediatric OSA.

Methods: De-identified $nSpO_2$ 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 $nSpO_2$ recordings from 589 patients referred for suspected OSA.

Results: The automatically estimated apnea-hypopnea index (AHI) showed high agreement with AHI from conventional polysomnography (intra-class correlation coefficient 0.785) when tested in 3,602 additional subjects. Further assessment on the widely-used AHI cut-off points 1, 5 and 10 events/hour 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 characteristics curve, respectively).

Conclusions: Neural network-based automated analyses of $nSpO_2$ recordings provide accurate identification of OSA severity among habitually snoring children with a high pre-test probability of OSA. Thus, nocturnal oximetry may enable a simple and effective diagnostic alternative to NPSG 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;

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

Introduction

Since its initial description, pediatric obstructive sleep apnea-hypopnea syndrome (OSA) has become recognized not only as a prevalent condition in children, but also has 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 healthcare utilization 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 inter-scorer 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 remarkably improve 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 very 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.

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

Methods

Patients

A total of 4,191 habitually snoring children (2,517 boys and 1,674 girls) ranging in age between 2 and 18 years who were referred for clinical suspicion of OSA and underwent NPSG composed the population under study. De-identified 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 due to habitual snoring and/or witnessed breathing pauses during sleep as reported by their parents or caretakers. The study was approved by the Ethical Review Committee of each of the 13 participating centers.

Sleep studies

All NPSG 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/hour (e/h). Table 1 shows the overall and each participating center patient demographics and OSA prevalence 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 1S in the supplemental data. The first one was a pre-processing 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 (AASM) (18). In addition, signals were all rounded to the second decimal place. Finally, artifacts were removed according to the automatic method suggested by Magalang *et al*. (19).

After pre-processing, 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 due to 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 adult patients (20-23).

Such comprehensive characterization of the $nSpO_2$ 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 (FCBF) was applied to the extracted features to evaluate the relevance and redundancy of their OSA-related information. In a first step, the FCBF algorithm conducts a relevancy analysis by comparing the information shared by each feature and a reference variable (26), i.e., 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, i.e., conducting a redundancy analysis. Features sharing more information with other higher-ranked ones than with AHI are discarded due to redundancy (26).

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

At this point of the processing, subjects under study were characterized by a vector, whose components are their corresponding values of the non-redundant extracted features. Thus, the fourth stage consisted in the training of a multi-layer perceptron (MLP) model with ability to automatically estimate AHI from these patterns. MLP is an artificial neural network 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, i.e., the number of non-redundant features. By contrast, the output is only composed of one 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

IBM SPSS Statistics version 20 software (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 non-parametric Mann-Whitney *U* test was applied to search for statistical significant differences between OSA negative and OSA positive groups. A p-value < 0.05 was considered significant.

Matlab R2016b was used to implement feature extraction, selection, and classification stages. Bland-Altman plot and intra-class 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 ≥ 10). In addition, the diagnostic performance for each cutoff (AHI = 1, 5, and 10 e/h) was assessed by means of sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR-), accuracy (Acc), and area under the receiver-operating characteristics (ROC) 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: *i*) selection of relevant and non-redundant features; *ii*) optimization of

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

MLP neural network parameters (N_H, and α); and *iii*) 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 in order 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 1000 bootstrap repetitions exhibited substantial redundancy (Supplemental Figure 2S), such that only two features were retained as the optimum set for the MLP training model: *ODI3* (995 times selected) and $Mf3_{BOI}$ (621 times selected).

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 Supplemental Figure 3S, 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, since 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, i. e., 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 95% confidence interval within [-13.80, 14.26]. In addition, high ICC is reached (0.785). A scatter plot that faces NPSG and our

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

AHI estimation can be found as the figure 4S of the supplemental data. In addition, the lack of any significant differences across either subject ages or BMI are also provided in Supplemental Tables 1 and 2).

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 Se, Sp, Acc, PPV, NPV, LR+, and LR- for the AHI = 1, 5, and 10 e/h cutoffs, derived from the confusion matrix. Our AHI estimation showed 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 Sp: 94.1%; NPV: 94.3%, and LR+: 11.64. Figure 2 displays the ROC curves for the three AHI cutoffs for 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 $nSPO_2$ recordings are capable of reliably identifying children with OSA using any of 3 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 utilization 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 $nSpO_2$, a careful automatic signal analysis was conducted, since sampling rates, resolution, and averaging time settings may impose significant influence on the collected data, which could affect time response and reproducibility of $nSpO_2$ (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 $nSpO_2$ 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 SpO_2 measures, which avoided inter-center 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 only used 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

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

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 $nSpO_2$ 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 lead to many of the adult-related features being suboptimal for pediatrics, and should prompt future exploration of $nSpO_2$ features that enable improved specific ability to analyze $nSpO_2$ 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 75-90% 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), that corresponds to the point of upward inflexion in morbidity risks associated with sleep-disordered breathing in children (30-33). In other words, the accuracy of the surrogate diagnostic method using the $nSpO_2$-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 with an estimated AHI under 1 e/h would have an actual NPSG AHI no greater than 5 e/h. In addition, 94.4% of children predicted with an estimated AHI greater or equal to 5 e/h would actually have a NPSG AHI above 1 e/h.

A diagnostic protocol could then be derived from these results as follows: i) if our neural network model predicts an estimated AHI < 1 e/h, OSA would be discarded since most probably (94.4%) patients will have a NPSG AHI value < 5 e/h; ii) if our neural network model predicts an estimated AHI ≥ 5 e/h, consider referring for treatment since most probably (94.4%) will have a NPSG AHI value ≥ 1 e/h in a context of snoring children; and iii) if our neural network model predicts an estimated AHI ≥ 1 e/h but < 5 e/h, we would recommend referral for NPSG, since doubts arise about the actual definitive diagnosis. Such protocol would potential reduce the need for 52.6% of NPSGs, while only treating 5.7% of the snoring children with an actual NPSG AHI < 1 e/h; in addition, this approach would not be treating 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 that would be potentially missed using $nSpO_2$ might be minimized by repeating the oximetry-based test within a short period of time (weeks to 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 $nSpO_2$ assessments were negative using the methodology proposed above, could be then referred for NPSG, thereby markedly reducing the overall need for more costly testing in resource-constrained settings. Since a large number of sleep laboratories appears 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 $nSpO_2$ 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 approach may be preferred by

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

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 $nSpO_2$ neural network-derived AHI would offer similar options, i.e., 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 effect, 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 (RP) is becoming increasingly accepted as a surrogate diagnostic approach in adults and children, even if their 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 $nSpO_2$ is at least as accurate as RP in the diagnosis of OSA in children, further confirming the adequacy and potential limitations of such approaches as reported in other recent studies (9,24).

The obvious simplicity of $nSpO_2$ recordings has prompted efforts to evaluate their potential diagnostic properties, which to date have examined commonly used oximetric indices such as the number of desaturations, clusters of events within a particular timeframe, or percentage of time spent with a SpO_2 below a particular threshold (9,38-43). However, such conventional $nSpO_2$ analytical approaches have not achieved the desirable diagnostic performance for detecting OSA in children. Conversely, as shown

by Garde *et al.* (14), more sophisticated mathematical analyses of the oximeter signal achieved 88.4% sensitivity and 83.6% specificity for a 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, i.e. to perform binary classification. Although three commonly used OAHI 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 pre-test symptomatic pediatric patients into the 4 common categories of severity (no disease, mild, moderate, and severe) as performed here, but further enable estimates of the actual PSG-derived AHI of each patient.

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

Conclusions

This study provides extensive validation on 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 $nSpO_2$ signal provides an accurate and widely implementable diagnostic tool for childhood OSA, particularly in resource constrained environments.

References

1. Marcus CL, Brooks LJ, Ward SD, Draper KA, Gozal D, Halbower AC, Jones J, Lehmann C, Schechter MS, Sheldon S, Shiffman RN, Spruyt K. Diagnosis and management of childhood obstructive sleep apnea syndrome. Pediatrics 2012;130:e714-55.
2. Gileles-Hillel A, Philby M., Lapping-Carr G. Insights into Selected Aspects of Pediatric Sleep Medicine. Am J Respir Crit Care Med 2015;191:1459-61.
3. Tarasiuk A, Simon T, Tal A, Reuveni H. Adenotonsillectomy in children with obstructive sleep apnea syndrome reduces health care utilization. Pediatrics 2004;113:351-6.
4. Mukherjee S, Patel SR, Kales SN, Ayas NT , Strohl KP, Gozal D, Malhotra A. 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.
5. Alonso-Álvarez ML, Canet T, Cubell-Alarco M, Estivill E, Fernandez-Julian E, Gozal D, Jurado-LuqueMJ, Lluch- Roselló A, Martínez-Pérez F, Merino-Andreú M, Pin-Arboledas G, Roure N, Sanmartí FX, Sans-Capdevila O, Segarra-Isern F, Tomás-Vila M, Terán-Santos J. Consensus document on sleep apnea-hypopnea syndrome in children. Arch Bronconeumol, 2011; 47 :1-18.
6. Kaditis AG, Alonso Alvarez ML, Boudewyns A, Alexopoulos EI, Ersu R, Joosten K, Larramona H, Miano S, Narang I, Trang H, Tsaoussoglou M, Vandenbussche N, Villa MP, Van Waardenburg D, Weber S, Verhulst S. Obstructive sleep disordered breathing in 2- to 18-year-old children: diagnosis and management. Eur Respir J 2016;47:69-94.

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

7. 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(4):297-307.
8. Kaditis AG, Kheirandish-Gozal L, Gozal D. Algorithm for the diagnosis and management of pediatric OSA: A proposal. Sleep Medicine 2012; 13(3):217-27.
9. Alonso-Álvarez ML, Terán-Santos J, Ordax-Carbajo E, Cordero-Guevara JA, Navazo-Egüia AI, Kheirandish-Gozal L, Gozal D. Reliability of home respiratory polygraphy for the diagnosis of sleep apnea in children. Chest 2015;147:1020-8.
10. Kaditis AG, Kheirandish-Gozal L, Gozal D. Pediatric OSAS: oximetry can provide answers when polysomnography is not available. Sleep Med. Rev. 2015; 27:96-105.
11. 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;105:405-12.
12. Ayas NT, Drager LF, Morrell MJ, Polotsky VY. Update in Sleep Disordered Breathing 2016. Am J Respir Crit Care Med, 2017;195:1561-6.
13. 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;77:1286-90.
14. 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 OximeterTM. PLoS One 2014;9: e112959.

15. Gutierrez-Tobal GC, Kheirandish-Gozal L, Alvarez 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.

16. Alvarez D, Kheirandish-Gozal L, Gutierrez-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.

17. Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, et al; 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. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. J Clin Sleep Med. 2012;8(5):597-619.

18. 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.

19. 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.

20. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of biological signals. Physical Review E, 2005;71(2):021906.

21. Alvarez D, Hornero R, Marcos JV, Wessel N, Penzel T, Glos M, Del Campo F. (2013). Assessment of feature selection and classification approaches to enhance

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

information from overnight oximetry in the context of apnea diagnosis. Int. J. Neural Systems 2013;23:1350020.

22. Álvarez D, Hornero R, García M, del Campo F, Zamarrón C. (2007). Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive sleep apnea detection by means of central tendency measure. Artificial Intelligence in Medicine 2007;41:13-24.
23. Gutiérrez-Tobal GC, Hornero R, Álvarez D, Marcos JV, del Campo F. Linear and nonlinear analysis of airflow recordings to help in sleep apnoea–hypopnoea syndrome diagnosis. Physiol. Meas. 2012;33(7),:1261.
24. 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 Mar 14. pii: jc-00341-16. [Epub ahead of print] PubMed PMID: 28356177.
25. Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal Of Machine Learning Research 2003;3:1157-1182.
26. Yu L, Liu H. (Feature selection for high-dimensional data: A fast correlation-based filter solution. ICML 2003;3:856-863.
27. Bishop CM. Neural networks for pattern recognition. Oxford University Press. 1995.
28. Bishop CM. Pattern recognition. Machine Learning 2006;128:1-58.
29. 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;31:875-88.

30. Gozal D, Kheirandish-Gozal L. Obesity and excessive daytime sleepiness in prepubertal children with obstructive sleep apnea. Pediatrics. 2009;123(1):13-8.

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;141:682-91.

32. Gozal D, Kheirandish-Gozal L, Bhattacharjee R, Spruyt K. Neurocognitive and endothelial dysfunction in children with obstructive sleep apnea. Pediatrics. 2010;126(5):e1161-7.

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(6):739-47.

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;141:990-6.

35. Spruyt K, Gozal D. Pediatric sleep questionnaires as diagnostic or epidemiological tools: a review of currently available instruments. Sleep Med Rev. 2011;15(1):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(6):1508-15.

37. Collop NA, Anderson WD, Boehlecke B, Claman D, Goldberg R, Gottlieb DJ, Hudhel D, Sateia M, Schwab R. Clinical guidelines for the use of unattended

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Journal of Clinical Sleep Med 2007;3:737-47.

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;48:754e60.
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;19:1409-14.
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(12):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;2015:7692-5.
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(11):1409-12.

Table 1. Demographic and clinical data for each participating site.

Center	Participants	Age	Male (%)	BMI (kg/m^2)	AHÍ (/hrTST)	OSA for AHI = 1 e/h (%)	OSA for AHI = 5 e/h (%)	OSA for AHI = 10 e/h (%)
UofC[1]	981	6.1 ± 3.4	61.4	19.7 ± 7.3	9.3 ± 17.2	82.2	41.3	23.3
UofTn[2]	611	7.2 ± 4.6	54.6	23.3 ± 10.1	5.8 ± 11.3	68.1	29.6	17.7
HUBU[3]	578	4.1 ± 2.2	61.8	17.1 ± 4.2	5.9 ± 11.3	64.5	26.3	15.2
BCH[4]	558	6.3 ± 5.3	66.3	17.8 ± 3.7	5.8 ± 11.7	65.1	27.4	17.0
MSU[5]	499	6.5 ± 5.0	55.5	17.8 ± 11.2	6.2 ± 9.3	85.8	22.0	14.8
CGMH[6]	283	9.9 ± 3.2	72.4	19.5 ± 4.6	4.3 ± 10.0	72.4	21.9	8.1
UofHK[7]	202	10.0 ± 2.4	62.9	18.7 ± 4.6	4.9 ± 7.5	70.3	26.2	10.4
PUCC[8]	183	5.4 ± 4.8	52.5	17.8 ± 4.2	3.7 ± 9.1	60.1	18.6	7.6
UofA[9]	130	11.7 ± 3.1	37.7	30.3 ± 5.7*	3.2 ± 7.1	63.1	22.3	10.0
SJDCH[10]	60	8.4 ± 4.8	58.3	19.5 ± 5.2	4.2 ± 6.1	76.7	25.0	11.6
ASCH[11]	51	7.0 ± 3.4	66.7	20.6 ± 6.4	10.6 ± 13.8	90.2	54.9	37.2
UofTu[12]	36	10.4 ± 3.5	61.1	21.0 ± 8.0	6.9 ± 12.9	72.2	27.8	16.7
HSM[13]	19	6.5 ± 3.8	47.4	19.1 ± 6.8	11.0 ± 15.2	73.7	47.4	36.8
ALL	4191	6.7 ± 4.4	60.0	20.0 ± 7.0	6.4 ± 12.5	72.9	29.6	16.8

[+]BMI: body mass index; AHI: apnea-hypopnea index; OSA: obstructive sleep apnea; [1]UofC: University of Chicago (USA); [2]UofTn: University of Tennessee (USA); [3]HUBU: Hospital Universitario de Burgos (Spain); [4]BCH: Beijing Children's Hospital (China); [5]MSU: Michigan State University (USA); [6]CGMH: Chang Gung Memorial Hospital (Taiwan); [7]UofHK: University of Hong Kong (Hong Kong, China); [8]PUCC: Pontificia Universidad Católica de Chile (Chile); [9]UofA: University of Antwerp (Belgium); [10]SJDCH: San Joan de Deu Children's Hospital (Spain); [11]ASCH: Aghia Sophia Children's Hospital (Greece); [12]UofTu: University of Tuebingen (Germany); [12]HSM: Hospital de Santa Maria (Portugal).

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

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

Table 2. Time and frequency domain features extracted from the $nSpO_2$ recordings.

Features	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	Wootter's 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

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

Table 3. Four-class confusion matrix showing the classification agreement of neural network- based AHI estimate and the NPSG-derived AHI.

	Estimated severity→	MLP-AHI			
		AHI < 1	1 ≤ AHI< 5	5 ≤ AHI< 10	10 ≤ AHI
NPSG-AHI	AHI < 1	551	427	44	14
	1 ≤ AHI< 5	356	892	206	63
	5 ≤ AHI< 10	51	193	149	104
	10 ≤ AHI	3	87	83	379

AHI: apnea-hypopnea index; MLP: multi-layer perceptron; NPSG: nocturnal polysomnography.

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

Table 4. Diagnostic performance of neural network estimated AHI for the cutoffs 1 e/h, 5 e/h, and 10 e/h.

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

AHI: apnea-hypopnea index; Se: sensitivity; Sp: specificity; PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR-: negative likelihood ratio; Acc: accuracy.

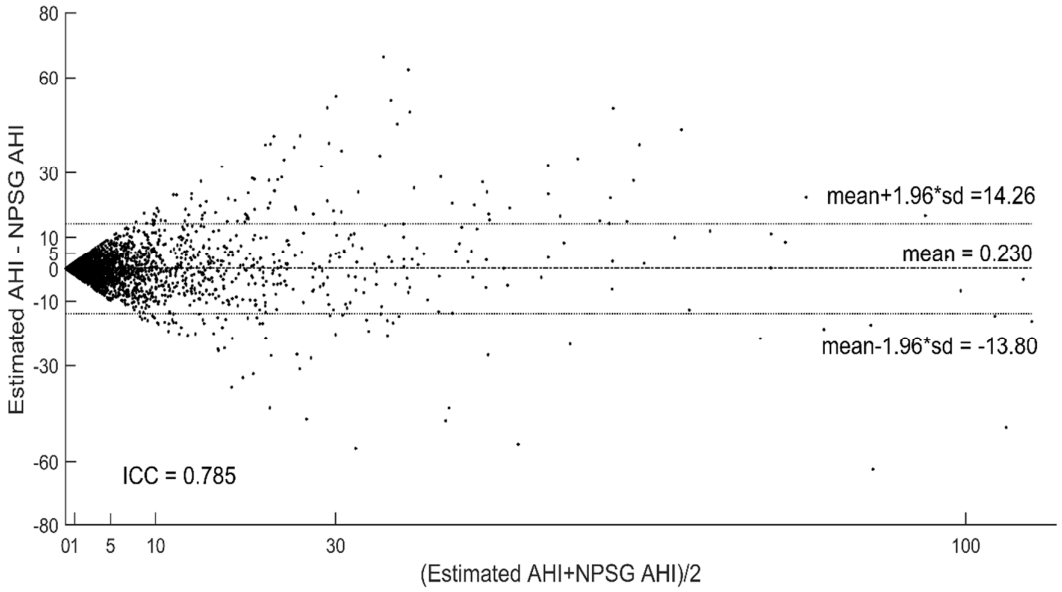


Figure 1. Bland-Altman plot comparing NPSG AHI with the estimated AHI using the neural network-based algorithm.

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

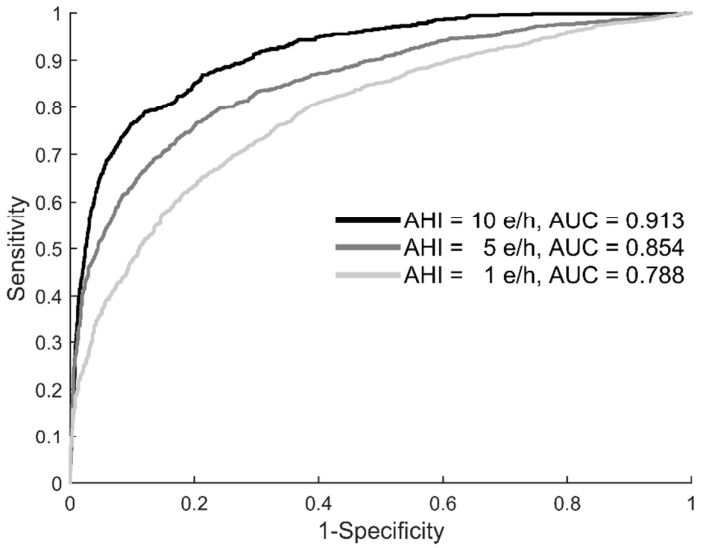
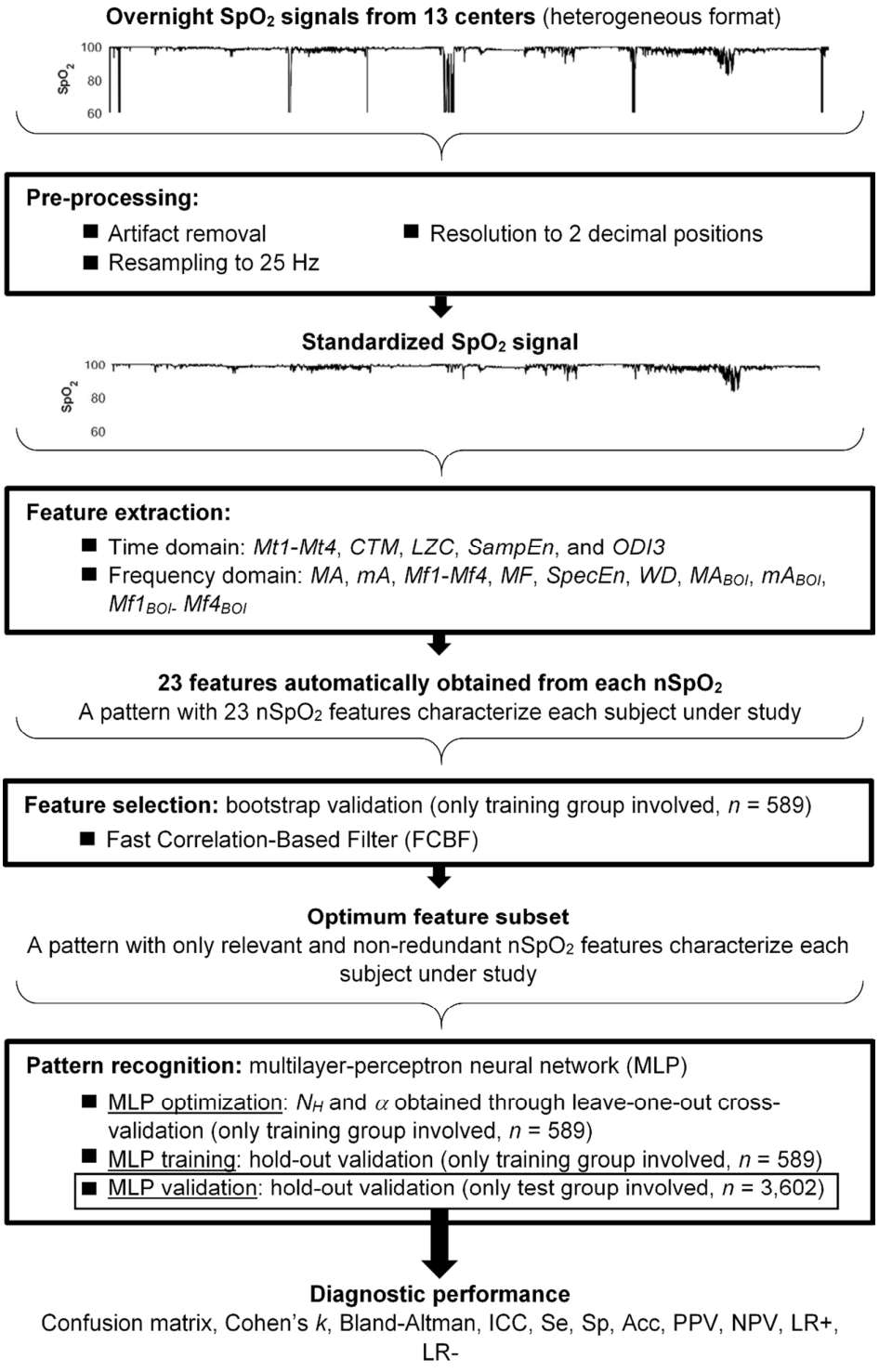


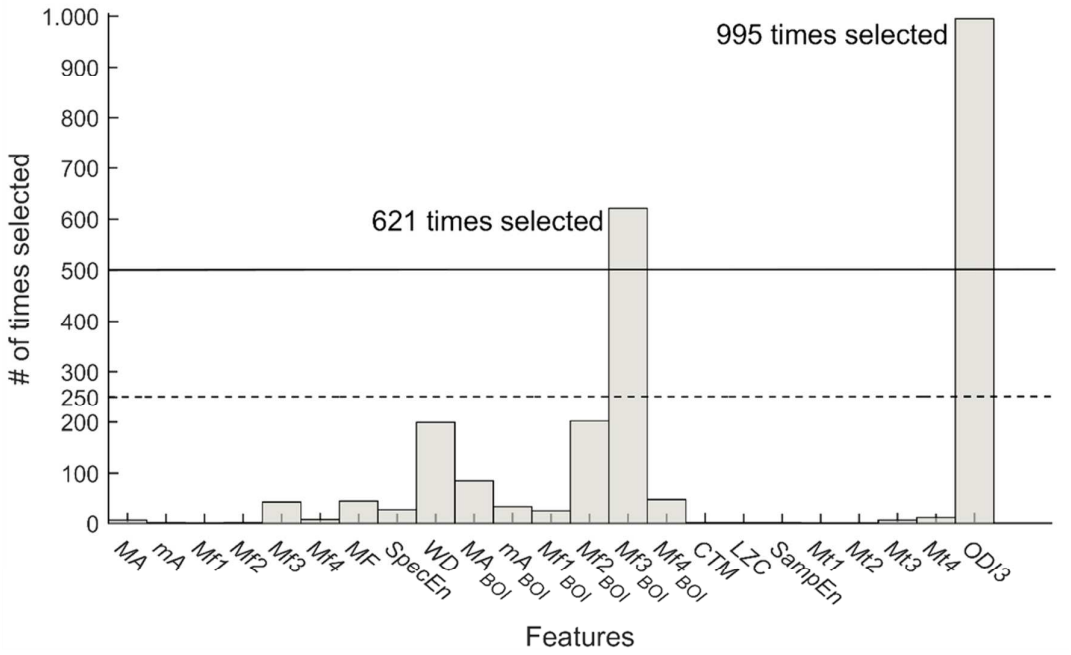
Figure 2. ROC curves of neural network-based AHI estimate in the test set for 1 e/h, 5 e/h, and 10 e/h.

Supplemental Data

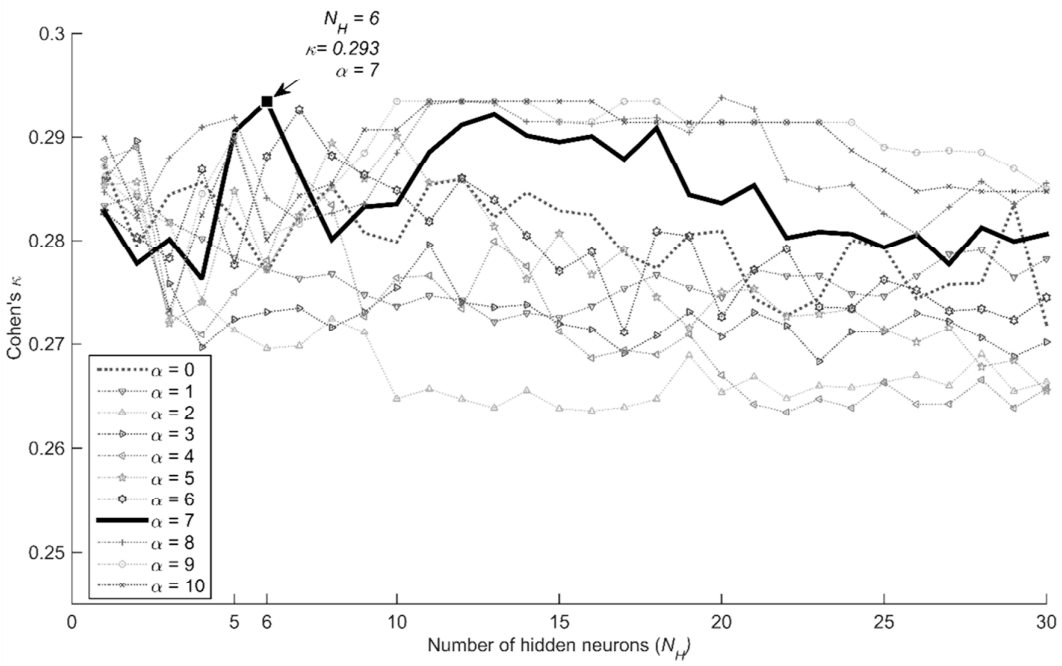


Supplemental Figure 1S. Scheme of the automatic methodology approaches applied to the $nSpO_2$ signals.

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC

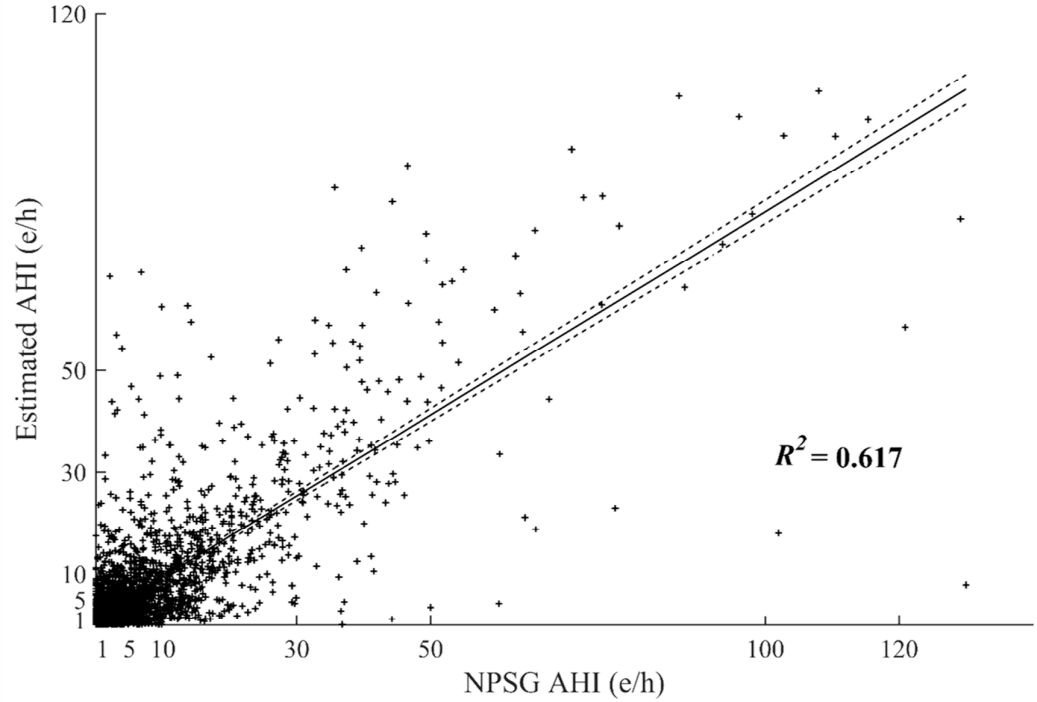


Supplemental Figure 2S. Histogram with the number of times each feature was selected after the bootstrap method.



Supplemental Figure 3S. Optimization of the MLP parameters: number of neurons in the hidden layer (N_H) and regularization (α).

AJRCCM Articles in Press. Published on 31-July-2017 as 10.1164/rccm.201705-0930OC



Supplemental Figure 4S. Scatter plot of the NPSG AHI and estimated AHI. The regression line and corresponding 95% confidence interval and R^2 statistic are also included.

Table 1 Supplement: Comparison between the ages of the subjects rightly classified and wrongly classified within the same severity group (rows). The ages of the subjects in each cell corresponding to rightly classified subjects (main diagonal) have been compared with the ages of the subjects wrongly classified in the same severity group (same row) by the use of the p-value of the non-parametric Mann-Whitney U test.

	Estimated severity→	Age in years: median [interquartile range] (# of subjects)			
		AHI < 1	**1 ≤ AHI< 5**	**5 ≤ AHI< 10**	**10 ≤ AHI**
NPSH-AHI	**AHI < 1**	7 [5,10] (551)	7 [4,10] (427)	**5* [3, 8] (44)**	5 [0, 10] (14)
	1 ≤ AHI< 5	6 [5,10] (356)	6 [4,9] (892)	5 [2,10] (206)	**2* [2,7] (63)**
	5 ≤ AHI< 10	**7* [5,12] (51)**	6 [4,10] (193)	5 [2,8] (149)	3 [1,10] (104)
	10 ≤ AHI	9 [8,13] (3)	7 [3,11] (87)	5 [3,9] (83)	5 [3,9] (379)

*p-value of the Mann-Whitney U test < 0.01, after Bonferroni correction.

Table 2 Supplement: Comparison between the BMIs of the subjects rightly classified and wrongly classified within the same severity group (rows). The BMIs of the subjects in each cell corresponding to rightly classified subjects (main diagonal) have been compared with the BMIs of the subjects wrongly classified in the same severity group (same row) by the use of the p-value of the non-parametric Mann-Whitney U test.

	Estimated severity→	BMI: median [interquartile range] (# of subjects)			
		AHI < 1	**1 ≤ AHI< 5**	**5 ≤ AHI< 10**	**10 ≤ AHI**
NPSH-AHI	**AHI < 1**	17.0 [15.1,20.6] (551)	**18.0* [15.9,22.2] (427)**	17.8 [16.0,19.9] (44)	17.5 [16.0,19.1] (14)
	1 ≤ AHI< 5	17.0 [15.2,20.9] (356)	17.7 [15.7,22.1] (892)	**19.4* [16.6,28.5] (206)**	17.0 [15.5,18.8] (63)
	5 ≤ AHI< 10	17.9 [15.2,22.2] (51)	17.3 [15.0,22.7] (193)	17.2 [15.0,22.5] (149)	18.1 [16.2,27.6] (104)
	10 ≤ AHI	24.9 [21.5,34.2] (3)	18.4 [15.5,25.0] (87)	17.3 [15.8,23.1] (83)	18.9 [15.8,26.1] (379)

*p-value of the Mann-Whitney U test < 0.01, after Bonferroni correction.