

ORIGINAL ARTICLE

Endotyping Sleep Apnea One Breath at a Time

An Automated Approach for Separating Obstructive from Central Sleep-disordered Breathing

✉ Ankit Parekh¹, Thomas M. Tolbert¹, Anne M. Mooney¹, Jaime Ramos-Cejudo², Ricardo S. Osorio², Marcel Trembl³, Simon-Dominik Herkenrath³, Winfried J. Randerath³, Indu Ayappa¹, and David M. Rapoport¹

¹Division of Pulmonary, Critical Care, and Sleep Medicine, Icahn School of Medicine at Mount Sinai, New York, New York; ²Center for Sleep and Brain Health, Department of Psychiatry, New York University Grossman School of Medicine, New York, New York; and ³Institute of Pneumology at the University of Cologne, Department of Pneumology and Allergology, Center of Sleep Medicine and Respiratory Care, Bethanien Hospital, Solingen, Germany

Abstract

Rationale: Determining whether an individual has obstructive or central sleep apnea is fundamental to selecting the appropriate treatment.

Objectives: Here we derive an automated breath-by-breath probability of obstruction, as a surrogate of gold-standard upper airway resistance, using hallmarks of upper airway obstruction visible on clinical sleep studies.

Methods: From five nocturnal polysomnography signals (airflow, thoracic and abdominal effort, oxygen saturation, and snore), nine features were extracted and weighted to derive the breath-by-breath probability of obstruction (P_{obs}). A development and initial test set of 29 subjects (development = 6, test = 23) (New York, NY) and a second test set of 39 subjects (Solingen, Germany), both with esophageal manometry, were used to develop P_{obs} and validate it against gold-standard upper airway resistance. A separate dataset of 114 subjects with 2 consecutive nocturnal polysomnographies (New York, NY) without esophageal manometry was used to assess the night-to-night variability of P_{obs} .

Measurements and Main Results: A total of 1,962,229 breaths were analyzed. On a breath-by-breath level, P_{obs} was strongly correlated with normalized upper airway resistance in both test sets (set 1: cubic adjusted [adj.] $R^2 = 0.87$, $P < 0.001$, area under the receiver operating characteristic curve = 0.74; set 2: cubic adj. $R^2 = 0.83$, $P < 0.001$, area under the receiver operating characteristic curve = 0.7). On a subject level, median P_{obs} was associated with the median normalized upper airway resistance (set 1: linear adj. $R^2 = 0.59$, $P < 0.001$; set 2: linear adj. $R^2 = 0.45$, $P < 0.001$). Median P_{obs} exhibited low night-to-night variability [intraclass correlation(2, 1) = 0.93].

Conclusions: Using nearly 2 million breaths from 182 subjects, we show that breath-by-breath probability of obstruction can reliably predict the overall burden of obstructed breaths in individual subjects and can aid in determining the type of sleep apnea.

Keywords: sleep apnea; esophageal pressure swings; airflow limitation; upper airway resistance; machine learning

Effective treatment of sleep apnea requires characterization of the disorder by establishing the presence, type, and severity of the sleep apnea (1, 2). Currently, the

presence of sleep apnea is documented through the observation of apneas and hypopneas on routine nocturnal polysomnography (NPSG). Severity is

assessed by the frequency of apneas and hypopneas, constituting the apnea-hypopnea index (AHI), as well as the physiological consequences associated with respiratory

(Received in original form November 2, 2020; accepted in final form August 25, 2021)

✉ This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0. For commercial usage and reprints, please e-mail Diane Gern (dgern@thoracic.org).

Supported by National Institute on Aging grants R01AG056531, R01AG056031, and R21AG049348; NHLBI grant K25HL151912; American Academy of Sleep Medicine Foundation Bridge Award BS-233-20; and Foundation for Research in Sleep Disorders.

Author Contributions: Conception and design: A.P., I.A., and D.M.R.; data collection/management: A.M.M., T.M.T., I.A., D.M.R., R.S.O., M.T., S.-D.H., and W.J.R.; algorithm development and statistical analysis: A.P., T.M.T., D.M.R., and I.A.; data interpretation: A.P., T.M.T., S.-D.H., W.J.R., I.A., and D.M.R.; critical revision of the article: A.P., T.M.T., A.M.M., J.R.-C., R.S.O., M.T., S.-D.H., W.J.R., I.A., and D.M.R.; final approval of the version to be published: A.P., T.M.T., A.M.M., J.R.-C., R.S.O., M.T., S.-D.H., W.J.R., I.A., and D.M.R.

Correspondence and requests for reprints should be addressed to Ankit Parekh, Ph.D., Division of Pulmonary, Critical Care, and Sleep Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029. E-mail: ankit.parekh@mssm.edu.

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 204, Iss 12, pp 1452–1462, Dec 15, 2021

Copyright © 2021 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.202011-4055OC on August 27, 2021

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Differentiating central from obstructive sleep apnea is critical in guiding treatment. This differentiation is largely dependent on classifying apneas and hypopneas using an assessment of inspiratory effort. Together with flow, effort determines upper airway resistance. Noninvasive signals that are surrogates of inspiratory effort are sufficient to classify apneas. However, for hypopneas, the gold standard for quantifying upper airway resistance is invasive esophageal manometry, which is not well tolerated and results in sleep disruption. As such, noninvasive surrogates of upper airway resistance are imperative to classify hypopneas, and thus, separate central from obstructive sleep apnea.

What This Study Adds to the

Field: Our study shows that a probability of obstruction derived using a feature-engineered machine learning approach is a reliable and noninvasive surrogate of upper airway resistance and can successfully distinguish central from obstructive sleep apnea both on a breath-by-breath level and on a subject level. Our probability of obstruction, which is derived within a matter of minutes, can determine the primary type of a subject's sleep apnea and aid in determining risks associated with untreated disorder and informing treatment approaches.

events (e.g., oxygen desaturation) (3). Establishing the type of sleep apnea as obstructive (OSA) or central (CSA) sleep apnea largely involves characterizing the type of apneas and hypopneas using an assessment of inspiratory effort; together with flow, effort determines whether upper airway resistance is elevated (likely obstructive) or not (likely central). For characterizing apneas, ventilatory effort visible on any of the noninvasive measurement modalities is considered

sufficient. On the other hand, definitive hypopnea classification requires invasive measurement using esophageal manometry to determine inspiratory effort. Noninvasive surrogates of effort, such as respiratory inductance plethysmography, do not quantify ventilatory effort, but rather qualify potential consequences of obstruction (e.g., paradoxical motion of the chest and abdomen) and thus incompletely characterize hypopneas. Esophageal manometry is not acquired during routine NPSGs as it is not well tolerated and may affect overall sleep (4). As such, alternative noninvasive methods that separate OSA from CSA are needed (1, 5, 6).

Routine NPSG signals contain subtle patterns that are potentially helpful in characterizing the etiology of sleep apnea. These patterns are often used in combination with noninvasive surrogates of effort to classify respiratory events as either obstructive or central (3, 7–10). We and others have previously shown that several of these patterns can predict obstructive (e.g., flattening of the flow/time tracing during inspiration, prolongation of inspiratory time, or paradoxical movement of the abdominal and thoracic effort signals) or central disturbances (e.g., a stepwise increase of flow and effort at the end of a hypopnea) (6, 10–13). Although current respiratory scoring rules by the American Academy of Sleep Medicine (AASM) recommend employing such patterns when esophageal manometry is not available (3), most clinical labs have refrained from discriminating hypopneas because of the subjective and laborious nature of the rules. Furthermore, events often contain both obstructive and central components, and currently there is no consensus on their identification or reporting. To circumvent these limitations, several studies have used automated identification of various signal patterns to distinguish OSA from CSA. Thomas and colleagues (14) used automated analysis of electrocardiogram to identify presence of CSA, while Morgenstern and colleagues (15) used automated analysis of the nasal cannula airflow signal to distinguish obstructive from central hypopneas. More recently, Mann and colleagues (16) used flow-shape analysis to deduce the severity of upper airway obstruction but did not specifically address the separation of OSA and CSA.

In the present study we develop a breath-by-breath probability of obstruction

using distinctive patterns visible on a routine NPSG and validate this as a noninvasive surrogate of gold-standard upper airway resistance during sleep. We show the utility of our probability of obstruction in classifying individual breaths as obstructed or nonobstructed, as well as in classifying individual subjects as with (likely OSA) or without (likely CSA) predominantly elevated upper airway resistance.

Methods

Subjects

A total of three datasets, comprising data from a total of 182 subjects, were used in this study (Figure 1). Initial data (used for development and initial testing) consisted of 29 patients (Table 1) seen at the New York University Sleep Disorders Center who previously underwent routine NPSGs with esophageal manometry as part of other research protocols (*see* Reference 12 for inclusion and exclusion criteria). The study was approved by the Institutional Review Board at the New York University School of Medicine, and all subjects provided written informed consent. A second set of data, consisting of 39 subjects (Table 1) who underwent routine NPSGs with esophageal manometry at Solingen, Germany, was used for further independent validation of the proposed algorithm. This study was approved by the ethical committee of University Witten/Herdecke, Germany and all subjects signed informed consent. A separate set of 114 patients who underwent 2 consecutive NPSGs without esophageal manometry was used to assess night-to-night variability of the derived probability of obstruction. Details on this data are provided in the online supplement.

NPSG Protocol

All NPSGs were collected using standard clinical equipment (Sandman; Embla Systems Inc. at New York University and SOMNOLab in Solingen, Germany). Each NPSG included signals for electroencephalography, electromyography, electrooculography, airflow using a nasal cannula/pressure transducer system, thoracic and abdominal effort, snoring, and oxygen saturation. In NPSGs with esophageal manometry, pressure (Pes) swings were measured with an esophageal catheter

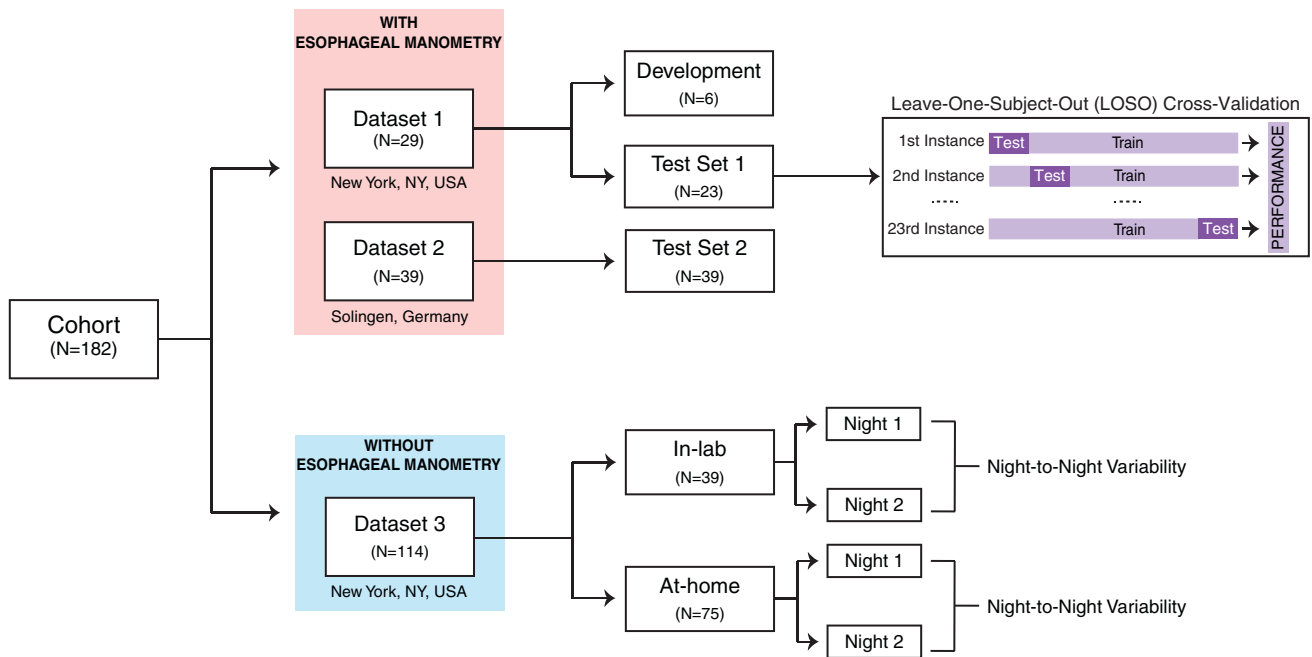


Figure 1. Datasets used in the study. Demographics and characteristics of the subjects are detailed in Table 1. Demographics for dataset 3 are detailed in the online supplement. For the leave-one-subject-out cross-validation, at each instance, $n=22$ subjects are used as a training set, whereas $n=1$ subject is used as a test set. The performance metrics from each instance are then tabulated and used to determine the performance of the model.

consisting of a thin catheter ending in a 10-cm latex balloon (Ackrad Labs) that was placed transnasally following lidocaine anesthesia and positioned in the lower third of the esophagus. Esophageal pressure measurements were made with a 100 cm H_2O pressure transducer (Validyne in data from New York University and UniTip catheters, UNISENSOR AG, in data from Solingen, Germany) (17).

Novel Breath-by-Breath Method for Endotyping Obstruction

Figure 2 depicts our overall approach of constructing the breath-by-breath probability of obstruction (P_{obs}). We adopted a feature-engineered, simplified machine learning approach to transform distinct patterns on routine NPSG to a breath-by-breath probability indicative of upper airway obstruction. Although the patterns were initially conceived on the basis of visual inspection, their identification in our algorithm was automated. Our primary assumptions driving the transformation were that 1) increased effort associated with a reduction in flow (i.e., elevated upper airway resistance) is indicative of obstruction of the upper airway (high P_{obs}); 2) sufficiently

reduced effort associated with a reduction in flow (i.e., low upper airway resistance) is indicative of preserved airway patency or decreased inspiratory effort in harmony with reduced airway patency (low P_{obs}); and 3) distinctive patterns on routine NPSG are representative of immediate physiologic consequences of either increased or reduced effort. All analyses were automated and written in MATLAB (MathWorks) and C++ and are publicly available (github.com/aparek/pobs).

Signal and feature selection. A total of five signals (airflow, thoracic and abdominal effort, oxygen saturation, and snoring) collected during routine NPSG were selected as signals of interest. All signals other than O_2 saturation were uncalibrated. Periods of NPSG where signals were invalid (e.g., disconnect) were excluded in an automated fashion. On the basis of the experience of the authors in research and clinical scoring, we extracted a total of nine features from the five signals, some of which are reported in previous studies (3, 18–20). A description of the selected features is given in Table 2 (see online supplement for detailed derivation of each feature and signal specific automated preprocessing). Scored sleep stages and marked respiratory events (e.g., apneas and

hypopneas) were not used for the proposed approach.

Development of breath-by-breath P_{obs}

The six development studies from the 29 studies acquired at NYU were visually assessed by four experts and were unanimously agreed to have predominantly either OSA (three studies) or CSA (three studies) pathophysiology and hence considered to be extremes in this set of data. Sleep apnea manifests as periodic reductions in airflow observed on a nasal cannula/pressure transducer system. As such, the first step in the diagnosis of sleep apnea is the identification of relatively small breaths. To this end, breaths within a study were identified, segmented, and labeled as either small (normalized amplitude $< 85\%$) or normal (normalized amplitude between 85 and 200%); see detailed methodology in online supplement. Each small breath in the six development set studies was assigned two weights based on the presence/absence of a given feature listed in Table 2: a weight suggesting the breath was part of a sequence of breaths during an ongoing upper airway obstruction event (likely OSA), and a weight suggesting the breath was part of a sequence of breaths during an ongoing event with preserved airway patency (likely CSA). For

Table 1. Subject Characteristics

	Comparison with Gold-Standard Resistance		
	Development (n = 6)	Test Set 1 (n = 23)	Test Set 2 (n = 39)
Demographics			
Age, yr	56 ± 19	50 ± 17	51 ± 16
Sex, M/F	5/1	21/2	31/8
BMI, kg/m ²	35.9 ± 10.5	34.3 ± 8.7	28.7 ± 5.0
Sleep-disordered breathing			
OAI, h ⁻¹	10.6 ± 9.8	7.8 (23.4)	3.0 ± 8.0
CAI, h ⁻¹	0.8 (11)	0.6 (3.1)	0.5 ± 1.6
AHI3A, h ⁻¹	68.2 ± 34.4	55.2 ± 26.4	13.2 ± 12.0
Breath size distribution*			
Imputed breaths	5,161 (14.6%)	15,689 (11.5%)	5,573 (0.1%)
Small breaths	8,932 (25.2%)	40,317 (29.5%)	59,636 (21.5%)
Normal breaths	21,305 (60.2%)	80,907 (59.0%)	211,594 (76.4%)
Total number of breaths	35,398	136,913	276,803

Definition of abbreviations: AHI3A = apnea–hypopnea index (hypopnea associated with 3% desaturation and/or arousal); BMI = body mass index; CAI = central apnea index; imputed breaths = artificial breaths during apneas; normal breaths = breaths with an amplitude of greater than 85% of normalized breath amplitude (all breaths during sleep/wake are included except for breaths during invalid periods of the study [e.g., equipment disconnects]); OAI = obstructive apnea index; small breaths = breaths with an amplitude of less than 85% of normalized breath amplitude. Values are presented as mean ± SD. In cases of non-normally distributed data, values are represented as median (interquartile range). Test set 1: New York, NY. Test set 2: Solingen, Germany.
*For the breath size distribution, values in parentheses represent percent of total number of breaths.

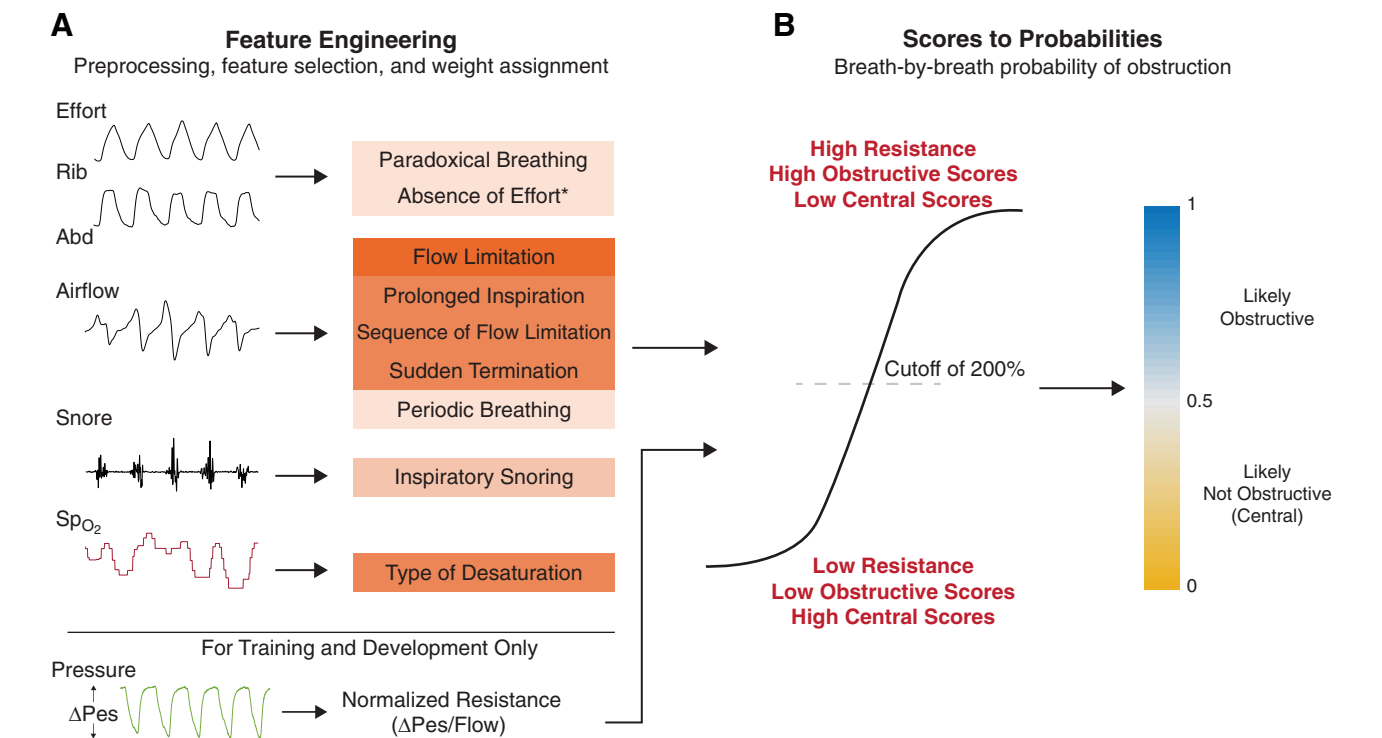


Figure 2. The proposed approach for estimating breath-by-breath probability of obstruction using patterns from routine nocturnal polysomnograms. (A) Input and feature engineering: raw and uncalibrated signals from routine nocturnal polysomnograms are preprocessed, and each breath is given weights according to the presence or absence of selected features. The color for each feature indicates our perceived importance of the feature used in assigning the weights (see Table 2). (B) Scores to probabilities: obstructive and central scores for each breath are determined using the sum of all weights. A logistic model is then learned to transform the raw scores to breath-by-breath probability of obstruction. Note that pressure is used only for comparison of our estimated breath-by-breath probabilities with gold-standard resistance (ΔPes/flow) and is not used developing the obstructive/central scores. *"Absence of effort" feature is used only for imputed breaths during apneas. Abd = abdomen; Pes = esophageal pressure; Rib = ribcage; SpO₂ = oxygen saturation.

Table 2. Feature Definition and Weight Assignment

Input	Feature Name	Description	Feature Weights			
			Obstructive		Central	
			Present	Absent	Present	Absent
Effort	Paradoxical breathing (19)	Paradoxical motion of the ribcage and/or abdomen during a sequence of small breaths	1	−1	−1	2
Airflow	Flow limitation (18)	Inspiratory flow limitation	5	−2	−5	3
Airflow	Prolonged inspiration (12)	Inspiratory time is prolonged relative to baseline inspiratory time	3	−1	−1	2
Airflow	Sequence of flow limitation	Flow limitation on adjacent breaths	3	−1	−1	0
Airflow	Sudden termination of hypopnea (27)	Sudden increase in flow after a sequence of small breaths	3	−1	−1	2
Airflow	Periodic breathing (10, 28)	Crescendo pattern of airflow	−1	0	3	0
Snore	Inspiratory snoring (3)	Inspiratory snoring on reduced-flow breaths that disappears with resumption of flow	2	0	−1	2
SpO ₂	Type of desaturation	Symmetric desaturation ≥ 3%	−1	0	3	0
		Asymmetric desaturation ≥ 3%	3	0	−1	0
Effort	Absence of effort* (3)	Absence of effort during apnea	−10	15	15	−10

Definition of abbreviations: Small breaths = breaths with an amplitude of less than 85% of normalized breath amplitude; SpO₂ = oxygen saturation using finger pulse oximetry.

None of the features required manual scoring of either sleep/wake or respiratory events. For each breath, two scores are obtained (obstructive and central).

*“Absence of effort” feature is only used in case of imputed breaths during apneas. Effort is absent if Δ Effort during apnea is less than 30% of Δ Effort preceding the apnea (see online supplement for detailed derivation). Δ Effort is defined as maximum effort – minimum effort during a breath (start of inspiration to end of expiration).

example, breaths that have inspiratory flow limitation are assigned the weights obstructive = +5 and central = −5, as it was deemed that inspiratory flow limitation is indicative of an ongoing obstructive event. Finally, for each small breath, the sum of all assigned weights was calculated resulting in an obstructive and a central score.

As a reference measurement of airway obstruction for comparison, we calculated breath-by-breath upper airway resistance using esophageal pressure swings (Δ Pes) and inspiratory flow. Δ Pes was defined as the change in esophageal pressure from the start of inspiration to the maximally negative peak within a breath. Flow was defined as the breath size (peak inspiratory flow; see online supplement for derivation). Resistance was defined as Δ Pes/Flow for each breath. For each NPSG, a value for overnight “normal” resistance was defined as the mean of the resistance of all normal breaths (breath size of 85–200%, as above). The resistance of each small breath was expressed as a percentage of this normalized resistance.

The two breath-by-breath scores in the six studies were used as predictors, and the corresponding breath-by-breath normalized resistance was used as the response to train a logistic regression model. The aim of the logistic regression model was to provide breath-by-breath probabilities that could indicate likelihood of obstruction. The learned logistic regression model transforms breath-by-breath scores to a breath-by-breath probability of obstruction, i.e., P_{obs} , (0 = most likely not obstructive; 1 = most likely obstructive). A cutoff of 200% for normalized resistance was used to define low versus high normalized resistance, on the basis of our previous study as well as on the observation by our group and others that airway resistance in normal subjects can double during sleep onset (12, 21).

Internal and external validation of breath-by-breath P_{obs} . We applied the learned logistic regression model on the 23 remaining studies in the initial test set from

New York and the 39 studies in the second test set from Solingen, Germany. The learned logistic regression model was used to transform breath-by-breath scores to P_{obs} on each small breath across the test sets and compared against the corresponding normalized resistance. Logistic regression model estimated probabilities (e.g., P_{obs}) can either be used as continuous values (i.e., 0 to 1) or as classes by dichotomizing (e.g., low vs. high) them using a predefined threshold (usually 0.5). Here, we assessed the predictive value of the learned logistic model using both approaches: using continuous values for normalized resistance and P_{obs} , as well as using dichotomized P_{obs} and normalized resistance, that is, low ($P_{obs} < 0.5$; normalized resistance < 200%) and high ($P_{obs} \geq 0.5$; normalized resistance > 200%)

Subject-level validation of breath-by-breath P_{obs} . We further tested the utility of the P_{obs} in classifying subjects as either those with predominantly low or high resistance, indicating a subject likely to have either CSA or OSA pathophysiology,

respectively. Using the six development NPSGs, a second binary logistic regression model was trained with median P_{obs} as the predictor and median normalized resistance (dichotomized as low vs. high with a cutoff of 200%) as the outcome. Recall that for each subject, both P_{obs} and normalized resistance are calculated for every small breath, and thus using their median values is a crude way of comparing P_{obs} and normalized resistance on a subject level. As previously, we assessed the subject level relationship between median P_{obs} and normalized resistance using median P_{obs} as a continuous value and as dichotomized into two classes (i.e., “likely central” as those with median $P_{\text{obs}} < 0.5$, and “likely obstructive” as those with median $P_{\text{obs}} \geq 0.5$).

Statistical Analyses

All statistical analyses were performed using IBM SPSS 24 and MATLAB R2020a (MathWorks). Normality of data was tested using the Shapiro-Wilk test. Spearman's rank correlation was used to assess correlations. Binomial logistic regression with tenfold cross validation was used to learn the logistic model from the development set ($n = 6$). Normalized resistance (%) was log transformed. We assessed the relationship between P_{obs} and normalized resistance, both treated as continuous values, using median-fit lines. For these median–median plots, P_{obs} values were divided into 100 equal-width bins ranging from 0 to 1 (i.e., bin width of 0.01). Bins with fewer than 10 breaths were discarded. Strength of relationships on the median–median plots was determined using the coefficient of determination (R^2) from linear and cubic fit lines obtained using linear and polynomial regression, respectively. When both P_{obs} and normalized resistance were dichotomized into classes, the primary metrics for assessing classification performance of the learned logistic regression model were area under the receiver operating characteristic curve (AUC-ROC), accuracy (ACC = % of breaths correctly classified), and Cohen's kappa scores. We used paired t tests to assess significant differences in classification metrics between the random forest classification models and the simplified feature-engineered model. Intraclass correlations (ICC, one-way random effects, absolute agreement, single measurement) and Bland-Altman (22) plots were used to

assess the night-to-night variability in median P_{obs} . A two-tailed P value of less than 0.05 was considered indicative of statistical significance for all tests.

Results

A combined total of 1,962,229 breaths from data in 182 subjects were analyzed. The computational runtime of deriving c was 3.1 ± 1.3 min (mean \pm SD) for a single study with between 5,000 and 7,000 breaths.

Efficacy of Learned Logistic Regression Model

The six NPSGs in the development set had a roughly equal distribution of breaths with low ($n = 3,360$ breaths) versus high ($n = 3,502$ breaths) resistance. The logistic model fit for the development set was statistically significant (Nagelkerke $R^2 = 21.4$; see Table E1 in the online supplement). Both obstructive and central scores were significant predictors of normalized resistance. However, compared with the model with only the obstructive score, the model with both obstructive and central scores reduced the model deviance by less than 1% (data not shown).

Internal and External Validation of Breath-by-Breath P_{obs}

Example tracings and the corresponding P_{obs} values for three representative subjects from the initial test set are shown in Figure 3. Tracings for three subjects who had a predominance of apneas over hypopneas are shown in Figure E1. An example tracing for a representative subject from the second test set is shown in Figure E2.

On a breath-by-breath level, P_{obs} was significantly correlated with normalized resistance (Figure 4). Visually, the median–median data in Figure 4 appeared to be polynomial in nature, which was confirmed by the higher R^2 with the cubic fit as opposed to the linear fit (cubic adjusted [adj.] $R^2 = 0.87$ vs. linear adj. $R^2 = 0.79$). Similarly, across the 39 subjects in the second test set, the P_{obs} was significantly correlated with normalized resistance (Figure 4B). In addition, as in the first test set, the relationship between P_{obs} and normalized resistance in the second test set appeared to be polynomial (cubic adj. $R^2 = 0.83$ vs. linear adj. $R^2 = 0.78$). It should be noted that although we excluded apneas when assessing the relationship between the estimated

probabilities and gold-standard resistance to avoid a 0/0 calculation, P_{obs} appeared to separate obstructive from central apneas (Figure E1). Further, P_{obs} appeared to correctly identify the likely obstructive versus likely central apneic components within mixed apneas (details in the online supplement).

Dichotomizing P_{obs} using a threshold of 0.5, (i.e., $P_{\text{obs}} < 0.5$ denoting low resistance and $P_{\text{obs}} \geq 0.5$ denoting high resistance) resulted in the classification performance shown in Figure 5. The accuracy (ACC) and kappa (κ) values indicate moderate to strong classification performance (first test set: ACC = 0.71 ± 0.08 , $\kappa = 0.43 \pm 0.13$; second test set: ACC = 0.69 ± 0.08 ; $\kappa = 0.39 \pm 0.15$; mean \pm SD). The AUC-ROC values were similar across both the test sets (first test set: AUC-ROC = 0.74, second test set: AUC-ROC = 0.70).

Subject-Level Validation of Breath-by-Breath P_{obs}

On a subject level, across two test sets with esophageal manometry, median P_{obs} was significantly correlated with median normalized resistance (Figure 6). Further, dichotomizing median P_{obs} (i.e., “likely central” as those with median $P_{\text{obs}} < 0.5$, or “likely obstructive” as those with median $P_{\text{obs}} \geq 0.5$), across both the test sets there was “substantial” to “almost perfect” agreement (first test set: ACC = 91.3%, AUC-ROC = 0.94, $\kappa = 0.81$; second test set: ACC = 89.7%, AUC-ROC = 0.92, $\kappa = 0.72$). In addition, it should be noted that median P_{obs} exhibited low night-to-night variability (ICC(2,1) = 0.93) (Figure E7; see online supplement).

Comparison with Data-driven Random Forest

Using feature weighting, the classification performance of dichotomized P_{obs} was significantly higher than the conventional random forest model ($P < 0.01$) (Figure E5; see online supplement for methodological details).

Discussion

Upper airway resistance, determined using invasive esophageal manometry, is the gold-standard measurement for classifying respiratory events as obstructive or nonobstructive, and therefore for distinguishing OSA from CSA. Using a

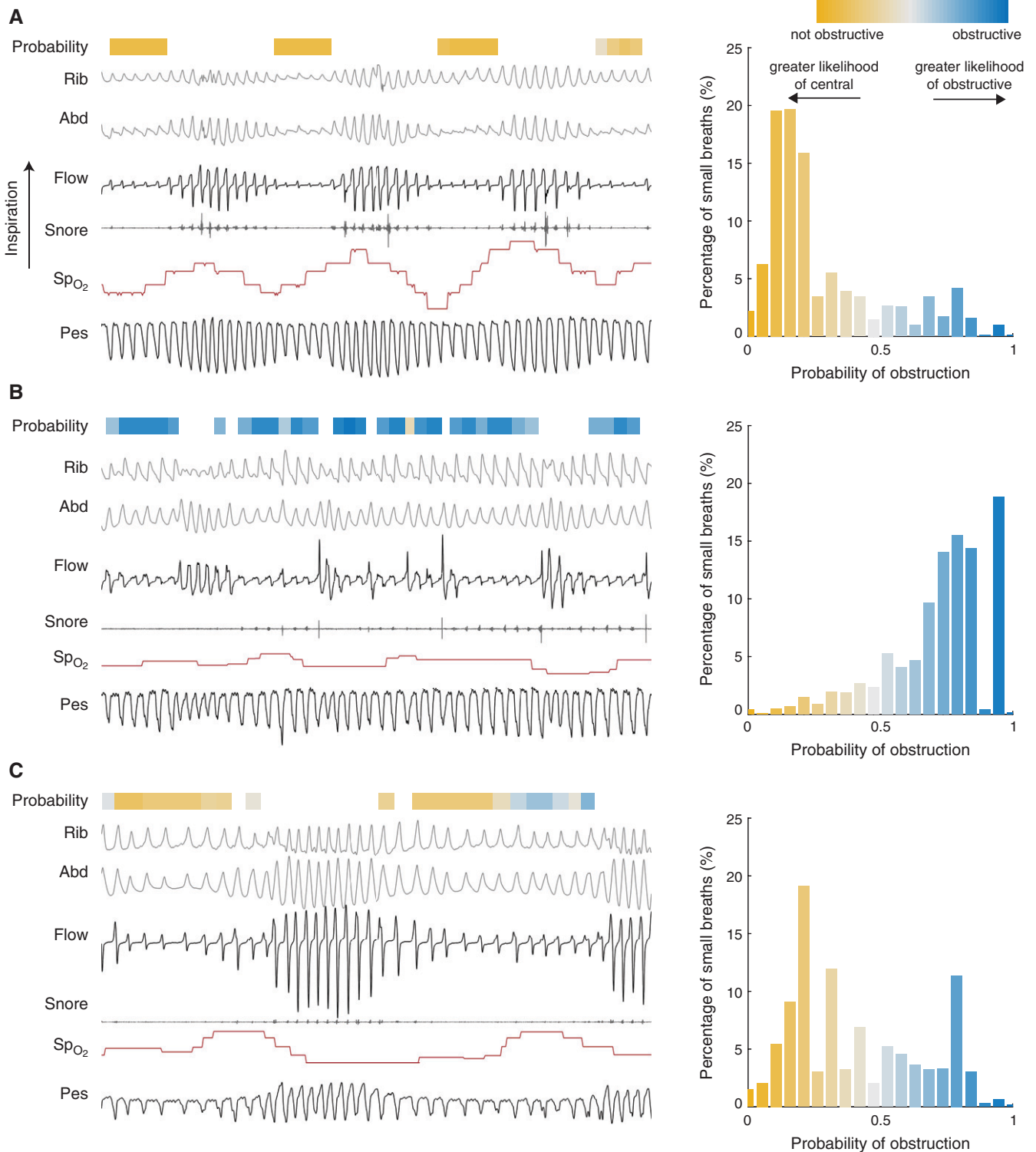


Figure 3. Example tracings and estimated probabilities from three representative subjects in our internal test set. Histograms on the right indicate the distribution of probability of obstruction (P_{obs}) throughout the night for each subject. (A) Pressure swings indicate a pattern of reduced effort on small breaths, which is also reflected in P_{obs} values (near 0, majority $P_{\text{obs}} < 0.5$). (B) Pressure swings indicate a pattern of increased effort on small breaths, which is also reflected in P_{obs} values (near 1). (C) Subject has patterns of reduced resistance; however, in certain cases the inspiratory flow shape indicates flow limitation. This ambiguity is reflected in distribution of P_{obs} (P_{obs} between 0.4 and 0.6). Overnight, this subject has a mix of obstructive and central physiology on small breaths. Abd = abdomen; Pes = esophageal pressure; Rib = ribcage; SpO₂ = oxygen saturation.

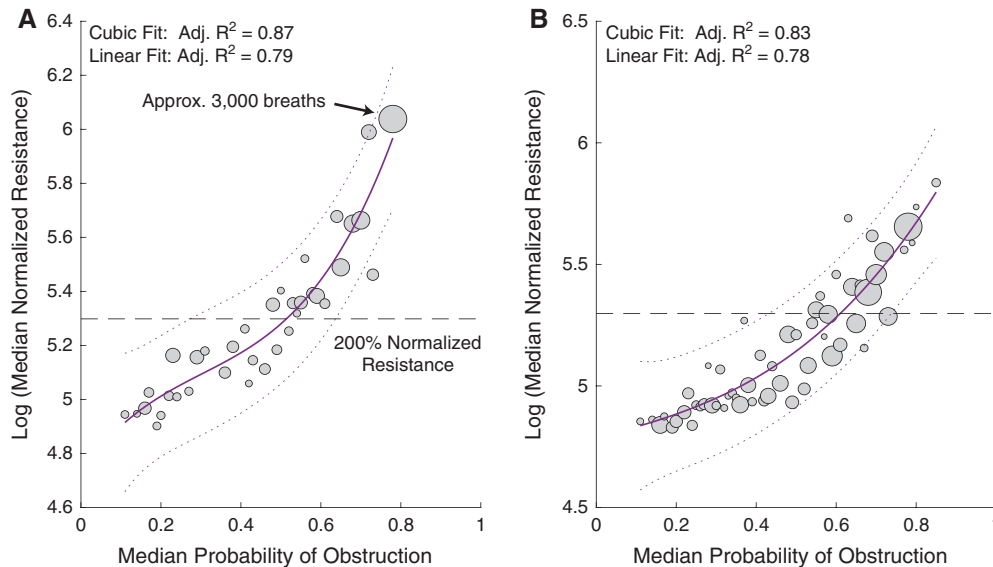


Figure 4. Median–median plots depicting the association of normalized resistance with median probability of obstruction in (A) test set 1 (New York, NY) and (B) test set 2 (Solingen, Germany). The size of each circle represents the number of breaths in a particular group. As an example, the circle indicated by the arrow represents approximately 3,000 breaths across the 23 studies in the internal test set that had probability scores between 0.78 and 0.79. Solid lines represent the cubic fit lines, and the dashed lines represent the 95% confidence intervals. Dashed horizontal line indicates the 200% normalized resistance. Adj. = adjusted.

combined total of 1,996,629 breaths during sleep, assessed in a simplified machine learning and feature-engineered approach, we demonstrate that a breath-by-breath probability of obstruction obtained from distinctive patterns observed on routine NPSGs is a noninvasive surrogate for gold-standard upper airway resistance. We show that the proposed breath-by-breath probability of obstruction exhibits low night-to-night variability and aids in classifying a subject's type of sleep apnea.

To our knowledge, our study uses one of the largest datasets consisting of NPSGs with concurrent esophageal manometry. We observed a strong correlation between P_{obs} and gold-standard upper airway resistance that held true for test sets acquired from two different sites, neither of which were used for development of P_{obs} . Our findings are consistent with two previously published related approaches: a visual schema using ensemble of polysomnographic patterns to distinguish obstructive from central hypopneas (3), and a breath-by-breath airflow shape-based algorithm that quantified the severity of upper airway obstruction (16, 17). The visual approach used known physiological features in a stepwise fashion to aid human scorers in classifying respiratory events. In a recent study by

Dupuy-McCauley and colleagues, the performance of this visual approach was similar to the AASM recommended rules, and the accuracy of either method capped at 69.3% (23). In contrast, by using additional features and subjectively weighting them, the proposed P_{obs} obtains a greater performance both on a breath-by-breath level and on a subject level (17). The relative importance of features using the data-driven random forest further confirms the validity of our subjectively weighing the features. Although our flow limitation feature uses similar airflow shape features as those used by Mann and colleagues in their study (16), our approach expands by including multiple NPSG signals. Finally, compared with the two approaches described above, our approach does not require human scoring of respiratory events or sleep stages.

Clinical Implications

Developing validated tools to better characterize CSA, especially in patients with heart failure, is one of the research priorities of the American Thoracic Society (2). A recent international taskforce report advised that the principal step in the approach to treating CSA is establishing the presence of an underlying CSA pathophysiology in an individual; establishing the presence of CSA

is fundamental to its treatment (6). As such, the primary step toward effective treatment of CSA is establishing presence of CSA, or absence thereof, which the proposed P_{obs} achieves reliably within a matter of minutes from a routine clinical NPSG. Further, identifying presence of a central ventilatory disturbance that is not based on the presence of central apnea alone is clinically meaningful as it could predict failure to respond to therapies that target upper airway obstruction (14).

Current AASM rules and practices do not adequately operationalize classifying respiratory events (apnea/hypopneas). Although some events may be clearly central (or obstructive), many events have a combined etiology, and a lack of consensus on their identification and reporting makes it difficult for labs to operationalize any event-based rules. As recently reported, the classification performance based on these rules is relatively low (23). Here, we have shown how our breath-by-breath probability may overcome these limitations. Breaths are the smallest division using which one can measure an individual's primary type of sleep apnea. With each sleep study containing several thousand breaths, breath-by-breath analyses are robust to both physiological and nonphysiological noise. In addition, breath-by-breath analyses, such as P_{obs} , can break

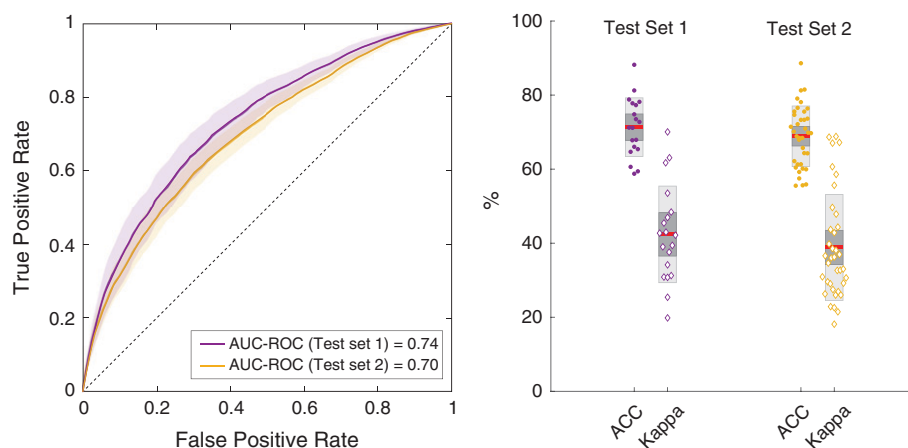


Figure 5. Classification performance of the learned logistic model using dichotomized probability of obstruction (two classes: low vs. high, with a threshold of 0.5). The left plot shows the mean area under the receiver operating characteristic curve (AUC-ROC) across all subjects in the two sets. Confidence interval is represented by the shaded area. The dashed reference line indicates the AUC-ROC for a randomized classifier (i.e., classifies breaths as low vs. high resistance by chance). Box plots on the right show classification accuracy (ACC) and kappa for each individual subject in the internal and the external test sets. Dark red lines indicate the mean ACC values, with the light and dark shaded sections of the boxes indicating one and two standard deviations.

down and effectively characterize mixed events for which there is no consensus currently. Furthermore, rudimentary techniques can be put in place that can effectively combine P_{obs} to generate an event-level probability, or as we have shown here, to a subject-level probability. Our results of near-perfect agreement of the probability scores on a subject level, when compared with gold-standard resistance, further indicate the clinical utility of breath-by-

breath analyses and P_{obs} in deducing subject-level outcomes.

Our approach of using probabilities as continuous values, rather than dichotomized into discrete categories, can provide noteworthy physiological insight while offering an intuitive clinical interpretation. As an example, a subject whose median P_{obs} is around 0.5, say between 0.4 and 0.6, could indicate several possibilities including 1) the subject has underlying OSA physiology, but

their upper airway resistance is not as elevated as others with more severe disease, 2) the subject has a mix of OSA and CSA physiology, or 3) presence of other comorbid conditions that alter the ventilatory stability. A dichotomized approach for P_{obs} would obscure these physiological insights, and as such, we recommend future studies in the direction where the entire histogram of overnight P_{obs} is used rather than in a binary form (yes vs. no, obstructive vs. central).

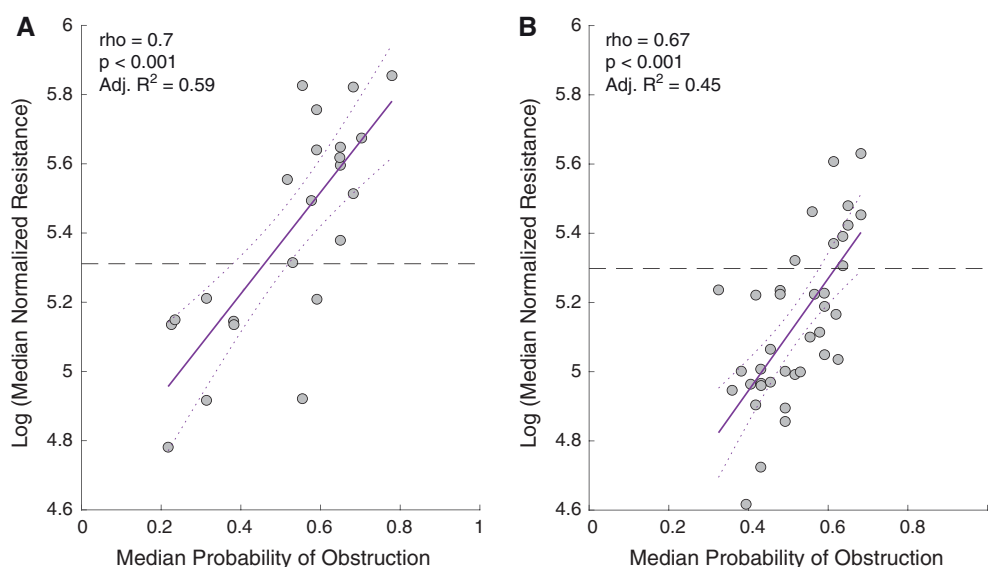


Figure 6. Association of median probability of obstruction with median normalized resistance across all small breaths on a subject level in (A) test set 1 (New York, NY) and the (B) test set 2 (Solingen, Germany). Each dot represents an individual subject. Solid lines represent the fit lines, and the dashed lines represent the 95% confidence intervals. The dashed horizontal line indicates a normalized resistance of 200%, which was used as a cutoff to determine low versus high resistance. Adj. = adjusted.

Furthermore, continuous values can also be used as a decision aid to human scorers, should such a need arise.

Limitations

We used 6 of 29 studies for the development and training of the binary logistic model that could transform raw obstructive/central scores to probabilities. Although we had a larger set of studies for training at our disposal, traditional machine learning theory and empirical evidence show that performance saturates as more training data are added (24). Accordingly, we chose the six studies that, on a clinical reading, had unambiguously OSA or CSA physiology overall to optimize the feature weights and features. Deep learning models that could be developed in the future, however, could benefit from using our large dataset for training if additional studies can be used to validate the results. It should be noted that the proportion of female subjects in the development and the test sets (internal and external) was low. Since it is observed in routine clinical practice that sleep apnea prevalence is greater among men than women, such a skewed distribution is not surprising. However, balanced datasets may be used in a future study to assess the role of sex in the performance of our automated approach.

We validated and tested the utility of our estimated breath-by-breath probability of obstruction in two ways: 1) “as is”, i.e., using probabilities as continuous values and 2) dichotomizing them into two classes (low vs. high with a cutoff of 0.5).

For both the approaches we observed a performance that is greater than the performance by a previously published manual visual algorithm (17). Logistic models provide probabilities as an outcome, and dichotomizing them is a forced choice and diminishes their overall utility (25). Thus, it is not surprising that the strength of the classification performance using the leave-one-subject-out cross-validated approach, which dichotomized the probabilities, was slightly less than when using the “as is” continuous value probabilities.

We compared our approach, which is based on a weighting scheme influenced by our perceived importance of features (feature engineered), with conventional machine learning (random forest), wherein feature importance is data driven. Although using custom weighted features may be thought of as subjective in nature, they offer translatability and are built upon years of clinical knowledge. We observe that the feature-engineered approach outperformed the conventional random forest model, which is not surprising as it has been long known that machine learning performance can be improved using as much domain knowledge as possible. We observed that both the approaches resulted in similar relative importance of the features. For example, the presence of an inspiratory shape suggesting flow limitation was considered one of the most important features. The relative importance of features that we observe here (see online supplement for details) was also echoed in the study by Randerath and colleagues (17). Further, as

pointed out by Iber in his corresponding commentary (26), snoring did appear to be a relatively important feature.

Conclusions

Upper airway resistance quantified using invasive esophageal manometry is the gold standard for deducing the type of sleep apnea as obstructive or central and is key in determining the appropriate course of treatment. However, esophageal manometry is not well tolerated during sleep, and a lack of consensus on events with combined obstructive and central components can make characterizing the type of sleep apnea difficult. It has been long known that respiratory disturbances are associated with distinctive patterns observed during routine nocturnal polysomnograms. In most clinical scoring schemes these patterns influence, either directly or subliminally, the visual characterization of respiratory events. In this study, we develop, validate, and test a probabilistic approach that amalgamates known physiologic features of respiratory disturbances into a breath-by-breath probability of obstruction. We show that the probability of obstruction is a reliable noninvasive surrogate for the invasive gold-standard measurement of upper airway resistance and can aid in identifying a subject's type of sleep apnea within a matter of minutes from a standard clinical sleep study. ■

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

References

1. Malhotra A, Ayappa I, Ayas N, *et al*. Metrics of sleep apnea severity: beyond the apnea-hypopnea index *Sleep* 2021;44:zsab030.
2. Orr JE, Ayappa I, Eckert DJ, Feldman JL, Jackson CL, Javaheri S, *et al*. Research priorities for patients with heart failure and central sleep apnea. an official American Thoracic Society research statement. *Am J Respir Crit Care Med* 2021;203:e11–e24.
3. Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, *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.
4. Chervin RD, Aldrich MS. Effects of esophageal pressure monitoring on sleep architecture. *Am J Respir Crit Care Med* 1997;156:881–885.
5. Rapoport DM. On beyond zebra (and the apnea-hypopnea index) in obstructive sleep apnea. *Am J Respir Crit Care Med* 2018;197:1104–1106.
6. Randerath W, Verbraecken J, Andreas S, Arzt M, Bloch KE, Brack T, *et al*. Definition, discrimination, diagnosis and treatment of central breathing disturbances during sleep. *Eur Respir J* 2017;49:1600959.
7. Montserrat JM, Badia JR. Upper airway resistance syndrome. *Sleep Med Rev* 1999;3:5–21.
8. Montserrat JM, Farré R, Ballester E, Felez MA, Pastó M, Navajas D. Evaluation of nasal prongs for estimating nasal flow. *Am J Respir Crit Care Med* 1997;155:211–215.
9. Montserrat JM, Kosmas EN, Cosio MG, Kimoff RJ. Mechanism of apnea lengthening across the night in obstructive sleep apnea. *Am J Respir Crit Care Med* 1996;154:988–993.
10. Yumino D, Bradley TD. Central sleep apnea and Cheyne-Stokes respiration. *Proc Am Thorac Soc* 2008;5:226–236.
11. Gibson GA. On Cheyne-Stokes respiration. *Trans Med Chir Soc Edinb* 1889;8:193–196.
12. Mooney AM, Abounasr KK, Rapoport DM, Ayappa I. Relative prolongation of inspiratory time predicts high versus low resistance categorization of hypopneas. *J Clin Sleep Med* 2012;8:177–185.

13. Mansour KF, Rowley JA, Badr MS. Noninvasive determination of upper airway resistance and flow limitation. *J Appl Physiol* . 2004;97:1840–1848.
14. Thomas RJ, Mietus JE, Peng CK, Gilmartin G, Daly RW, Goldberger AL, et al. Differentiating obstructive from central and complex sleep apnea using an automated electrocardiogram-based method. *Sleep* 2007;30:1756–1769.
15. Morgenstern C, Schwaibold M, Randerath W, Bolz A, Jane R. Automatic non-invasive differentiation of obstructive and central hypopneas with nasal airflow compared to esophageal pressure. *Annu Int Conf IEEE Eng Med Biol Soc* 2010;2010:6142–6145.
16. Mann DL, Terrill PI, Azarbarzin A, Mariani S, Franciosini A, Camassa A, et al. Quantifying the magnitude of pharyngeal obstruction during sleep using airflow shape. *Eur Respir J* 2019;54:1802262.
17. Randerath WJ, Treml M, Priegnitz C, Stieglitz S, Hagmeyer L, Morgenstern C. Evaluation of a noninvasive algorithm for differentiation of obstructive and central hypopneas. *Sleep (Basel)* 2013;36:363–368.
18. Condos R, Norman RG, Krishnasamy I, Peduzzi N, Goldring RM, Rapoport DM. Flow limitation as a noninvasive assessment of residual upper-airway resistance during continuous positive airway pressure therapy of obstructive sleep apnea. *Am J Respir Crit Care Med* 1994;150:475–480.
19. Staats BA, Bonekat HW, Harris CD, Offord KP. Chest wall motion in sleep apnea. *Am Rev Respir Dis* 1984;130:59–63.
20. Loube DI, Andrada T, Howard RS. Accuracy of respiratory inductive plethysmography for the diagnosis of upper airway resistance syndrome. *Chest* 1999;115:1333–1337.
21. Hudgel DW, Martin RJ, Johnson B, Hill P. Mechanics of the respiratory system and breathing pattern during sleep in normal humans. *J Appl Physiol* 1984;56:133–137.
22. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–160.
23. Dupuy-McCauley KL, Mudrakola HV, Colaco B, Arunthari V, Slota KA, Morgenthaler TI. A comparison of two visual methods for classifying obstructive versus central hypopneas. *J Clin Sleep Med* 2021;17.
24. Zhu X, Vondrick C, Fowlkes C, Ramanan D. Do we need more training data? *Int J Comput Vis* 2016;119:76–92.
25. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. *Stat Med* 1986;5:421–433.
26. Iber C. Are we ready to define central hypopneas? *Sleep* 2013;36:305–306.
27. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The report of an American Academy of Sleep Medicine task force. *Sleep* 1999;22:667–689.
28. Bradley TD, Phillipson EA. Central sleep apnea. *Clin Chest Med* 1992;13:493–505.