Endotyping sleep apnea one breath at a time: An automated approach for separating obstructive from central sleep disordered breathing

SUPPLEMENTARY MATERIAL

Authors:

Ankit Parekh, PhD[1],

Thomas M Tolbert, MD[1],

Anne M. Mooney, MD[1],

Jaime Ramos-Cejudo, PhD[2]

Ricardo S. Osorio, MD[2],

Marcel Treml, PhD[3],

Simon-Dominik Herkenrath, MD[3],

Winfried J. Randerath, MD[3],

Indu Ayappa, PhD[1],

David M. Rapoport, MD[1]

1. Division of Pulmonary, Critical Care, and Sleep Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

2. Center for Sleep and Brain Health, Department of Psychiatry, NYU Grossman School of Medicine, New York, NY, USA

3. Institute of Pneumology at the University of Cologne, Department of Pneumology and Allergology, Center of Sleep Medicine and Respiratory Care, Bethanien Hospital, Solingen, Germany.

Corresponding Author:

Ankit Parekh, PhD

Division of Pulmonary, Critical Care and Sleep Medicine

Icahn School of Medicine at Mount Sinai

New York, NY – 10029

Ph: 212 – 241 – 3884

Email: ankit.parekh@mssm.edu

SUPPLEMENT METHODS

Input Signals and Feature Engineering

Airflow

We used a nasal cannula/pressure transducer system as our airflow signal. Our in-house algorithm segmented each breath from start of inspiratory to end of expiration, derived a breath baseline, and classified the shape of inspiratory flow from 1=definitely not flow limited, to 3=definitely flow limited. The rules for flow limitation identification were similar to those reported in a recent American Thoracic Society (ATS) workshop guidelinesE[1]. Inspiration and expiration parts of each breath were also segmented, and inspiration time (Ti) was calculated. Inspiration time was considered prolonged if Ti was greater than 110% of baseline Ti. Breath baselines from ten prior breaths, to a given breath, were used to derive a moving average baseline.

We used the mean mid-breath (middle 1/3rd) flow during inspiration as our metric of breath amplitude. This metric of breath amplitude avoids issues associated with signal sampling as well as other artifactual errors. An example of this breath amplitude measure is shown in Fig. E3. The amplitude (size) of each breath is expressed as a % of the moving average of prior breaths, excluding from the average all definitely flow-limited breaths and all hyperventilatory breaths (>1.5 times the current average). The breath amplitude as a % of the moving average is termed “normalized amplitude”. Small breaths were defined as those with a normalized amplitude < 85% and normal breaths were defined as those with normalized amplitude > 85%. Since current AASM rules for hypopnea scoring require a reduction in flow of >= 30%, we observed empirically that the >= 30% cutoff coincided with approximately a normalized breath size of 85%. As such, we chose the cutoff of 85% to determine breaths as “normal” or “small”. Further, “normal” breaths only included breaths that had their normalized amplitude < 200%. . During apneas, “zero-flow” breaths were imputed based on the current average respiratory rate (Fig. E4).

The relative increase and/or decrease in breath amplitudes (sizes) was used to determine a) sudden termination of a sequence of reduced-flow breaths (Feature #5) and b) Cheyne-Stokes breathing patterns (Feature #6). In case of signal clipping, a

validated sparse optimization based approach was used to estimate the clipped signal.[2] In our internal set, 12% of the normal breaths were clipped.

Effort

Effort signals acquired were uncalibrated. Paradoxical breathing was quantified by estimating the phase angle between the rib and abdominal movement signals. Paradoxical breathing was considered significant only when it disappeared with normalization of airflow. In case of an apnea, imputed breaths were assigned scores as likely obstructive if effort was present and as likely central otherwise. Effort was considered to be present if ΔEffort (max. effort – min. effort during a breath) during an imputed breath exceeded 30% of the average ΔEffort during the two normal breaths preceding the apnea. Both rib and abdominal movement signals were used to derive ΔEffort and effort present on either of the signals was considered.

Snore

Inspiratory snoring on reduced-flow breaths was determined using a Hilbert envelope. Only snoring that disappeared with resumption of flow was considered as indicative of obstructive pathophysiology and snoring on small breaths that continued throughout the normal arousal breaths was not considered.

Oxygen Saturation (SpO_2)

Small (< 0.5%) non-physiological changes in the oxygen saturation signal were removed using a validated detrending algorithm.[3] Desaturations of more than 3% were quantified as either symmetric (similar time from steady-state to desaturation and from desaturation to resaturation) or asymmetric (quick resaturation) and scores were assigned toward likelihood of a preceding obstructive or central event. Preceding breaths associated with the desaturation were identified automatically and given the scores listed in Table 2.

Night-to-Night Variability of Estimated Probability of Obstruction (P_{obs})

In order to assess the night-to-night variability of P_{obs}, we utilized consecutive night data already available from an ongoing prospective longitudinal study. Briefly, breath-by-breath scores were first derived for all small breaths in the study and then transformed to a probability score using the learned logistic model. For each subject, we derived a single probability score by using the median of all overnight estimated breath-by-breath probabilities, i.e., median P_{obs}. Since these data do not contain concurrent esophageal manometry, here we only assess the night-to-night variability P_{obs} on a subject-level. The primary metric we used was the two-way random effect, absolute agreement, single measurement intraclass correlation coefficient (ICC(2,1)) according to the McGraw and Wong convention.[4] Bland-Altman plots were also used to describe the agreement between the estimated probabilities across the two nights.

Subjects

Data from a total of 114 subjects was analyzed to assess the night-to-night variability of our estimated probability of obstruction. These data were derived from an ongoing prospective longitudinal study that sampled community-dwelling healthy elderly volunteers to examine sleep, aging and Alzheimer's disease biomarkers from 2013 to 2019 at NYU Center for Brain Health. The study protocol was approved by the IRB at Icahn School of Medicine at Mount Sinai (16-00398; 18-00108). All participants signed a written informed consent.

NPSG Protocol

In-lab:

Full in-lab NPSG data from a total of 39 subjects was assessed. The NPSG included an EEG, EOG, EMG, airflow using nasal cannula/pressure transducer system, finger pulse oximetry for oxygen saturation, respiratory effort using respiratory inductance plethysmography, ECG, snoring (acoustic microphone) and body position measurements. Data were used as-is and neither sleep nor respiratory event scoring was utilized for the present study. Our in-house algorithm detected apneas.

At-home:

At-home polysomnographic data from a total of 75 subjects was assessed. Sleep monitoring consisted of 2-night of unsupervised home sleep testing (HST) using the using the Embletta MPR (Natus Medical Inc., San Carlos, CA) HST equipment. Airflow was recorded from a nasal cannula/pressure transducer system, oxygen saturation from finger (Embletta) and snoring from acoustic microphone. Effort was measured using respiratory inductance plethysmography. Only subjects who had valid data (more than 3 hours of recording with valid signals) were analyzed in this study.

Benefits of a feature-weighted approach as opposed to conventional random forest

We used the 23 studies in the initial test set to assess the benefits of weighting features based on our perceived importance as compared to conventional machine learning wherein features are not selectively weighted, rather used as present/absent. To this end, we used a leave-one-subject-out (LOSO) cross-validation approach in which 23 separate models are generated consisting of 22 studies as the test set and 1 study as the test set (see Fig. 1). As a result, 23 random forest models are built (1000 learners with 10 splits per tree for each random forest model) using the same set of features as listed in Table 2. The random forest models used 50 trees with a max depth of 10 and max bin of 32.

It should be noted that the features were used as-is and not selectively weighted. For example, flow limitation and periodic breathing are seen by the random forest model as equally important during training. The LOSO cross-validation approach provides an unbiased estimate of model performance and is strongly recommended in clinical use as opposed to a k-fold cross validation.[5] The primary metric for comparing the feature engineered logistic and conventional random forest model was accuracy (% correctly classified breaths). For comparison, we dichotomized P_{obs} as low ($P_{obs} < 0.5$) and high ($P_{obs} \geq 0.5$).

SUPPLEMENT RESULTS

Classifying type of apneas using P_{obs}

Although we excluded apneas for comparisons with gold standard resistance, it should be noted that the P_{obs} is able to distinguish between the two type of apneas (see Fig. E1). Moreover, as noted previously[6], apneas are relatively easy to discriminate and the ability to discriminate apneas and hypopneas using a unified method, as is proposed here, provides for a more robust and comprehensive way to characterize the type of sleep apnea.

Benefits of a feature-weighted approach as opposed to conventional random forest

We observe that the weighting the features provides better performance than the conventional random forest model across the 23 instances using the LOSO cross validation approach (Fig. E5). The accuracy was significantly higher when using the logistic model as compared to the random forest model (logistic ACC=0.72±0.11 vs. random forest ACC=0.61±0.05; mean±std). Using the curvature test[7] for each of the 23 random forest models, the relative feature importance is depicted in Fig. E6. Similar to our perceived importance of the features (Table 2), the random forest models show that flow limitation and prolongation of inspiratory time were considered to be the most relevant feature, followed by the sudden increase in flow after a sequence of flow limitation (i.e., sudden termination of a hypopnea), inspiratory snoring and sequence of flow limited breaths. These features were the most relevant for discriminating low against high normalized resistance (i.e., obstructive vs. central). The feature that was considered the least important was the periodic breathing.

Night-to-night Variability

The demographics and characteristics of the distribution of breaths across the 114 subjects is shown in Table E2. Across the two nights using data from both the in-lab and at-home PSG's, P_{obs} demonstrated low night-to-night variability (Fig. E8). For the combined dataset (at-home and in-lab), the ICC value was 0.93 with a $R^2 = 0.81$. Separating the data into in-lab and at-home PSG's we still observe low night-to-night variability ($ICC_{in\text{-}lab} = 0.89$, $R^2 = 0.73$, $ICC_{at\text{-}home} = 0.79$, $R^2 = 0.69$). The mean difference in P_{obs} for each type of PSG (in-lab vs. at-home) was 0 ± 0.07 (mean ± SD).

SUPPLEMENT TABLES

Table E1: Logistic model fit statistics for development set (N = 6,862 breaths across 6 development studies).

Variables	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for Exp(B)	
							Lower	Upper
Constant	-1.1	0.07	255.7	1	<0.001	0.34	-	-
Obstructive Score	0.1	0.01	30.7	1	<0.001	1.1	1.03	1.07
Central Score	-0.1	0.01	47.8	1	<0.001	0.91	0.88	0.93

Table E2:

	Night to Night Variability	
	In-lab NPSG N = 39	**At-home PSG N = 75**
Demographics		
Age (yrs.)	68 ± 7	71 ± 11
Gender (M/F)	14 / 25	30 / 45
BMI (kg/m^2)	27.4 (9.6)	26.8 (10.1)
Sleep Disordered Breathing		
OAI (/hr.)	1.5 (4.1)	2.6 (7.4)
CAI (/hr.)	0.3 (0.9)	0.1 (0.4)
AHI3A (/hr.)	13.1 (21.4)	16.5 (10.9)[#]
Breath size distribution**		
Imputed Breaths	29,012 (5.3%)	87,024 (9.0%)
Small Breaths (% of total)	143,662 (26.4%)	425,441 (43.8%)
Normal Breaths	370,749 (68.3%)	457,227 (47.2%)
Total number of breaths	543,423	969,692

[#]For at-home PSG AHI3A is the respiratory event index (REI) that is the sum of respiratory events divided by the time for which there was valid airflow.

SUPPLEMENT FIGURES

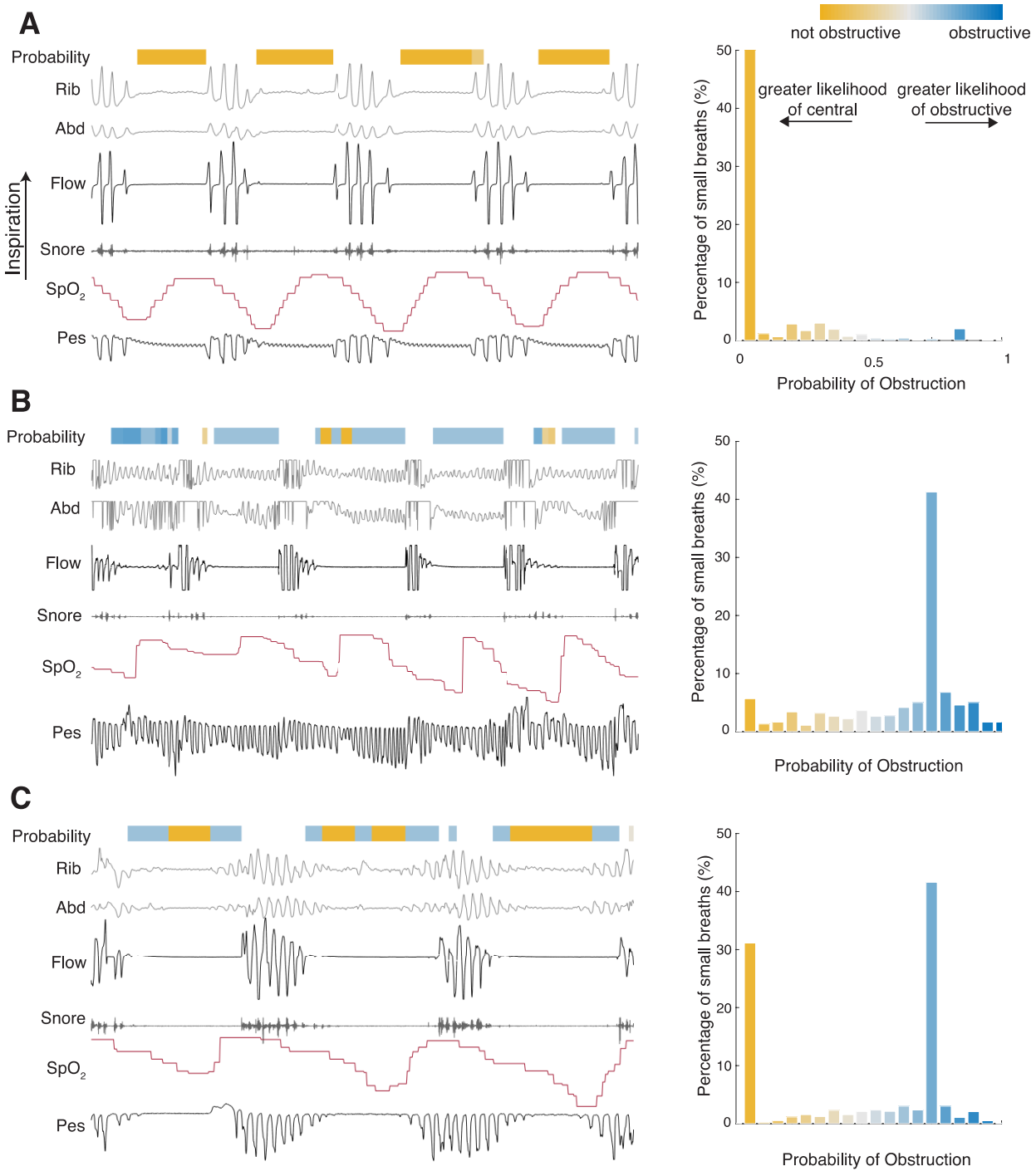

Figure E1: Example tracings and estimated probabilities for three representative subjects in our internal test set that had predominantly apneas and relatively few

hypopneas. Histogram of overnight estimated probabilities is shown in the plots on right. A) Subject had predominantly central apnea overnight, confirmed by the absence of effort on the pressure signal during apneas and reflected in our estimated probabilities. B) Subject had obstructive apneas overnight, confirmed by the presence of effort on the pressure signal during apneas and also reflected in our estimated probabilities. C) Subject had mixed apneas throughout the night. The histogram highlights the mixed apnea nature of the study and indicates that the duration of apnea with presence of effort was greater than the duration of apnea without effort.

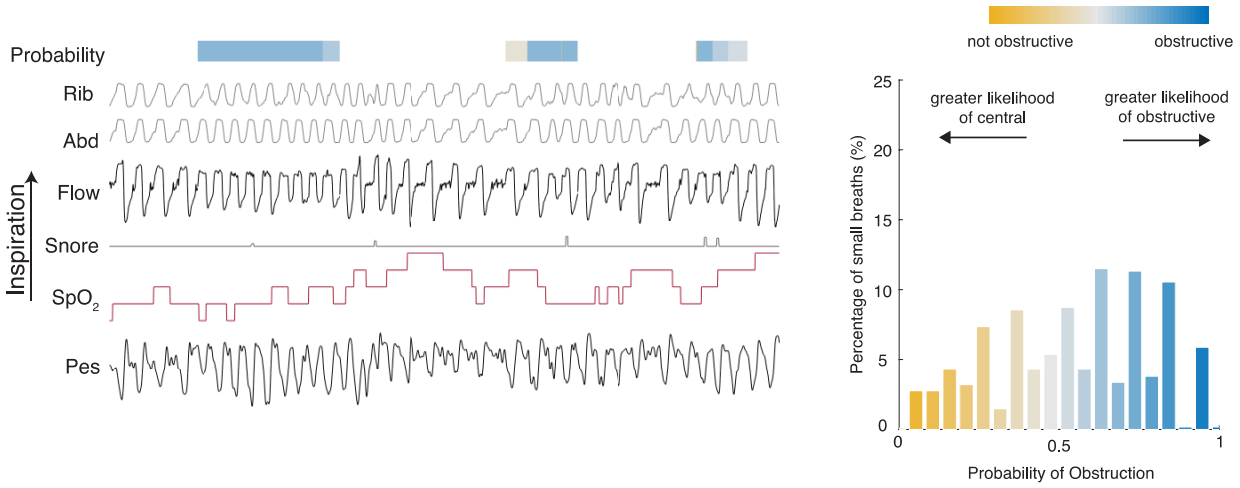

Figure E2: Example tracing from a representative subject in the external data set is shown on the left with a histogram of overnight estimated probabilities on the right. It can be seen that the majority of small breaths (reduced-flow) were deemed to have a greater likelihood of obstruction (estimated probability > 0.5). During the small breaths, esophageal pressure (Pes) swings indicate an increase in resistance (effort), confirming that the estimated probabilities appear to be accurate. Overnight, the subject appeared to have predominantly obstructive small breaths with a few breaths that were deemed to be associated with likely central sleep apnea pathophysiology.

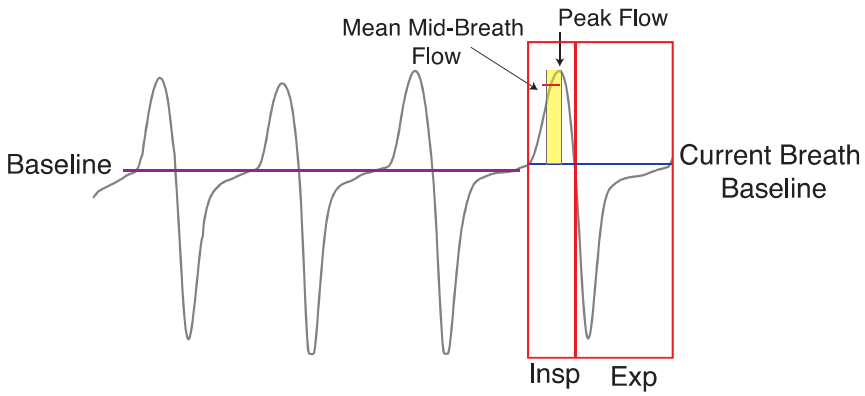

Figure E3: Derivation of our breath amplitude measure (mean mid-breath flow). For an example breath highlighted in the red rectangular area, inspiration and expiration periods are segmented by our in-house algorithm. Mean flow during the middle 1/3rd of inspiration is then calculated. Purple line shows the moving average baseline using ten prior breaths and the blue line within the highlighted breath indicates the baseline of the current breath.

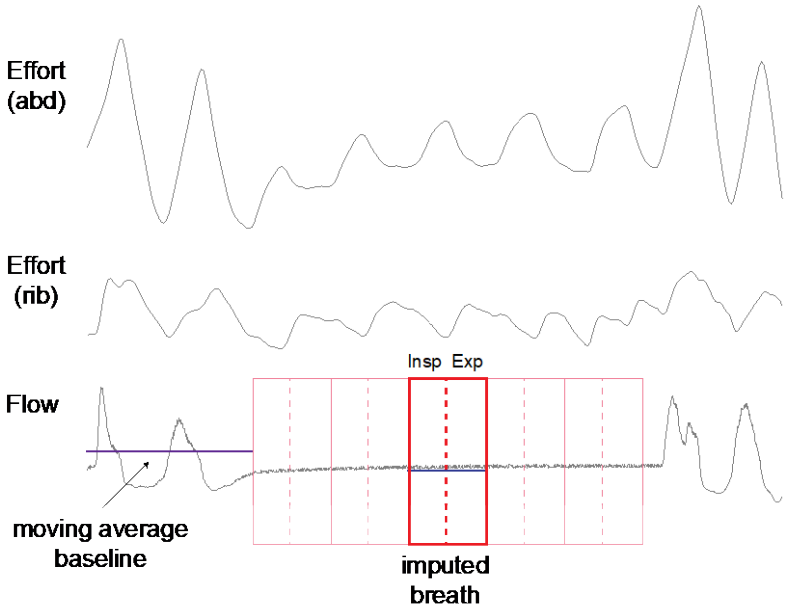

Figure E4: Example tracings showing the methodology for imputing zero-flow breaths during an apnea as well as deducing presence/absence of effort. During the entire apnea (zero-flow), five breaths are imputed, each with a fixed inspiration and expiration time. The decision for imputing five breaths was arrived at using the duration of the apnea in question and the instantaneous respiratory rate during the two normal breaths (non reduced-flow and non flow-limited) preceding the apnea. For each of the imputed breaths, corresponding ΔEffort, defined as the max. effort – min. effort during the entire breath, was calculated and compared with average ΔEffort during the two normal breaths preceding the apnea. ΔEffort calculations were done on both the rib and the abdomen signals. If ΔEffort during the imputed apneas was greater than 30% of ΔEffort preceding the apnea (either rib or abdomen), effort was considered to be present. In this example, for all the imputed breaths, effort was considered to be present.

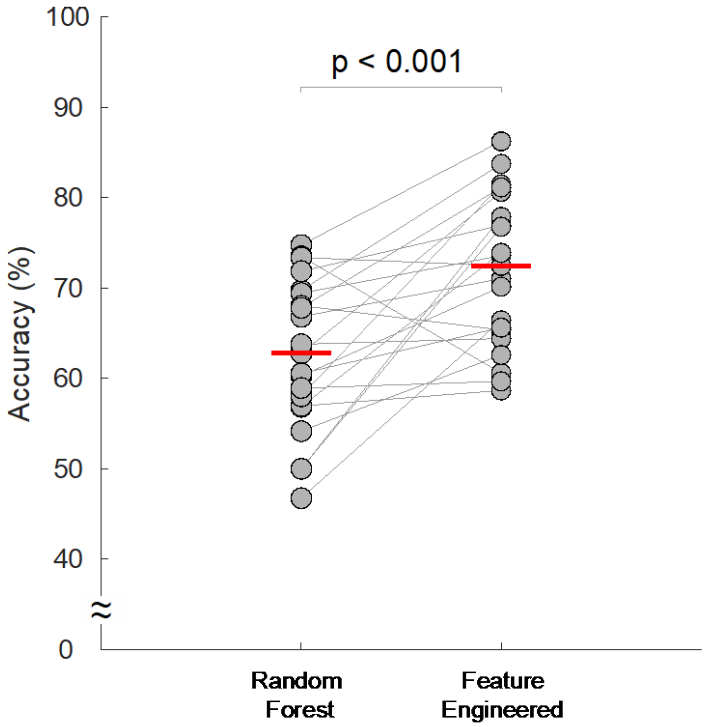

Figure E5: Comparison of conventional random forest model to our feature engineered logistic model using a leave-one-subject-out approach on our internal test set (N = 23). The logistic model reweights features according to our perceived importance, whereas the random forest model uses a data-driven approach to calculate the importance of the features. Accuracy (% correctly classified breaths) is significantly higher using the feature engineered logistic model as compared to the random forest model.

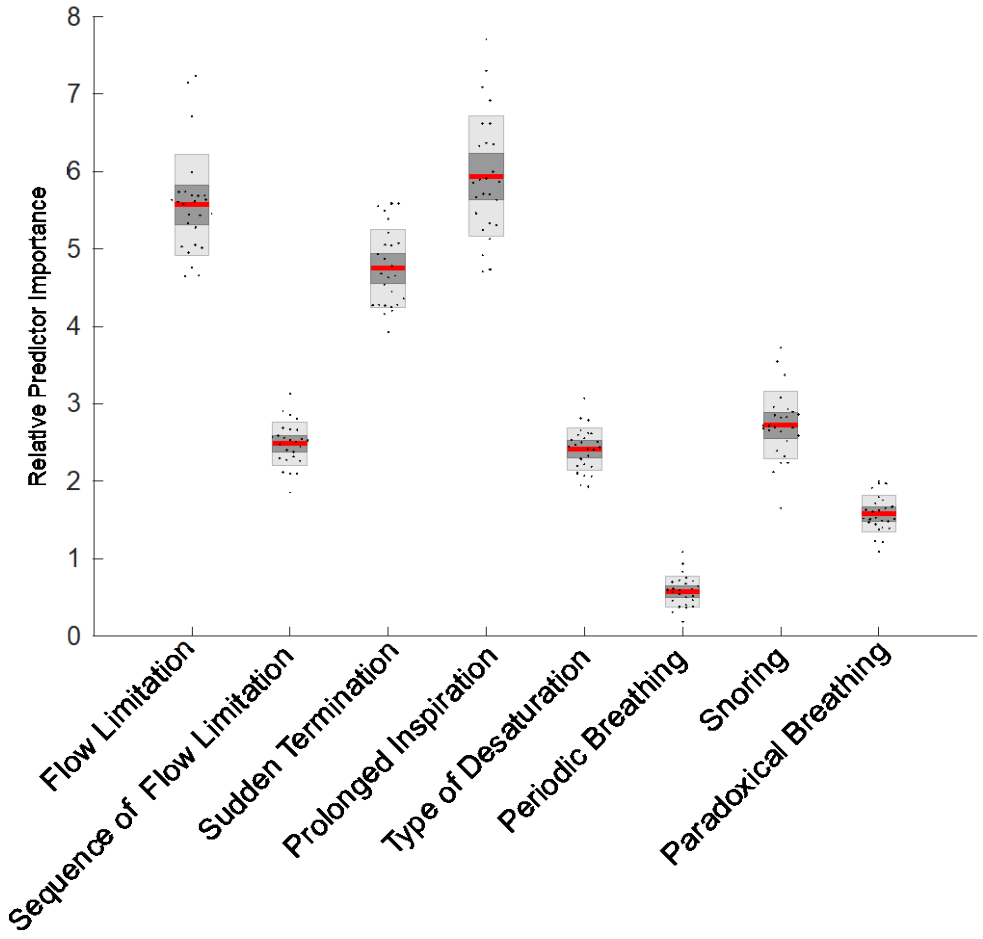

Figure E6: Unbiased feature importance estimates using the random forest model with Curvature Test.[7] Individual dots represent the predictor importance across 23 random forest models. Dark red lines indicate the mean predictor importance and the dark and light patches indicate 1 and 2 standard deviations respectively.

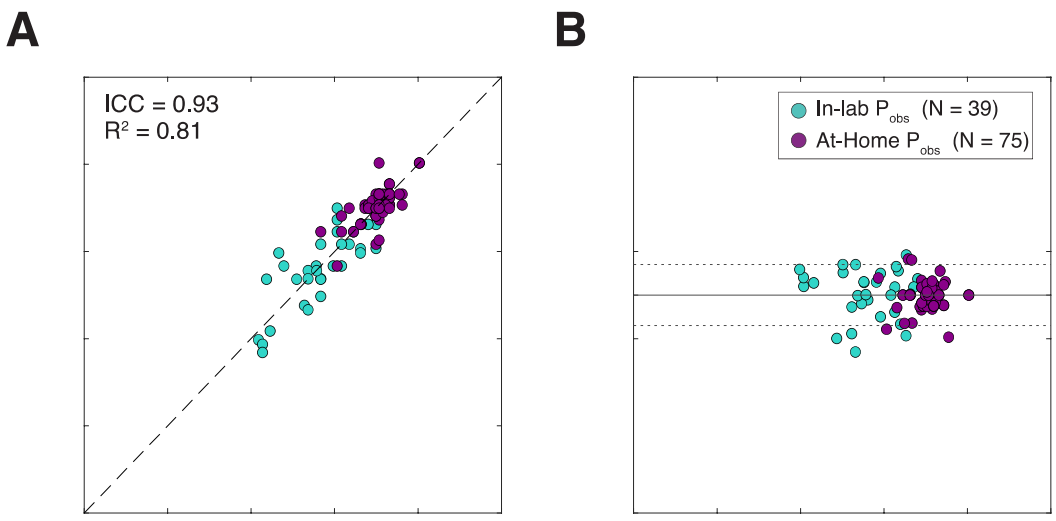

Figure E7: Night to night variability of median estimated probability of obstruction. A) Night 1 vs. Night 2 scatterplot of estimated probability of obstruction. Dashed line indicates line of identity. B) Bland-Altman plot for assessing the difference in nightly median estimated probability of obstruction. Solid line indicates the mean (0), and the dashed lines indicate the SD (0.06).

REFERENCES

E1. Pamidi S, Redline S, Rapoport D, et al. An Official American Thoracic Society Workshop Report: Noninvasive Identification of Inspiratory Flow Limitation in Sleep Studies. *Annals of the American Thoracic Society*. Jul 2017;14(7):1076-1085. doi:10.1513/AnnalsATS.201704-318WS

2. Parekh A, Burschtin O, Rapoport D, Ayappa I. Doris: Airflow signal de-clipper based on sparse optimization. *Am J Respir Crit Care Med*. 2020;201(A2419)

3. Parekh A, Selesnick IW. Convex Fused Lasso Denoising with Non-Convex Regularization and its use for Pulse Detection. 2015:

4. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996;1(1):30-46.

5. Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. The need to approximate the use-case in clinical machine learning. *Gigascience*. May 1 2017;6(5):1-9. doi:10.1093/gigascience/gix019

6. 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*. Mar 1 2013;36(3):363-8. doi:10.5665/sleep.2450

7. Loh WY. Regression Trees with Unbiased Variable Selection and Interaction Detection. *Statistica Sinica*. 2002;12:361-386.