A scalable method of determining physiological endotypes of sleep apnea from a polysomnographic sleep study.

Eysteinn Finnsson¹, Guðrún H. Ólafsdóttir¹, Dagmar L. Loftsdóttir¹, Sigurður Æ. Jónsson³, Halla Helgadóttir¹, Jón S. Ágústsson¹, Scott A. Sands², Andrew Wellman².

¹ Nox Research, Nox Medical, Reykjavík, Iceland

² Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Corresponding author:

Eysteinn Finnsson, Katrínartún 2 105 Reykjavík, Email: eysteinnf@noxmedical.com

© Sleep Research Society 2020. Published by Oxford University Press on behalf of the Sleep Research Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Abstract

Sleep apnea is caused by several endophenotypic traits, namely pharyngeal collapsibility, poor muscle compensation, ventilatory instability (high loop gain), and arousability from sleep (low arousal threshold). Measures of these traits have shown promise for predicting outcomes of therapies (e.g. oral appliances, surgery, hypoglossal nerve stimulation, CPAP, and pharmaceuticals), which may become an integral part of precision sleep medicine. Currently the methods Sands et al. [1] developed for endotyping sleep apnea from polysomnography (PSG) are embedded in the original authors’ code, which is computationally expensive and requires technological expertise to run. We present a re-implementation and validation of the integrity of the original authors’ code by reproducing the endo-Phenotype Using Polysomnography (PUP) method of Sands et al. [1, 2] The original MATLAB methods were reprogrammed in Python; efficient methods were developed to detect breaths, calculate normalized ventilation (moving time-average), and model ventilatory drive (intended ventilation). The new implementation (PUPpy) was validated by comparing the endotypes from PUPpy with the original PUP results. Both endotyping methods were applied to 38 manually scored polysomnographic studies. Results of the new implementation were strongly correlated with the original (p<10^{-6} for all): ventilation at eupnea $V_{passive}$ (ICC=0.97), ventilation at arousal onset $V_{active}$ (ICC=0.97), loop-gain (ICC=0.96), and arousal threshold (ICC=0.90). We successfully implemented the original method by Sands et.al. [1, 2] providing further evidence of its integrity. Additionally, we created a cloud-based version for scaling up sleep apnea endotyping that can be used more easily by a wider audience of researchers and clinicians.

**Keywords**: sleep apnea, endotype, phenotype, loop gain, collapsibility, arousal threshold, upper airway anatomy, personalized medicine
Statement of Significance

It has been assumed that accurate endo-phenotyping of sleep apnea is integral to developing precision medicine in the field of sleep medicine. A recent review of the literature, however, suggested that sleep laboratory endotyping would not become available anytime soon. [3] In this paper we present and validate a cloud-based re-implementation of the previously published method for endo-Phenotyping Using Polysomnography (PUP). [1, 2] The new cloud-based implementation confirms the reproducibility the PUP method and could be made available to researchers who are interested in endotyping but do not have the resources or expertise required to use the previously published method. This validation and improved access could allow scientists to further investigate the clinical relevance of sleep apnea endotypes.
Introduction
Sleep apnea is a chronic disorder where breathing is periodically interrupted during sleep. Diagnosis of sleep apnea is made via polysomnography (PSG), where a variety of physiological signals are measured during a night of sleep, including electroencephalogram (EEG), peripheral capillary oxygen saturation (SpO2) and ventilation. The sleep study is subsequently analyzed by a technician that labels sleep stages and events according to a standard rubric. [4] By counting the number of both partially and totally obstructed breathing events and dividing it by total sleep time the Apnea-Hypopnea Index (AHI) is calculated. The AHI can be interpreted as the average number of apneas and hypopneas per hour of sleep and has been traditionally used as the main indicator for sleep apnea severity. [5] A major limitation of current diagnostics techniques is that they do not provide information regarding the underlying cause or impact of sleep apnea in different individuals. [6]

Respiratory endotyping is a methodology for identifying the pathophysiological traits of sleep apnea by better utilizing the wealth of data collected in a PSG. The endotypes of sleep apnea are loop-gain, upper airway collapsibility, arousal threshold, and upper airway dilator muscle response (compensation). [7] These parameters can be estimated by examining the characteristics of a patient’s ventilation during obstructed breathing periods. [7, 8, 1, 2, 9, 10]

The respiratory endotypes correspond directly to the pathophysiological mechanisms underlying sleep apnea, making it an attractive method for guiding treatment options. [1, 10] For instance, patients with poor pharyngeal muscle responsiveness may respond well to hypoglossal nerve stimulation [11] and/or drugs that increase the upper airway dilator muscle activity. [12] Similarly, patients with an abnormally low arousal threshold may respond better to sedative drugs. [13] Other treatment options that have been shown to target specific endotypes are: supplemental oxygen, [14] upper airway surgery, [15, 16] and oral appliances. [17] Endotyping could further help to guide combination therapy by identifying which traits to target with drugs or devices, for instance, individuals with a collapsible airway and a high loop gain might respond to the combination of oral
appliance and acetazolamide. [18] A more complete review of endotype driven OSA treatment options can be found in a recent paper by Edwards et al. [19]

There are two established methods of determining the endotypes of sleep apnea during sleep. The first involves dropping CPAP pressure, creating a controlled apnea or hypopnea, and measuring the ventilation response when CPAP pressure is re-established. [7, 8, 18, 20] The second method requires measuring the respiratory drive response to obstruction using esophageal manometry [10] or diaphragm electromyogram (EMG). [9] These methods have several shortcomings: they are not scalable, cause discomfort to the patient and use equipment that is not a part of a standard PSG. To mitigate these problems Sands et al. [1, 2] proposed a method for polysomnographic endotyping (implemented in a tool called “Phenotyping Using Polysomnography” or PUP) where the endotypes can be estimated from a standard PSG. The PUP method relies on an uncalibrated estimate of minute ventilation during sleep, derived from the measured flow signal, and it uses inverse modelling of the respiratory control system to estimate respiratory drive.

In this paper we present a python implementation (PUPpy) of the PUP method. With the reimplementation we aim to reproduce and validate the PUP method with the goal of building a foundation for cloud-based software making the method accessible for a broader audience. The PUP method is re-implemented in a different programming language, building on its theoretical basis, with improvements with regards to efficiency. The cloud-based platform was chosen for its scalability, making it possible to run these computationally intensive methods at scale. This implementation represents an important step in making endotyping more accessible for both research and the clinic and lays a solid foundation for further developments of polysomnographic endotyping. Moreover, re-implementing the method from first principles and comparing the endotype results between the two versions serves as an independent validation of the integrity of the method itself since implementation errors are not likely to be replicated.
Methods

The control and regulation of ventilation can be described by a feedback system whereby a reduction in minute ventilation ($V_t$) raises the pCO$_2$ in the blood and causes a corresponding change in “chemical drive” to breathe ($V_{chem}$). This feedback system can be formulated as a first-order linear model with a transport delay, Equation 1. [7, 18, 21] which captures the magnitude ($LG_0$, steady-state loop gain), response time ($\tau$, time constant) and latency ($\delta$, delay) of the chemical drive response to a drop in ventilation. The characteristics of the response depend on the model parameters: steady-state loop-gain, the respiratory time constant ($\tau$) and the feedback delay ($\delta$).

Engineers (or researchers using appropriate software) can analyze this model of the ventilatory control system by examining the governing equation for the feedback system, given by:

$$V_{chem}(s) = \frac{-LG_0}{1 + s\tau} e^{-s\delta} V_t(s) \tag{1}$$

For example, with known parameters, Equation 1 is used to calculate the “loop gain” (magnitude of chemical drive response for any ventilatory disturbance) that has been used to predict responses to therapies [15, 16, 18]. Polysomnographic endotyping employs Equation 1 to generate a continuous chemical drive $V_{chem}$ signal (output of the ventilatory control system) based on the measured ventilatory fluctuations ($V_t$). Parameters are adjusted through least squares regression (below).

Once ventilatory drive is estimated, the remaining endotypes can be quantified: upper airway collapsibility, upper airway compensation and the arousal threshold.

Figure 1 shows a simulation of the chemical drive response ($V_{chem}$) to a loss of ventilation ($V_t$) when a spontaneous hypopnea (blue area) interrupts normal (eupneic) breathing. Before the obstruction, the measured ventilation and chemical drive are identical. During the hypopnea, the loss of ventilation yields a gradual compensatory rise in chemical drive. When this drive reaches the arousal threshold, an arousal occurs (red area) and the obstruction is terminated. The open airway reveals the underlying elevation in chemical drive and yields a period of hyperventilation, which
eventually converges back to the baseline quiet breathing. During arousal (red area), a non-chemical drive to breath (wakefulness drive, parameter $\Upsilon$) also raises the overall ventilatory drive independently of the chemical drive. Chemical drive plus wakefulness drive is referred to as the ventilatory drive. The arousal threshold is defined as the chemical drive at arousal onset, expressed as a percentage of eupnea.

A. Minute Ventilation
The model in Equation 1 can be interpreted as an input-output system where the input ($V_d$) is the minute ventilation and the output ($V_{chem}$) is the chemical drive which is thought of as the intended minute ventilation and has the same units as $\dot{V}_c$. The minute ventilation is calculated from the measured flow signal on a breath-by-breath basis. The breath detection is carried out by integrating the flow trace, correcting for drift, and detecting the troughs and peaks of the resulting signal. When breaths have been identified, minute ventilation ($V_d$) is calculated by integrating the inspired flow for each individual breath, yielding a volume estimate, and dividing by the breath duration.

The PSG data used for the validation of PUPpy includes a flow signal recorded with a pneumotachograph equipped with a sealed oronasal mask. The pneumotachograph was calibrated at the start of the sleep study and the signal amplification set to provide a signal with a large amplitude without clipping. Although the pneumotachograph is calibrated at the beginning of the night the eupneic ventilation can drift over the span of a sleep study (i.e. the data is non-stationary). To account for drift in signal amplitude the continuous minute ventilation trace is normalized using a moving average window of 7 minutes. It is assumed that the amount of hyper- and hypoventilation even out over long periods and therefore the average is close to eupnea. This means that ventilation at 100% eupnea can be sustained indefinitely; values below 100% eupnea are interpreted as hypoventilation and above 100% eupnea as hyperventilation. All ventilation and drive estimates are expressed in the same way, as a percentage of eupnea. An additional upside of normalizing the
minute ventilation signal is that semi-quantitative, uncalibrated flow measurements can also be used.

In a clinical PSG study, changes in ventilation are typically measured using either a nasal cannula or two RIP belts measuring thoracoabdominal breathing movements. Both sensors have associated complications; the nasal cannula is prone to dislocations and confounded by oral ventilation while the RIP needs to be calibrated and is sensitive to movement artifacts. Previous research has proposed scaling the nasal cannula by an exponent; this transformation improves the correlation between endotypes derived from a sealed oronasal mask and nasal cannula [2, 1]. To the best of our knowledge no research has yet been done on endotyping with RIP using different calibration techniques.

B. Inverse modeling

The model in Equation 1 is algorithmically tuned such that it reflects the ventilatory patterns of the minute ventilation data for each subject. The implementation details of the model fitting process are detailed in the supplementary materials. The model is fit for an interval of 7 minutes at a time yielding both the model endotypes as well as the chemical drive estimate. The final product of the model fitting procedure for a single window of PSG data can be seen in Figure 2. Figure 2 (a) shows the measured flow for a period of repeated apneas (blue areas) and recovery breaths during cortical arousal (red areas). Figure 2 (b) shows the corresponding blood oxygen saturation. Figure 2 (c) shows the normalized minute ventilation and the estimated chemical drive, expressed as a percentage of eupnea.

Due to the physiological changes that occur during REM it is reasonable to assume that endotypes are different between REM and NREM sleep. Loop gain has been shown to decrease during REM [22]. Similarly, upper airway muscle compensation is diminished during REM due to reduced muscle tone. Furthermore, unlike NREM sleep, during REM sleep the ventilation is not under the sole regulation of the chemical control system so it is not clear how well the model in
Equation 1 applies. To minimize the confounding effects of REM sleep, epochs scored as REM are traditionally omitted from the endotype analysis. More research is needed to explore the difference between REM and NREM sleep apnea endotypes.

Although the goal was to replicate PUP, some architectural decisions were made to make the code fast, testable, and maintainable. In addition, changes were made to the data processing and parameter fitting procedure. Firstly, a sliding window was used for eupnea normalization to account for slow changes in signal amplitude (gradual changes in average ventilation). Secondly, in the original paper a modified least-squares regression algorithm fit a third order polynomial to the error and subtracted this prior to the mean squared error being calculated. This effectively high-pass filters the error, accounting for drift and non-zero-mean noise. To improve performance and avoid edge effects (outlier extremes at the start and end of the window) we opted to subtract a double-exponential moving average from the error instead of the polynomial, see implementation details in supplementary materials.

C. Deriving endotypes from model simulated drive

From the model fitting step, we directly get the loop-gain endotype from the best-fit parameters, Equation 1. The remaining endotypes are all based on chemical drive which can be estimated using the identified model. To derive the ventilatory endotypes of compensation and collapsibility, a ventilatory versus ventilatory drive ("endogram") plot is generated, see Figures 3(a)-(b). Chemical drive is binned into quantile bins and the median ventilation is plotted for each bin, black line, and the inter quartile range, gray area indicating the variance of the prediction. Figures 3(a)-(b) show the results from two subjects with different compensation endotype. From the figures we can read the collapsibility ($V_{\text{passive}}$) as the ventilation at 100% eupnic drive and $V_{\text{active}}$ as the ventilation at the
arousal threshold which is defined as the chemical drive at arousal onset as shown in Figure 1. [9]

The upper airway compensation $\dot{V}_{\text{comp}}$ is calculated using Equation 2. [1]

$$\dot{V}_{\text{comp}} = \dot{V}_{\text{active}} - \dot{V}_{\text{passive}}$$ (2)

D. Loop gain

Loop gain is the magnitude component of the system in Equation 1 and indicates the strength of the response to ventilatory disturbance (i.e. how high and fast the ventilatory drive increases following obstruction). The loop-gain derived in the CPAP-drop method is the steady-state loop-gain ($LG_0$). For reasons detailed in the supplementary materials $LG_0$ cannot be accurately derived from naturally occurring apneas. Therefore, $LG_1$, or the loop gain at 1 cycle per minute is used as a surrogate.

While elevated $LG_1$ (elevated chemoreflex sensitivity) will yield higher levels of drive for any reduction in ventilation, the chemoreflex delay also determines whether spontaneous periodic breathing (central sleep apnea) might occur. From the perspective of control theory, $LG_n$ or the loop gain at the natural frequency of the system (incorporates delay) is the parameter that determines ventilatory instability. Thus, we also examined $LG_n$ in the current analysis.

E. Deployment

The original PUP implementation was written in MATLAB as prototype software and has been made available online by the original authors. [2] The implementation documented in this paper was written entirely in Python 3.7. Python was selected as a programming language due to its mature data science stack and strong cross-platform support. The system was deployed to a fully managed cloud architecture for high scalability and fault-tolerance with minimal operation burden and will be made available to researchers and clinicians as a closed source cloud service. The goal of the cloud implementation is to make the PUP method accessible to a larger audience of scientists to facilitate the clinical validation and acceptance of the method.
F. Cohort

The comparison between methods is done using retrospective data collected by Brigham and Women’s Hospital. The dataset contained 38 measurements, 23 males and 15 females, Table 1. 16 of these had simultaneous measurements of nasal pressure and oronasal mask with a pneumotachograph. The data collection including polysomnographic setup and scoring criteria has been described elsewhere. [1]

G. Statistics

The agreement between the PUP and PUPpy methods was analyzed using intra-class correlation (ICC), Pearson correlation coefficient (PCC), and Bland-Altman (BA) mean error and agreement. All statistical analysis was carried out using Python 3.7, using the SciPy 1.2.3 and NumPy 1.17.5 libraries. ICC was calculated based on two-way mixed effects, single rater, and absolute agreement (ICC(2,1)). [23] Confidence intervals for ICC and PCC were calculated via bootstrapping where studies were randomly sampled with replacement over 10,000 iterations and the ICC and PCC coefficients calculated.

Results

We identify the endotypes in a dataset of 38 patients using both the MATLAB (PUP) and the new Python (PUPpy) implementation. This validates the entire pipeline, from data processing, breath detection, and minute ventilation calculations to model fitting and finally parameter derivation.

The validation was threefold: 1) Comparing the two implementations using oronasal pneumotachograph NREM sleep only. Using a gold standard measurement of ventilation and omitting REM sleep minimizes the confounding effects of sensor noise and model uncertainty. 2) The two implementations compared include all sleep stages; the clinical relevance of lumping together REM and NREM endotypes is not clear but is done here in order to verify that the re-implementation
is valid for all sleep stages. 3) To test feasibility for usage in clinical setting, endotypes derived from nasal cannula using PUPpy were compared to endotypes derived from oronasal pneumotachograph using PUPpy.

**A. NREM sleep using oronasal mask**

A list of summary metrics for the comparison of each endotype can be found in Table 2. All parameters derived using the PUPpy method correlate strongly with the PUP method ($p<10^{-6}$ for all) with ICC ≥ 0.90 (Pearson’s $r>0.90$) and are unbiased compared to the original PUP method for each endotype. Scatterplots are shown in Figure 4 and the corresponding Bland-Altman plots in Figure 5. In each subfigure of Figure 4 the Pearson correlation coefficient of the respective parameter is shown. Figure 4 shows the range in the endotype values as calculated by the PUP and PUPpy methods. The figures show that the endotype values cover broad ranges and the high correlations are not caused by single or few outliers or leverage points. Furthermore, the figures show that the endotype values sit around the 1:1 line plotted in a red dash line. Figure 5 shows the Bland-Altman plots for the endotypes. The figures illustrate the bias and agreement intervals. A summary of the endotype results for the cohort for each implementation are presented in Tables 3 and 4.

**B. All sleep stages, oronasal mask**

To validate the integrity of the re-implementation for all sleep stages (including REM sleep) we calculated the endotype values using the PUP and PUPpy methods for all sleep periods. Scatter plots and Bland-Altman plots comparing the endotype values from the PUP and PUPpy methods can be found in the supplementary materials Figures s2 and s3. Table 5 shows the summary statistics comparing the two methods. The table shows that the correlation is still very high for all the derived parameters, ICC > 0.90 (Pearson’s $r>0.90$). The table shows that including REM sleep does not change the agreement between the two implementations meaningfully compared to NREM only.
C. NREM sleep Nasal cannula

In clinical PSGs, nasal cannulas are used rather than oronasal mask and pneumotach. Here we compare the endotypes derived from the flow signals measured by the oronasal pneumotach and nasal cannula using the PUPpy method. In the dataset 16 sleep studies included a simultaneous measurement of breathing using the oronasal pneumotach and nasal cannula. The nasal cannula signal was transformed using a scaling exponent of 0.67 as suggested by Sands et al. [1] and Mann et al. [24]. The comparison was made in the same fashion as reported previously and the results are reported in Table 6 in a manner that is directly comparable to the same comparison done by Sands et al. [1]. The results show that the Pearson correlation between the pneumotach and cannula endotypes using the PUPpy method is high, $r>0.9$. Furthermore, the mean error (i.e. bias) and mean absolute error are small or comparable to what was reported by Sands et al. [1].

Discussion

Sleep apnea endotyping is a promising method of guiding OSA treatment but its exploration is still in early stages. Making sleep apnea endotyping widely-available to a broader range of researchers is the next major step limiting the determination of the clinical importance of the endotypes. Currently, due to the lack of large general population studies, there are no definitions of what constitutes clinically high or low endotype values, how the endotypes interact, and how to tailor treatment for a patient presenting with a combination of endotypic traits to alleviate the cause of their sleep apnea. Intra-subject endotype variability should also be explored, both for a single night and multiple nights.

It should be noted that even with excellent aggregate results between the PUP and PUPpy implementations ($r > 0.9$) there are individuals who deviate from the 1:1 line. This may be due to an individual demonstrating variability in endotype over a single night study, which would result in the endotype estimate being sensitive to slight algorithmic changes. Being able to provide a certainty estimate (i.e. a confidence interval) for each endotype may help in such cases. Confidence interval
calculations could help detect misclassified endotypes due to sensor faults, such as oronasal mask leak, poor signal quality (including: filter distortions, clipping and sensor dislocation), incorrect RIP belt placement or oral ventilation when flow is measured by a nasal cannula. It is, however, not clear how best to calculate these confidence intervals (simple measures of dispersion/variance versus state/position dependence) making it an interesting avenue for future research.

Since the model fitting relies on manual respiratory event scoring to mask out periods of ventilatory disturbance it can be assumed that accurate scoring of apneas and hypopneas is important. The same is true for arousals and sleep stages as these annotations are also used as inputs to the fitting routine. This observation merits further investigation and scoring guidelines for endotyping should be subsequently created. Some new evidence indicates that scoring based on airflow that relies less on desaturation may be preferable. [25]

When implementing complicated algorithms there is always a risk of systematic biases and errors. This can be especially difficult to handle in medical signal analysis software, since biological signals have inherent variations that may either be interpreted as a physiological characteristic or a calculation error. In the case of complex algorithms, such as the one described in this paper, these errors can be very subtle and hard to detect. An effective method to reveal such biases and errors is to re-implement the algorithm from a conceptual level and replicate previous experimental results. This is especially true when the re-implementation is done using a different programming languages and software libraries, since the odds of repeating the errors or biases in the same way are low.

The re-implementation of the PUP method was done as a cloud-enabled service rather than as locally run software. Cloud-enabled services have many benefits, especially for computationally demanding algorithms where the processing time of a local computer quickly becomes a limiting factor. The scalability of a cloud computing makes it possible to quickly batch-process larger datasets than would be practical running in the local computer environment. In addition, this approach simplifies issues of algorithm version control and deployment to other researchers. It is our hope
that the PUPpy implementation can serve as a platform for scientists who want to further investigate the clinical importance of the endotypes without having the expertise required to use the PUP implementation.

Conclusions
Respiratory endotyping is a promising method for differentiating physiological etiologies of sleep apnea and potential treatment guidance.

In this paper we have successfully re-implemented the endo-Phenotyping Using Polysomnography (PUP) method by Sands et al. [1] in a different software environment, validating the integrity of the original method. The outcome of the new implementation, PUPpy, was compared to the outcome of the original PUP and demonstrated an absence of systematic biases or errors. The primary validation was performed using flow measured by oronasal pneumotachography. Although this is not the standard method of measuring flow during a sleep study, the oronasal pneumotachography allowed us to validate the performance of the re-implementation without the influence of external factors caused by less reliable sensors. A thorough validation of the PUP and PUPpy implementations using breathing sensors such as nasal cannulas and RIP belts deserves its own publication is beyond the scope of the current work. Further investigation is needed to investigate how best to apply these methods on data from a standard sleep study with the presence of oral and nasal breathing, sensor movement, and other sources of signal disturbances found in routine sleep studies.

PUPpy was implemented in Python which allows for the deployment of respiratory endotyping in a scalable cloud environment. The re-implementation and its validation serve as a first step in developing a cloud service to provide access to the PUP method. As a result, sleep apnea endotyping can be offered to researchers who may not have the resources or expertise required to run the previously published MATLAB PUP method. This improved access will allow scientists to further investigate the clinical relevance of the endotypes.
Acknowledgements

None

Disclosure Statement

This work was supported by the Icelandic Centre for Research RANNÍS (153523-613), the European Union’s Horizon 2020 SME Instrument (733461), and the American Heart Association (15SDG25890059).

The study was approved by the Brigham and Women’s Hospital Institutional Review Board, and all participants provided informed, written consent before participation in the study.

Scott A. Sands and Andrew Wellman are consultants to Nox Medical. Dr. Sands also serves as a consultant for Merck and Apnimed, and has project grant support from Apnimed, Prosomnus, and Dynaflex.

Andrew Wellman works as a consultant for Apnimed, Nox, Inspire, and Somnifix. He has received grants from Sanofi and Somnifix. He also has a financial interest in Apnimed Corp., a company developing pharmacologic therapies for sleep apnea. Dr. Wellman’s interests were reviewed and are managed by Brigham and Women’s Hospital and Partners HealthCare in accordance with their conflict of interest policies.

Non-financial Disclosure: None
References

1. Sands SA, Edwards BA, Terrill PI, et al. Phenotyping pharyngeal pathophysiology using polysomnography in patients with obstructive sleep apnea. American journal of respiratory and critical care medicine. 2018;197:1187-1197.

2. Terrill PI, Edwards BA, Nemati S, et al. Quantifying the ventilatory control contribution to sleep apnoea using polysomnography. European Respiratory Journal. 2015;45:408-418.

3. Bosi M, De Vito A, Kotecha B, et al. Phenotyping the pathophysiology of obstructive sleep apnea using polygraphy/polysomnography: a review of the literature. Sleep and Breathing. 2018;22:579-592.

4. Berry RB, Albertario CL, Harding SM, et al. The AASM Manual for the Scoring of Sleep and Associated Events; Rules, Terminology and Technical Specifications, version 2.5: American Academy of Sleep Medicine; 2018.

5. Rapoport DM. POINT: Is the apnea-hypopnea index the best way to quantify the severity of sleep-disordered breathing? Yes. Chest. 2016;149:14-16.

6. Punjabi N. COUNTERPOINT: Is the Apnea-Hypopnea Index the Best Way of Quantifying the Severity of Sleep-Disordered Breathing? No. Chest. January 2016;149(1):16-19.

7. Wellman A, Eckert DJ, Jordan AS, et al. A method for measuring and modeling the physiological traits causing obstructive sleep apnea. Journal of applied physiology. 2011;110:1627-1637.

8. Wellman A, Edwards BA, Sands SA, et al. A simplified method for determining phenotypic traits in patients with obstructive sleep apnea. Journal of Applied Physiology. 2013;114:911-922.

9. Sands SA, Terrill PI, Edwards BA, et al. Quantifying the arousal threshold using polysomnography in obstructive sleep apnea. Sleep. 2017;41:zsx183.

10. Eckert DJ, White DP, Jordan AS, Malhotra A, Wellman A. Defining phenotypic causes of obstructive sleep apnea. Identification of novel therapeutic targets. American journal of respiratory and critical care medicine. 2013;188:996-1004.

11. Eastwood PR, Barnes M, Walsh JH, et al. Treating obstructive sleep apnea with hypoglossal nerve stimulation. Sleep. 2011;34:1479-1486.

12. Taranto-Montemurro L, Messineo L, Sands SA, et al. The Combination of Atomoxetine and Oxybutynin Greatly Reduces Obstructive Sleep Apnea Severity. A Randomized, Placebo-controlled, Double-Blind Crossover Trial. American journal of respiratory and critical care medicine. 2019;199:1267-1276.

13. Eckert DJ, Owens RL, Kehlmann GB, et al. Eszopiclone increases the respiratory arousal threshold and lowers the apnoea/hypopnoea index in obstructive sleep apnoea patients with a low arousal threshold. Clinical science. 2011;120:505-514.

14. Sands SA, Edwards BA, Terrill PI, et al. Identifying obstructive sleep apnoea patients responsive to supplemental oxygen therapy. European Respiratory Journal. 2018;52:1800674.
15. Joosten SA, Leong P, Landry SA, et al. Loop gain predicts the response to upper airway surgery in patients with obstructive sleep apnea. Sleep. 2017;40:zsx094.

16. Li Y, Ye J, Han D, et al. The Effect of Upper Airway Surgery on Loop Gain in Obstructive Sleep Apnea. Journal of Clinical Sleep Medicine. 2019;15:907-913.

17. Bamagoos AA, Cistulli PA, Sutherland K, et al. Polysomnographic endotyping to select patients with obstructive sleep apnea for oral appliances. Annals of the American Thoracic Society. 2019;16:1422-1431.

18. Edwards BA, Sands SA, Eckert DJ, et al. Acetazolamide improves loop gain but not the other physiological traits causing obstructive sleep apnoea. The Journal of physiology. 2012;590:1199-1211.

19. Edwards BA, Redline S, Sands SA, Owens RL. More than the sum of the respiratory events: personalized medicine approaches for obstructive sleep apnea. American journal of respiratory and critical care medicine. 2019;200(6):691-703.

20. Messineo L, Magri R, Corda L, Pini L, Taranto-Montemurro L, Tantucci C. Phenotyping-based treatment improves obstructive sleep apnea symptoms and severity: a pilot study. Sleep and Breathing. 2017;21:861-868.

21. Khoo MCK. Physiological Control Systems: Analysis, Simulation and Estimation.: IEEE Press Series in Biomedical Engineering; 2000.

22. Messineo L, Taranto-Montemurro L, Azarbarzin A, et al. Loop gain in REM versus non-REM sleep using CPAP manipulation: A pilot study. Respirology. 2019;24(8):805-808.

23. Koo T, Li M. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. Jun 2016;15(2):155-163.

24. Mann DL, Terrill PI, Azarbarzin A, et al. Quantifying the magnitude of pharyngeal obstruction during sleep using airflow shape. European Respiratory Journal. 2019;54(1).

25. Landry SA, Joosten SA, Thomson LD, et al. Effect of Hypopnea Scoring Criteria on Noninvasive Assessment of Loop Gain and Surgical Outcome Prediction. Annals of the American Thoracic Society. 2020;17(4):484-491.
Figure captions

**Figure 1:** A simulated hypopnea for illustrating key concepts of the model underlying the endo-
Phenotyping Using Polysomnography (PUP) method.

(a) The blue trace shows the simulated flow during an hypopnea event and the black trace shows a
simulation of the flow that the chemical drive would result in if there were no obstruction (intended
flow). The hypopnea starts at around 40 seconds where the airflow is reduced. As the flow decreases
the chemical drive increases, attempting to compensate for the reduced ventilation. The hypopnea
terminates at the 75 second mark in an arousal, opening the airway. The buildup of intended flow
results in large recovery breaths before stabilizing at eupnea.

(b) The ventilation during the simulated hypopnea in (a). The hypopnea event, labelled by the blue
square, terminates in an arousal, labelled by the red square. The blue trace shows the change in
ventilation, calculated from the simulated flow signal in (a). The black and green traces show the
chemical and ventilatory drives, respectively. Due to the blood circulatory delay the drive only starts
increasing 12 seconds after the ventilation is reduced and similarly continues to build up 12 seconds
into the recovery period. The chemical drive is the $V_{chem}$ from Equation 1 and the ventilatory drive is
$V_{chem}$ plus an added drive contribution during the arousal (wakefulness drive, $\gamma$), simulated here as
20% *eupnea*. The arousal threshold can be read directly from the chemical drive estimate as the
drive at arousal onset.

**Figure 2:** Example of the model fitting on PSG data during an episode of obstructed breathing.

(a) The measured flow signal (blue trace) with scored obstructive apneas (blue areas) and arousals
(red areas). Flow has been normalized to the [-1, 1] range where inspiration takes a positive value.

(b) The oxygen saturation corresponding to the flow.

(c) The normalized minute ventilation ($V_E$, blue trace) is derived from the measured flow. Chemical
drive ($V_{chem}$, black trace) was estimated using the identified feedback model in Equation 1. The
chemical drive increases during the obstructed periods resulting in large recovery breaths when obstruction is terminated.

**Figure 3:** Drive-ventilation graph with derived endotypes. The black solid line shows the median ventilation as a function of chemical drive for a single sleep study. The grey areas indicate the interquartile range of ventilation as a function of drive. $V_{\text{passive}}$ is the ventilation at eupnea drive and $V_{\text{active}}$ is the ventilation at the arousal threshold. This representation of drive and ventilation helps illustrate negative effort dependence (NED) as well as the endotype of upper airway compensation $V_{\text{comp}} = V_{\text{active}} - V_{\text{passive}}$, labelled by an arrow to the right of the figure.

(a) The figure shows a patient with effort dependent reduction of airflow with low collapsibility ($V_{\text{passive}}$ is close to eupnea) and low compensation ($V_{\text{active}} < V_{\text{passive}}$).

(b) The figure shows a patient with a similar collapsibility as the patient in (a) ($V_{\text{passive}}$) but better compensation (since $V_{\text{active}} > V_{\text{passive}}$). This is a sign of a good functional response of airway muscles since the airway remains open as drive increases.

**Figure 4:** Scatter plots of the relevant endotype parameters identified using the two endotype Phenotype Using Polysomnography (PUP) implementations. (a) loop gain at 1 cycle/min, (b) loop gain at the system’s natural frequency n cycle/min, (c) $V_{\text{active}}$, (d) $V_{\text{passive}}$, (e) circulatory delay ($\delta$), (f) arousal threshold.

**Figure 5:** Bland-Altman plots of the relevant endotype parameters identified using the two endotype Phenotype Using Polysomnography (PUP) implementations. The dotted centerline is the mean of the difference (PUPpy - PUP) and the red dotted lines signify the ±1.96 SD agreement interval. (a) loop gain at 1 cycle/min, (b) loop gain at the system’s natural frequency n cycle/min, (c) $V_{\text{active}}$, (d) $V_{\text{passive}}$, (e) circulatory delay ($\delta$), (f) arousal threshold.
Tables and captions

Patient cohort metadata

Table 1 Summary statistics of the patient cohort.

|        | Age [years] | BMI [kg / m²] | AHI [events / hour] |
|--------|-------------|---------------|---------------------|
| Mean   | 55.18       | 33.74         | 31.08               |
| Std    | 11.26       | 7.66          | 28.34               |
| Min    | 22.25       | 21.83         | 5.28                |
| Max    | 70.05       | 53.51         | 117.82              |
NREM sleep stages endotype comparison

Table 2 Summary of results using NREM sleep only and an oronasal mask for flow. Pearson correlation coefficient (PCC), intra-class correlation (ICC), Bland-Altman (BA) mean error, and agreement intervals (±1.96 SD). 95% confidence intervals in parenthesis. Confidence intervals are calculated using a bootstrap method.

|                | ICC          | PCC          | Mean error   | 95% agr. interv. |
|----------------|--------------|--------------|--------------|------------------|
| LG₁            | 0.96 (0.92, 0.98) | 0.96 (0.93, 0.98) | 0.01 (-0.01, 0.03) | ±0.10 (0.08, 0.12) |
| LGₙ            | 0.95 (0.86, 0.97) | 0.95 (0.86, 0.98) | 0.01 (-0.01, 0.02) | ±0.07 (0.06, 0.08) |
| Delay          | 0.91 (0.83, 0.95) | 0.91 (0.84, 0.96) | -0.08 (-0.37, 0.19) | ±1.73 (1.31, 2.06) |
| Ar.Thresh.     | 0.90 (0.82, 0.95) | 0.92 (0.90, 0.96) | 4.76 (1.91, 8.05)  | ±19.23 (12.25, 25.74) |
| V.passive      | 0.97 (0.92, 0.99) | 0.98 (0.94, 0.99) | 1.73 (0.35, 3.13)  | ±8.59 (4.49, 11.77) |
| V.active       | 0.97 (0.92, 0.99) | 0.97 (0.93, 0.99) | -0.17 (-2.92, 2.77) | ±17.49 (11.65, 23.20) |
Endotype results for cohort using PUP

Table 3 Summary of the endotype results for the cohort using the original PUP implementation, NREM sleep only and an oronasal mask for flow.

|    | LG₁ | LGₙ | Delay [s] | Ar.Thresh. [% eupnea] | V_{passive} [% eupnea] | V_{active} [% eupnea] |
|----|-----|-----|-----------|-----------------------|-----------------------|----------------------|
| Mean | 0.65 | 0.42 | 10.05     | 119.86                | 83.58                 | 86.86                |
| Std  | 0.19 | 0.12 | 2.12      | 21.96                 | 21.14                 | 34.01                |
| Median | 0.61 | 0.39 | 10.32     | 116.50                | 92.49                 | 94.76                |
| Min  | 0.36 | 0.25 | 5.07      | 89.85                 | 0.00                  | 0.00                 |
| Max  | 1.13 | 0.86 | 13.52     | 201.33                | 100.68                | 173.55               |
Endotype results for cohort using PUPpy

Table 4 Summary of the endotype results for the cohort using the re-implemented PUPpy, NREM sleep only and an oronasal mask for flows.

|          | LG₁  | LG₂  | Delay [s] | Ar.Thresh. [% eupnea] | V\textsubscript{passive} [% eupnea] | V\textsubscript{active} [% eupnea] |
|----------|------|------|-----------|-----------------------|-----------------------------------|----------------------------------|
| Mean     | 0.66 | 0.42 | 9.97      | 124.62                | 85.31                             | 86.69                            |
| Std      | 0.18 | 0.11 | 2.12      | 20.65                 | 20.65                             | 34.71                            |
| Median   | 0.65 | 0.40 | 9.77      | 117.64                | 93.54                             | 94.60                            |
| Min      | 0.40 | 0.29 | 5.36      | 99.60                 | 1.65                              | 1.73                             |
| Max      | 1.16 | 0.84 | 13.89     | 201.24                | 101.76                            | 169.47                           |
All sleep stages endotype comparison

Table 5 Summary of results using all sleep stages and an oronasal mask for flow. Pearson correlation coefficient (PCC), intra-class correlation (ICC), Bland-Altman (BA) mean error, and agreement intervals (±1.96 SD). 95% confidence intervals in parenthesis. Confidence intervals are calculated using a bootstrap method.

|                | ICC      | PCC      | Mean error  | 95% agr. interv. |
|----------------|----------|----------|-------------|------------------|
| LG1            | 0.96 (0.93, 0.98) | 0.96 (0.93, 0.98) | 0.00 (-0.01, 0.02) | ±0.10 (0.08, 0.11) |
| LGn            | 0.96 (0.93, 0.98) | 0.94 (0.85, 0.98) | 0.00 (-0.01, 0.02) | ±0.07 (0.06, 0.08) |
| Delay          | 0.93 (0.85, 0.97) | 0.93 (0.86 0.97) | -0.08 (-0.35, 0.19) | ±1.65 (1.20, 2.04) |
| Ar.Thresh.     | 0.92 (0.86, 0.96) | 0.94 (0.91, 0.97) | 3.87 (1.26, 6.87)   | ±17.35 (11.09, 23.14) |
| V_{passive}    | 0.98 (0.91, 0.99) | 0.98 (0.93, 0.99) | 1.60 (0.19, 2.98)   | ±8.52 (4.88 11.53)  |
| V_{active}     | 0.96 (0.91, 0.98) | 0.95 (0.91, 0.98) | 0.83 (-2.47, 4.17)  | ±20.39 (13.70, 26.04) |
Ventilatory endotypes, comparing nasal cannula to oronasal mask

Table 6 Summary of results when comparing the oronasal mask endotypes to the nasal cannula using the PUPpy implementation. Error is calculated for each of the three ventilatory endotypes as the endotype value using nasal cannula minus the endotype value when using oronasal pneumotachograph. Pearson correlation coefficient (PCC) is reported with 95% confidence interval, calculated using a bootstrap method.

|        | Mean error (±SD) [ % eupnea] | Mean absolute error [% eupnea] | PCC (95% CI) |
|--------|-----------------------------|-------------------------------|--------------|
| \(V_{\text{passive}}\) | -0.81 ± 4.98 | 3.72 | 0.93 (0.7, 0.98) |
| \(V_{\text{active}}\) | -0.24 ± 11.07 | 7.51 | 0.92 (0.73, 0.98) |
| \(V_{\text{comp}}\) | 0.58 ± 10.65 | 8.59 | 0.94 (0.42, 0.99) |
(a) Arousal Threshold = 184
    $V_{\text{passive}} = 88$
    $V_{\text{active}} = 3$

(b) Arousal Threshold = 122
    $V_{\text{passive}} = 91$
    $V_{\text{active}} = 116$
(a) Loop gain at 1 cycle/min
(b) Loop gain at the system's natural frequency n cycle/min
(c) $V_{active}$
(d) $V_{passive}$
(e) Circulatory delay ($\delta$)
(f) Arousal threshold
(a) Loop gain at 1 cycle/min

(b) Loop gain at the system’s natural frequency n cycle/min

(c) $V_{active}$

(d) $V_{passive}$

(e) Circulatory delay ($\delta$)

(f) Arousal threshold