



ORIGINAL ARTICLE

Within-night repeatability and long-term consistency of sleep apnea endotypes: the Multi-Ethnic Study of Atherosclerosis and Osteoporotic Fractures in Men Study

Raichel M. Alex, Tamar Sofer^o, Ali Azarbarzin^o, Daniel Vena^o, Laura K. Gell, Andrew Wellman, David P. White, Susan Redline and Scott A. Sands*

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

*Corresponding author. Scott A. Sands, Division of Sleep Medicine, Brigham and Women's Hospital, 221 Longwood Ave, Boston, MA 02115, USA. Email: sasands@bwh.harvard.edu.

Abstract

Study Objectives: Obstructive sleep apnea (OSA) is characterized by multiple “endotypic traits,” including pharyngeal collapsibility, muscle compensation, loop gain, and arousal threshold. Here, we examined (1) within-night repeatability, (2) long-term consistency, and (3) influences of body position and sleep state, of endotypic traits estimated from in-home polysomnography in mild-to-severe OSA (apnea-hypopnea index, AHI > 5 events/h).

Methods: Within-night repeatability was assessed using Multi-Ethnic Study of Atherosclerosis (MESA): Traits derived separately from “odd” and “even” 30-min periods were correlated and regression (error vs. *N* windows available) provided a recommended amount of data for acceptable repeatability ($R_{\text{threshold}} = 0.7$). Long-term consistency was assessed using the Osteoporotic Fractures in Men Study (MrOS) at two time points 6.5 ± 0.7 years apart, before and after accounting for across-year body position and sleep state differences. Within-night dependence of traits on position and state (MESA plus MrOS data) was estimated using bootstrapping.

Results: Within-night repeatability for traits ranged from $R = 0.62$ – 0.79 and improved to $R = 0.69$ – 0.83 when recommended amounts of data were available (20–35 7-min windows, available in 94%–98% of participants); repeatability was similar for collapsibility, loop gain, and arousal threshold ($R = 0.79$ – 0.83), but lower for compensation ($R = 0.69$). Long-term consistency was modest ($R = 0.30$ – 0.61) and improved ($R = 0.36$ – 0.63) after accounting for position and state differences. Position/state analysis revealed reduced loop gain in REM and reduced collapsibility in N3.

Conclusions: Endotypic traits can be obtained with acceptable repeatability. Long-term consistency was modest but improved after accounting for position and state changes. These data support the use of endotypic assessments in large-scale epidemiological studies.

Clinical Trial Information: The data used in the manuscript are from observational cohort studies and are not a part of the clinical trial.

Statement of Significance

Obstructive sleep apnea is caused by pathophysiological mechanisms or “endotypic traits,” including upper airway collapsibility, arousal threshold, ventilatory stability (loop gain), and upper airway muscle compensation. Quantifying these traits using routine polysomnography may help to disentangle sources of disease heterogeneity. Our study is the first to quantify within-night measurement error, longitudinal (over 6.5 years) consistency, and the impact of the changes in body position and sleep states for these key endotypic traits in large community-based studies. Trait consistency across years is comparable to that of the apnea-hypopnea index. This information supports the utility of using these traits for clinical and research purposes and provides guidance on how to optimize trait reliability.

Key words: sleep apnea; endotype; reliability; pathophysiology; precision medicine

Submitted: 28 January, 2022; Revised: 20 May, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Sleep Research Society. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Introduction

Obstructive sleep apnea (OSA) is a heterogeneous disorder caused by variable combinations of anatomic risk factors, obesity, and deficits in multiple endotypic traits, namely greater pharyngeal collapsibility, reduced upper airway dilator muscle compensation, unstable respiratory control (high loop gain), and a low arousal threshold [1–4]. Notably, an estimated 69% of OSA patients exhibit deficits in one or more endotypic traits (upper airway collapsibility, low arousal threshold, and high loop gain) [3]. Accumulating evidence indicates that individual differences in pathophysiological endotypic traits, estimated using polysomnographic airflow and an estimated ventilatory drive signal [5–7], may provide predictive insight into how well patients respond to multiple non-CPAP treatments for OSA. For example, elevated loop gain is associated with reduced efficacy of pharyngeal surgery, oral appliances, hypoglossal nerve stimulation, and atomoxetine-plus-oxybutynin, but is associated with increased efficacy of supplemental oxygen [8–13]. Greater pharyngeal muscle compensation is associated with increased efficacy of supplemental oxygen, hypoglossal nerve stimulation, and atomoxetine-plus-oxybutynin but reduced oral appliance efficacy [10–12]. A lower arousal threshold is a risk factor for reduced efficacy of hypoglossal nerve stimulation [13]. More severe collapsibility is likely to be a risk factor for reduced efficacy for a range of CPAP alternatives [10]. Estimation of endotypic traits may be a promising approach to characterize subgroups of patients with different underlying risk factors for OSA, who potentially could benefit from alternative treatments.

In physiology laboratories, endotypic traits can be obtained by invasive measurements such as intraesophageal diaphragm EMG [6, 14] or using specialized CPAP manipulations [15, 16]. To facilitate widespread use of endotypic traits beyond the physiology laboratory, we developed a method to non-invasively estimate the endotypic traits from a diagnostic polysomnography (PSG) using a nasal cannula signal, without the need for CPAP manipulation or specialized measurements [5–7]. Briefly, the method involves estimating ventilation and ventilatory drive throughout the night to calculate the traits; median values for each trait are then taken to represent the pathophysiology of each individual. However, several practical questions must be addressed before these trait measures will be ready for future clinical or epidemiological applications: how much data are needed to obtain a repeatable measurement within a night? Are some trait measures more reliable than others? Are trait measurements consistent across multiple years? To what extent do changes in body position and sleep state influence the estimates?

Accordingly, in this study, we leveraged existing in-home PSG data collected in large community-based samples of middle-aged or older individuals to determine (1) the within-night repeatability (measurement error) of the endotypic traits, using cross-sectional data from the Multi-Ethnic Study of Atherosclerosis (MESA), (2) the long-term consistency (approximately 6.5 years) of endotypic traits, which reflects both physiological variability plus measurement error, using polysomnographic data from the longitudinal Osteoporotic Fractures in Men Study (MrOS), and (3) influences of body position and sleep state on endotypic traits measurements using both MESA and MrOS.

Methods

Participants and study design

Multi-Ethnic Study of Atherosclerosis (MESA). The MESA is a U.S.-based prospective cohort study designed to assess risk factors for cardiovascular disease (CVD) in white, black/African American, Hispanic, and Chinese adults [17, 18]. At baseline (Exam 1; 2000–2002) 6814 men and women, ages 45–84 years without clinically apparent CVD, were recruited and followed longitudinally with additional assessments. A subgroup participated in the MESA Sleep Ancillary Study (MESA Exam 5, 2010–2013, N = 2237) which included in-home PSG (Compumedics Somte System, Compumedics Ltd., Abbotsville, AU) that recorded electroencephalography, electrooculography, chin electromyography, electrocardiography, thoracoabdominal movements, finger pulse oximetry, and airflow (assessed by thermistor and nasal pressure cannula). Sleep, arousals, and respiratory events were scored by research technicians blinded to all other data according to standard criteria as summarized before [19, 20]. In this analysis, apneas were identified by >90% airflow reduction from pre-event baseline for ≥ 10 s; hypopneas were defined as $\geq 30\%$ airflow reduction with $\geq 3\%$ desaturation or arousal. In total, 2060 participants provided polysomnographic recordings for potential analysis.

The Osteoporotic Fractures in Men Study (MrOS). Between 2000 and 2002, 5994 community-dwelling men, 65 years or older, were enrolled at six clinical centers in a baseline examination of the Osteoporotic Fractures in Men Study (MrOS) [21] to assess risk factors for falls, fractures, and mortality. An ancillary sleep study was conducted between 2003 and 2005 to understand the relationship between sleep disorders and adverse outcomes. Unattended PSG (Sleep Visit 1) was conducted in 2907 participants using a similar protocol and equipment as in MESA [22, 23]. Between 2009 and 2012, a second sleep study assessment (Sleep Visit 2) was completed in 1055 participants who also participated in Sleep Visit 1, of whom 1026 provided polysomnographic recordings for potential analysis. During both sleep study visits, PSG was performed using a multichannel portable monitor (Compumedics Safiro Sleep Monitoring System, Compumedics Ltd., Abbotsville, Australia) that recorded electroencephalography, electrooculography, chin electromyography, electrocardiography, thoracoabdominal movements, finger pulse oximetry, and airflow assessed by thermistor and nasal pressure cannula. Sleep, arousals, and respiratory events were scored by the same Sleep Reading Center as for MESA; apneas and hypopneas were similarly defined in these analyses [19, 20]. Mean follow-up time between sleep visit 1 and 2 was approximately 6.5 years (SD = 0.7).

For both MESA and MrOS, inter- and intra-scoring reliability of the AHI were high (inter/intraclass correlations > 0.94).

Estimation of endotypic traits

From both MESA and MrOS polysomnograms, endotypic traits were estimated for non-REM and REM sleep using established automated methods [5–7]. In this method, the nasal pressure signal was linearized to provide a surrogate of ventilatory flow, that in turn was integrated to provide an uncalibrated breath-to-breath ventilation signal (tidal volume \times respiratory rate). The ventilation signal was normalized such that 100% approximates eupneic ventilation. Subsequently, windows of polysomnographic data containing sleep were identified (7-min

duration, 2 min steps; see [5]). “Ventilatory drive”—defined as the intended ventilation that would be seen if the pharyngeal airway was unobstructed—was estimated for each data window. Theoretically, the ventilatory drive becomes evident when the airway is reopened following an obstructive respiratory event. Ventilatory drive was calculated using the ventilation signal input to a simplified chemoreflex feedback control model; model parameters (gain, response time, delay) were adjusted to best fit the ventilatory drive signal to the ventilation signal between scored obstructive events [5, 24, 25]. Models were fit separately for each available 7-min window containing sleep. Using the derived ventilation and ventilatory drive signals, the following endotypic traits can be estimated.

Loop gain. Two measures of loop gain were estimated for each 7-min window. LG_1 represents the sensitivity of the ventilatory feedback loop, i.e. the magnitude of the ventilatory drive response to a specific 1 cycle/min disturbance. This reflects the product of chemoreflex sensitivity (hypoxic/hypercapnic ventilatory response) and plant gain (e.g. lung volume). We also estimated the overall ventilatory instability (LG_n , predisposition to cyclic central sleep apnea) that is the combined effect of sensitivity and delay (latency between reduced ventilation and increased ventilatory drive). Median values were used to summarize multiple windows.

Arousal threshold. The ventilatory drive immediately prior to each scored EEG arousal (e.g. at the termination of a respiratory event) was identified, and the arousal threshold was calculated as the mean value of these ventilatory drive values for each 7-min window. Lower values reflect greater arousability from sleep [7]. Median values were used to summarize multiple windows.

Pharyngeal collapsibility and compensation. Breath-by-breath values of ventilation and the ventilatory drive were tabulated (excluding breaths during arousals and within two breaths after arousals/sleep onset) from all available windows. Next, breaths were sorted into 10 bins (deciles) based on ventilatory drive; for each bin, median values of ventilation and ventilatory drive were obtained and plotted. Linear interpolation between bins was then used to find the median ventilation at normal drive (“ $V_{passive}$ ”) [26, 27]; a lower $V_{passive}$ reflects a greater collapsibility. To calculate *compensation*, we first calculate V_{active} (median ventilation when drive is at the arousal threshold, i.e. maximal); V_{active} minus $V_{passive}$ is the pharyngeal muscle compensation.

Analysis was automated and executed using custom software (Phenotyping Using Polysomnography; MATLAB, Mathworks, Natick, MA).

Artifact rejection for nasal pressure signal. A major challenge for estimation of ventilation using nasal pressure is that the respiratory airflow signal can occasionally be contaminated by noise, e.g. when the intranasal pressure swings are small, when the cannula prongs are no longer in the nares, when there is nasal congestion, or in the presence of severe mouth breathing (or when the noise floor of the device is high). To automatically reject 7-min windows with low signal-to-noise ratio (SNR) and/or unusually small airflow signals, we (1) calculated SNR as the nasal pressure signal power in the respiratory range (0.1–1 Hz)

versus noise in the higher frequency range (1–10 Hz), and (2) calculated signal magnitude (again power in 0.1–1 Hz range) compared with typical power for the night (mean signal magnitude on a log scale of all data with SNR > 10 units) [28]. If 10% or more of a window exhibited small and/or noisy signals (product of SNR and signal magnitude < 0.125) that window was excluded from further analysis. Based on this approach, ~80% of the nasal pressure data during sleep was considered noise free and suitable for further analysis (MrOS visit 1: $80 \pm 26\%$, MrOS visit 2: $72 \pm 30\%$, MESA: $87 \pm 22\%$, mean \pm SD).

Since only windows containing respiratory events were subsequently used, the proportion of the night used for trait calculation was ~50% (MrOS visit 1: $53 \pm 25\%$, MrOS visit 2: $50 \pm 27\%$, MESA: $60 \pm 25\%$)

Statistical analysis

Quantification of the within-night repeatability and the long-term consistency of endotypic traits were obtained for participants with OSA (AHI > 5) in MESA and MrOS sleep visit datasets. Five endotypic traits variables were analyzed: collapsibility ($V_{passive}$), compensation, loop gain (LG_1), ventilatory instability (LG_n), and arousal threshold (defined above). Collapsibility ($V_{passive}$) and arousal threshold data were square-root transformed [12] (centered at $100\%_{eupnea}$) for normality; transformed variables are presented throughout the manuscript.

Within-night repeatability. To describe within-night measurement error (in MESA), two fully independent endotypic sets of trait measurements were obtained for each PSG by dividing the sleep data into “odd” and “even” 30-min periods that were analyzed separately. Any 7-min endotypic trait calculation windows that overlapped successive odd and even periods were removed to obtain independent odd and even trait estimates. To quantify repeatability, we calculated the coefficient of repeatability (CR) and Pearson correlation coefficient (R) between even and odd estimates. (Intraclass correlation coefficients were also calculated and provided near identical estimates, see Online Supplementary Table S1). CR provides an upper bound (95th centile) on the absolute difference between two repeated measurements and can be used to quantify repeatability in a manner that includes both mean bias as well as uncertainty [29]; if there is no mean bias the CR is simply equal to the width of the 95% limits of agreement (upper minus lower divided by 2). In addition, we modeled how repeatability (R-value) varied with the amount of available data (number of windows N): mean squared error for each pair of odd-even data (squared difference between “even” and predicted even values [based on odd values], normalized by mean squares) were plotted against window number N, and a hyperbolic link function was fit to the data (squared error falls as N rises). Model equation was given by: $error = 1/(\beta_0 + \beta_1[N \text{ windows}])$. R was calculated from $[1 - error]^{0.5}$ for any given N. From these models, we calculated the recommended number of windows required for obtaining a minimally acceptable “repeatable estimate” using $R_{threshold} = 0.7$. The coefficient of repeatability and correlation coefficients were then recalculated for the subset of participants who met the minimum data requirements to describe repeatability. We also calculated the percentage of participants who would have enough data from the whole night to provide a repeatable trait measure. To provide a comparison, an

analysis of AHI was performed using total sleep time to describe the amount of data required.

Long-term consistency. To describe long-term consistency (in MrOS), we compared measurements of endotypic traits obtained at sleep visit 1 and sleep visit 2 (mean follow-up time of 6.5 years). Because our goal was to assess the long-term consistency in physiological mechanisms, or “trait like” phenomena, our analytic approach was to minimize the impact of measurement noise. Hence, we analyzed participants with a sufficient number of windows to provide reliable trait estimates ($R_{\text{threshold}} = 0.7$, obtained from MESA above). The same approach was taken for AHI, to provide a benchmark for interpreting consistency. The similarity between the traits at sleep visit 1 vs. 2 were obtained by correlation analysis, the coefficient of repeatability and Bland-Altman limits of agreement (Intraclass correlation coefficients were also calculated and provided near identical estimates, see Online [Supplementary Table S2](#)). Furthermore, correlation analysis was also repeated after adjusting traits and AHI for between-study changes in (1) body position, (2) REM sleep duration, and finally (3) non-REM stages N1 and N3 duration (expressed as fraction of sleep time). Specifically, multivariable regression analysis modeled the difference between traits at visit 1 and visit 2 as a function of differences in position and state ($\Delta\text{trait} = \Delta\text{position} + \Delta\text{REM} + \Delta\text{N1} + \Delta\text{N3}$). The predicted difference (Δtrait) was then added to trait 1 values to provide a position and state-corrected trait 1 value for comparison with trait 2 values.

Influences of body position and sleep state. To describe the dependence of each trait on differences in time spent in different body positions (supine vs. lateral) or sleep states, we pooled data from MESA and MrOS sleep visit 1. Bootstrapping was performed to provide 50 modified “versions” of each sleep study made from the original study by resampling windows at random (with replacement). Due to random resampling, data in each version contained slightly different proportions of body positions and sleep states, which were tabulated with recalculated trait values. Mixed-model analysis (subject as a random effect) quantified associations between each trait and body position and sleep states. AHI was similarly analyzed for comparison. A meaningful association was considered as a change in trait values of above 0.5 % SD per %time (SD for each trait is taken from the across-sample population).

Results

Within-night repeatability of endotypes in MESA

From the initial 2060, MESA Sleep participants, 27 studies were of insufficient quality for scoring (did not provide AHI values), 215 participants had AHI < 5, a further 44 participants did not generate any endotype traits (e.g. noisy flow signal/ no good quality nasal pressure during respiratory events), leaving $N = 1774$ available for analysis ([Table 1](#)). A further 24 participants did not have 3 windows with respiratory events for any endotypic analysis in both odd and even 30-min periods leaving $N = 1750$ available for within-night repeatability assessment.

[Figure 1](#) shows illustrative plots of ventilation vs. ventilatory drive and endotypic traits during odd and even periods for

individual examples. Note that, visually, plots are very similar within individuals, but large differences across individuals are apparent. The average values of endotypic traits and AHI for the entire night (Trait Averages), odd period (“Odd” Estimate), and even period (“Even” Estimate) are given in [Table 2](#). On average, the number of 7-min analysis windows (N Windows in [Table 2](#)) available for deriving the odd and even estimates of endotypic traits ranged from 31 ± 14 for arousal threshold to 53 ± 21 for collapsibility and compensation (*nb.* windows for arousal threshold require presence of scored arousals). An average total sleep time of 180 ± 40 min was available for assessing the odd and even AHI estimates. The correlation coefficients between odd and even estimates of endotypic traits ([Table 2](#), [Figure 2 top panel](#)) ranged from 0.62–0.79; collapsibility (V_{passive}) exhibited the highest within-night repeatability ($R = 0.79$), and compensation exhibited the lowest value ($R = 0.62$). Within-night repeatability of AHI was very high, $R = 0.92$. To illustrate the agreement between odd and even estimates and their variability over time, we have included the Bland-Altman plot ([Figure 2](#), middle panel) and the coefficient of repeatability ([Table 2](#)). Limits of the agreement are also shown on the “correlation” X-Y plots ([Figure 2 top panel](#)) to facilitate interpretation of agreement.

Repeatability (assessed as measurement error) was improved in association with higher numbers of available analysis windows ([Table 2](#), right). Based on these associations, the recommended number of windows needed to obtain $R \geq 0.7$ ranged from 20 for arousal threshold to 35 for compensation. These numbers are equivalent to 26%–27% of the available number of windows over a whole night (77 ± 35 and 131 ± 53 respectively, Total N Windows in [Table 2](#)), illustrating that at least a quarter of a typical night is needed. By contrast, 45 min of sleep time was required for a repeatable AHI measure (~13% of the available 359 ± 81 min). Overall, the majority of the participants (94%, [Figure 2 bottom panel](#)) had the recommended number of windows for all endotypic traits (> 99% for AHI). After restricting analysis of repeatability to participants who had enough data to meet the $R = 0.7$ threshold, correlation coefficients ranged from 0.69–0.83; collapsibility (V_{passive}) was $R = 0.83$ and compensation was $R = 0.69$. Furthermore, accounting for differences in body position and state (odd vs. even) had a negligible impact on repeatability ($\leq 2\%$, results not shown). Coefficient of repeatability was calculated again after restricting the analysis to participants who had enough data ([Table 2](#), right), which showed improvement (smaller CR) for endotypic traits.

Long-term consistency of endotypes in MrOS

Out of the 2906 MrOS participants at Sleep Visit 1, 39 studies were of insufficient quality for scoring, 268 participants had AHI < 5209 participants did not provide any endotype traits (noisy flow signal/ no good quality nasal pressure during respiratory events), leaving $N = 2390$ available for analysis ([Table 1](#)).

A subset of visit 1 participants ($n = 1026$) had a follow-up sleep visit 2 (mean follow-up = 6.5 ± 0.7 years). Of these, 775 participants had OSA (AHI ≥ 5) and provided endotypic traits, 665 of whom also had paired measurements (AHI ≥ 5 and endotypic traits calculated) at visit 1. Based on the “minimum reliable” thresholds for the amount of data required for a valid measurement ($R_{\text{threshold}} = 0.7$), 595 (89.5%) participants provided “reliable” endotypic trait data for consistency analysis.

Table 1. Participant characteristics

Characteristic	Median [IQR]	N*
MESA		
Demographics and body habitus		
Age (years)	68 [61–76]	1774
Sex (N, M:F)	859:915	1774
Body mass index (kg/m ²)	28.1 [25.0–32.0]	1770
Polysomnography		
OSA severity (N, mild:moderate:severe)	594:592:588	1774
Apnea-hypopnea index, total (events/h)	21 [12.2–35.0]	1774
AHI non-REM (events/h)	17.4 [8.8–33.5]	1774
AHI REM (events/h)	34.1 [19.1–53.2]	1755
Arousal index (events/h)	20.8 [15.1–29.0]	1774
Total sleep time (min)	368 [312–414]	1774
Non-REM 1 (% total sleep time)	12.7 [8.8–18.5]	1774
Non-REM 2 (% total sleep time)	57.7 [51.0–64.0]	1774
Non-REM 3 (% total sleep time)	8.0 [2–15.3]	1774
REM (% total sleep time)	18.2 [13.6–22.3]	1774
Supine (% total sleep time)	35.5 [13.0–64.3]	1774
MrOS Sleep Visit 1		
Demographics and body habitus		
Age (years)	76 [72–80]	2390
Body mass index (kg/m ²)	27.0 [25.0–30.0]	2389
Polysomnography		
OSA severity (N, mild:moderate:severe)	864:854:672	2390
Apnea-hypopnea index, total (events/h)	19.0 [11.0–32.0]	2390
AHI non-REM (events/h)	17.0 [9.0–32.0]	2390
AHI REM (events/h)	25.0 [14.0–39.0]	2381
Arousal index (events/h)	22.3 [16.2–30.2]	2384
Total sleep time (min)	361 [317–401]	2390
Non-REM 1 (% total sleep time)	6.1 [4.1–8.7]	2390
Non-REM 2 (% total sleep time)	63.0 [56.5–69.4]	2390
Non-REM 3 (% total sleep time)	10.0 [3.8–16.7]	2390
REM (% total sleep time)	19.5 [14.7–23.6]	2390
Supine (% total sleep time)	31.6 [12.0–59.4]	2361

*Sample size N shows the number of participants available out of the 1774 participants with AHI >5 and endotypic traits for MESA. A subset of these participants who had a minimum number of windows (3 windows with respiratory events) for any endotypic analysis in both odd and even 30-min periods were used for within-night repeatability analysis (N = 1750, with AHI >5 for both odd and even 30-min periods and endotypic traits calculated); the subset had similar characteristics to the larger sample shown here (age = 68 [61–76], BMI = 28.2 [25.0–32.0], AHI = 21 [12.3–35.1]). For MrOS, sample size N represents the number of participants available out of the 2390 participants who had AHI >5 and endotypic traits at sleep visit 1. A subset of MrOS participants who had both sleep visit 1 and visit 2 and met criteria for minimum amount of data for “reliable” trait measurement were used for longitudinal consistency analysis (N = 595, with AHI >5 for both visits and endotypic traits calculated for both visits); characteristics of the subset are: At visit 1: age = 74 [71–78], BMI = 27 [25.0–29.0], AHI = 19 [11.0–30.0]; At visit 2: age = 80 [77–84], BMI = 27 [25.0–29.0], AHI = 21.0 [13.0–35.0]. OSA severity was defined based on AHI as: Mild (5 ≤ AHI < 15), Moderate (15 ≤ AHI < 30) and Severe (AHI ≥ 30).

Changes over time. Compared to visit 1 (mean age 74.7 years), at Visit 2 (mean age 81.0 years) there was an increase in compensation (+2.5 [0.3, 4.6] %_{eupnea}, estimate[95%CI]), a decrease in arousal threshold (−4.5 [−6.8, −2.2] %_{eupnea}), and increase in AHI (4.1 [2.9, 5.2] events/h, see Table 3). Changes in collapsibility and loop gain were not detected. In addition to this mean bias, we also included Bland-Altman’s limits of agreement (Figure 3, Table 3) to provide the agreement interval within which 95% of the differences between visit 1 and visit 2 fell for the same subjects. Coefficient of repeatability metric (Table 3) combined the mean bias and limits of agreement into a single variable to summarize

the across-year uncertainty in original units. As expected, the CR values for endotypic traits measured after 6.5 years were higher (~2 times or more) compared to the CR values obtained from within-night calculations.

The correlation coefficients for endotypic traits and AHI across time periods ranged from 0.30–0.61 (Table 3, Figure 3): collapsibility (V_{passive}) showed the lowest long-term consistency (R = 0.30), and loop gain displayed the highest value (R = 0.61). For comparison, long-term consistency of AHI was R = 0.47.

Consistency analysis was repeated after accounting for changes in body position and sleep states between the two visits: incorporating changes in body position improved the long-term consistency for collapsibility (R = 0.35) and arousal threshold (R increased from 0.37 to 0.43). Incorporating changes in REM sleep duration (% total sleep time) slightly improved long-term consistency of loop gain (R = 0.63). Incorporating changes in REM, N1, and N3 sleep durations further improved the consistency of collapsibility (R = 0.41).

Influences of body position and sleep state

Overall, the effects of variability in position and state on traits in MESA and MrOS were similar to or smaller than the effects on AHI, with the exception of a stronger effect of REM on loop gain, and stronger effect of N3 on collapsibility (Table 4). As defined *a priori*, meaningful influences (> 0.5 %SD/%time) were observed for: lateral (vs. supine) position on Arousal Threshold and AHI, but not the other traits. REM had a meaningful effect on loop gain (and ventilatory instability), Arousal Threshold, and AHI. Lighter sleep (greater N1) had a meaningful association with collapsibility, compensation, loop gain, instability, and AHI. Deeper sleep (greater N3) had a meaningful association with collapsibility and AHI, but not the other traits. Note, however, that even the largest effects were relatively small compared with the across-subject variability. For example, additional REM sleep by 10% of total sleep time would yield a reduction in estimated loop gain by just 0.03 points (0.175 SD). Likewise, additional N3 sleep by 10% of total sleep time would yield an improvement in collapsibility (increase in V_{passive}) by just 1.4%_{eupnea} (0.102 SD).

Discussion

The current study examined (1) within-night repeatability, (2) long-term consistency, and (3) dependence on sleep state and body position, of polysomnographic endotype measurements in individuals with mild-to-severe OSA. We found moderate to high levels of within-night repeatability for all endotypic traits (R ≥ 0.62), suggesting that measurement error—modest in magnitude—is unlikely to substantively reduce the utility of these measurements. As expected, the within-night repeatability of endotypic traits was contingent on the amount of data available. Whereas within-night repeatability of the trait measurements was lower than that observed for AHI, restricting analyses to studies that met minimum data requirements (20–35 windows) increased repeatability to R ≥ 0.69. Moreover, even with use of in-home PSG, sufficient data were obtained in the majority of participants (94%–98% subjects). Also, as expected, long-term consistency of the endotypic traits was lower compared to within-night analysis. However, consistency of the traits compared well with AHI, and likely reflected multiple

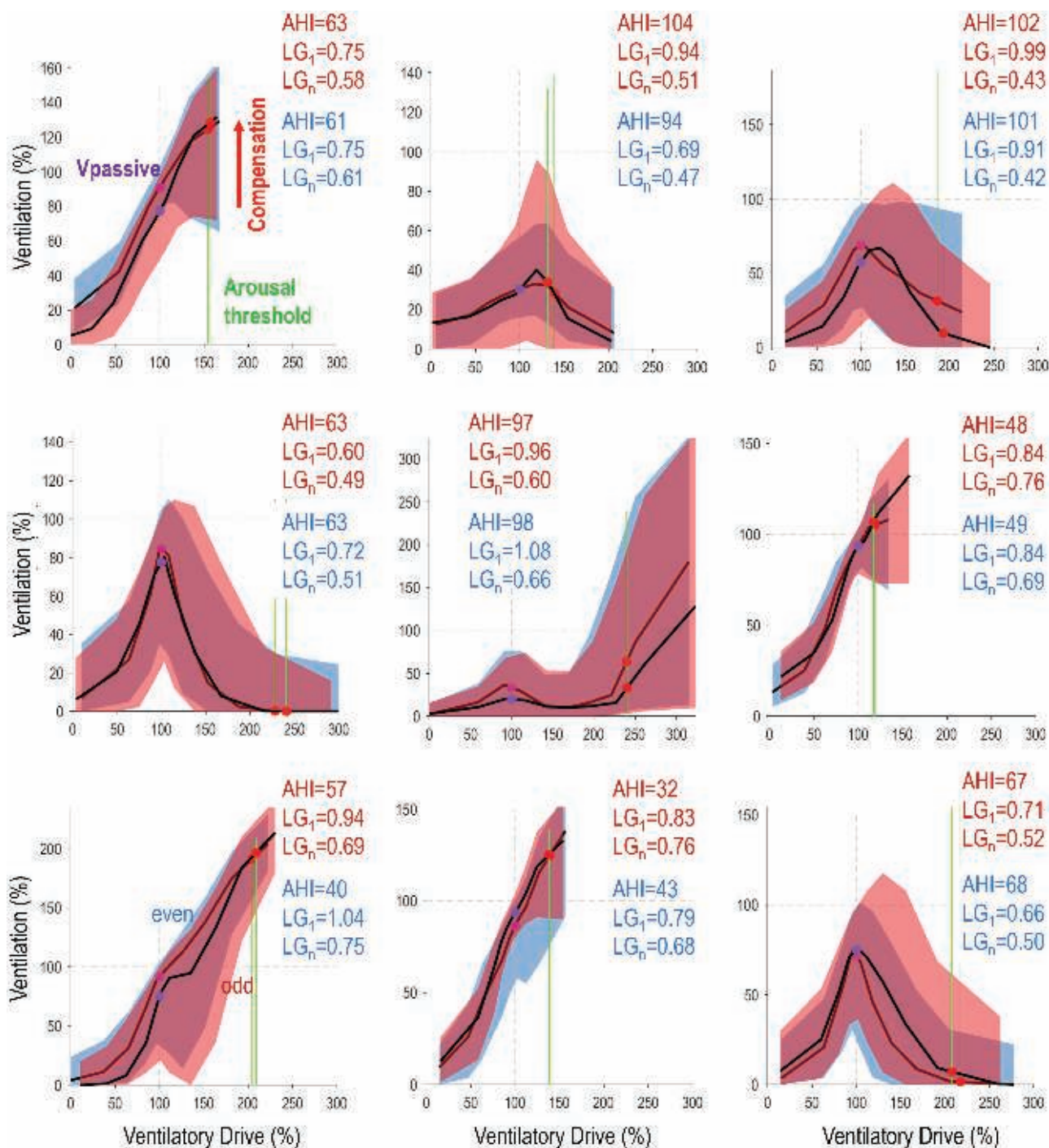


Figure 1. Examples illustrating within-night repeatability of endotypic traits (Ventilation-versus-ventilatory drive) and apnea-hypopnea index in 9 MESA participants. Two independent estimates are provided for each individual: Odd periods in red (40% of the night) are shown superimposed over even periods in blue (40% of the night). Note that, visually, plots are similar within individuals, but large differences across individuals are apparent. Accompanying values for apnea-hypopnea index (AHI), loop gain (LG_1) and ventilatory instability (LG_n) are also shown for odd and even periods (red and blue respectively). Values for $V_{passive}$ (purple dots, lower values indicate lower ventilation i.e. greater collapsibility), compensation (the increase in ventilation achieved when ventilatory drive rises to the arousal threshold), and arousal threshold (green vertical lines) are shown graphically. Arousal threshold and $V_{passive}$ values shown here are not transformed.

factors influencing sleep disordered breathing in an aging cohort studied over 6.5 years. Finally, we showed that differences in body position and state modestly influence the measured traits; yet effects of position and state on AHI was typically greater. Such an understanding is needed to appropriately interpret trait estimates and understand sources of bias for any future application in research and therapeutic decision-making.

Within-night repeatability

Prior to the current study, little was known about the repeatability and consistency of the polysomnographic endotypic traits. Our original paper described the within subject average standard error of the loop gain metric as $6 \pm 2\%$ [5]. Available data from supplemental oxygen therapy studies [12] also suggested that arousal threshold, collapsibility, and compensation

Table 2. Within-night repeatability in MESA (measurement error)

Trait averages (Mean ± SD)	"Odd" Estimate (Mean ± SD)	"Even" Estimate (Mean ± SD)	N windows" (Mean ± SD)	Coefficient of repeatability	Correlation (R)	Repeatability model $\beta_{N\text{windows}} \pm SE$	Recommended N windows" (% subjects exceeding criterion)	Coefficient of Repeatability for Subjects with Recommended N Windows	Correlation (R) for Subjects with Recommended N Windows	Total N Windows" (Mean ± SD)	
Endotypes											
Collapsibility (Vpassive, %)*	75.9 ± 13.0	75.6 ± 13.8	75.0 ± 14.3	53 ± 21 win	17.7	0.79	0.07 ± 0.005	28 win (98.1%)	15.4	0.83	131 ± 53 win
Compensation (%)	4.9 ± 17.4	4.9 ± 18.1	5.2 ± 17.6	53 ± 21 win	30.3	0.62	0.06 ± 0.01	35 win (96.7)	29.0	0.69	131 ± 53 win
Loop gain, LG ₁	0.52 ± 0.15	0.52 ± 0.16	0.52 ± 0.16	39 ± 17 win	0.26	0.68	0.06 ± 0.004	34 win (94.1)	0.20	0.79	97 ± 42 win
Ventilatory instability, LGn	0.42 ± 0.10	0.42 ± 0.10	0.42 ± 0.10	39 ± 17 win	0.16	0.67	0.07 ± 0.003	30 win (95.5)	0.13	0.77	97 ± 42 win
Arousal threshold (%)*	144.4 ± 21.9	144.1 ± 24.0	144.7 ± 23.6	31 ± 14 win	33.7	0.74	0.1 ± 0.01	20 win (97.4)	27.6	0.82	77 ± 35 win
Apnea-hypopnea index (events/h)	27.6 ± 0	27.7 ± 19.4	27.7 ± 19.4	180 ± 40 min	14.9	0.92	0.04 ± 0.002	45 min (99.9%)	14.9	0.92	359 ± 81 min

Odd and even estimates were obtained for endotypic traits and apnea-hypopnea index by dividing the sleep study into 30-min periods; "odd" and "even" numbered periods were aggregated to yield two independent measures per individual.

*Denotes transformed values (for Vpassive and Arousal Threshold, see Methods).

"For endotypic traits, the average number of 7-min windows ("N Windows") for analysis of odd periods and for even periods is shown (i.e. 39 "odd" windows and 39 "even" windows were typically available for loop gain measurements); for apnea-hypopnea index, total sleep time analyzed is reported. Coefficient of Repeatability is also shown as an estimate of the upper 95% confidence limit for absolute error. Correlation indicates the pairwise linear correlation coefficient between each pair of odd and even window traits using Pearson correlation method.

"Associations between measurement error (test-retest "error") and the number of windows used for analysis were significant for each trait ($p < 10^{-4}$). Test-retest repeatability was modeled (right hand side, "Repeatability Model") to examine dependence on N windows. Error was defined by the mean squared difference between Even Estimate and model-predicted Even Estimate (from Odd Estimate, regression analysis), normalized by total mean squares (Even Estimate minus mean of Even Estimate, estimate of trait variance). Thus, the mean value of Error across subjects equals $1-R^2$, thus repeatability $R = [1 - \text{mean}(\text{Error})]^{0.5}$. Error was modeled using a hyperbolic link function, whereby error falls with increasing N windows. Model results provided an estimate for the number of windows recommended as required for a repeatable estimate (based on $R = 0.7$ threshold). Using $R = 0.7$ threshold, we recalculated coefficient of repeatability and correlation coefficients based on pairs of odd and even windows where the recommended number of windows were available for both measurements. Total N windows denote the actual number of windows available for the whole night (odd, plus even, plus overlap periods). Sample size for this analysis was $N = 1750$.

were repeatable within subjects across separate nights before and after supplemental oxygen therapy (1-week apart)—an intervention that has no known effects on these traits [30]. Repeatability across nights for loop gain is also evident from the plots provided in a study exploring the effects of donepezil on AHI [31], but did not provide a quantitative analysis. Recent preliminary data from Tolbert et al [32] ($N = 44$) also supports short term night-to-night repeatability of the traits (R values: collapsibility 0.81, compensation 0.71, loop gain 0.90, arousal threshold 0.92) which compared well with repeatability for AHI ($R = 0.87$). In the current study, within-night repeatability in patients with sufficient windows also supported adequate repeatability (collapsibility 0.83, compensation 0.69, loop gain 0.79, arousal threshold 0.82). In our analysis, collapsibility, loop gain, and arousal threshold all had similar repeatability ($R \sim 0.8$), with somewhat lower repeatability for compensation ($R \sim 0.7$). Of note, the lower repeatability for compensation was also observed by Tolbert et al.

Within-night assessments also allowed us to compare measurement error associated with endotypic traits to the AHI. Since the derivation of endotypes depends on scored annotations (respiratory events, arousals, stages) it is not surprising that there is a greater measurement error for the endotypes than with events based on single annotations (apnea occurrence) or two annotations (hypopneas, also using associated arousals). However, overall measurement error was estimated to be modest, particularly when restricting the analysis to those meeting criteria for the minimal numbers of windows. These findings suggest that endotypes can serve as reasonable proxies for important physiological traits, although care should be taken to ensure sufficient data are available for optimal reliability.

Long-term consistency

The current study is also the first to assess long-term consistency of endotypic trait characteristic of OSA. Notably, the long-term consistency of the traits was comparable to the long-term consistency of AHI, despite our concerns about the potential amplifying effect on measurement error associated with multiple manually annotated components used in the derivation of endotypes. This may reflect the greater specificity of endotypes to reflect physiological processes that change less over time (and are calculated by inter-relationships among multiple physiological features) than the cruder AHI, which does not provide information on trait-like processes influencing OSA pathogenesis. Of all endotypes, loop gain measurements are the most consistent when measured across the years ($R = 0.63$) and was found to be more consistent than AHI ($R = 0.47$ – 0.52) yet measures of upper airway physiology were not as consistent (0.30–0.41). Accounting for body position and sleep state improved the consistency measures, particularly for collapsibility, but did not bring its long-term consistency up to that of loop gain. Furthermore, investigation is needed to determine whether long-term consistency for the upper airway traits may be higher in younger individuals or might be improved with advances in the techniques for estimating these traits or with improvements in raw signal quality used for estimating ventilation.

Quantitative interpretation of coefficient of repeatability results

One way to interpret repeatability and consistency based on the coefficient of repeatability (CR) begins with considering that 95%

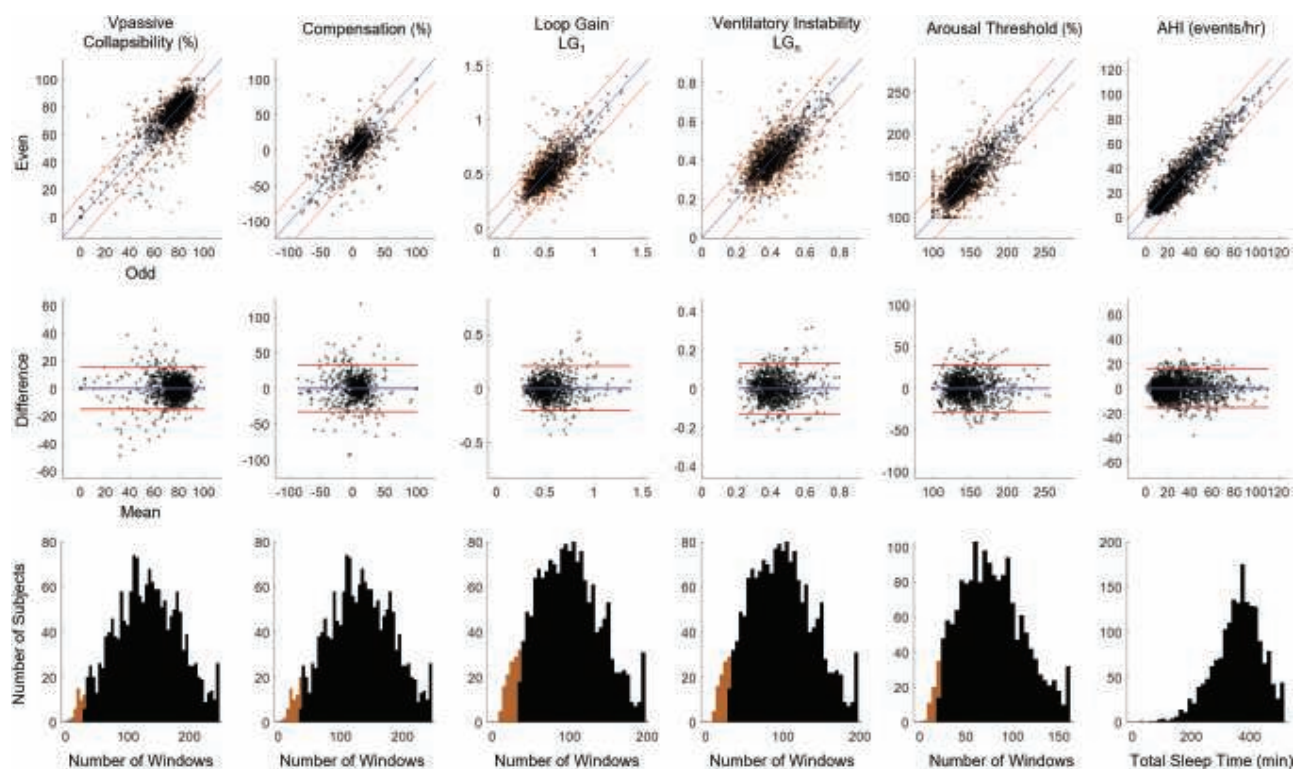


Figure 2. Illustration of within-night repeatability of endotypic traits and apnea-hypopnea index in MESA participants with OSA (per AHI>5, N = 1750). Top panels illustrate the concordance between two independent measures of the endotypic traits (odd on x-axis, even on y-axis); apnea-hypopnea index is also shown to provide a benchmark for comparison. Orange dots represent participants who did not have sufficient windows or total sleep time to obtain a reliable estimate (see Table 2); black dots represent participants who had sufficient windows to providing a reliable estimate. Blue lines denote the line of unity and red lines denote limits of agreement (shown for participants with sufficient windows). Middle panels illustrate the difference (even minus odd) versus the mean (Bland-Altman plot). Bottom panels illustrate the amount of data that is typically available across the whole night (total number of windows or TST across the MESA OSA participants). Black (orange) shading represents the participants with sufficient (insufficient) windows across the night to provide reliable (unreliable) estimates. Vpassive and arousal threshold data are square root transformed as described in the Methods.

of the population spread of a variable lies across 4SD (i.e. $\pm 2SD$). For example, for loop gain (LG_1), which has a population SD of 0.15 in MESA, 95% of the samples lie within 0.22–0.82 ($\pm 2SD$); nominal “low” and “high” values are 0.37 ($-1SD$) and 0.67 ($+1SD$) respectively. Within-night repeatability per CR was 0.2, which indicates that a single measurement in an individual with a “high” loop gain value (0.67) should yield repeat within-night measurements above the low loop gain level (0.37) > 95% of the time. This was true for each of the traits (i.e. $CR < 2SD$ of the population spread). Thus, an individual with a high measured value will only rarely exhibit a low measured value upon repeated within-night assessment.

Across multiple years (MrOS), however, the CR results for long-term consistency illustrated greater ambiguity. For example, the CR for collapsibility rose from 15% within-night to 32% across years, thus a “high” collapsibility measure (60.8%, $-1SD$ for visit 1) would not be above the low collapsibility value (83.8%, $+1SD$) more than 95% of the time with a future measurement; likewise, CR for compensation and arousal threshold exceeded 2SD of the population. On the other hand, the long-term consistency for loop gain still permitted discrimination between a higher measured value and lower level threshold ($CR < 2SD$ of the population) with 95% confidence. On an optimistic note, however, we emphasize that for all of the traits CR was below 3SD of the population, i.e. a high measured value for each of the traits would still exceed the lower threshold value more than 87% of the time.

Longitudinal change in endotypes

In the assessment of long-term consistency, we also identified systematic differences in the traits over the 6.5-year time frame: with aging, compensation increased, and arousal threshold declined, whereas AHI increased. Although an increase in compensation was unexpected based on a prior cross-sectional physiology study (reduced genioglossus sensitivity to negative pressures during wakefulness [33]), we note that there have been reports in older patients of a pattern of breathing characterized by crescendo-decrescendo periodic obstruction that is consistent with high compensation (flow rises when drive rises) [34]. Likewise, available data have also suggested that increased muscle responsiveness could be an acquired trait to protect against pharyngeal obstruction [35]. Effects of age on the arousal threshold are more consistent with expectations, given that older individuals have more frequent arousals and lighter sleep than younger individuals [36]. It is plausible that the lower arousal threshold contributes to increased OSA severity with aging among older men. A longer duration follow-up time may be needed to detect changes in collapsibility and loop gain.

Dependence on body position and sleep state

The current study also assessed the extent to which changes in body position and sleep state influence endotypic trait estimates.

Table 3. Long-term consistency in MrOS

	Visit 1 (Mean±SD)	Visit 2 (Mean±SD)	Difference (Mean [95% confidence interval])	Limits of agreement	Coefficient of repeatability	Correlation R	Adjusted correlation R		
							+ΔLateral	+ΔREM	+ΔN1,ΔN3
Endotypes									
Collapsibility (Vpassive,%)	72.3 ± 11.5	72.6 ± 15.8	0.2 [-2.5,0.4]	[32.16, -31.51]	31.8	0.30	0.35	0.38	0.41
Compensation(%)	4.1 ± 22.3	6.6 ± 23.1	2.5 [0.3,4.6]*	[54.11, -49.15]	51.8	0.33	0.33	0.33	0.36
Loop gain, LG ₁	0.57 ± 0.14	0.56 ± 0.14	-0.004 [-0.016,0.007]	[0.26, -0.27]	0.26	0.57	0.57	0.63	0.63
Ventilatory instability, LGn	0.48 ± 0.10	0.49 ± 0.12	0.006 -0.002,0.014]	[0.19, -0.18]	0.19	0.61	0.61	0.62	0.63
Arousal threshold (%)	152.7 ± 22.8	148.2 ± 25.6	-4.5 [-6.8, -2.2]*	[48.84, -57.91]	54.1	0.37	0.43	0.43	0.43
Apnea- hypopnea index (events/h)	18.0 ± 13.0	22.1 ± 15.5	4.1[2.9,5.2]*	[33.03, -24.91]	30.0	0.47	0.49	0.51	0.52

Traits differences, limits of agreement, coefficient of repeatability and correlation coefficients were based on comparison of Visit 2 versus Visit 1 (using All Sleep i.e. pooled nREM and REM measures). Crude across-visit correlations are shown (left hand side) and are interpreted as longitudinal consistency; R values were all significant to $p < 1 \times 10^{-13}$. Correlations were then repeated after adjusting for between-study changes in (1) body position (Δ Lateral), then additionally (2) REM sleep (Δ REM), finally also (3) nREM stages N1 and N3 (Δ N1, Δ N3; all expressed as fraction of sleep time). Specifically the difference between traits at visit 1 and visit 2 were modeled as a function of differences in position and state; the predicted difference was then added to trait 1 values to provide a position and state-corrected trait 1 value for comparison with trait 2 values. Most notably, long-term consistency improved for collapsibility and arousal threshold when accounting for changes in body position, and improved for loop gain when accounting for changes in REM sleep. Bold denotes a potentially meaningful increase in R (≥ 0.03) after accounting for across-year differences in position/state (for each bold case highlighted, associations between the across-year change in trait and changes in position/state were significant per $p < 0.01$). Sample size for this analysis was $N = 595$. *denotes $p < 0.025$.

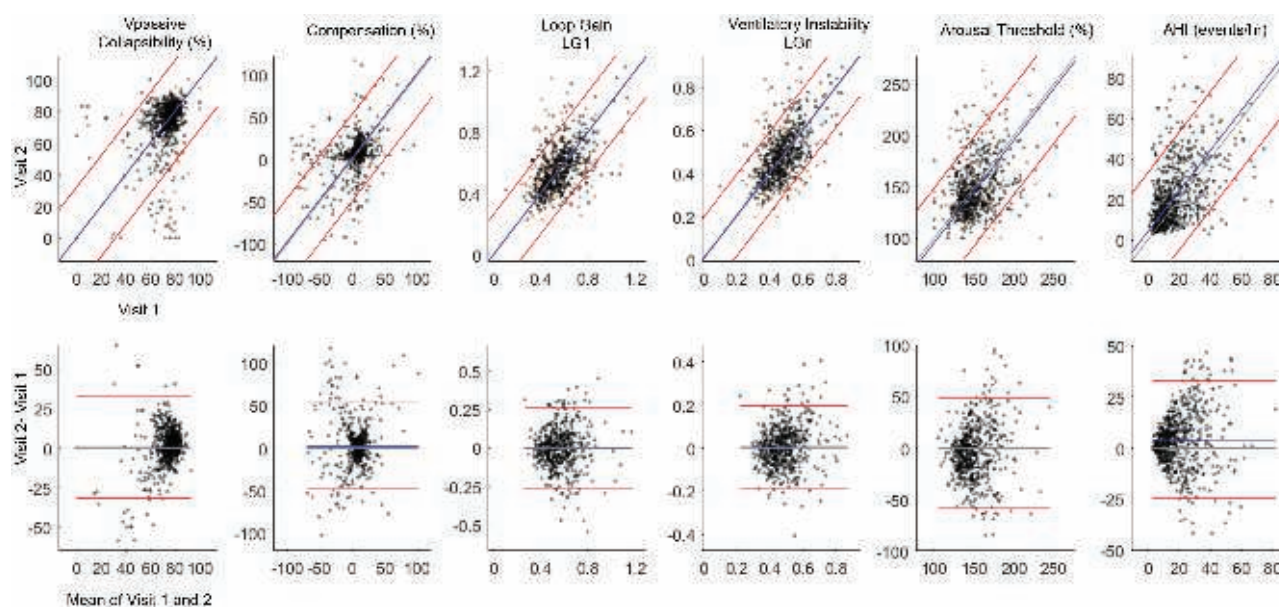


Figure 3. Long-term consistency of endotypic traits in MrOS; consistency of the apnea-hypopnea index (AHI) is shown for comparison. Top panels illustrate the second time point plotted against the first time point. Orange dots represent participants who did not have sufficient windows or total sleep time to obtain a reliable estimate; black dots represents participants who had sufficient windows to providing a reliable estimate (and provided data for Table 3 analysis). Gray lines denote the line of unity, blue lines denote mean bias, and red lines denote limits of agreement. Bottom panels are Bland-Altman plots. Note the substantial variability from one time point to the next in both the traits and AHI. Although variability over time is considered a combination of authentic physiological variability and measurement noise, we note that measurement noise was minimized since we only analyzed participants who had sufficient number of windows to provide reliable trait estimates using $R = 0.7$ criteria. Vpassive and arousal threshold data are square root transformed as described in the Methods. *denotes $p < 0.025$.

Previous physiological studies indicate that supine position affects collapsibility [37], but has minimal effect on the other traits. REM sleep has a profound effect on lowering loop gain [38, 39]. Whereas REM also increases collapsibility via a reduction in muscle activity, there is also a well-described loss of ventilatory

drive to both pump and upper airway muscles during REM sleep [40]; however, compensation was preserved. The arousal threshold is also reduced in REM [40]. Based on gold standard physiological signals, the greatest effect of deeper non-REM sleep appears to be a higher arousal threshold [41]; reduced collapsibility has also

Table 4. Effect of body position and sleep states on endotypic traits (MESA and MrOS sleep visit 1)

	$\beta \pm SE$ (% _{SD} / _{%time})			
	F _{Lateral}	F _{REM}	F _{N1}	F _{N3}
Endotypes				
Collapsibility (V _{passive} ,%)	0.45 ± 0.01	-0.36 ± 0.01	-0.78 ± 0.03	1.02 ± 0.02
Compensation(%)	0.37 ± 0.02	-0.15 ± 0.02	-0.79 ± 0.05	0.04 ± 0.04 ⁺
Loop gain, LG ₁	-0.28 ± 0.02	-1.75 ± 0.02	-0.70 ± 0.05	-0.06 ± 0.04 ⁺
Ventilatory instability, LG _n	-0.04 ± 0.02 ⁺	-0.89 ± 0.02	-0.80 ± 0.06	0.09 ± 0.05 ⁺
Arousal threshold (%)	-0.86 ± 0.02	0.91 ± 0.02	-0.06 ± 0.05 ⁺	-0.44 ± 0.04
Apnea-hypopnea index (events/h)	-1.00 ± 0.02	0.91 ± 0.02	1.54 ± 0.02	-0.95 ± 0.02

Within subject effects of differences in sleep time spent in different body positions (F_{Lateral}), non-REM sleep (F_{N1}, F_{N3}) and REM sleep (F_{REM}) on endotypic traits were assessed by using bootstrapping randomly sampled windows of data from each participant to generate 50 measurements per participant. Model equation: Trait ~ F_{Lateral} + F_{REM} + F_{N1} + F_{N3} + Subject. AHI was also included to provide a benchmark. F_{Lateral}, F_{N1}, F_{N3}, F_{REM} are expressed as fraction of total sleep time. Data from MESA (N = 1774) and MrOS (N = 2390, visit 1) were combined for this analysis. Beta coefficients and SE ($\beta \pm SE$ (%_{SD}/_{%time})) shown describe the effect on each trait estimate (in percentage of a SD) per 1% increase in sleep time in the specified position or state. For example, an increase in REM (F_{REM}) of 1% of the night is associated with a 0.0175 SD reduction in the estimated loop gain (i.e. reduction in loop gain by 0.0026). Intersubject SD for the above analyses were: V_{passive}* = 14.2, Compensation = 20.9, Loop gain = 0.151, Ventilatory Instability = 0.106, Arousal Threshold* = 23.9 (*denotes transformed). AHI was square root transformed prior to analysis; SD for transformed AHI = 1.55, SD for untransformed AHI = 17.2 events/h. Shading denotes meaningful effects based on $\beta > 0.5$ %_{SD}/_{%time}. Bold indicates an effect that is greater than that observed for AHI. All effects are significant except those denoted by + where $p \geq 0.05$.

been described [42]. In the current study, we specifically examined the effects of body position and state on the estimates of the OSA traits: as expected, position and state differences affected multiple endotypic trait measurements. In general, effects on traits were smaller than for AHI, with two exceptions: we observed a strong effect of N3 on collapsibility, and a strong effect of REM on loop gain (which supports the previous focus on endotypes calculated only within non-REM [5–8, 10, 13, 31, 43]). Unexpectedly we observed a reduction in compensation and loop gain with greater proportion of stage 1 sleep, the physiological underpinnings of which are unclear. We also observed that the currently described analysis did not detect the known effects of deeper sleep on the arousal threshold. This could be secondary to our noninvasive estimate of ventilatory drive being relative and may not recognize the known steady-state rise in drive seen with deeper sleep. The same concept may explain the increase in arousal threshold observed in REM (which is contrary to the known physiological direction of effects). Advances in the methods to capture absolute levels of ventilatory drive (rather than estimating local changes in drive) may improve the ability of the method to detect these known physiological changes.

Clinical and research implications

Endotypic traits have been proposed for use in selecting patients that might benefit preferentially from a range of non-CPAP alternative therapies including oral appliances, hypoglossal nerve stimulation, supplemental oxygen, and pharmacotherapy. Our data support the potential utility of these measurements for physiologically directed therapeutic intervention. However, appropriate use of these data requires an understanding of the limitations, and particularly the need to obtain sufficient data to derive reliable estimates and consider influences of position and sleep states.

Our findings also support the use of estimated endotypes in large-scale epidemiological studies that aim to characterize OSA physiological subtypes—something not possible by simply using the AHI. For example, we recently described differences in arousal threshold, collapsibility, and loop gain in men compared

to women [43], which underscored the need to consider sex/gender as important factors influencing OSA pathogenesis. Given that polysomnography is available in several population studies, endotypes can be derived to further understand the impact of differences in OSA subtypes in disease outcomes as well as in efforts to identify genetic markers of susceptibility (adjusting for age and sex variation). The modest within-night measurement error and long-term consistency for the panel of endotypes suggest that these measures provide fairly reliable trait estimation, with long-term reliability similar to that of many other complex disorders shown to have a strong genetic basis and have prognostic value [44, 45].

Limitations

The current study has several limitations. First, repeatability was assessed within a single night using odd and even windows to provide fully independent measures; this analysis provides an upper limit of repeatability for a half-night measurement. Across nights, physiological changes may be greater, although preliminary data from Tolbert et al using in-laboratory assessments in a small sample showed similar or slightly greater reliability across nights than our larger sample studied at home [32]. Second, the endotypic traits calculated are estimates and depend on several validated assumptions: (1) changes in the (linearized) nasal pressure signal amplitude reflect the changes in inspired ventilation, (2) the mean ventilation during sleep approximates the eupneic level, (3) the airway is largely patent following respiratory events, and (4) ventilatory drive fluctuations can be estimated meaningfully using airflow measurement and a delayed feedback model [46]. Importantly, a nasal cannula yields near identical traits compared with gold standard pneumotachography with sealed full-face mask ($R > 0.95$ for all traits simultaneously recorded) [6]. Direct measurement of ventilatory drive would require additional, potentially invasive signals that are not part of routine polysomnography, and thus would prevent widespread use for precision sleep medicine. Third, our long-term consistency analysis was based on 6.5-year follow-up data from a cohort of older individuals—a group in whom co-morbidities are common and

may promote heterogeneity in OSA pathophysiology over time. Hence care must be taken although applying these long-term consistency values to younger individuals. It is also possible that a shorter follow-up time may have improved consistency measures; our study was not able to directly address this question since the follow-up range was narrow (range: 4.79–7.95 years). Available evidence, albeit limited, currently leads us to contend that there is greater night-to-night consistency when the duration between measurements is shorter. Fourth, our study was a community-based sample of men; thus, care should be taken extrapolating our results to women or a clinical sample. Fifth, in this study traits were measured using in-home polysomnography rather than a more controlled in-laboratory PSG environment, which may adversely affect repeatability and consistency. Sixth, our main analysis described repeatability and consistency of traits as continuous variables, but future application may include separating participants into endotype-based subgroups (i.e. low vs. high [2 classes], or low, medium, and high [3 classes]); additional analysis (Supplemental Table S3) showed that 2-class repeatability/consistency for traits was 70%–79% within-night and 59%–72% across 6.5 years. 3-class repeatability/consistency for traits was 58–64% exact (92–97% within 1 class) within-night and 44%–56% (84%–93% within 1 class) across 6.5 years. We note that it was relatively uncommon for individuals to jump between low and high classes in 3-class analysis.

Conclusions

Noninvasive estimation of endotypic traits from standard PSG may help predict individual responses to OSA therapy in clinical settings, and phenotype individuals in research studies. The utility of these endotypic traits depend on the magnitude and sources of measurement error. This study has shown that in a large community-based research population with mild-to-severe OSA, repeatable endotypic traits can be estimated from a single night in-home PSG recording, and the consistency of these measurements across years are comparable with that of AHI. We also demonstrated generally modest influences of body position and sleep states, which, once accounted for, appeared to improve the consistency of the trait calculation, and should be considered during the interpretation of endotypic traits. Given the large numbers of routinely collected PSG for clinical and research purposes, these results support the value of generating these metrics to advance knowledge and use of OSA subtypes.

Supplementary Material

Supplementary material is available at SLEEP online.

Funding

This work was supported financially by the American Heart Association (15SDG25890059, PI Sands) and the National Institute of Health NHLBI (R35HL135818, PI Redline). The parent Multi-Ethnic Study of Atherosclerosis (MESA) Sleep Ancillary study was funded by NIH-NHLBI Association of Sleep Disorders with Cardiovascular Health Across Ethnic Groups (R01HL098433). MESA is supported by NHLBI funded contracts HHSN268201500003I, N01HC95159, N01HC95160, N01HC95161, N01HC95162, N01HC95163, N01HC95164, N01HC95165, N01HC95166, N01HC95167, N01HC95168, and N01HC95169, and by cooperative agreements UL1TR000040,

UL1TR001079, and UL1TR001420 funded by The National Center for Advancing Translational Sciences. The National Heart, Lung, and Blood Institute provided funding for the ancillary MrOS Sleep Study, “Outcomes of Sleep Disorders in Older Men,” under the following grant numbers: R01 HL071194, R01 HL070848, R01 HL070847, R01 HL070842, R01 HL070841, R01 HL070837, R01 HL070838, and R01 HL070839. The National Sleep Research Resource was supported by the NHLBI (R24HL114473 and 75N92019R002).

Acknowledgments

The authors thank the investigators, the staff, and the participants of the MESA and MrOS study for their valuable contributions. A full list of participating investigators and institutions can be found at <https://www.mesa-nhlbi.org> (MESA) and <https://mrosonline.ucsf.edu/> (MrOS).

Disclosure Statement

AA receives personal fees as a consultant for Somnifix, Apnimed, and Respicardia, and receives grant support from Somnifix. AW has a financial interest in Apnimed, a company developing pharmacologic therapies for sleep apnea. AW also consults for Somnifix and Nox Medical and received grants from Somnifix and Sanofi. DPW is a part-time employee of Apnimed and also consults for Philips Respironics and Neumora Therapeutics. SR receives personal fees from Jazz Pharma, Eli Lilly, and Apnimed. SAS receives personal fees as a consultant for Nox Medical, Merck, Apnimed, Inspire, Respicardia, and Eli Lilly outside the submitted work and receives grant support from Apnimed, ProSomnus, and Dynaflex for unrelated studies. The industry interactions of AA, DPW, and SAS are managed by Brigham and Women’s Hospital and Mass General Brigham in accordance with their conflict of interest policies. RMA, TS, DV, and LKG have no conflict of interest to disclose.

Data Availability

The sleep polysomnography data underlying this article are available in National Sleep Research Resource Repository, at www.sleepdata.org/datasets. MESA data access is managed by the MESA Coordinating Center and requires completion of a Data Distribution Agreement, review, and approval of a detailed proposal by MESA Publications and Steering Committees (<https://www.mesa-nhlbi.org>), and subsequent investigator approval. Information on accessing MrOS dataset is available at <https://mrosonline.ucsf.edu/>. For requests for statistical analysis code please contact the corresponding author.

References

1. Eckert DJ. Phenotypic approaches to obstructive sleep apnoea—new pathways for targeted therapy. *Sleep Med Rev*. 2018;37:45–59.
2. Subramani Y, et al. Understanding phenotypes of obstructive sleep apneas applications in anesthesia, surgery, and perioperative medicine. *Anesth Analg*. 2017;124(1):179–191.

3. Eckert DJ, et al. Defining phenotypic causes of obstructive sleep apnea. Identification of novel therapeutic targets. *Am J Respir Crit Care Med.* 2013;**188**(8):996–1004.
4. Wellman A, et al. Ventilatory control and airway anatomy in obstructive sleep apnea. *Am J Respir Crit Care Med.* 2004;**170**(11):1225–1232.
5. Terrill PI, et al. Quantifying the ventilatory control contribution to sleep apnoea using polysomnography. *Eur Respir J.* 2015;**45**(2):408–418.
6. Sands SA, et al. Phenotyping pharyngeal pathophysiology using polysomnography in patients with obstructive sleep apnea. *Am J Respir Crit Care Med.* 2018;**197**(9):1187–1197.
7. Sands SA, et al. Quantifying the arousal threshold using polysomnography in obstructive sleep apnea. *Sleep.* 2018;**41**(1). doi:[10.1093/sleep/zsx183](https://doi.org/10.1093/sleep/zsx183)
8. Joosten SA, et al. Loop gain predicts the response to upper airway surgery in patients with obstructive sleep apnea. *Sleep.* 2017;**40**(7). doi:[10.1093/sleep/zsx094](https://doi.org/10.1093/sleep/zsx094)
9. Edwards BA, et al. Upper-airway collapsibility and loop gain predict the response to oral appliance therapy in patients with obstructive sleep apnea. *Am J Respir Crit Care Med.* 2016;**194**(11):1413–1422.
10. Bamagoos AA, et al. Polysomnographic endotyping to select patients with obstructive sleep apnea for oral appliances. *Ann Am Thorac Soc.* 2019;**16**(11):1422–1431.
11. Taranto-Montemurro L, et al. Effects of the combination of atomoxetine and oxybutynin on OSA endotypic traits. *Chest.* 2020;**157**(6):1626–1636.
12. Sands SA, et al. Identifying obstructive sleep apnoea patients responsive to supplemental oxygen therapy. *Eur Respir J.* 2018;**52**(3).
13. Op de BS, et al. Endotypic mechanisms of successful hypoglossal nerve stimulation for obstructive sleep apnea. *Am J Respir Crit Care Med.* 2021;**203**(6):746–755.
14. Gell L, Vena D, Alex RM, et al. Neural ventilatory drive decline as a predominant mechanism of obstructive sleep apnea events. *Thorax.* 2022;**77**:707–716.
15. Eckert DJ, et al. Defining phenotypic causes of obstructive sleep apnea. Identification of novel therapeutic targets. *Am J Respir Crit Care Med.* 2013;**188**(8):996–1004.
16. Wellman A, et al. A method for measuring and modeling the physiological traits causing obstructive sleep apnea. *J Appl Physiol* (1985). 2011;**110**(6):1627–1637.
17. Chen X, et al. Racial/Ethnic differences in sleep disturbances: the Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep.* 2015;**38**(6):877–888. doi:[10.5665/sleep.4732](https://doi.org/10.5665/sleep.4732)
18. Bild DE, et al. Ethnic differences in coronary calcification: the Multi-Ethnic Study of Atherosclerosis (MESA). *Circulation.* 2005;**111**(10):1313–1320.
19. Rechtschaffen A, Kales A. *A Manual of Standardized Terminology, Techniques and Scoring Systems for Sleep Stages of Human Subjects.* Los Angeles, CA: UCLA Brain Information Service/Brain Research Institute; 1968.
20. Berry RB, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. *J Clin Sleep Med.* 2012;**8**(5):597–619.
21. Blank JB, et al. Overview of recruitment for the osteoporotic fractures in men study (MrOS). *Contemp Clin Trials.* 2005;**26**(5):557–568.
22. Stone KL, et al. Sleep disturbances and risk of falls in older community-dwelling men: the outcomes of Sleep Disorders in Older Men (MrOS Sleep) Study. *J Am Geriatr Soc.* 2014;**62**(2):299–305.
23. Blackwell T, et al. Associations between sleep architecture and sleep-disordered breathing and cognition in older community-dwelling men: the Osteoporotic Fractures in Men Sleep Study. *J Am Geriatr Soc.* 2011;**59**(12):2217–2225.
24. Farre R, et al. Noninvasive monitoring of respiratory mechanics during sleep. *Eur Respir J.* 2004;**24**(6):1052–1060.
25. Thurnheer R, et al. Accuracy of nasal cannula pressure recordings for assessment of ventilation during sleep. *Am J Respir Crit Care Med.* 2001;**164**(10 Pt 1):1914–1919.
26. Wellman A, et al. A method for measuring and modeling the physiological traits causing obstructive sleep apnea. *J Appl Physiol.* 2011;**110**(6):1627–1637.
27. Wellman A, et al. A simplified method for determining phenotypic traits in patients with obstructive sleep apnea. *J Appl Physiol.* 2013;**114**(7):911–922.
28. Vena D, et al. Clinical polysomnographic methods for estimating pharyngeal collapsibility in obstructive sleep apnea. *Sleep.* 2022;**45**(6):zsac050. doi:[10.1093/sleep/zsac050](https://doi.org/10.1093/sleep/zsac050)
29. Vaz S, et al. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One.* 2013;**8**(9):e73990.
30. Edwards BA, et al. Effects of hyperoxia and hypoxia on the physiological traits responsible for obstructive sleep apnoea. *J Physiol.* 2014;**592**(20):4523–4535.
31. Li Y, et al. The effect of donepezil on arousal threshold and apnea-hypopnea index. A randomized, double-blind, crossover study. *Ann Am Thorac Soc.* 2016;**13**(11):2012–2018.
32. Tolbert T, et al. Variability of physiologic traits determined by phenotyping using polysomnography on consecutive nights [Abstract]. *Am J Respir Crit Care Med.* 2022;**205**:A4816.
33. Malhotra A, et al. Aging influences on pharyngeal anatomy and physiology: the predisposition to pharyngeal collapse. *Am J Med.* 2006;**119**(1):72.e79–72.e14.
34. Hudge DW, et al. Pattern of breathing and upper airway mechanics during wakefulness and sleep in healthy elderly humans. *J Appl Physiol* (1985). 1993;**74**(5):2198–2204.
35. Sands SA, et al. Enhanced upper-airway muscle responsiveness is a distinct feature of overweight/obese individuals without sleep apnea. *Am J Respir Crit Care Med.* 2014;**190**(8):930–937.
36. Bonnet MH, et al. EEG arousal norms by age. *J Clin Sleep Med.* 2007;**3**(3):271–274.
37. Joosten SA, et al. Evaluation of the role of lung volume and airway size and shape in supine-predominant obstructive sleep apnoea patients. *Respirology.* 2015;**20**(5):819–827.
38. Douglas NJ, et al. Hypercapnic ventilatory response in sleeping adults. *Am Rev Respir Dis.* 1982;**126**(5):758–762.
39. Messineo L, et al. Loop gain in REM versus non-REM sleep using CPAP manipulation: a pilot study. *Respirology.* 2019;**24**(8):805–808.
40. Messineo L, et al. Ventilatory drive withdrawal rather than reduced genioglossus compensation as a mechanism of obstructive sleep apnea in REM. *Am J Respir Crit Care Med.* 2022;**205**(2):219–232. doi:[10.1164/rccm.202101-0237OC](https://doi.org/10.1164/rccm.202101-0237OC)
41. Ratnavadivel R, et al. Upper airway function and arousability to ventilatory challenge in slow wave versus stage 2 sleep in obstructive sleep apnoea. *Thorax.* 2010;**65**(2):107–112.
42. Carberry JC, et al. Upper airway collapsibility (Pcrit) and pharyngeal dilator muscle activity are sleep stage dependent. *Sleep.* 2016;**39**(3):511–521. doi:[10.5665/sleep.5516](https://doi.org/10.5665/sleep.5516)

43. Won CHJ, et al. Sex differences in obstructive sleep apnea phenotypes, the Multi-Ethnic study of atherosclerosis. *Sleep*. 2020;**43**(5). doi:[10.1093/sleep/zsz274](https://doi.org/10.1093/sleep/zsz274)
44. Chia YC, et al. Long-Term visit-to-visit blood pressure variability and renal function decline in patients with hypertension over 15 years. *J Am Heart Assoc*. 2016;**5**(11):e003825. doi:[10.1161/jaha.116.003825](https://doi.org/10.1161/jaha.116.003825)
45. Li H, et al. Variability of Type 2 inflammatory markers guiding biologic therapy of severe asthma: a 5-year retrospective study from a single tertiary hospital. *World Allergy Organ J*. 2021;**14**(9):100547.
46. Mann DL, et al. Quantifying the magnitude of pharyngeal obstruction during sleep using airflow shape. *Eur Respir J*. 2019;**54**(1):1802262. doi:[10.1183/13993003.02262-2018](https://doi.org/10.1183/13993003.02262-2018)