



ORIGINAL ARTICLE

Clinical polysomnographic methods for estimating pharyngeal collapsibility in obstructive sleep apnea

Daniel Vena^{1,*}, Luigi Taranto-Montemurro¹, Ali Azarbarzin¹, Sara Op de Beeck^{2,3},
Melania Marques^{1,4}, Olivier M. Vanderveken^{2,3,5}, Bradley A. Edwards^{6,7}, Laura Gell¹,
Nicole Calianese¹, Lauren B. Hess¹, Reza Radmand¹, Garun S. Hamilton^{8,9},
Simon A. Joosten^{8,9}, Johan Verbraecken^{3,5}, Marc Braem^{3,10}, David P. White¹, Susan Redline¹,
Scott A. Sands^{1,†} and Andrew Wellman^{1,†}

¹Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, ²Faculty of Medicine and Health Sciences, University of Antwerp, Wilrijk, Antwerp, Belgium, ³Department of ENT, Head and Neck Surgery, Antwerp University Hospital, Edegem, Antwerp, Belgium, ⁴Laboratório do sono, Instituto do Coração (InCor), Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil, ⁵Multidisciplinary Sleep Disorder Center, Antwerp University Hospital, Edegem, Antwerp, Belgium, ⁶Department of Physiology, Biomedicine Discovery Institute, Monash University, Melbourne, Victoria, Australia, ⁷School of Psychological Sciences and Turner Institute for Brain and Mental Health, Monash University, Melbourne, Victoria, Australia, ⁸Monash Lung and Sleep, Monash Health, Clayton, Victoria, Australia, ⁹School of Clinical Sciences, Monash University, Clayton, Victoria, Australia and ¹⁰Division of Special Care Dentistry, Department of ENT, Head and Neck Surgery, Antwerp University Hospital, Edegem, Antwerp, Belgium

[†]These authors contributed equally to this work.

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

Abstract

Study Objectives: Obstructive sleep apnea has major health consequences but is challenging to treat. For many therapies, efficacy is determined by the severity of underlying pharyngeal collapsibility, yet there is no accepted clinical means to measure it. Here, we provide insight into which polysomnographic surrogate measures of collapsibility are valid, applicable across the population, and predictive of therapeutic outcomes.

Methods: Seven promising polysomnography-derived surrogate collapsibility candidates were evaluated: *V_{passive}* (flow at eupneic ventilatory drive), *V_{min}* (ventilation at nadir drive), *event depth* (depth of the average respiratory event), *oxygen desaturation slope* and *mean oxygen desaturation* (events-related averages), *Fhypopneas* (fraction of events scored as hypopneas), and *apnea index*. Evaluation included (1) validation by comparison to physiological gold-standard collapsibility values (critical closing pressure, *P_{crit}*), (2) capacity to detect increased collapsibility with older age, male sex, and obesity in a large community-based cohort (Multi-Ethnic Study of Atherosclerosis, MESA), and (3) prediction of treatment efficacy (oral appliances and pharmacological pharyngeal muscle stimulation using atomoxetine-plus-oxybutynin).

Results: *P_{crit}* was significantly correlated with *V_{min}* ($r = -0.54$), *event depth* ($r = 0.49$), *V_{passive}* ($r = -0.38$), *Fhypopneas* ($r = -0.46$), and *apnea index* ($r = -0.46$; all $p < .01$) but not others. All measures detected greater collapsibility with male sex, age, and obesity, except *Fhypopneas* and *apnea index* which were not associated with obesity. *Fhypopneas* and *apnea index* were associated with oral appliance and atomoxetine-plus-oxybutynin efficacy (both $p < .05$).

Conclusions: Among several candidates, *event depth*, *Fhypopneas*, and *apnea index* were identified as preferred pharyngeal collapsibility surrogates for use in the clinical arena.

Statement of Significance

Pharyngeal collapsibility can be derived from routine polysomnography and could be used in clinical decision-making for obstructive sleep apnea treatment

Key words: obstructive sleep apnea; pharyngeal collapsibility; precision medicine

Submitted: 9 August, 2021; Revised: 22 February, 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 highly prevalent disorder with major implications for health and quality of life. Alternative OSA therapies to continuous positive airway pressure (CPAP) are limited primarily by unpredictable efficacy. Recent evidence has demonstrated that treatment efficacy for oral appliances and pharmacological therapies are dependent on the severity of underlying pharyngeal collapsibility in individual patients. For example, patients with less-severe collapsibility are amenable to efficacious treatment with oral appliances [1–5] and atomoxetine-plus-oxybutynin [6]. Unfortunately, gold-standard measurement of pharyngeal collapsibility—based on manipulation of CPAP during sleep to assess the critical pressure at which the airway completely collapses (Pcrit) [7]—is limited to specialized physiology laboratories, such that widespread clinical use is currently not feasible. Valid, broadly-applicable, and predictive measures of collapsibility are needed in the clinical arena to open the doors for precision sleep medicine.

To allow estimation of collapsibility in clinical practice, our team developed a method to estimate collapsibility from routine polysomnography [8]: the metric, known as V_{passive} , represents the airflow at a normal (“eupneic”) level of respiratory drive. More recently, several new concepts have emerged for estimating passive pharyngeal collapsibility that could improve on V_{passive} or further simplify collapsibility assessment. The potential clinical value of these methods remains untested.

In the current study, we evaluated the clinical potential of candidate novel and existing polysomnographic surrogate measures of collapsibility. We focused on measures that reflect airflow reduction during respiratory events. Three phases of evaluation were followed: first, validity of each measure was based on comparison with the gold standard measure of collapsibility (Pcrit). Second, applicability across the population was assessed in a large cohort study (Multi-Ethnic Study of Atherosclerosis, MESA) to determine whether measures had the capacity to detect the known increase in collapsibility with older age, male sex, and obesity. Last, the predictive value of the measures was evaluated based on their ability to predict response to representative therapies in the oral appliance and pharmaceutical domains where efficacy is known to be dependent on the severity of pharyngeal collapsibility: oral appliances and atomoxetine-plus-oxybutynin [1–6]. Any measure that performed well across all phases of evaluation was considered to have strong potential for widespread clinical use.

Methods

Summary of candidate estimates of collapsibility

Seven promising surrogate collapsibility candidates were evaluated: V_{passive} (flow at eupneic ventilatory drive), V_{min} (ventilation at nadir drive), event depth (i.e. mean decrease in minute ventilation for all events), $\text{oxygen desaturation slope}$ and $\text{mean oxygen desaturation}$ (events-related averages), Fhypopneas (fraction of events scored as hypopneas compared to total events), and apnea index (apnea frequency). Measures are described in detail below.

Validation with critical pharyngeal closing pressure

To quantify a gold-standard measure of collapsibility, we assessed the passive Pcrit in 57 patients. Forty-six patients were

part of previous studies [1, 8, 9]; the remaining 11 were newly acquired. Included patients had an apnea-hypopnea index (AHI) ≥ 15 events/h. Our method for measuring passive Pcrit has been described in detail elsewhere [1, 5, 8, 10–15] and is based on the techniques first described by Schwartz et al. [7]. Briefly, Pcrit was assessed during the first approximately 2 h of sleep with the patient in the supine position. Patients fell asleep on 4 cmH₂O of CPAP delivered via a modified device (Pcrit 2000, Philips-Respironics, Murrysville, PA). Once asleep, CPAP was adjusted to the minimum level required to eliminate snoring, hypopneas, and flow-limited breathing (known as “holding pressure”). Once stable N2 or N3 sleep was achieved at the holding pressure, CPAP was dropped to subtherapeutic levels for five breaths, and then increased back to the holding pressure. This procedure was repeated for increasingly lower subtherapeutic levels, with at least 1 min between pressure drops, until complete obstructive apnea was observed. If an arousal occurred during a CPAP drop, the pressure was increased back to the holding pressure and the CPAP drop was repeated once the subject was stably asleep again. To calculate passive Pcrit, we used custom Matlab software (Mathworks, Natick, MA) to fit a line to peak flow (average for breaths 3 and 4 during the pressure drop) and the corresponding CPAP level. Pcrit is determined as the zero-flow intercept of this best-fit line [7]. In the corresponding polysomnography data, patients slept at least 75% of the night supine.

If a surrogate collapsibility metric was not significantly associated with Pcrit, it did not move on for further assessment because it undermined the physiological basis of subsequent analyses.

Applicability across the population in a large cohort study

The relationship of each surrogate collapsibility metric with age, sex, and body mass index (BMI) was assessed in participants from the MESA who underwent a sleep study that included full overnight unattended polysomnography [16]. The MESA dataset was collected through six centers to investigate factors associated with the development of subclinical cardiovascular disease and the progression to clinical cardiovascular disease in Black, White, Hispanic, and Chinese-American men and women initially aged 54–93 years at their sleep study visits in 2013 to 2015. As described previously [17, 18], 15-channel polysomnography (Somté PSG, Compumedics Ltd., Abbotsford, AU) was successfully collected in 2057 participants, 1916 (93%) of whom could have collapsibility metrics computed (supine non-rapid eye movement sleep [NREM] sleep only). Notably, airflow was collected via nasal pressure from which all flow-based measures of collapsibility were derived. A priori coefficients were calculated for context from a previously published study of 108 sleep apnea patients, evaluating male sex, age, and obesity on mechanical upper-airway instability [19]. From this study, we established that change in sex and a 2SD change in age and BMI is associated with SD changes in Pcrit of 0.5, 0.25, and 0.25, respectively. For collapsibility surrogates, associations at least half these levels were considered acceptable.

Predicting treatment response

Predictive value of each surrogate collapsibility measure was assessed using representative OSA treatments from the oral appliance and pharmacotherapy domains: oral appliances [3]

and atomoxetine-plus-oxybutynin [6]. Treatment response was measured as the percent reduction in AHI from baseline to treatment. Oral appliance efficacy was determined in 81 patients from three separate studies that included polysomnography data from one night off oral appliance treatment and another night on oral appliance treatment. Data from these same 81 patients has been published previously [3]. The three studies are: (1) “MADOX” (NCT03189173), an ongoing clinical trial from which we analyzed 20 OSA patients as part of the interim analysis; (2) “SSPO” (NCT02489591), a published study with 25 OSA patients [1]; and (3) “PROMAD” (NCT01532050), a published study with 36 moderate to severe OSA patients. Studies were selected based on the availability of unfiltered airflow signals [20, 21]. Efficacy of atomoxetine-plus-oxybutynin “ato-oxy” was determined in 15 OSA patients with polysomnography data from one night on drug combination and another night on placebo [6].

In the protocols for all four studies (MADOX, SSPO, PROMAD, ato-oxy), standard techniques and criteria were used to score sleep stages and arousals. Polysomnograms for three of the studies included airflow measurement with sealed nasal (SSPO baseline and treatment nights) or oronasal mask (MADOX baseline night only; ato-oxy baseline and treatment nights) connected to a pneumotachometer (Hans-Rudolph, Kansas City, MO; Validyne, Northridge, CA). Airflow was measured using nasal pressure cannula for PROMAD baseline and treatment nights, as well as a MADOX treatment night. All studies scored hypopneas requiring a 3% oxygen desaturation or an arousal from sleep [20, 21]. All patients available from these studies were included in the present analysis. Written informed consent was provided by all patients and approval was granted from their respective Institutional Review Boards

Inclusion criteria for included studies

Inclusion and exclusion for the studies included in the present paper were as follows. Age criteria was 21–79 years for Pcrit [1, 8, 9], MADOX, and ato-oxy [6] study data; 21–70 years for SSPO [1]; and > 18 years for PROMAD [20, 21]. Only PROMAD had a BMI cutoff < 35 kg/m². All the above studies included participants with diagnosed or suspected OSA. Oral appliance studies excluded patients with contraindications for oral appliances. MADOX included a baseline study, such that patients with AHI < 20 events/h were excluded and did not continue with the rest of the study. Similarly, in PROMAD patients with AHI < 15 events/h were excluded and did participate in the follow-up PSG. SSPO and ato-oxy included all patients in both baseline and treatment study conditions, but the analysis in the parent paper was limited to AHI > 10 events/h (25/30 and 15/20 patients, respectively). We followed the same criteria in the present analysis. All studies excluded patients with comorbid sleep disorders and use of drugs that depress respiration. The main relevant criteria for inclusion in the MESA study was men and women who identified themselves as White, Black/African American, Hispanic, or Chinese aged 45–84 years and free of clinically apparent cardiovascular disease [16]. The MESA ancillary sleep study (Exam 5) invited back all patients that participated in the original MESA study that did not report regular use of positive airway pressure (PAP) devices, oral devices, or nocturnal oxygen [17]. Our analysis only included patients with AHI > 10 events/h.

Signal processing

The methods to quantify the 7 surrogate measures of collapsibility are summarized below (see [Supplemental Methods](#) for details and a link to the source code for measuring surrogate collapsibility measures). All measures were computed during NREM sleep. To validate Pcrit, collapsibility was calculated from supine sleep data (analysis with all positions data included in [Supplemental Results](#)). Similarly, to evaluate the relationship of each collapsibility metric with age, sex, and BMI, collapsibility was computed from supine sleep. This was determined to be most appropriate for comparisons to published relationships of age, sex, and BMI with Pcrit [19], which was measured in the supine position. For prediction of treatment response, given the impracticality of using collapsibility computed from supine sleep, we computed collapsibility from sleep in all positions and included fraction of time in the non-supine position during sleep as a covariate (analysis without covariates and with supine-only data included in [Supplemental Results](#)). For flow-derived metrics, it is important to point out the quality of the airflow signal was assessed, and low-quality airflow signals were automatically rejected by our software based on Fourier analysis of the airflow signal (see [Supplementary Materials](#)).

Event depth. The average depth of the respiratory event (i.e. event depth) was calculated from the ventilation profile of the average respiratory event [3] ([Figure 1](#)). Briefly, uncalibrated ventilation (tidal volume × rate) was computed and expressed as a percentage of the eupneic ventilation (local mean). Ventilation signals for each respiratory event were then aligned to the terminal breath of the event and then ensemble averaged. The *event depth* is then calculated as the mean reduction in ventilation (from eupnea) during the event.

V_{passive} and V_{min}. V_{passive} and V_{min} describe the (median) ventilation observed at eupneic drive and at minimal drive (lowest decile), respectively ([Figure 1](#)). Ventilatory drive was estimated using a polysomnographic method described in detail previously [8]. V_{min} was proposed as a new metric to overcome the limitations of V_{passive}, which is often skewed towards milder collapsibility and appears to underestimate collapsibility. V_{min} is quantified at the lowest observed drive (lowest decile rather than eupneic levels), which may better approximate collapsibility under truly hypotonic conditions of the pharyngeal muscles.

Oxygen desaturation slope and mean oxygen desaturation. In principle, the greater the severity of event-related obstruction, the faster and deeper oxygen saturation should fall during an event. *Desaturation slope* was calculated using the ensemble-averaging method [22, 23], the average event-related saturation profile was calculated ([Figure 1](#)). *Desaturation slope* was measured as the change in oxygen saturation from the 90th (i.e. event start) to the 10th (i.e. event end) percentile of the saturation profile of the average respiratory event, divided by time between these points. *Mean desaturation* was calculated on an event-by-event basis.

Apnea index and fraction of hypopneas. Apnea index was calculated as the number of apneas per hour of sleep. The fraction of hypopneas was calculated as the total number of hypopneas divided by the total number of respiratory events during sleep.

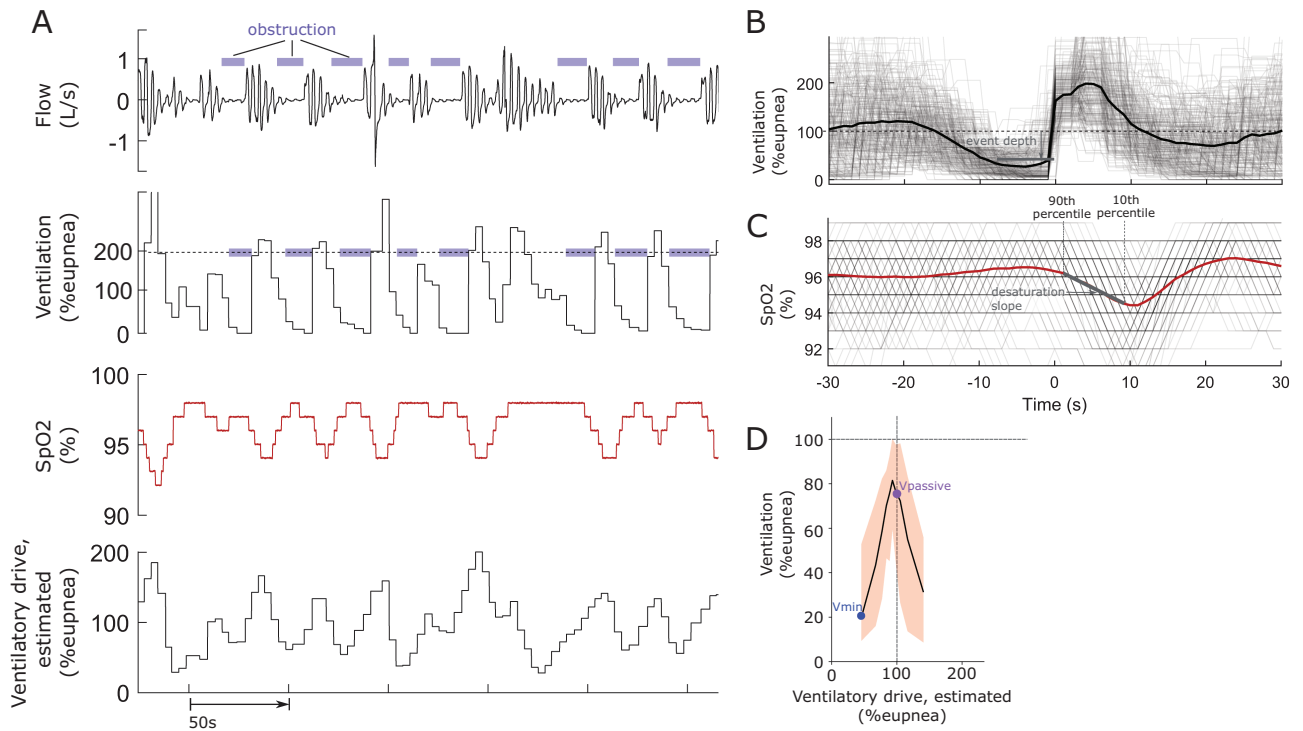


Figure 1. (A) Raw sample traces of flow, ventilation, SpO₂, and modeled (estimated) ventilatory drive during a series of respiratory events. (B) To compute event depth, ventilation profiles for each respiratory event are aligned to the terminal breath of the event and then ensemble averaged. The event depth is then calculated as the mean reduction (from eupnea) in ensemble averaged ventilation during the event. (C) To compute desaturation slope, saturation profiles for each respiratory event were synchronized at event termination and ensemble averaged. Desaturation slope was measured as the change in oxygen saturation from the 90th (i.e. event start) to the 10th (i.e. event end) percentile of the ensemble averaged saturation profile during the event, divided by time between these points. (D) The plot of ventilation as a function of ventilatory drive illustrating the derivation of $V_{passive}$ and V_{min} as the (median) ventilation observed at eupneic drive and at minimal drive (lowest decile), respectively.

Statistical analysis

Validation of each metric with the gold standard Pcrit measurement was performed using Pearson's correlation. To identify associations with demographics in the MESA cohort, multiple linear regression models were developed with age, sex, and BMI as predictors and the collapsibility metrics as the outcome. Prediction of treatment response was assessed with a multiple linear regression model with the collapsibility metric as the predictor, fraction of sleep time in the non-supine position as a covariate, and percent reduction in AHI from baseline to treatment as the outcome variable. For all analyses and variables, distributions were tested for normality (Shapiro-Wilk test) and transformed if necessary. Data analysis was performed using MATLAB (Statistics and Machine Learning Toolbox, Mathworks, Natick, MA).

While the present analysis includes multiple comparisons of seven different collapsibility metrics (increasing risk for type I error), this risk was more than counterbalanced by the requirement for each metric to pass multiple evaluations. In quantifying this balance, we determined that a collapsibility metric was considered statistically significant if it passed validation against Pcrit, was associated with two of age, sex, and BMI, and demonstrated a significant association with treatment response to oral appliance and/or ato-oxy (at an alpha level of $p < .05$ for each test). Criteria were designed with the use of simulations and permutation tests to minimize risk for type I error, while maintaining statistical power (for each measure) above 85% (see [Supplemental Materials](#) for details).

Results

The patient population for each of the datasets included are described in [Table 1](#).

Correlations with critical pharyngeal closing pressure

Pcrit, the gold-standard measure of collapsibility, was significantly correlated with all flow-derived measures of collapsibility ($V_{passive}$, $r = -0.38$; V_{min} , $r = -0.56$; and *event depth*, $r = 0.49$) and conventionally-derived metrics of collapsibility (*fraction of hypopneas*, $r = 0.46$ and *apnea index*, $r = 0.49$) plus AHI ($r = 0.32$) ([Figure 2](#)). Neither of the oxygen desaturation-based measures of collapsibility were correlated with Pcrit ([Figure 2](#)) and were thus not considered in subsequent evaluations. As a benchmark, we note a weaker correlation between AHI and Pcrit. Notably, among the new metrics, V_{min} and *event depth* performed better than the current standard polysomnographic surrogate of collapsibility ($V_{passive}$).

Association with age, sex, and obesity

All remaining collapsibility metrics ($V_{passive}$, V_{min} , *event depth*, *fraction of hypopneas*, *apnea index*) were significantly associated with sex and age. Effects of age and sex on all collapsibility metrics were comparable with the published effects of age and sex on Pcrit (0.5 SD/ Δ sex and 0.25 SD/2SD age, respectively, [Table 2](#)),

Table 1. Baseline values for patient demographics, apnea-hypopnea index (AHI), collapsibility metrics for comparison across the included datasets (median interquartile range).

	Pcrit	MESA	Oral appliance	Ato-oxy
Sex (men/total)	31/61	895/1916	61/81	11/15
Age (years)	56.4 [28.1, 70.2]	67.0 [61.0, 75.0]	50.0 [27.6, 67.0]	55.0 [36.3, 68.3]
BMI (kg/m ²)	32.8 [23.4, 44.2]	27.8 [24.6, 31.7]	30.3 [23.8, 42.8]	35.2 [23.8, 47.9]
Total sleep time (min)	200 [148,261]	369 [313, 416]	268 [155, 298]	212 [158, 263]
AHI (/h)	46.8 [16.0, 91.4]	32.7 [20.2, 50.9]	32.8 [16.3, 84.7]	36.7 [11.7, 86.9]
% Reduction AHI	NA	NA	54.9 [-28.9, 92.1]	73.8 [38.9, 98.2]
Treatment AHI (/h)	NA	NA	13.2 [2.6, 72.0]	8.5 [0.3, 33.4]
V _{passive} (% eupnea)	76.0 [0.0, 96.8]	96.5 [93.7, 98.1]	90.1 [30.4, 97.1]	88.9 [43.9, 99.3]
V _{min} (% eupnea)	38.3 [0.0, 68.5]	66.1 [52.9, 74.5]	51.7 [0.0, 68.9]	43.1 [2.2, 80.0]
Event depth (% eupnea)	46.5 [28.8, 61.6]	32.0 [25.4, 41.5]	40.9 [29.0, 66.0]	40.6 [26.3, 56.5]
Fraction of hypopneas	0.5 [0.0, 1.0]	1.0 [0.9, 1.0]	0.8 [0.1, 1.0]	0.8 [0.3, 1.0]
Apnea index (/h)	15.9 [0.0, 54.0]	1.3 [0.3, 4.7]	11.4 [0.1, 72.5]	6.4 [0.5, 65.3]

MESA data only includes patients from whom collapsibility metrics could be computed from supine NREM sleep.

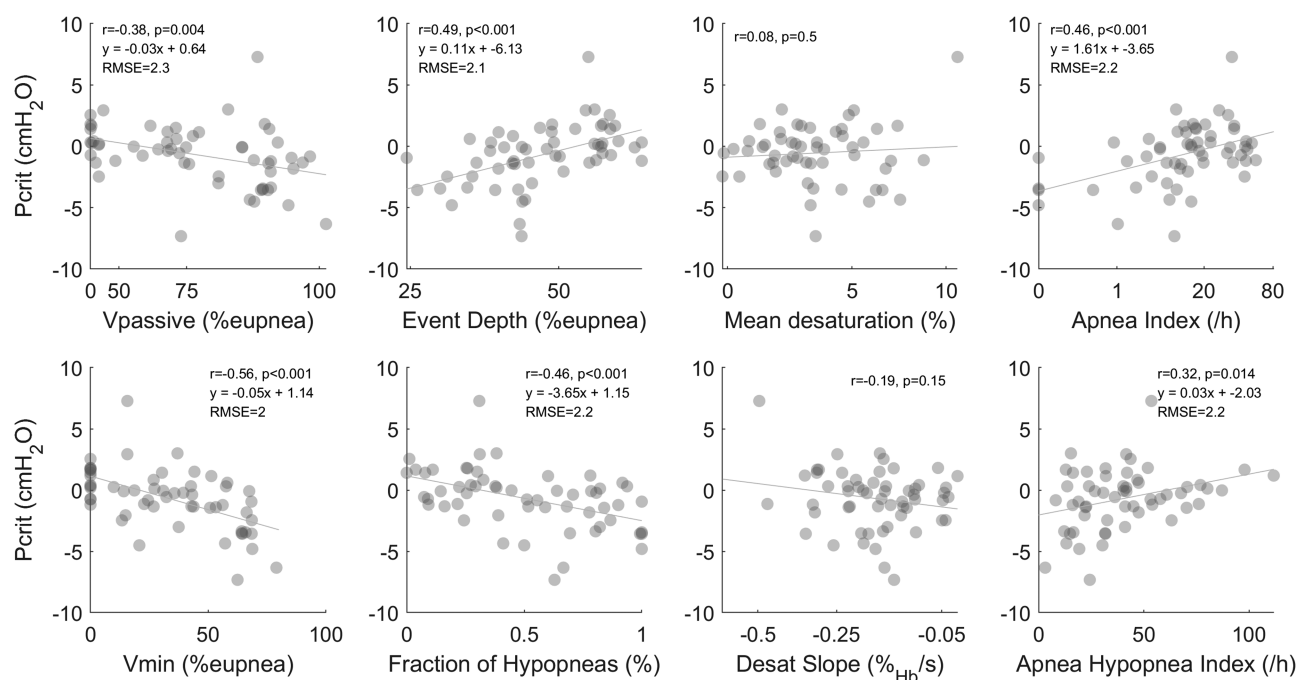


Figure 2. The gold-standard measure of collapsibility correlates with flow-based and clinically derived measures of collapsibility. Equations of the line represent the formula for calculating an estimated Pcrit (y) based on an input collapsibility surrogate (x). Root-mean square error (RMSE) provides a measure of variability around the estimated Pcrit measurement.

described in methods. The effect of obesity (BMI) on surrogate collapsibility metrics was within the acceptable margin of the published effect of BMI on Pcrit, but lower than expected (0.25 SD/2SD BMI, Table 2) for V_{passive}, V_{min}, and event depth. For apnea index and fraction of hypopneas, the association with obesity was outside the acceptable margin.

Predicting treatment response

Event depth, fraction of hypopneas, and apnea index were significantly associated with both oral appliance and ato-oxy efficacy (percent reduction in AHI, Table 3), after adjusting for fraction of time in non-supine position. Specifically, patients with shallower events, a greater fraction of hypopneas, and lower apnea

index (i.e. less collapsible pharynx) experienced a greater treatment response. In addition, V_{min} was associated with ato-oxy efficacy, while V_{passive} was not associated with treatment efficacy (Table 3).

Discussion

The current study identified event depth, fraction of hypopneas, and apnea index as preferred means to estimate pharyngeal collapsibility noninvasively in the clinical context. These measures were moderately correlated with the gold standard (Pcrit, $r = 0.49, 0.46, \text{ and } 0.46$, respectively), successfully detected increases in collapsibility with age, BMI (event depth only) and male sex, and importantly predicted treatment response (i.e.

Table 2. Associations between collapsibility measures and population variables in MESA (median [95% CI])

	Age	BMI	Sex
Collapsibility			
Vpassive	-0.23 [-0.32 to -0.14]	-0.14 [-0.24 to -0.05]	-0.68 [-0.77 to -0.59]
Vmin	-0.42 [-0.33 to -0.51]	-0.14 [-0.06 to -0.23]	-0.72 [-0.64 to -0.81]
Event depth	0.31 [0.23 to 0.40]	0.13 [0.04 to 0.21]	0.73 [0.64 to 0.81]
Fraction of hypopneas	-0.26 [-0.35 to -0.17]	0.07 [-0.02 to 0.16]	-0.42 [-0.51 to -0.33]
Apnea index	0.23 [0.15 to 0.32]	0.05 [-0.04 to 0.13]	0.50 [0.41 to 0.58]
Disease severity			
Apnea-hypopnea index (AHI)	0.26 [0.18 to 0.34]	0.66 [0.58 to 0.74]	0.53 [0.45 to 0.61]

Beta coefficients from multiple linear regression describe the change in collapsibility metrics with sex, age, and BMI (known population-level determinants of pharyngeal collapsibility). Beta coefficients represent SD change in collapsibility per unit change in sex (female = 0, male = 1) and per 2 SD change in age and BMI. Bolded values represent statistically significant Beta coefficients. Each row represents a single multivariable regression model.

Table 3. Associations between collapsibility measures and treatment efficacy (median [95% CI]).

	Vpassive	Vmin	Event depth	Fraction of hypopneas	Apnea Index	Apnea-hypopnea index
Oral appliance	-2.1 [-15.9 to 11.7]	10.1 [-3.0 to 23.2]	-15.0* [-28.1 to -1.9]	13.9* [0.3 to 27.5]	-15.4* [-29.6 to -1.1]	-0.2 [-13.9 to 13.5]
Ato-oxy	6.2 [-19.8 to 32.1]	32.5** [13.1 to 51.8]	-27.6* [-49.7 to -5.5]	33.0** [13.9 to 52.0]	-31.2** [-51.1 to -11.3]	-19.6 [-43.5 to 4.3]

Beta coefficients represents the % reduction in apnea-hypopnea index per 2SD change in collapsibility, after adjusting for fraction of sleep time in non-supine position. Each entry represents a single bivariate regression model result.

* $p < .05$; ** $p < 0.01$.

percent reduction in AHI) to both oral appliances and drug therapy. Vmin also passed our study-wide criteria for successful evaluation (successfully predicting responses to one of two therapies) and is expected to have utility in various circumstances: Vmin had the strongest association with Pcrit (also older age) and thus may be particularly useful when the strongest physiological link to collapsibility is required (i.e. to test the effect of an intervention on collapsibility). Finally, oxygen desaturation-based measures of collapsibility did not pass our evaluation and have limited evidence of utility at this stage. Overall, our study provides the field with three clinical surrogates of pharyngeal collapsibility, obtained from routine polysomnography, that have the potential to provide clinicians with the means to assess collapsibility and make treatment decisions that are physiologically informed.

Apnea index and fraction of hypopneas. The discovery of *apnea index* and *fraction of hypopneas* as surrogate measure of collapsibility provides a simple approach to estimating pharyngeal collapsibility from the routine polysomnography report. This finding builds on previous research showing that patients with more apneas have a more collapsible pharynx than patients with a majority of hypopneas [24]. This also aligns with Genta *et al.* who described a model that discriminated patients with high and low Pcrit using anthropometric and polysomnographic indices, which included the apnea index [25].

Event depth. The discovery of *event depth* as a surrogate measure of collapsibility provides a simpler approach than computing Vpassive and Vmin, while also requiring less data. Specifically, event depth only requires airflow data, while the Vpassive and Vmin require airflow and electroencephalography, along with carefully scored arousals and a modeled ventilatory drive. This makes the *event depth* metric amenable to computation by other research groups and manufacturers that develop sleep diagnostic technology. Incorporating the event depth metric into

polysomnography reports would be valuable information to aid in OSA treatment decision making.

Interpreting the size of the effect of *event depth*, *apnea index* and *fraction of hypopneas* on treatment response to each therapy is important in understanding the potential for pharyngeal collapsibility to select either treatment. For ato-oxy, a 2SD reduction in collapsibility results in an additional 30% reduction in AHI from baseline, an effect size that is clinically relevant on its own. By contrast, for oral appliance therapy, the same reduction in pharyngeal collapsibility results in an additional 15% reduction in AHI, suggesting that additional physiology is needed to fully inform therapeutic decision making (e.g. site of pharyngeal collapse, loop gain, and arousal threshold) [1-5]. Nonetheless, knowledge of the underlying collapsibility provides the backbone of any assessment of sleep apnea pathophysiology and prediction of therapeutic response [1-5, 26, 27].

Vmin. The present study reports a new flow-based collapsibility metric, Vmin, that appeared to be superior to the published Vpassive [8] in multiple tests. As a modification of Vpassive, Vmin describes the ventilation at the minimal (lowest decile) of (modeled) ventilatory drive rather than eupneic ventilatory drive (Figure 1D). This difference is designed to capture more hypotonic conditions [28], obviating the issue that actual ventilatory drive generally exceeds our estimated levels [8, 29]. Thus, we suspected that Vmin may be a better reflection of the anatomical (i.e. passive) contribution to pharyngeal collapsibility (i.e. the passive Pcrit). Indeed, Vmin yielded the highest correlation with passive Pcrit. A weakness of Vmin was that it was not associated with oral appliance treatment response, possibly due to a weaker correlation between nasal pressure and pneumotach derivations (see Supplemental Materials, e-Figure 3); notably after excluding the 36/81 patients in whom flow was recorded with nasal pressure cannula, there was a trend to suggest Vmin was associated with oral appliance treatment response with

a similar effect to event depth (every 2SD increase in V_{min} results in an additional 18% reduction in AHI, see [Supplementary Materials](#)). Thus, V_{min} may be less translatable to the clinical arena than *event depth*.

$V_{passive}$. In contrast to V_{min} , $V_{passive}$ had a weaker correlation with P_{crit} . It should still be noted, however, that the metric was predictive of response to oxygen therapy [26] and therefore may be useful in predicting response to select therapies. Furthermore, in multivariate OSA endotype-based models predicting response to oral appliance therapy, $V_{passive}$ was independently associated, albeit weakly, with treatment efficacy [2, 20]. However, the present study suggests that event depth, fraction of hypopneas, and apnea index are stronger metrics for characterizing pharyngeal collapsibility and responses to therapies.

Neither oxygen desaturation metric—oxygen desaturation slope and mean oxygen desaturation—was associated with P_{crit} , despite a theoretical basis (deeper events lead to faster alveolar oxygen depletion and desaturation) and a previous link between deeper respiratory events and faster desaturation [30]. Further investigation ([Supplemental Materials](#)) showed that *event depth* was associated with both oxygen desaturation measures (adjusting for BMI and pre-event oxygen saturation as appropriate), confirming the expected theoretical relationship. Thus, while we may provide a new explanation for heterogeneity of OSA-related oxygen desaturation, our data do not support the utility of these measures when airflow data are available.

Limitations

We considered several limitations of the current study. First, the advanced collapsibility metrics evaluated (*event depth*, V_{min}) are not currently available in sleep diagnostic technology and can only be computed with third-party software. As a result, hurdles remain for these measures to become available. However, ongoing research highlighting the clinical need is the first step toward their adoption in standard diagnostic technology and polysomnography reports. Fortunately, collapsibility surrogates that can be derived from the polysomnography report (fraction of hypopneas and apnea index) appear to be of equal utility in describing pharyngeal collapsibility and predicting treatment efficacy. Second, the computed surrogates of collapsibility aimed to characterize collapsibility in the passive (hypotonic) state. However, we acknowledge it is a great challenge to achieve a condition of a purely hypotonic airway on a sleep study without the use of CPAP and/or anesthesia. As such, each surrogate represents the upper airway in a state that is somewhere between passive and maximally activated. To explore this in more detail the associations between surrogates and P_{crit} were reexamined after adjusting for a surrogate of “active collapsibility” (V_{active} [8]). Associations for V_{min} and event depth with P_{crit} were only slightly attenuated, whereas associations for $V_{passive}$ were substantially attenuated and was better explained by V_{active} than P_{crit} (see [Supplementary Materials](#)). Mechanistically, V_{min} is expected to be maximally passive, given it is the ventilation calculated at the lowest level of estimated respiratory drive. Event depth is also designed to be as passive as possible, as respiratory events are a time of typically lower than average respiratory drive [31]. The consequences of including active collapsibility in surrogates of passive collapsibility are not entirely clear, but may not

necessarily make estimates less useful; measuring collapsibility at the very time the patient is experiencing obstructive events may be the most appropriate measure for characterizing the improvement in collapsibility needed to treat OSA. An inherent tradeoff is that a metric that more closely reflects active collapsibility may be less appropriate for describing tissue properties for investigating the genetic causes of OSA. Third, our study used unfiltered (direct current-coupled) airflow signals (best practice per American Academy of Sleep Medicine GUIDELINES), which is optimal for various aspects of our analyses. Nasal pressure signals that are high pass filtered (e.g. optional high-pass “baseline removal” filtering of 0.03 Hz) could distort the signal enough to negatively impact flow-derived measures of collapsibility. Fourth, reliability of event-based metrics diminishes with fewer events. In the current study, all patients had an AHI of at least 10/h to protect against this issue. Therefore, for patients with very low AHI (i.e. <5/h), these collapsibility measures may be unreliable. However, for low AHI (<10/h) we can safely assume mild collapsibility [32]. Fifth, we explored treatment responses to oral appliances and atomoxetine-plus-oxybutynin as representative of oral device and drug interventions more generally. Further testing is needed to assess (1) the repeatability of these clinical measures and (2) the applicability to a wider variety of additional non-PAP alternatives (e.g. hypoglossal nerve stimulation treatment, surgical treatment [e.g. palatopharyngoplasty and maxillo-mandibular advancement, and others]). Sixth, the present study does not link collapsibility metrics to other clinical outcomes of therapy such as sleep quality, blood pressure, or adherence to therapy. However, we expect that benefits in these domains will stem from efficacious interventions. Lastly, predictive capability of collapsibility surrogates was established retrospectively on data collected for other research aims. Therefore, while we report an important developmental step in providing surrogate measures of collapsibility for clinical decision-making, validation on prospectively collected data is a necessary next step.

Conclusion

The present study identifies *event depth*, *fraction of hypopneas*, and *apnea index* as preferred surrogate measures of collapsibility. While prospective validation is needed, *fraction of hypopneas* and *apnea index* are two collapsibility surrogates that could be readily calculated in clinical context. This research represents an important step toward simplifying advanced metrics that characterize OSA disease mechanisms for use in clinical decision-making on OSA patients.

Supplementary material

Supplementary material is available at *SLEEP* online.

Funding

The research project received funding from the National Institutes of Health (R01HL128658 and HL102321). DV is supported by the American Heart Association Postdoctoral Fellowship Award (20POST35210530); BAE is supported by a Heart Foundation of Australia Future Leader Fellowship (101167). MM

was supported by the Sao Paulo Research Foundation (FAPESP). SAS was supported by the NIH (R01HL146697) and the American Academy of Sleep Medicine Foundation (228-SR-20). SAS, LT-M, and AA were supported by the American Heart Association (15SDG25890059, 17POST33410436, 19CDA34660137). SAJ is supported by NHMRC Fellowship (1139745).

Disclosure Statement

OMV reports grants from Government of Flanders (Belgium)-IWT, during the conduct of the study; grants at Antwerp University Hospital from Philips and Somnomed, research support from Inspire Medical Systems, member of Advisory Board for Zephyr and Liva Nova, speaker's fees from Somnomed and Inspire Medical Systems, OMV holds a Senior Clinical Fellowship Grant (Fundamenteel Klinisch Mandaat) from Research Foundation—Flanders—Vlaanderen (FWO); LT-M works as a consultant for Apnimed and received personal fees as a consultant for Cambridge Sound Management outside the submitted work. LT-M is CSO of Apnimed, a company developing pharmacologic therapies for sleep apnea; AW works as a consultant for Apnimed, Somnifix, and Nox and he has received grants from Somnifix and Sanofi. AW has a financial interest in Apnimed, a company developing pharmacologic therapies for sleep apnea; DPW works as Distinguished Scientist for Apnimed; SAS received grants from Apnimed, Prosomnus, and Dynaflex, and personal fees from Apnimed, Merck, and Nox Medical; AA receives personal fees as a consultant for Somnifix and Apnimed and receives grant support from American Heart Association, American Academy of Sleep Medicine, and Somnifix; BE receives grant support from Apnimed and NHMRC; GSH has received equipment for research from Resmed, Philips Respironics and Air Liquide Healthcare; JV receives personal fees as a consultant or speaker from Ectosense, Philips, ResMed Narval, Sanofi, AstraZen, Daiichi-Sankyo, Zambon, Bioprojet, Schwabe Pharma, Metagenics, Agfa-Gevaert, SomnoMed, Springer, Inspire Medical Systems, Boehringer Ingelheim, Plastiflex, Jansen-Cilag, Vivisol, and Air Liquide outside the submitted work. MB reports governmental grant from Government of Flanders (Belgium)-IWT during the conduct of the study; MB is member of the advisory board of Somnomed and ResMed and receives as such travel funding and speaker fees outside the submitted work. SR reports grants from NIH, during the conduct of the study, and a grant from Jazz Pharma; consulting fees from Jazz Pharma, Eisai Pharma, Respicardia Inc, and Apnimed. DV, MM, SODB, RR, SJ, LG have no conflict of interest to disclose. Nonfinancial disclosure: None.

Author Contributions

DV: contributed to study design, data analysis and interpretation, and drafting and review of the manuscript for important intellectual content. LT-M: contributed to the study design, and interpretation, and review of the manuscript. MM, SODB, OMV, BAE, RR, GSH, SAJ, JV, and MB: contributed to data collection and review of the manuscript. AA, LG, SR, and DPW: contributed to data interpretation and review of the manuscript. AW: contributed to the study design, data analysis and interpretation, and drafting and review of the manuscript. SAS: contributed to the study design, data collection, data analysis and interpretation, and drafting and review of the manuscript.

References

- Marques M, et al. Structure and severity of pharyngeal obstruction determine oral appliance efficacy in sleep apnoea. *J Physiol*. 2019;597:5399–5410.
- 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.
- Vena D, et al. Predicting sleep apnea responses to oral appliance therapy using polysomnographic airflow. *Sleep*. 2020;43(7). doi:10.1093/sleep/zsaa004.
- Op de Beeck S, et al. Mandibular advancement device treatment efficacy is associated with polysomnographic endotypes. *Ann Am Thorac Soc*. 2021;18(3):511–518.
- Edwards BA, S, 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.
- Taranto-Montemurro L, et al. The combination of atomoxetine and oxybutynin greatly reduces obstructive sleep apnea severity. A randomized, placebo-controlled, double-blind crossover trial. *Am J Respir Crit Care Med*. 2019;199(10):1267–1276.
- Patil SP, et al. A simplified method for measuring critical pressures during sleep in the clinical setting. *Am J Respir Crit Care Med*. 2004;170(1):86–93.
- 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.
- Mann DL, et al. Quantifying the magnitude of pharyngeal obstruction during sleep using airflow shape. *Eur Respir J*. 2019;54(1):1802262.
- Taranto-Montemurro L, et al. Desipramine improves upper airway collapsibility and reduces OSA severity in patients with minimal muscle compensation. *Eur Respir J*. 2016;48(5):1340–1350.
- 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.
- Taranto-Montemurro L, et al. Desipramine increases genioglossus activity and reduces upper airway collapsibility during non-REM sleep in healthy subjects. *Am J Respir Crit Care Med*. 2016;194(7):878–885.
- 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.
- Edwards BA, et al. Acetazolamide improves loop gain but not the other physiological traits causing obstructive sleep apnoea. *J Physiol*. 2012;590(5):1199–1211.
- 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.
- Bild DE, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156(9):871–881.
- 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.
- Won CH, et al. Sex differences in obstructive sleep apnea phenotypes, the multi-ethnic study of atherosclerosis. *Sleep*. 2020;43(5). doi:10.1093/sleep/zsz274.
- Kirkness JP, et al. Contribution of male sex, age, and obesity to mechanical instability of the upper airway during sleep. *J Appl Physiol*. 2008;104(6):1618–1624.

20. Op de Beeck S, et al. Phenotypic labelling using drug-induced sleep endoscopy improves patient selection for mandibular advancement device outcome: a prospective study. *J Clin Sleep Med*. 2019;**15**(08):1089–1099.
21. Verbruggen AER, et al. Predicting therapeutic outcome of mandibular advancement device treatment in obstructive sleep apnoea (PROMAD): study design and baseline characteristics. *J Dent Sleep Med*. 2016;**03**(04):119–138.
22. Azarbarzin A, et al. The hypoxic burden of sleep apnoea predicts cardiovascular disease-related mortality: the osteoporotic fractures in men study and the sleep heart health study. *Eur Heart J*. 2019;**40**(14):1149–1157.
23. Azarbarzin A, et al. The sleep apnea-specific hypoxic burden predicts incident heart failure. *Chest*. 2020;**158**(2):739–750.
24. Gleadhill IC, et al. Upper airway collapsibility in snorers and in patients with obstructive hypopnea and apnea. *Am Rev Respir Dis*. 1991;**143**(6):1300–1303.
25. Genta PR, et al. Discriminating the severity of pharyngeal collapsibility (Pcrit) using anthropometric and polysomnographic indices. *J Clin Sleep Med*. 2020;**16**(9):1531–1537.
26. Sands SA, et al. Identifying obstructive sleep apnoea patients responsive to supplemental oxygen therapy. *Eur Respir J*. 2018;**52**(3):1800674.
27. Taranto-Montemurro L, et al. Effects of the combination of atomoxetine and oxybutynin on OSA endotypic traits. *Chest*. 2020;**157**(6):1626–1636.
28. Onal E, et al. Dynamics of respiratory drive and pressure during NREM sleep in patients with occlusive apneas. *J Appl Physiol*. 1985;**58**(6):1971–1974.
29. Sands SA, et al. Quantifying the arousal threshold using polysomnography in obstructive sleep apnea. *Sleep*. 2017;**41**(1). doi:[10.1093/sleep/zsx183](https://doi.org/10.1093/sleep/zsx183).
30. Vena D, et al. Slope of the oxygen desaturation reflects the pharyngeal collapsibility in OSA [abstract]. *Sleep Med*. 2019;**64**(Suppl 1):S403.
31. Gell L, et al. Neural ventilatory drive decline as a predominant mechanism of obstructive sleep apnoea events. *Thorax*. 2022.
32. 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.