

# Polysomnographic Endotyping to Select Patients with Obstructive Sleep Apnea for Oral Appliances

Ahmad A. Bamagoos<sup>1,2,3,4</sup>, Peter A. Cistulli<sup>1,2</sup>, Kate Sutherland<sup>1,2</sup>, Melanie Madronio<sup>2</sup>, Danny J. Eckert<sup>4,5</sup>, Lauren Hess<sup>6</sup>, Bradley A. Edwards<sup>7,8</sup>, Andrew Wellman<sup>6</sup>, and Scott A. Sands<sup>6</sup>

<sup>1</sup>Sleep Research Group, Charles Perkins Centre, University of Sydney, Sydney, New South Wales, Australia; <sup>2</sup>Centre for Sleep Health and Research, Department of Respiratory and Sleep Medicine, Royal North Shore Hospital, Sydney, New South Wales, Australia; <sup>3</sup>Department of Physiology, Faculty of Medicine in Rabigh, King Abdulaziz University, Jeddah, Saudi Arabia; <sup>4</sup>Sleep and Breathing Lab, Neuroscience Research Australia (NeuRA), Randwick, New South Wales, Australia; <sup>5</sup>Adelaide Institute for Sleep Health, Flinders University, Bedford Park, South Australia, Australia; <sup>6</sup>Division of Sleep and Circadian Disorders, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; <sup>7</sup>Sleep and Circadian Medicine Laboratory, Department of Physiology, and <sup>8</sup>Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Melbourne, Victoria, Australia

ORCID IDs: 0000-0001-7142-081X (A.A.B.); 0000-0003-3503-2363 (D.J.E.); 0000-0001-9383-1644 (B.A.E.).

## Abstract

**Rationale:** Oral appliance therapy is efficacious in many patients with obstructive sleep apnea (OSA), but prediction of treatment outcome is challenging. Small, detailed physiological studies have identified key OSA endotypic traits (pharyngeal collapsibility and loop gain) as determinants of greater oral appliance efficacy.

**Objectives:** We used a clinically applicable method to estimate OSA traits from routine polysomnography and identify an endotype-based subgroup of patients expected to show superior efficacy.

**Methods:** In 93 patients (baseline apnea-hypopnea index [AHI],  $\geq 20$  events/h), we examined whether polysomnography-estimated OSA traits (pharyngeal: collapsibility and muscle compensation; nonpharyngeal: loop gain, arousal threshold, and ventilatory response to arousal) were associated with oral appliance efficacy (percentage reduction in AHI from baseline) and could predict responses to treatment. Multivariable regression (with interactions) defined endotype-based subgroups of “predicted” responders and nonresponders (based on 50% reduction in AHI). Treatment efficacy was compared between the predicted subgroups (with cross-validation).

**Results:** Greater oral appliance efficacy was associated with favorable nonpharyngeal traits (lower loop gain, higher arousal threshold, and lower response to arousal), moderate (nonmild, nonsevere) pharyngeal collapsibility, and weaker muscle compensation (overall  $R^2 = 0.30$ ; adjusted  $R^2 = 0.19$ ;  $P = 0.003$ ). Predicted responders ( $n = 54$ ), compared with predicted nonresponders ( $n = 39$ ), exhibited a greater reduction in AHI from baseline (mean [95% confidence interval], 73% [66–79] vs. 51% [38–61];  $P < 0.0001$ ) and a lower treatment AHI (8 [6–11] vs. 16 [12–20] events/h;  $P = 0.002$ ). Differences persisted after adjusting for clinical covariates (including baseline AHI, body mass index, and neck circumference).

**Conclusions:** Quantifying OSA traits using clinical polysomnography can identify an endotype-based subgroup of patients that is highly responsive to oral appliance therapy. Prospective validation is warranted.

**Keywords:** sleep-disordered breathing; precision medicine; targeted therapy; phenotype; mandibular advancement splints

(Received in original form March 1, 2019; accepted in final form July 15, 2019)

Supported by funding for doctoral studies from the Saudi Arabian Government (Department of Physiology, Rabigh Medical School, King Abdulaziz University) (A.A.B.); the American Heart Association (grant 15SDG25890059), the American Thoracic Society Foundation, and the National Institutes of Health (grant R01HL102321) (A.W. and S.A.S.); a National Health and Medical Research Council Australia Senior Research Fellowship (1116942) (D.J.E.); and a Heart Foundation of Australia Future Leader Fellowship (101167) and grants from the National Health and Medical Research Council Australia (B.A.E.).

**Author Contributions:** Conception and design: A.A.B., P.A.C., K.S., B.A.E., A.W., and S.A.S. Parent study data collection: P.A.C. and K.S. Data collation and analysis: A.A.B., P.A.C., K.S., M.M., L.H., and S.A.S. Arousal scoring for the present study: A.A.B., L.H., and M.M. All authors interpreted results, edited the manuscript for important intellectual content, and approved the final draft.

Correspondence and requests for reprints should be addressed to Scott A. Sands, Ph.D., Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, 221 Longwood Avenue, Boston, MA 02115. E-mail: sasands@bwh.harvard.edu.

Ann Am Thorac Soc Vol 16, No 11, pp 1422–1431, Nov 2019

Copyright © 2019 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201903-1900C

Internet address: www.atsjournals.org

Oral appliances, intraoral devices worn during sleep to protrude the mandible, are increasingly used as a treatment alternative for obstructive sleep apnea (OSA) (1). The literature indicates that oral appliance therapy reduces OSA severity (indicated by the apnea-hypopnea index [AHI]) by an average of 50–70% (2–6). Although not as efficacious as continuous positive airway pressure (CPAP) at ameliorating OSA, oral appliance therapy has a proven positive impact on sleepiness, blood pressure, and quality of life (7–11). In addition, studies suggest that treatment outcomes of oral appliances and CPAP are similar (12), reflecting superior adherence to oral appliance therapy. Efficacy of oral appliance therapy is variable across patients with OSA. Without experimental testing in each patient (13, 14), there is currently no clinically applicable means to predict the likelihood of oral appliance therapy success before treatment prescription (15).

The variability in oral appliance efficacy across patients with OSA may be attributed to the extent to which OSA endotypic traits (pharyngeal: collapsibility and muscle compensation; nonpharyngeal: loop gain, arousal threshold, and ventilatory response to arousal) contribute to the pathogenesis of the condition (16–25). Small, detailed physiological studies have revealed two key traits associated with reduced oral appliance efficacy: namely, greater pharyngeal collapsibility, whereby the severity is beyond the scope of treatment (26), and higher loop gain (i.e., ventilatory control instability) that cannot be resolved with anatomical interventions (20). Notably, these detailed studies were performed in specialized laboratories using invasive instrumentation, which is out of reach for clinical sleep laboratories.

Recently, our team has developed a method for estimating the key endotypic traits causing OSA from routine diagnostic polysomnography (27–29). The method is based on automated analysis of a surrogate uncalibrated ventilation signal (derived from nasal pressure) from which ventilatory drive is estimated and OSA endotypic traits are characterized. In the present study, we applied this method to diagnostic polysomnography of patients treated with oral appliance therapy. We aimed to 1) determine whether greater oral appliance efficacy is associated with favorable nonpharyngeal OSA endotypes (i.e., lower loop gain, higher arousal threshold, and

lower ventilatory response to arousal) with the ultimate goal of 2) defining an endotype-based subgroup of patients with OSA (prediction model) who are most likely to benefit from oral appliance therapy (predicted responders).

## Methods

### Subjects

In this study, we performed a secondary analysis of polysomnographic data from previous oral appliance research studies (which included newly diagnosed patients with OSA with baseline AHI >10 events/h) (12, 30, 31). Patients were included in our analysis if they had a baseline polysomnography-derived AHI greater than or equal to 20 events per hour (prespecified), which was selected to minimize the influence of night-to-night variability (noise) on the percentage reduction in AHI with treatment (efficacy) (32). For example, a 75% reduction in AHI from 20 to 5 events per hour was considered more reliable than the same from 10 to 2.5 events per hour. (The latter is within the expected night-to-night variability; i.e., ~10 events/h.)

Polysomnographic data ( $n = 94$ ) were taken from three parent studies: a three-center randomized crossover trial (12) comparing health outcomes of CPAP therapy with oral appliance therapy in patients who were recommended both treatments ( $n = 108$ ; 80% males; mean  $\pm$  standard deviation (SD) age,  $50 \pm 11$  yr; body mass index [BMI],  $30 \pm 6$  kg/m<sup>2</sup>; AHI,  $26 \pm 12$  events/h), a single-center observational study (30) designed to examine awake-based predictors of oral appliance efficacy in patients to whom mandibular advancement splint therapy was recommended ( $n = 142$ ; 59% males; mean  $\pm$  SD age,  $56 \pm 11$  yr; BMI,  $30 \pm 5$  kg/m<sup>2</sup>; AHI,  $29 \pm 18$  events/h), and an ongoing dual-center observational study (31) that is using magnetic resonance imaging-based genioglossus dynamics to explain heterogeneity in oral appliance efficacy (at assessment,  $n = 40$ ; 72% males; mean  $\pm$  SD age,  $43 \pm 11$  yr; BMI,  $30 \pm 5$  kg/m<sup>2</sup>; AHI,  $27.7 \pm 16.9$  events/h). Although some of the parent studies were multicenter, all polysomnographic data included in our analysis were derived from a single sleep clinic. Key exclusion criteria for the original research studies were previous

oral appliance use, oral appliance contraindications (including insufficient number of teeth, periodontal disease, and severe daytime sleepiness requiring urgent intervention), and predominance of central sleep apnea at baseline. Oral appliance efficacy was determined using in-laboratory polysomnography with the oral appliance *in situ*. Because all data were deidentified, the present analysis was deemed to be exempt from the need for consent by the Human Research Ethics Committee at North Sydney Local Health District, New South Wales, Australia, and the Partners Institutional Review Board, Boston, Massachusetts.

### Study Protocol

Patients first attended in-laboratory diagnostic polysomnography (electroencephalography, electrocardiography, electrooculography, chin and leg electromyography, thoracoabdominal plethysmography, pulse oximetry, body position, nasal pressure airflow, and thermistor signals), which was scored according to the 2012 American Academy of Sleep Medicine (AASM) criteria (12, 33) (30% reduction in airflow with either 3% oxygen desaturation or cortical arousal) or 2007 AASM criteria (30, 31, 34) (30% reduction in airflow with 4% oxygen desaturation). Models were not adjusted for scoring type, because no effect was evident (*see RESULTS*). All patients were treated with a custom-made oral appliance (SomnoDent; SomnoMed Ltd.) implemented for individual patients under the supervision of a treating dentist. Devices were initially set to 70% of the maximal mandibular protrusion from habitual bite. Patients were instructed to incrementally advance the protrusive level of the device until the maximum comfortable limit was reached (~6–8 wk), which was then confirmed by the treating dentist. A second in-laboratory polysomnography was performed to determine response to therapy.

### Oral Appliance Efficacy

Oral appliance efficacy was described by the percentage reduction in AHI with treatment relative to baseline (primary outcome, continuous variable). This measure was selected (over absolute reduction or treatment AHI) to maximize statistical power (typically largest mean/SD ratio [35], correlates least with baseline AHI [36]).

## Endotypic Trait Analysis

**Raw data.** We identified 94 patients who met our prespecified eligibility criteria for analysis ( $n = 50$  excluded for  $\text{AHI} < 20$  events/h). For optimal endotyping analyses, a thorough manual check of cortical arousal onset and end times was performed for baseline polysomnography by three experienced scorers. Adjustment of arousal timing (if required) was performed blinded to treatment outcome. Raw data and accompanying annotations (staging, arousals, and respiratory events) were exported for each patient. Data from one participant were excluded because of poor nasal pressure signal (automated quality control algorithm, verified visually), leaving 93 patients for analysis. Analysis of the traits was restricted to non-rapid eye movement (non-REM) sleep for consistency with previous validation studies using our methods and to avoid the influence of night-to-night variability in REM duration on the measurements (27–29). Data from supine and lateral positions were pooled, given the interest in predicting changes in the total AHI with treatment, regardless of position.

### Quantifying OSA endotypic traits.

These methods have been described in detail previously (27–29, 37). Each diagnostic polysomnogram was automatically segmented into 7-minute overlapping windows containing non-REM sleep. The analyses were performed for each window separately, and median values across windows were used to represent each individual. First, nasal pressure (linearized, square root) provided an uncalibrated breath-to-breath ventilation signal (volume  $\times$  rate), calibrated such that the mean eupneic ventilation for the window being analyzed was 100% (29). “Ventilatory drive” was defined as the intended ventilation that would be observed if the airway was completely open (i.e., immediately after a scored cortical arousal). Ventilatory drive was estimated by least squares fitting of a regression model that seeks to predict ventilation (i.e., overshoot between obstructive events) on the basis of previous values of ventilation. This chemoreflex model is physiologically constrained such that three key parameters are identified (gain, time constant, and delay [27]). These parameters were used to calculate the loop gain (ventilatory drive response to an oscillatory disturbance at the

natural frequency, which captures the combined influence of chemoreflex sensitivity, plant gain, and circulatory delay; a value of 1 would predict central sleep apnea) (27, 37). The ventilatory response to arousal (additional ventilatory drive response that accompanies arousal, independent of the chemoreflex contribution) was found by including the presence of a scored electroencephalographic arousal on any breath as a covariate. The arousal threshold was calculated as the mean ventilatory drive on the breath immediately preceding scored arousals (29).

To calculate collapsibility and muscle compensation, an overnight endotype plot (28) was generated, whereby all breath-by-breath values of ventilation and ventilatory drive for the whole night (except breaths in awake, arousals, and REM) are tabulated and plotted against each other. The median value of ventilation at eupneic ventilatory drive was taken as a measure of passive collapsibility ( $V_{\text{PASSIVE}}$ ). A lower  $V_{\text{PASSIVE}}$  value reflects greater collapsibility. The median value of ventilation at maximal ventilatory drive (at arousal threshold) was taken as a measure of active collapsibility ( $V_{\text{ACTIVE}}$ ). The difference between  $V_{\text{ACTIVE}}$  and  $V_{\text{PASSIVE}}$  was used as a measure of pharyngeal muscle compensation. Analysis was fully automated using custom in-house software (MATLAB; MathWorks; interested users are directed to contact the authors) and visually verified.

### Model Development and Statistical Analyses

The goals of the statistical analyses were twofold. First, we sought to describe the associations between oral appliance efficacy and OSA endotypic traits (in combination) at baseline to provide insight into mechanistic causes of variability in oral appliance efficacy. A multivariable regression model approach (37) was employed (outlined below). Second, we sought to use the same endotype-based regression model as the basis of a prediction model to examine the extent to which these endotypes could be employed to identify an endotype-based subgroup of patients with OSA who would exhibit greater oral appliance efficacy (“predicted responders”) than other patients (“predicted nonresponders”).

**Statistical power.** Ninety patients were estimated to provide 86% power to

detect significant independent associations with  $R^2 > 0.1$  ( $\alpha = 0.05$ ) and 92% power to detect differences in efficacy between endotypic subgroups of at least  $20 \pm 40\%$ .

**Data transformation.** Several variables were not normally distributed (Shapiro-Wilk test) and were transformed accordingly:  $V_{\text{PASSIVE}}$  and arousal threshold were square root transformed via the equations  $y = 1 + (x - 1)^{0.5}$  and  $y = 1 - (1 - x)^{0.5}$ , respectively, where  $x = 1$  describes 100% (37). The percentage reduction in AHI with treatment versus baseline (primary outcome variable) was transformed to avoid left skewness via the equation  $y = x/(2 - x)$ , which is equivalent to  $y = (\text{baseline AHI} - \text{treatment AHI})/(\text{baseline AHI} + \text{treatment AHI})$ , where  $x$  ranges between  $-1$  and  $+1$ . This transformation was made so that, for instance, halving or doubling baseline AHI with oral appliance therapy would produce equal and opposite effects on the transformed outcome ( $y$  is noted as  $\Delta\text{AHI}$  from now on). Thus, halving baseline AHI ( $x = 0.5$ ) yields  $\Delta\text{AHI} = 33\%$ , and doubling baseline AHI ( $x = -1$ ) yields  $\Delta\text{AHI} = -33\%$ . Baseline and treatment AHI were also leftward skewed and transformed using  $y = x^{1/3}$ . All variables were back transformed for presentation.

**Bivariate analyses.** Simple linear regression analyses were initially performed to evaluate the relationships between  $\Delta\text{AHI}$  and each OSA endotypic trait individually.

### Multivariable regression analysis.

Initial visual inspection of the data (plots showing responders and nonresponders against combinations of collapsibility, loop gain, and arousal threshold; data not shown) suggested complex interactions between multiple traits and oral appliance efficacy that could not be captured appropriately using bivariate regression. We therefore employed a quadratic regression analysis in which the total possible variable terms (i.e., 20 terms) were included. These 20 terms were the 5 individual OSA endotypic traits, the 5 endotypic traits after square transformation (1 square transformation per trait), and 10 interaction terms. Significant square-transformed terms would indicate nonlinear relationships with efficacy (e.g., U-shaped curve), and interaction terms would imply that the relationship between efficacy and an endotypic trait varies depending on the level of another trait. An example quadratic

model expression with just two traits would be  $\beta_0 + \beta_1(\text{loop gain}) + \beta_2(V_{\text{PASSIVE}}) + \beta_3(\text{loop gain})^2 + \beta_4(V_{\text{PASSIVE}})^2 + \beta_5(\text{loop gain} \times V_{\text{PASSIVE}})$ . To determine which terms should be included, we employed a backward elimination method. We started with a model that included all 20 terms. Subsequently, backward stepwise elimination iteratively removed each term with the highest  $P$  value if  $P > 0.157$  (Wald test, equivalent to Akaike information criterion, indicating that the relative quality of the model was not improved with the inclusion of the term [37–41]). This approach was used on the basis that removal of very weak predictors reduces uncertainty of the remaining model coefficients. Terms were accepted as significant at  $P < 0.05$ . For interpreting associations, we did not adjust the  $P$ -value threshold for multiple comparisons (e.g., five traits: conservative  $P$  threshold = 0.01).

**Defining endotypic subgroups.** Defining the endotypic subgroups of predicted responders and predicted nonresponders was based on the regression model above and the following steps. True responders and true nonresponders were defined by percentage reduction in AHI with treatment (true efficacy cutoff, 50%). Predicted responders and predicted nonresponders were defined by determining the optimal cutoff, derived from the multivariable regression model output, that maximized sensitivity plus specificity (a model-predicted efficacy cutoff of 60% was found; see RESULTS); note that model-predicted efficacy and true efficacy are not equal. Predicted subgroups were allocated using a “leave one patient out” cross-validation procedure to avoid overestimating predictive performance. This procedure ensured that the outcome status of a given patient was predicted on the basis of a model that included all patients’ data except his/her own. Thus, cross-validated results are more conservative (more likely indicative of future retest performance). For example, allocation of patient 1 to a predicted responder subgroup or a predicted nonresponder subgroup was determined by building a modified model (via rerunning backward elimination regression described above) without the data of patient 1 and then using this modified model to predict patient 1’s response. This process was then repeated for all other patients from patient 2 to patient 93. The primary statistical comparison for the study was the difference

in percentage reduction in AHI (primary outcome variable) between the predicted endotypic subgroups.

**Adjusting for clinical covariates.** Multivariable linear regression was used to determine whether predicted response status (being a predicted responder vs. a predicted nonresponder) could predict oral appliance efficacy ( $\Delta$ AHI) independently of clinical covariates (i.e., baseline AHI, BMI, age, sex, and neck circumference; baseline REM/NREM AHI and change in REM sleep duration with treatment were also assessed). To perform this test, predicted response status (1 or 0, respectively) was included as an independent variable, and clinical covariates were sequentially included and then removed from the model.

**Data presentation.** Data are presented as mean  $\pm$  SD for descriptive variables and mean  $\pm$  SEM for comparisons. Back-transformed data were presented as mean (95% confidence interval [CI]). Data were described as median [25th–75th centile] for nonnormally distributed data as appropriate. Significance was accepted at  $P < 0.05$ . Figures were created using custom MATLAB software (MathWorks).

## Results

### Baseline Characteristics

Data from 93 participants (56% males) were analyzed. Baseline versus treatment characteristics for the overall group are presented in Table 1. On average, participants were middle-aged ( $56.2 \pm 11.0$  yr), obese ( $30.5 \pm 5.3$  kg/m<sup>2</sup>), and had moderate to severe OSA ( $30.6$  [24.4–43.5] events/h).

### Oral Appliance Therapy

The final protrusion provided by the oral appliance was, on average, 89% (range, 44–100%) of the maximal mandibular protrusion. Overall, treatment lowered AHI by a median of 67% and had favorable effects on arousal frequency and oxygenation (Table 1). Sixty-three patients were responders (>50% reduction in AHI).

### Bivariate Analyses

Using simple linear regression analyses, we observed no bivariate associations between oral appliance efficacy (percentage reduction in AHI transformed [ $\Delta$ AHI])

and any of the individual endotypic traits at baseline ( $R^2 < 0.01$  for all). There were also no associations between oral appliance efficacy and baseline AHI, BMI, age, sex, or neck circumference.

### Multivariable Regression Analysis

When endotypic traits were considered in combination (multivariable regression), we found that greater oral appliance efficacy was associated with moderate  $V_{\text{PASSIVE}}$  (nonsevere and nonmild), lower pharyngeal compensation, and more favorable nonpharyngeal traits (i.e., lower loop gain, higher arousal threshold, and lower response to arousal) (see Table 2 and Figure 1). Several interaction variables were also associated with treatment efficacy (see Table 2 and Figure 1 for interpretation of each of the 12 terms included in the model).

### Defining Endotypic Subgroups

Use of the multivariable regression model above to define endotype subgroups of predicted responders and predicted nonresponders revealed the findings reported in the subsections below.

**Before cross-validation.** Predicted responders ( $n = 57$ ), compared with predicted nonresponders ( $n = 36$ ), exhibited a greater reduction in AHI from baseline (mean [95% CI], 76% [70–80] vs. 42% [28–55];  $P < 0.0001$ ) and had lower treatment AHI (8 [6–10] vs. 18 [14–23] events/h;  $P < 0.0001$ ). Positive and negative predictive values were 83% and 56%, respectively; accuracy was 72%.

**After cross-validation (main results).** Differences in responses between subgroups remained clinically significant after cross-validation: Predicted responders ( $n = 54$ ), compared with predicted nonresponders ( $n = 39$ ), exhibited a greater reduction in AHI from baseline (mean [95% CI], 73% [66–79] vs. 51% [38–61];  $P < 0.0001$ ) and had lower treatment AHI (8 [6–11] vs. 16 [12–20] events/h;  $P = 0.002$ ) (see Figure 2). Positive and negative predictive values were 78% (42:12) and 46% (18:21), respectively ( $P = 0.02$ , Fisher’s exact test); accuracy was 65%.

**Further analyses.** Adjusting for covariates (baseline AHI, BMI, age, sex, neck circumference, baseline REM/NREM AHI, and change in REM sleep duration with treatment) did not attenuate the differences between groups. Notably, baseline AHI was similar between groups (predicted responders vs. predicted nonresponders,

**Table 1.** Patient characteristics

Characteristic	Baseline	Oral Appliance	P Value
Sex, M:F		52:41	
Age, yr	56.2 ± 11.0		
BMI, kg/m <sup>2</sup>	30.5 ± 5.3		
Neck circumference, cm	40.2 ± 4.1		
Maximum possible advancement, mm	10.4 ± 3.4		
Final advancement, mm	9.2 ± 3.0		
Mandibular advancement, % maximum	88.5 ± 14.8		
AHI, total events/h	30.6 [24.4–43.5]	11.3 [5.5–19.1]	<0.001
% reduction ( $\Delta$ AHI)	67.4 [42.5–83.1]		
AHI, non-REM, events/h	31.4 [21.5–48.6]	7.3 [3.1–16.9]	<0.001
AHI, REM, events/h	48.8 [29.1–66.1]	25.6 [6.1–45.3]	<0.001
AHI, supine*, events/h	52.9 [33.0–72.4]	16.7 [8.7–38.4]	<0.001
Arousal index, events/h	43.5 [35.7–54.6]	9.9 [4.4–16.5]	<0.001
Minimum oxygen saturation, %	80 [76–84]	87 [81–90]	<0.001
TST, min	356 ± 65	363 ± 65	0.43
REM sleep time, % TST	16.8 ± 6.2	17.6 ± 6.9	0.45
Supine sleep time*, % TST	40.1 [26.3–69.3]	37.0 [21.2–80.6]	0.83

*Definition of abbreviations:* AHI = apnea-hypopnea index; BMI = body mass index; F = female; M = male; REM = rapid eye movement; TST = total sleep time.

On average, participants were typical patients with obstructive sleep apnea, middle aged, predominantly obese with moderate to severe obstructive sleep apnea. Continuous variables are presented as mean ± standard deviation or median [25th–75th centile].

\*Data available in  $n = 62$ .

mean [95% CI], 34 [30–38] vs. 33 [29–37] events/h;  $P = 0.5$ ). In addition, none of the clinical covariates above were significantly associated with oral appliance efficacy ( $\Delta$ AHI) when considered individually (linear regression) or in combination (multivariable regression, total  $R^2 = 0.08$ ).

Adjusting for scoring type had no impact (<1% change in model coefficient) on the association between endotypic subgroup and oral appliance efficacy and was not associated with efficacy ( $P = 0.9$ ). Altering the cutoff of true responder from greater than 50% to greater than 70% reduction in AHI yielded similar results, with group differences in efficacy of 22% (cross-validated,  $P = 0.0006$ ) becoming 20% ( $P = 0.0011$ ). Positive and negative predictive values became 65% (30:16) and 72% (34:13), respectively ( $P = 0.0004$ , Fisher's exact test); accuracy was 69%.

## Discussion

The present study is, to our knowledge, the first to demonstrate that the endotypic traits causing OSA, estimated from routine diagnostic polysomnography, have utility in defining a subgroup of patients who are more likely to respond to oral appliance therapy. Our study shows that a greater treatment efficacy is associated with

favorable nonpharyngeal traits (lower loop gain, higher arousal threshold, and lower ventilatory response to arousal), moderate collapsibility (not mild or severe, U-shaped), and weaker pharyngeal muscle compensation. Using measurements of the traits alone, predicted responders, on average, exhibited half the residual AHI (8 events/h, approximately one-fourth of baseline) compared with predicted nonresponders (16 events/h, approximately half of baseline), despite similar baseline AHI. Moreover, 78% of patients in the predicted responders group exhibited at least a 70% reduction in AHI. These results provide a basis for future identification of patients who could potentially be prioritized for personalized therapy with oral appliances based on the OSA endotypic traits estimated from diagnostic polysomnography.

### Consistency with Available Literature and Novel Physiological Insights

Our findings confirm previous work in that OSA endotypes can be estimated from routine diagnostic polysomnography and provide insight into therapeutic outcomes (26–29, 42). In concordance with physiological principles and our recent small, detailed physiology study, we confirmed the finding in a larger dataset that lower loop gain contributes significantly to

greater oral appliance efficacy (20). We emphasize, however, that in the present study, unlike our prior work, we did not find a strong bivariate relationship between loop gain and oral appliance efficacy. However, the requirement for multiple interacting endotypic predictors to be considered in combination is consistent with our previous study (37).

Previous studies have also found that severe collapsibility is associated with reduced oral appliance efficacy (15, 20, 26, 43–45). Oral appliance therapy typically reduces critical collapsing pressure by 3–5 cm H<sub>2</sub>O (26, 46–48) and therefore is unlikely to resolve OSA in patients with severe collapsibility at baseline. Greater collapsibility (lower  $V_{PASSIVE}$ ), higher BMI, nonpositional OSA (a marker of greater collapsibility), and higher CPAP requirement have each been shown to predict poor response to oral appliance therapy (15, 20, 26, 43–45), although these are not robust predictors individually. The present study found a U-shaped relationship between collapsibility and response to oral appliances. As expected, more severe collapsibility predicted reduced responses to oral appliance therapy. Milder collapsibility, unexpectedly, also predicted a reduced response to treatment. We consider that these individuals, rather than being easier to treat, have primarily nonpharyngeal mechanisms underpinning their sleep apnea. We emphasize that although we initially considered that the U-shaped relationship could be spurious, we noted that a large proportion of patients were nonresponders with mild collapsibility (and high loop gain or low arousal threshold; see Figure 1), such that this unexpected U-shaped effect at the mild end of the spectrum was unlikely to be attributable to low sample size.

We also found that elevated loop gain, lower arousal threshold, and greater ventilatory response to arousal also contributed to a reduced oral appliance efficacy. These nonpharyngeal factors contributing to breathing instability are unlikely to be corrected by mandibular advancement (20). Indeed, it was precisely this subgroup of patients that responded preferentially to supplemental oxygen in our recent study (37). Furthermore, we found that reduced pharyngeal compensation was associated with a higher oral appliance efficacy. According to physiological principles, a stronger pharyngeal dilator

**Table 2.** Traits associated with oral appliance efficacy: multiple regression

Variable	$\beta$ -Value	SEM	$\beta$ -SD	P Value	Interpretation
Constant	47.1	5.1	1.4	<0.0001	
Pharyngeal traits					
$V_{PASSIVE}$	-0.771	0.325	-0.47	0.02	Not severe and not mild collapsibility → success
$V_{PASSIVE}^2$	-0.0293	0.0095	-1.1	0.003	
Compensation <sup>2</sup>	0.0215	0.0067	0.94	0.002	Higher compensation → failure
Compensation × arousal threshold	0.0486	0.0107	1.4	<0.0001	Particularly when arousal threshold is low or response to arousal is high
Compensation × response to arousal	-0.0171	0.0064	-0.41	0.009	
Nonpharyngeal traits					
Loop gain	-112	41	-0.37	0.008	Higher loop gain → failure
Loop gain × compensation	-6.95	2.23	-0.52	0.003	Particularly when compensation or response to arousal is high
Loop gain × response to arousal	-8.79	2.52	-0.70	0.0008	
Arousal threshold	0.420	0.233	0.32	0.076	Lower arousal threshold → failure
Arousal threshold <sup>2</sup>	0.0151	0.0055	0.51	0.007	
Response to arousal	-0.514	0.193	-0.34	0.009	Higher response to arousal → failure
Response to arousal × $V_{PASSIVE}$	-0.0212	0.0115	-0.50	0.069	

*Definition of abbreviations:* SD = standard deviation; SEM = standard error of the mean;  $V_{PASSIVE}$  = passive collapsibility.

Oral appliance efficacy is defined as the percentage reduction in apnea-hypopnea index with treatment compared with baseline (transformed; see METHODS). The table describes final results (12 of 20 terms) after backward stepwise elimination ( $P$ -to-remove = 0.157), which began with five traits, their squares, and all interaction terms. Note that the significance levels were  $P < 0.05$  in 10 of 20 terms and  $P < 0.01$  in 9 of 20 terms. Traits were mean subtracted before terms were generated and applied to the model (see below).  $\beta$ -SD describes the number of SDs of change in treatment efficacy per SD increase in each term (1.3 SD is needed to move a typical nonresponder to a typical responder). Mean values of the endotypic traits before mean subtraction:  $V_{PASSIVE} = 79.0 \pm 20.8$ ; loop gain =  $0.43 \pm 0.11$ ; compensation =  $-9.5 \pm 27.0\%$ ; arousal threshold =  $141.8 \pm 26.0\%$ ; response to arousal =  $36.3 \pm 22.6\%$ . A regression model cutoff of 60% (predicted reduction in apnea-hypopnea index, untransformed) was used to define predicted responders and predicted nonresponders (maximized sensitivity plus specificity). Overall  $R^2 = 0.30$ ; adjusted  $R^2 = 0.19$ ;  $P = 0.003$ .

muscle compensation will act to mask a more severely collapsible airway. Therefore, attempts to improve collapsibility via oral appliance therapy will be partially counteracted by attenuation of the pharyngeal dilator muscle activity as airway obstruction is mitigated. Thus, our finding that poor compensation is associated with a higher oral appliance efficacy is consistent with physiological principles.

Our study shows no significant predictive value of routine clinical variables (such as baseline AHI, BMI, neck circumference, age, or sex), whether individually or in combination with OSA endotypic traits. These data confirm the difficulty in using routine clinical variables to predict outcomes of oral appliance therapy (44). Our study also supports the concept that baseline severity of OSA (AHI) is not a useful predictor of response to therapy.

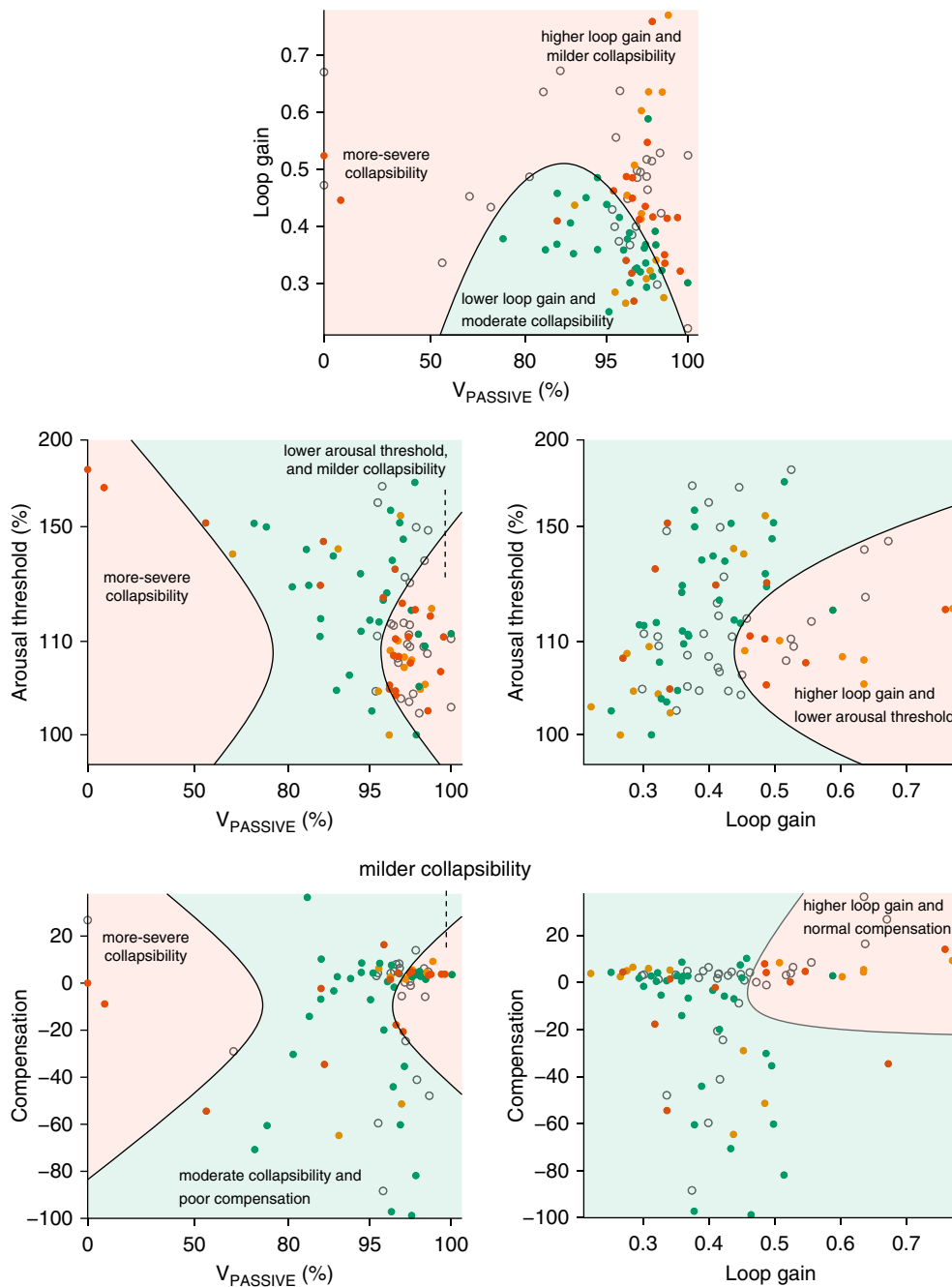
### Clinical Implications

The present study sought to advance knowledge for future precision sleep medicine. In the context of heterogeneous oral appliance efficacy in unselected patients, a major goal of our work was to enable the identification of a subgroup of

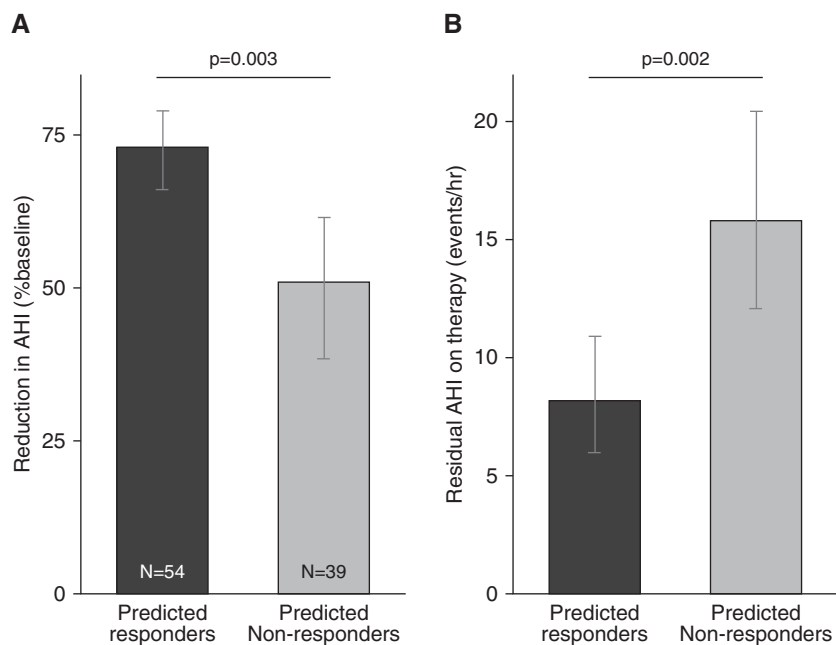
patients with (moderate to severe) OSA with a superior treatment efficacy compared with that of other patients with OSA whose average efficacy is more modest. We used an automated clinical tool to estimate the key endotypic traits causing OSA from routine diagnostic polysomnography and combined these traits to define two endotype-based subgroups of patients. On average, the predicted responders subgroup exhibited good treatment efficacy (~75% reduction in AHI), which, when coupled with the reported high adherence to therapy (12), appears sufficient to justify offering oral appliances as a first-line therapy in select patients with (moderate to severe) OSA (specifically those who have a preference for this intervention). Although our results were based on unseen holdout data (leave one patient out cross-validation), these findings require replication in a larger prospective study before this method can be adopted for routine clinical use. Notably, even the predicted nonresponders subgroup had, on average, 50% reduction in AHI (residual AHI, ~16 events/h). Although this level of efficacy seems unlikely to show superior health benefit compared with CPAP, the considerable improvement in nonresponders is likely

to confer benefit over no therapy, justifying prescription of oral appliance therapy as a second-line option even in this subgroup (e.g., in CPAP-intolerant patients).

Our automated method has several advantages as a clinical tool for predicting outcomes. It is based on OSA endotypes (e.g., rather than demographic factors) and therefore has a close connection with the underlying mechanisms. The approach used in the present study is inexpensive, is not dependent on specialized equipment or physiological interventions, and can produce results rapidly. The data used for analysis in the present study were also clinical in nature, supporting clinical generalizability and translatability of physiological endotypes. Data were extracted from standard clinical sleep studies (rather than research studies) acquired using a commercially available sleep recording system (Profusion PSG; Compumedics Ltd.). Because the analysis was retrospective, there was no opportunity to pay extraordinary attention to nasal pressure quality beyond AASM standards (unfiltered nasal pressure). Other challenges for widespread implementation of our tool in clinical practice include incorporation of endotyping methods into



**Figure 1.** Key aspects of the five-trait multivariable model (see Table 2) illustrating how combinations of traits may influence oral appliance efficacy. Each plot depicts a two-trait cross-section of the full model drawn at the mean values of the remaining three traits. Dots represent true response observations of individual patients: red for nonresponders (<50% reduction in apnea–hypopnea index [AHI] with treatment), orange and green for responders (50–70% reduction in AHI and >70% reduction in AHI, respectively). Background regions represent predicted response subgroups (light green for predicted responders and light red for predicted nonresponders). Top and left: A U-shaped relationship between collapsibility ( $V_{PASSIVE}$ ) and efficacy is evident. For example, in the top figure, the light green shading indicating predicted responders is only seen in a midrange of moderate collapsibility and at lower loop gain. Note that nonresponders with high  $V_{PASSIVE}$  (mild collapsibility) tend to have high loop gain, low arousal threshold, and higher compensation (see dense regions of red dots). Top and right: A higher loop gain is associated with reduced treatment efficacy, particularly in milder collapsibility (high  $V_{PASSIVE}$ ), but also in the presence of a lower arousal threshold and higher compensation. Open gray circles on each plot represent individual patients whose values for the three remaining traits were too far from the mean to be fairly represented in the simplified two-trait view (i.e., two-trait prediction differed from the full model prediction).



**Figure 2.** (A–B) On the basis of combined endotype traits, predicted responders (black), compared with predicted nonresponders (gray), exhibited a greater oral appliance efficacy, indicated by a greater reduction in apnea–hypopnea index (AHI) (untransformed) from baseline (A) and a lower residual AHI on treatment (B). Error bars illustrate 95% confidence in the mean. Results are based on cross-validated analysis, whereby the endotypic subgroup allocation for each individual patient was based on a modified regression model using data from all other patients. Thus, group differences are not guaranteed by definition, based on the regression model results in Table 2.

commercial systems and requirement for rescoring of arousal timing (not performed clinically). Neither obstacle is insurmountable.

### Methodological Considerations

There are several limitations of our work. First, the endotypic traits described are not based on gold standard measurements but rather estimated from a nasal pressure surrogate of ventilatory airflow and a mathematically estimated ventilatory drive signal. However, it would be a highly challenging endeavor to perform gold standard measurements of physiology (via CPAP decreases or esophageal catheterization) in such large numbers of patients undergoing a specific treatment regimen. Thus, a strength and novelty of our work is obtaining such measures in a sample size of more than 90 oral appliance–treated patients with OSA. Second, we studied patients with baseline AHI greater than 20 events per hour (average AHI, 30 events/h), and thus our results are relevant to those with similar OSA severity and may not apply to many patients with a milder condition who seek oral appliance therapy.

Indeed, a major goal of our work was to identify patients who might exhibit favorable outcomes of oral appliance therapy despite having more severe OSA. Further investigation is needed to identify those with milder OSA (AHI, <20 events/h) who might be suitable for oral appliance therapy regardless of their endotypic characteristics.

Third, the incomplete data nature of retrospective studies precluded full assessment of the impact of some other relevant variables. For example, we did not have systematic data at baseline and on therapy for supine sleep duration. Controlling for body position would likely reduce a source of undesirable variability. Nonetheless, the influence of endotypes on efficacy is unlikely to be confounded by differences in body position at baseline and on therapy (i.e., no plausible mechanism by which treatment-related changes in supine sleep duration could influence endotypic traits of OSA and thus oral appliance efficacy). We also did not have systematic measures of daytime sleepiness (e.g., Epworth Sleepiness Scale) and thus could not assess the role of daytime sleepiness in

the context of the endotypic traits. However, we found a relationship between lower arousal threshold and reduced oral appliance efficacy, suggesting that a higher propensity for arousal from sleep might render oral appliance treatment less efficacious. Further investigation along these lines is warranted.

Fourth, we used the percentage reduction in AHI as a continuous outcome measure and a single cutoff (i.e., 50% reduction in AHI) to define the true response subgroups. However, changing the cutoff (e.g., to a 60% or 70% reduction in AHI) did not alter the findings substantially. We also note that proportions of patients defined as complete responders ( $\geq 50\%$  reduction in AHI and residual AHI <10 events/h), partial responders ( $\geq 50\%$  reduction in AHI and residual AHI  $\geq 10$  events/h), and nonresponders (<50% reduction in AHI) were 30:12:12 in predicted responders and 10:11:18 in predicted nonresponders ( $P = 0.01$ , Fisher's exact test), respectively. Fifth, although subgroup differences in efficacy appear clinically relevant, the overall model accuracy is modest; as noted above, predicted nonresponders show an average of 50% reduction in AHI. Thus, currently, we are unable to identify a subgroup of patients with OSA who may exhibit negligible benefit. Incorporation of additional information on site/structure of pharyngeal obstruction (e.g., through coupling of our approach with other polysomnographic methods such as airflow shape [49]) may further improve the model precision and predictive performance.

Finally, we caution that the noninvasive measurements of endotypic traits were validated against gold standard values in relatively small samples ( $n = 28–41$ ) and would benefit from further refinement and validation studies, including efforts to improve reliability (e.g., incorporating respiratory inductance plethysmography to handle mouth leak) and make the measurements independent of manual scoring (e.g., quantitative electroencephalographic analysis [50]).

### Conclusions

In the largest study to date, we elucidated the relationships between the pathophysiological traits causing OSA and oral appliance treatment efficacy. Although bivariate linear associations between efficacy and endotypes were not evident, our



multivariable analyses showed that greater oral appliance efficacy is associated with favorable nonpharyngeal endotypic traits of OSA at baseline (including lower loop gain, higher arousal threshold, and lower ventilatory response to arousal). Greater efficacy was also associated with moderate (nonmild or nonsevere) collapsibility and weaker dilator muscle compensation. Combining endotypic traits

identified a predicted responders subgroup of patients who exhibited good treatment efficacy and could potentially be targeted judiciously for early oral appliance intervention compared with a predicted nonresponders subgroup. Further studies are needed to prospectively validate our predictive model for clinical use. We anticipate that identifying endotypes from routine diagnostic polysomnography will

allow patient selection for oral appliance therapy in OSA. ■

**Author disclosures** are available with the text of this article at <http://www.atsjournals.org>.

**Acknowledgment:** The authors are grateful to the staff and patients who were involved in the parent studies from which we collated data for the present study.

## References

- Ramar K, Dort LC, Katz SG, Lettieri CJ, Harrod CG, Thomas SM, *et al*. Clinical practice guideline for the treatment of obstructive sleep apnea and snoring with oral appliance therapy: an update for 2015. *J Clin Sleep Med* 2015;11:773–827.
- Petri N, Svanholt P, Solow B, Wildschjødzt G, Winkel P. Mandibular advancement appliance for obstructive sleep apnoea: results of a randomised placebo controlled trial using parallel group design. *J Sleep Res* 2008;17:221–229.
- Pitsis AJ, Darendeliler MA, Gotsopoulos H, Petocz P, Cistulli PA. Effect of vertical dimension on efficacy of oral appliance therapy in obstructive sleep apnea. *Am J Respir Crit Care Med* 2002;166:860–864.
- Blanco J, Zamarrón C, Abeleira Pazos MT, Lamela C, Suarez Quintanilla D. Prospective evaluation of an oral appliance in the treatment of obstructive sleep apnea syndrome. *Sleep Breath* 2005;9:20–25.
- Bloch KE, Iseli A, Zhang JN, Xie X, Kaplan V, Stoeckli PW, *et al*. A randomized, controlled crossover trial of two oral appliances for sleep apnea treatment. *Am J Respir Crit Care Med* 2000;162:246–251.
- Ferguson KA, Cartwright R, Rogers R, Schmidt-Nowara W. Oral appliances for snoring and obstructive sleep apnea: a review. *Sleep* 2006;29:244–262.
- Gotsopoulos H, Chen C, Qian J, Cistulli PA. Oral appliance therapy improves symptoms in obstructive sleep apnea: a randomized, controlled trial. *Am J Respir Crit Care Med* 2002;166:743–748.
- Gotsopoulos H, Kelly JJ, Cistulli PA. Oral appliance therapy reduces blood pressure in obstructive sleep apnea: a randomized, controlled trial. *Sleep* 2004;27:934–941.
- Andrén A, Hedberg P, Walker-Engström ML, Wahlén P, Tegelberg A. Effects of treatment with oral appliance on 24-h blood pressure in patients with obstructive sleep apnea and hypertension: a randomized clinical trial. *Sleep Breath* 2013;17:705–712.
- Gauthier L, Laberge L, Beaudry M, Laforge M, Rompré PH, Lavigne GJ. Mandibular advancement appliances remain effective in lowering respiratory disturbance index for 2.5–4.5 years. *Sleep Med* 2011;12:844–849.
- Naismith SL, Winter VR, Hickie IB, Cistulli PA. Effect of oral appliance therapy on neurobehavioral functioning in obstructive sleep apnea: a randomized controlled trial. *J Clin Sleep Med* 2005;1:374–380.
- Phillips CL, Grunstein RR, Darendeliler MA, Mihailidou AS, Srinivasan VK, Yee BJ, *et al*. Health outcomes of continuous positive airway pressure versus oral appliance treatment for obstructive sleep apnea: a randomized controlled trial. *Am J Respir Crit Care Med* 2013;187:879–887.
- Sutherland K, Ngiam J, Cistulli PA. Performance of remotely controlled mandibular protrusion sleep studies for prediction of oral appliance treatment response. *J Clin Sleep Med* 2017;13:411–417.
- Remmers J, Charkhandeh S, Grosse J, Topor Z, Brant R, Santosham P, *et al*. Remotely controlled mandibular protrusion during sleep predicts therapeutic success with oral appliances in patients with obstructive sleep apnea. *Sleep (Basel)* 2013;36:1517–1525, 1525A.
- Okuno K, Pliska BT, Hamoda M, Lowe AA, Almeida FR. Prediction of oral appliance treatment outcomes in obstructive sleep apnea: a systematic review. *Sleep Med Rev* 2016;30:25–33.
- Eckert DJ. Phenotypic approaches to obstructive sleep apnoea – new pathways for targeted therapy. *Sleep Med Rev* 2018;37:45–59.
- Eckert DJ, White DP, Jordan AS, Malhotra A, Wellman A. Defining phenotypic causes of obstructive sleep apnea: identification of novel therapeutic targets. *Am J Respir Crit Care Med* 2013;188:996–1004.
- Wellman A, Eckert DJ, Jordan AS, Edwards BA, Passaglia CL, Jackson AC, *et al*. A method for measuring and modeling the physiological traits causing obstructive sleep apnea. *J Appl Physiol (1985)* 2011;110:1627–1637.
- Wellman A, Edwards BA, Sands SA, Owens RL, Nemati S, Butler J, *et al*. A simplified method for determining phenotypic traits in patients with obstructive sleep apnea. *J Appl Physiol (1985)* 2013;114:911–922.
- Edwards BA, Andara C, Landry S, Sands SA, Joosten SA, Owens RL, *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;194:1413–1422.
- Eckert DJ, Owens RL, Kehlmann GB, Wellman A, Rahangdale S, Yim-Yeh S, *et al*. Eszopiclone increases the respiratory arousal threshold and lowers the apnoea/hypopnoea index in obstructive sleep apnoea patients with a low arousal threshold. *Clin Sci (Lond)* 2011;120:505–514.
- Edwards BA, Sands SA, Owens RL, Eckert DJ, Landry S, White DP, *et al*. The combination of supplemental oxygen and a hypnotic markedly improves obstructive sleep apnea in patients with a mild to moderate upper airway collapsibility. *Sleep (Basel)* 2016;39:1973–1983.
- Gleeson K, Zwillich CW, White DP. The influence of increasing ventilatory effort on arousal from sleep. *Am Rev Respir Dis* 1990;142:295–300.
- Younes M. Role of arousals in the pathogenesis of obstructive sleep apnea. *Am J Respir Crit Care Med* 2004;169:623–633.
- Ratnavadivel R, Stadler D, Windler S, Bradley J, Paul D, McEvoy RD, *et al*. Upper airway function and arousability to ventilatory challenge in slow wave versus stage 2 sleep in obstructive sleep apnoea. *Thorax* 2010;65:107–112.
- Marques M, Genta P, Sands SA, Taranto Montemurro L, Azarbarzin A, De Melo C, *et al*. Characterizing site and severity of upper airway collapse to guide patient selection for oral appliance therapy for obstructive sleep apnea [abstract]. *Am J Respir Crit Care Med* 2017;195:A2584.
- Terrill PI, Edwards BA, Nemati S, Butler JP, Owens RL, Eckert DJ, *et al*. Quantifying the ventilatory control contribution to sleep apnoea using polysomnography. *Eur Respir J* 2015;45:408–418.
- Sands SA, Edwards BA, Terrill PI, Taranto-Montemurro L, Azarbarzin A, Marques M, *et al*. Phenotyping pharyngeal pathophysiology using polysomnography in patients with obstructive sleep apnea. *Am J Respir Crit Care Med* 2018;197:1187–1197.
- Sands SA, Terrill PI, Edwards BA, Taranto Montemurro L, Azarbarzin A, Marques M, *et al*. Quantifying the arousal threshold using polysomnography in obstructive sleep apnea. *Sleep (Basel)* 2018;41:zsx183.
- Sutherland K, Chan ASL, Ngiam J, Dalci O, Darendeliler MA, Cistulli PA. Awake multimodal phenotyping for prediction of oral appliance treatment outcome. *J Clin Sleep Med* 2018;14:1879–1887.
- Lowth A, Juge L, Knapman F, Burke P, Brown E, Butler J, *et al*. Dynamic MRI tongue deformation patterns during mandibular advancement and associations with craniofacial anatomy in OSA [abstract]. *J Sleep Res* 2018;27(Suppl 2):e169\_12766.

- 32 White LH, Lyons OD, Yadollahi A, Ryan CM, Bradley TD. Night-to-night variability in obstructive sleep apnea severity: relationship to overnight rostral fluid shift. *J Clin Sleep Med* 2015;11:149–156.
- 33 Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, *et al.* Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. *J Clin Sleep Med* 2012;8:597–619.
- 34 Iber C, Ancoli-Israel S, Chessson A, Quan SF; American Academy of Sleep Medicine. The American Academy of Sleep Medicine manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Westchester, IL: American Academy of Sleep Medicine; 2007.
- 35 Akobeng AK. Understanding type I and type II errors, statistical power and sample size. *Acta Paediatr* 2016;105:605–609.
- 36 Zhang S, Paul J, Nantha-Aree M, Buckley N, Shahzad U, Cheng J, *et al.* Empirical comparison of four baseline covariate adjustment methods in analysis of continuous outcomes in randomized controlled trials. *Clin Epidemiol* 2014;6:227–235.
- 37 Sands SA, Edwards BA, Terrill PI, Butler JP, Owens RL, Taranto-Montemurro L, *et al.* Identifying obstructive sleep apnoea patients responsive to supplemental oxygen therapy. *Eur Respir J* 2018;52:1800674.
- 38 Dunkler D, Plischke M, Leffondré K, Heinze G. Augmented backward elimination: a pragmatic and purposeful way to develop statistical models. *PLoS One* 2014;9:e113677.
- 39 In Lee K, Koval JJ. Determination of the best significance level in forward stepwise logistic regression. *Commun Stat Simul Comput* 1997;26:559–575.
- 40 Hosmer DW, Lemeshow S. Applied logistic regression. Wiley series in probability and statistics. Vol. 10. New York: Wiley; 1989. pp. 1162–1163.
- 41 Heinze G, Dunkler D. Five myths about variable selection. *Transpl Int* 2017;30:6–10.
- 42 Azarbarzin A, Sands SA, Taranto-Montemurro L, Oliveira Marques MD, Genta PR, Edwards BA, *et al.* Estimation of pharyngeal collapsibility during sleep by peak inspiratory airflow. *Sleep (Basel)* 2017;40:zsw005.
- 43 Sutherland K, Phillips CL, Davies A, Srinivasan VK, Dalci O, Yee BJ, *et al.* CPAP pressure for prediction of oral appliance treatment response in obstructive sleep apnea. *J Clin Sleep Med* 2014;10:943–949.
- 44 Sutherland K, Takaya H, Qian J, Petocz P, Ng AT, Cistulli PA. Oral appliance treatment response and polysomnographic phenotypes of obstructive sleep apnea. *J Clin Sleep Med* 2015;11:861–868.
- 45 Ng AT, Qian J, Cistulli PA. Oropharyngeal collapse predicts treatment response with oral appliance therapy in obstructive sleep apnea. *Sleep* 2006;29:666–671.
- 46 Bamagoos AA, Cistulli P, Sutherland K, Ngiam J, Burke P, Bilston L, *et al.* Dose-dependent effects of mandibular advancement on upper airway collapsibility and muscle activity in obstructive sleep apnea. *Sleep* 2019;42:zsz049.
- 47 Ng AT, Gotsopoulos H, Qian J, Cistulli PA. Effect of oral appliance therapy on upper airway collapsibility in obstructive sleep apnea. *Am J Respir Crit Care Med* 2003;168:238–241.
- 48 Kato J, Isono S, Tanaka A, Watanabe T, Araki D, Tanzawa H, *et al.* Dose-dependent effects of mandibular advancement on pharyngeal mechanics and nocturnal oxygenation in patients with sleep-disordered breathing. *Chest* 2000;117:1065–1072.
- 49 Genta PR, Sands SA, Butler JP, Loring SH, Katz ES, Demko BG, *et al.* Airflow shape is associated with the pharyngeal structure causing OSA. *Chest* 2017;152:537–546.
- 50 Younes M, Ostrowski M, Soiferman M, Younes H, Younes M, Raneri J, *et al.* Odds ratio product of sleep EEG as a continuous measure of sleep state. *Sleep (Basel)* 2015;38:641–654.