**Simplified EEG Analysis for the Continuous Assessment of Wake-Sleep Balance**

Ali Azarbarzin, Laura Gell, Daniel Vena. Raichel Alex, Lauren Hess1, Nicole Calianese1, Sara Op de Beeck, Erik Smales, Andrey Zinchuk, John X, XYZ, Andrew Wellman1, Scott A. Sands1\*

1Division of Sleep and Circadian Disorders, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, USA.

\*Corresponding author: Scott Sands, PhD

Division of Sleep Medicine, Brigham and Women’s Hospital,

221 Longwood Ave, Boston 02115, MA, USA.

email: sasands@bwh.harvard.edu | T: +1 8579280341 | F: +1 6177327337

**Author Contributions.** Conception and Design: SS. Parent-Study Data Collection: XX. Sleep and arousal scoring: XX. Model development and statistics: SS. All authors interpreted results, edited the manuscript for important intellectual content, and approved the final draft.

# Abstract

**Study Objectives:** Currently there is no simple, accurate, and publicly-available method to quantify the intensity of wake/sleep/arousals using electroencephalography (EEG) on a continuous scale for the translatable pathophysiological phenotyping of sleep disordered breathing.

**Methods:**  , model wake versus sleep as a simple continuous function of EEG power (log-scaled, 6-s median-filtered) in beta, alpha, theta, and delta frequency bands using logistic regression. Powers were normalized per recording to obviate errors due to larger/smaller *global EEG magnitude*, modelled using overnight centiles (5th, 25th, 50th, 75th, 95th) of the four powers. Specialized manual scoring identified periods of sleep onset and arousals in wake and sleep with careful focus on start and end times. Sleep studies from XXX participants were examined retrospectively; odd and even numbers were used for development and validation respectively. Overall, X million samples (every 3 s) overall were used. The EEG power model (linear function of EEG powers) provided a continuous *wake-sleep intensity* *score* (−5, 0, +5 represent 1%, 50%, 99% probability of wake). A second model detected arousals based on local increases in *wake-sleep intensity* (arousal intensity). Arousal intensity was also compared against changes in respiratory drive (calibrated intraesophageal diaphragm EMG) and heart rate.

**Results:** A model using beta, alpha, theta, and delta powers, plus theta2, delta2 and theta-delta interaction explained 80% (pseudo-R2) of the variability in manually scored wake versus sleep (excluding arousal), and matched scores from the lookup table method (R2=0.99). In validation, overall accuracy was 93%. WSI fell progressively from wake (median: 5) to stages N1 (−3), REM (−3), N2 (−4.5), and N3 (−6); a typical arousal had an intermediate score of 0 (i.e. half way from sleep to wake). Arousal scoring based on WSI changes had an accuracy of 84%. More intense arousals exhibited progressively-greater increases in ventilatory drive and heart rate (data). Models are freely available from the authors.

**Conclusions:** Simple linear combinations of powers can provide an accurate continuous measure of wake versus sleep.

# Keywords

Wake sleep continuum | power spectrum | phenotype | automated scoring

# **Introduction**

Oral appliances, intraoral devices worn during sleep to protrude the mandible, are increasingly utilized as a treatment alternative for obstructive sleep apnea (OSA).[1](#_ENREF_1) The literature indicates that oral appliance therapy reduces OSA severity (indicated by the apnea-hypopnea index, AHI) by an average of approximately 50-70%, with variable efficacy across patients.[2-5](#_ENREF_2) Although not as efficacious as continuous positive airway pressure (CPAP) at ameliorating OSA, they have a proven positive impact on sleepiness, blood pressure and quality of life.[6-10](#_ENREF_6) Additionally, studies suggest that treatment outcomes of oral appliances and CPAP are similar,[11](#_ENREF_11) reflecting superior adherence to oral appliance therapy. Without experimental testing in each patient,[12](#_ENREF_12),[13](#_ENREF_13) there is currently no clinically-applicable means to predict the likelihood of oral appliance therapy success before treatment prescription.[14](#_ENREF_14)

The variability in oral appliance efficacy across OSA patients may be attributed to the extent to which OSA endotypic traits (pharyngeal: *collapsibility* andmuscle *compensation*; non-pharyngeal: *loop gain*, *arousal threshold* and ventilatory *response to arousal*) contribute to the pathogenesis of the condition.[15-24](#_ENREF_15) 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,[25](#_ENREF_25) and higher loop gain (i.e. ventilatory control instability) that cannot be resolved with anatomical interventions.[19](#_ENREF_19) Notably, these detailed studies were performed in specialized laboratories using invasive instrumentation, which are 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.[26-28](#_ENREF_26) 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 current study, we applied this method to diagnostic polysomnography of patients treated with oral appliance therapy. We aimed to: 1) determine whether oral appliance efficacy is associated with favorable non-pharyngeal 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 OSA patients (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 data from previous oral appliance research studies[11](#_ENREF_11" \o "Phillips, 2013 #200),[29](#_ENREF_29" \o "Sutherland, 2018 #743),[30](#_ENREF_30" \o "Lowth, 2018 #812). Patients were included in our analysis if they had a baseline polysomnography-derived AHI ≥20 events/hr (pre-specified), selected to minimize the influence of night-to-night variability (noise) on the percent reduction in AHI with treatment (efficacy).[31](#_ENREF_31" \o "White, 2015 #619) For example, a 75% reduction in AHI from 20 to 5 events/hr was considered more reliable than the same from 10 to 2.5 events hr (the latter is within the expected night-to-night variability, SD ~10 events/hr). All polysomnographic data from parent studies were from a single sleep clinic and included newly-diagnosed OSA patients (baseline AHI > 10 events/hr for original studies).Data from 14 patients were taken from the “MASPAP” study[11](#_ENREF_11) (ACTRN 12607000289415, n=108 (80% males), mean±SD age=50±11 years, body mass index [BMI]=30±6 kg/m2 and AHI=26±12 events/hr), which was a 3-center randomized cross-over trial comparing health outcomes of CPAP therapy versus oral appliances therapy in patients who were recommended both treatments. Data from 64 patients were taken from the “PhenoMAS” study[29](#_ENREF_29) (ACTRN12611000409976, n=142 (59% males), mean±SD age=56±11 years, BMI=30±5 kg/m2 and AHI= 29±18 events/hr), which was a single-center observational study designed to examine awake-based predictors of oral appliance efficacy in patients to whom MAS therapy were recommended. Data from 16 patients were taken from the ongoing “OSAMAS” study[30](#_ENREF_30) (at assessment: n=40 (72% males), mean±SD age=43±11 years, BMI=30±5 kg/m2 and AHI=27.7±16.9 events/hr), which is a dual-center observational study using MRI-based genioglossus dynamics to explain heterogeneity in oral appliance efficacy. Key eseOral appliance efficacy in-laboratory the

Since all data were de-identified, the current analysis was deemed to be exempt from consent by the Human Research Ethics Committee at North Sydney Local Health District, NSW, Australia and the Partners Institutional Review Board, Boston, MA, USA.

## 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 were scored according to the AASM criteria 2012[11](#_ENREF_11),[32](#_ENREF_32) (30% reduction in airflow with either 3% oxygen desaturation or cortical arousal) or AASM criteria 2007[29](#_ENREF_29),[30](#_ENREF_30),[33](#_ENREF_33) (30% reduction in airflow with 4% oxygen desaturation). Models presented were not adjusted for scoring type as no effect was evident (see Results).

All patients were treated with a common custom-made oral appliance (SomnoDent, SomnoMed Ltd., Australia), implemented by an experienced 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 (approximately 4-8 weeks), 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 efficacywas described by the percent reduction in AHI with treatment relative to baseline (primary outcome variable, continuous). This measure was selected (over absolute reduction or treatment AHI) to maximize statistical power (typically largest mean/SD ratio,[34](#_ENREF_34) correlates least with baseline AHI[35](#_ENREF_35)).

## Endotypic Trait Analysis

***Raw data.***We identified 94 patients who met our pre-specified eligibility criteria for analysis (N=50 excluded for AHI <20 events/hr). To optimize automated analyses, a thorough manual check of cortical arousal onset and end times was performed for baseline polysomnography by 3 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 due to poor nasal pressure signal (automated quality control algorithm, verified visually) leaving 93 for analysis. Analysis of the traits was restricted to non-REM sleep for consistency with previous validation studies using our methods[26-28](#_ENREF_26) 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.[26](#_ENREF_26),[28](#_ENREF_28),[36](#_ENREF_36),[37](#_ENREF_37) In brief, each diagnostic polysomnogram was automatically segmented into 7-min 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 × rate), calibrated such that the mean eupneic ventilation for the window being analyzed = 100%.[28](#_ENREF_28) *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) based on previous values of ventilation. This chemoreflex model is physiologically constrained such that 3 key parameters are identified (gain, time-constant and delay[26](#_ENREF_26)). These parameters were used to calculate the *loop gain* (LGn, 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 apnoea).[26](#_ENREF_26),[36](#_ENREF_36) 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 EEG arousal on any breath as a covariate. The *arousal threshold* was calculated as the mean ventilatory drive on the breath immediately preceding scored arousals.[28](#_ENREF_28)

To calculate collapsibility and muscle compensation, an overnight endotype plot[37](#_ENREF_37) was generated, whereby all breath-by-breath values of ventilation and ventilatory drive for the whole night (except breaths in wake, 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* (VPASSIVE). A lower VPASSIVE reflects a greater collapsibility. The median value of ventilation at maximal ventilatory drive (at arousal threshold) was taken as a measure of active collapsibility (VACTIVE). The difference between VACTIVE and VPASSIVE was used as a measure of pharyngeal muscle *compensation*. Analysis was fully-automated using in-house software (Matlab, Mathworks, Natick MA, USA; interested users are directed to contact the authors) and visually verified.

## Model Development and Statistical Analyses

The goals of the statistical analyses were twofold. 1) We sought to describe the associations between oral appliance efficacy and OSA endotypic traits (in combination) at baseline, to provide physiological insight into downstream mechanistic causes of variability in oral appliance efficacy. A multivariable regression model approach was employed (details below).[36](#_ENREF_36" \o "Sands, 2018 #3899) Second, we sought to use the same endotype-based regression model as the basis of a prediction model; we sought to examine the extent to which these endotypes could be employed to identify an endotype-based subgroup of patients who would exhibit greater oral appliance efficacy (“predicted responders”) compared with other patients (“predicted non-responders”).

***Statistical Power.*** Ninety patients were expected to provide 86% power to detect significant independent associations with R2>0.1 (alpha=0.05), and 92% power to detect differences in efficacy between endotypic subgroups of at least 20±40%.

***Data transformation.*** Several variables were not normally distributed (Shapiro-Wilk) and were transformed accordingly: VPASSIVE and arousal threshold were square-root-transformed via the equations *y*=1+(*x*−1)0.5 and y=1−(1−*x*)0.5 respectivelywhere *x*=1 describes 100%.[38](#_ENREF_38) The percent reduction in AHI with treatment vs. baseline (primary outcome variable) was transformed to avoid left skewness via the equation *y*=*x*/(2−*x*), which is equivalent to *y*=(BaselineAHI−TreatmentAHI) / (BaselineAHI+TreatmentAHI), 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 ΔAHI from now on). Thus, halving baseline AHI (*x*=0.5) yields a ΔAHI= 33% and doubling baseline AHI (*x*=−1) yields a ΔAHI= −33%. Baseline and treatment AHI were also left 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 ΔAHIand each OSA endotypic trait individually.

***Multivariable regression analysis.*** Initial visual inspection of the data (plots showing responders and non-responders against combinations of collapsibility, loop gain, arousal threshold; 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; the total possible terms include the 5 individual OSA endotypic traits, 5 squared terms (1 per trait), and 10 interaction terms, for a total of 20 possible terms. Significant square-transformed terms would indicate non-linear 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 β0 + β1(loop gain) + β2(VPASSIVE) + β3(loop gain)2+ β4(VPASSIVE)2 + β5(loop gain × VPASSIVE). 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)[36](#_ENREF_36),[39-41](#_ENREF_39); the approach was employed on the basis that removal of very weak predictors reduces uncertainty of remaining model coefficients. For interpreting associations, we did not adjust the P-value threshold for multiple comparisons (e.g. 5-traits: conservative p-threshold = 0.01). Terms were accepted as significant at p<0.05.

***Defining endotypic subgroups.*** Definingthe endotypicsubgroups of predicted responders and predicted non-responders were based on the above regression model and the following steps: 1) True responders and true non-responders were defined by percent reduction in AHI cutoff = 50% (true efficacy). Predicted responders and predicted non-responders were defined by determining the optimal cutoff from the multivariable regression model output that maximized sensitivity plus specificity (cutoff of 60% model-*predicted* efficacy was found, see Results; note model-predicted efficacy and true efficacy are not equal). Predicted subgroups were allocatedusing a “leave-one-patient-out cross-validation” procedure, to avoid overestimating predictive value. The goal of this procedure is to ensure that the outcome status of a given patient is predicted exclusively using a model made using all patient data except his/her own; thus cross-validated results are more conservative (more likely indicative of future re-test performance). Accordingly, individual patient subgroup allocation as a “predicted responder” vs “predicted non-responder” was determined—for patient 1—by building a modified model (re-running backwards elimination regression) without Patient 1’s data, and using this model to predict Patient 1’s response, then repeat for Patients 2-93.

The primary statistical comparison for the study was the difference in percent reduction in AHI (primary outcome variable, see above) between the predicted endotypic subgroups.

***Adjusting for clinical covariates.***Multiple linear regression was used to determine whether being a predicted responder (vs. non-responder) predicted oral appliance efficacy (ΔAHI) independently of clinical covariates (i.e. baseline AHI, BMI, age, gender and neck circumference; also baseline REM:NREM AHI and change in REM sleep proportion). To perform this test, predicted response status (1 or 0) was included as an independent variable and each candidate’s clinical covariates was sequentially included-then-removed from the model.

***Presentation.*** Data are presented as mean±SD for descriptive variables and mean±SEM for comparisons. Back transformed data were presented as mean [95% confidence interval]. Data were described as median [25th - 75th centile] for non-normally-distributed data as appropriate. Significance was accepted at p<0.05. Figures were created using custom MATLAB software (MathWorks, Natick, MA, USA).

# Results

**Baseline Characteristics**

Data from 93 participants (56% males) were analyzed. Baseline vs. treatment characteristics for the overall groups are presented in Table 1. On average, participants were middle-aged (56.2±11.0 years), obese (30.5 ±5.3 kg/m2) with moderate to severe OSA (30.6 [24.4 – 43.5] events/hr).

**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). Forty-three patients were considered responders according to our a priori treatment outcome definition.

**Bivariate Analyses**

Using simple linear regression analyses, we observed no bivariate associations between oral appliance efficacy (percent reduction in AHI transformed; ΔAHI) and any of the individual endotypic traits at baseline (R2<0.01 for all). There were also no associations between oral appliance efficacy and baseline AHI, BMI, age, gender 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 VPASSIVE (non-severe and non-mild), lower pharyngeal compensation and more favorable non-pharyngeal traits (i.e. lower loop gain, higher arousal threshold and lower response to arousal), see Table 3 and Figure 1. Several interaction variables were also associated with treatment efficacy (see Table 3 and Figure 1 for interpretation of each of the 12 terms included in the model).

**Defining Endotypic Subgroups**

Use of the above multivariable regression model to define endotype subgroups of predicted responders and predicted non-responders revealed the following:

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

***After cross-validation (main results)*.** Differences in responses between subgroups remained clinically significant after cross-validation: Predicted responders (N=54), compared with predicted non-responders (N=39), exhibited a greater reduction in AHI from baseline (73[66-79] vs. 51[38-61]%, mean[95%CI], p=0.0006) and had lower treatment AHI (8[6-11] vs. 16[12-20] events/hr, p=0.002), see Figure 2. Positive and negative predictive values were 78% (42:12) and 46% (18:21) respectively (p=0.02, Fisher exact test); accuracy = 65%.

***Further analysis.*** Adjusting for covariates (baseline AHI, BMI, age, gender, neck circumference, baseline REM AHI : NREM AHI, change in REM sleep proportion) did not attenuate the differences between groups. Notably, baseline AHI was similar between groups (predicted responders: 34[30-38] vs. predicted non-responders: 33[29-37] events/hr, mean[95%CI], p=0.5). Additionally, none of the above clinical covariates were significantly associated with the response to therapy (ΔAHI) when considered individually (linear regression) or in combination (multivariable regression, total R2=0.08).

We also note that 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 definition of “true responder” from >50% to >70% reduction in AHI yielded similar results, with group differences in efficacy (cross-validated) of 22% (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 exact test); accuracy = 69%.

# Discussion

The current study is 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 non-pharyngeal traits (lower loop gain, higher arousal threshold and lower ventilatory response to arousal), moderate collapsibility (not mild nor 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/hr, ~quarter of baseline) compared with “predicted non-responders” (16 events/hr, ~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.[25-28](#_ENREF_25),[42](#_ENREF_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.[19](#_ENREF_19) We emphasize, however, that in the current 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.[36](#_ENREF_36)

Previous studies have also found that severe collapsibility is associated with reduced oral appliance efficacy.[14](#_ENREF_14),[19](#_ENREF_19),[25](#_ENREF_25),[43-45](#_ENREF_43) Oral appliances typically reduce critical collapsing pressure by 3-5 cmH2O[25](#_ENREF_25),[46-48](#_ENREF_46) and, therefore, oral appliance therapy is unlikely to resolve OSA in those patients with severe collapsibility at baseline. Greater collapsibility (lower VPASSIVE), higher BMI, non-positional OSA (a marker of greater collapsibility) and higher CPAP requirement have each been shown to predict poor response to oral appliance therapy,[14](#_ENREF_14),[19](#_ENREF_19),[25](#_ENREF_25),[43-45](#_ENREF_43) although these are not robust predictors individually. Unexpectedly, the current study found a U-shaped relationship between collapsibility and response to oral appliances. As expected, severe collapsibility predicted reduced responses. However, mild collapsibility also predicted a reduced response. We consider that these individuals, rather than being an “easier to treat”, have a more “non-pharyngeal” mechanisms underpinning their sleep apnea. We emphasize that, while we initially considered that the U-shaped relationship could be spurious, we noted that a large proportion of patients were non-responders with mild collapsibility (and higher loop gain / 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 found that—in addition to elevated loop gain—lower arousal threshold and greater ventilatory response to arousal also contributed to a reduced oral appliance efficacy. These non-pharyngeal factors contributing to breathing instability are unlikely to be corrected by mandibular advancement.[19](#_ENREF_19) Indeed, it was precisely this subgroup of patients that responded preferentially to supplemental oxygen in our recent study.[36](#_ENREF_36) Furthermore, we found that reduced pharyngeal compensation was associated with a higher oral appliance efficacy. According to physiological principles, a stronger pharyngeal dilator 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 findings 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 gender) whether individually or in-combination with OSA endotypic traits. These data confirm the difficulty in using routine clinical variables to predict outcomes of oral appliances therapy.[44](#_ENREF_44) Our study also supports the concept that baseline severity of OSA (AHI) is not a useful predictor of responses to therapy.

## Clinical Implications

The current 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 (moderate-to-severe) OSA patients who have a superior treatment efficacy compared with other OSA patients 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 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,[49](#_ENREF_49) appears sufficient to justify offering oral appliances as a first-line therapy in selected (moderate-to-severe) OSA patients (specifically those who have a preference for this intervention). Although our results were based on unseen ‘hold-out’ data (leave-one-patient-out cross-validation), these findings nevertheless require replication in a larger prospective study for this method to be adopted for routine clinical use. Notably, even the “predicted non-responders” subgroup had, on average, 50% reduction in AHI (residual AHI ~16 events/hr). While this level of efficacy seems unlikely to show superior health benefit compared to CPAP, the considerable improvement in non-responders 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 here is inexpensive, not dependent on specialized equipment or physiological interventions and can produce results rapidly. The data used for analysis in the current 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 by a commercially-available sleep recording system (Profusion PSG, Compumedics Ltd., Australia). Since the analysis was retrospective, there was no opportunity to pay extraordinary attention to nasal pressure quality beyond AASM standards (unfiltered nasal pressure). Thus, the remaining obstacles for more widespread include 1) incorporation of endotyping methods into commercial systems, and 2) the need re-scoring of arousal timing (not performed clinically); neither are insurmountable.

## Methodological Considerations

There are several limitations of our work. We emphasize that the endotypic traits described here 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. We consider that it would be a highly challenging endeavor to perform gold standard measurements of physiology (via CPAP drops or esophageal catheterization) in such a 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 >90 oral-appliance-treated OSA patients. Second, we studied patients with baseline AHI >20 events/hr (median 30 events/hr) and, thus, our results are relevant to those with similar OSA severity and may not apply to many patients with 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 more severe OSA. to determine whether those with milder OSA (AHI <20 event/hr) 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 clinically-relevant variables. For example we did not have systematic data on the percentage of time spent supine at baseline and on therapy, which may have added a source of uncontrolled variability to the model; controlling for position would have likely reduced a source of undesirable variability. We also consider that the influence of endotypes on efficacy is unlikely to be confounded by treatment-vs-baseline differences in position (i.e. we do not have a plausible mechanism by which treatment-related changes in supine sleep duration could differ by baseline physiological traits). We also did not have systematic measures of daytime sleepiness (e.g. Epworth Sleepiness Scale), and thus could not assess the role of sleepiness in the context of the endotypic traits. We note, however that we did find a relationship between lower arousal threshold and reduced OA efficacy (Table 2), suggesting that a lower (physiological) drive for sleep might render OA treatment less efficacious; further investigation along these lines is warranted. Fourth, we used the percent reduction in AHI as a continuous outcome measure and a single cutoff (>50% reduction) to define the “true responders” (often used to define non-responders vs partial-to-complete responders[ref]); we note that changing the cutoff e.g. to a 60 or 70% reduction did not appreciably alter the findings. We also note that the proportions of patients defined as *complete responders* (≥50% reduction in AHI and residual AHI<10 events/hr), *partial responders* (≥50% reduction in AHI but residual AHI ≥10 events/hr) and *non-responders* (<50% reduction in AHI) were 30:12:12 in predicted responders and 10:11:18 in predicted non-responders (p=0.01, Fisher exact test), respectively. Fifth, we note that while subgroup differences in efficacy appear clinically-relevant, the overall model accuracy is modest (as noted above, predicted non-responders still average a 50% reduction in AHI; thus at present we are unable to identify a subgroup who may exhibit negligible benefit). Further incorporation of additional information on site/structure of pharyngeal obstruction, e.g. through coupling of our approach with other polysomnographic methods (airflow shape[50](#_ENREF_50)) may further improve the model precision and predictive performance. Finally, we caution that the non-invasive 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 EEG analysis).

## 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 non-pharyngeal contributions to OSA at baseline (including lower loop gain, higher arousal threshold and lower ventilatory response to arousal). Greater efficacy was also associated with moderate (non-mild or non-severe) 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 non-responders” 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 personalized oral appliance therapy in OSA.

# Acknowledgements

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

# Support

AB receives funding for his PhD studies from the Saudi Arabian Government (Department of Physiology, Rabigh Medical School, King Abdulaziz University). SS was supported by the American Heart Association (15SDG25890059), American Thoracic Society Foundation, and the National Institute of Health (R01HL102321). DJE is supported by a National Health and Medical Research Council Australia Senior Research Fellowship (1116942). BE is supported by a Heart Foundation of Australia Future Leader Fellowship (101167) and holds grants from the National Health and Medical Research Council Australia.

# Disclosure Statement

AW receives research support from Philips Respironics. SS and AW served as consultants for Cambridge Sound Management, Nox Medical, Inspire. SS also serves as a consultant for Merck. SS also receives grant support from Apnimed. PC has an appointment to an endowed academic Chair at the University of Sydney that was established from ResMed funding. He has received research support from ResMed, SomnoMed and Zephyr Sleep Technologies. He is a consultant / adviser to Zephyr Sleep Technologies, ResMed (Narval), and Bayer. He has a pecuniary interest in SomnoMed related to a previous role in R&D (2004). DJE has research grants from Bayer, Apnimed and a Cooperative Research Centre Project Grant (a joint Government, Academia and Industry collaboration, Industry partner: Oventus Medical). BE receives grant support from Apnimed.

# References

1. Ramar K, Dort LC, Katz SG, 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(7):773-827.

2. Petri N, Svanholt P, Solow B, Wildschiodtz 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(2):221-229.

3. 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(6):860-864.

4. Blanco J, Zamarron 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(1):20-25.

5. Bloch KE, Iseli A, Zhang JN, et al. A randomized controlled crossover trial of two oral appliances for sleep apnea treatment. *Am J Respir Crit Care Med.* 2000;162(1):246-251.

6. 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(5):743-748.

7. Gotsopoulos H, Kelly JJ, Cistulli PA. Oral appliance therapy reduces blood pressure in obstructive sleep apnea: a randomized controlled trial. *Sleep.* 2004;27(5):934-941.

8. Andren A, Hedberg P, Walker-Engstrom ML, Wahlen 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(2):705-712.

9. Gauthier L, Laberge L, Beaudry M, Laforte M, Rompre PH, Lavigne GJ. Mandibular advancement appliances remain effective in lowering respiratory disturbance index for 2.5-4.5 years. *Sleep Med.* 2011;12(9):844-849.

10. 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(4):374-380.

11. Phillips CL, Grunstein RR, Darendeliler MA, 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(8):879-887.

12. 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(3):411-417.

13. Remmers J, Charkhandeh S, Grosse J, et al. Remotely controlled mandibular protrusion during sleep predicts therapeutic success with oral appliances in patients with obstructive sleep apnea. *Sleep.* 2013;36(10):1517-1525.

14. 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.* 2015;30:25-33.

15. Eckert DJ. Phenotypic approaches to obstructive sleep apnoea - New pathways for targeted therapy. *Sleep Med Rev.* 2016.

16. 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(8):996-1004.

17. Wellman A, Eckert DJ, Jordan AS, et al. A method for measuring and modeling the physiological traits causing obstructive sleep apnea. *J Appl Physiol.* 2011;110(6):1627-1637.

18. Wellman A, Edwards BA, Sands SA, et al. A simplified method for determining phenotypic traits in patients with obstructive sleep apnea. *J Appl Physiol.* 2013;114(7):911-922.

19. Edwards BA, Andara C, Landry 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.

20. Eckert DJ, Owens RL, Kehlmann GB, et al. Eszopiclone increases the respiratory arousal threshold and lowers the apnoea/hypopnoea index in obstructive sleep apnoea patients with a low arousal threshold. *Clin Sci.* 2011;120(12):505-514.

21. Edwards BA, Sands SA, Owens RL, 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.* 2016;39(11):1973-1983.

22. Gleeson K, Zwillich CW, White DP. The influence of increasing ventilatory effort on arousal from sleep. *Am Rev Respir Dis.* 1990;142(2):295-300.

23. Younes M. Role of arousals in the pathogenesis of obstructive sleep apnea. *Am J Respir Crit Care Med.* 2004;169(5):623-633.

24. Ratnavadivel R, Stadler D, Windler S, et al. Upper airway function and arousability to ventilatory challenge in slow wave versus stage 2 sleep in obstructive sleep apnoea. *Thorax.* 2010;65(2):107-112.

25. Marques M, Genta P, Sands SA, 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-A2584.

26. Terrill PI, Edwards BA, Nemati S, et al. Quantifying the ventilatory control contribution to sleep apnoea using polysomnography. *Eur Respir J.* 2015;45(2):408-418.

27. Sands SA, Edwards BA, Terrill PI, et al. Phenotyping Pharyngeal Pathophysiology Using Polysomnography in Patients with Obstructive Sleep Apnea. *Am J Respir Crit Care Med.* 2018;197(9):1187-1197.

28. Sands SA, Terrill PI, Edwards BA, et al. Quantifying the Arousal Threshold Using Polysomnography in Obstructive Sleep Apnea. *Sleep.* 2018;41(1).

29. 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(11):1879-1887.

30. Lowth A, Juge L, Knapman F, et al. Dynamic MRI tongue deformation patterns during mandibular advancement and associations with craniofacial anatomy in OSA. 2018;27(S2):e169\_12766.

31. 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(2):149-156.

32. Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med.* 2012;8(5):597-619.

33. Iber. *The American Academy of Sleep Medicine Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification.* Westchester: American Academy of Sleep Medicine; 2007.

34. Akobeng AK. Understanding type I and type II errors, statistical power and sample size. *Acta Paediatr.* 2016;105(6):605-609.

35. Zhang S, Paul J, Nantha-Aree M, 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.

36. Sands SA, Edwards BA, Terrill PI, et al. Identifying obstructive sleep apnoea patients responsive to supplemental oxygen therapy. *Eur Respir J.* 2018;52(3).

37. Sands SA, Edwards BA, Terrill PI, et al. Phenotyping Pharyngeal Pathophysiology using Polysomnography in Patients with Obstructive Sleep Apnea. *Am J Respir Crit Care Med.* 2018;197(9):1187-1197.

38. Sands SA, Edwards BA, Terrill PI, et al. Identifying obstructive sleep apnea patients responsive to supplemental oxygen therapy. *Eur Respir J.* 2018.

39. Dunkler D, Plischke M, Leffondre K, Heinze G. Augmented backward elimination: a pragmatic and purposeful way to develop statistical models. *PLoS One.* 2014;9(11):e113677.

40. In Lee K, Koval JJ. Determination of the best significance level in forward stepwise logistic regression. *Commun Stat Simul Comput.* 1997;26(2):559-575.

41. Hosmer DW, Lemeshow S. *Applied logistic regression.* Vol 10: Wiley, New York, 1989; 1991.

42. Azarbarzin A, Sands SA, Taranto-Montemurro L, et al. Estimation of Pharyngeal Collapsibility During Sleep by Peak Inspiratory Airflow. *Sleep.* 2017;40(1).

43. Sutherland K, Phillips CL, Davies A, et al. CPAP pressure for prediction of oral appliance treatment response in obstructive sleep apnea. *J Clin Sleep Med.* 2014;10(9):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(8):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(5):666-671.

46. Bamagoos AA, Cistulli P, Sutherland K, et al. Dose-Dependent Effects of Mandibular Advancement on Upper Airway Collapsibility and Muscle Activity in Obstructive Sleep Apnea [abstract]. *Am J Respir Crit Care Med.* 2017: A6972-A6972.

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(2):238-241.

48. Kato J, Isono S, Tanaka A, et al. Dose-dependent effects of mandibular advancement on pharyngeal mechanics and nocturnal oxygenation in patients with sleep-disordered breathing. *Chest.* 2000;117(4):1065-1072.

49. Phillips CL, Grunstein RR, Darendeliler MA, 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(8):879-887.

50. Genta PR, Sands SA, Butler JP, et al. Airflow shape is associated with the pharyngeal structure causing obstructive sleep apnea. *Chest.* 2017.

# Tables

## Table 1: Patient characteristics

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic | Baseline | Oral Appliance | p-value |
| Sex (M:F) | 52:41 | |  |
| Age (years) | 56.2 ± 11.0 | |  |
| BMI (kg/m2) | 30.5 ± 5.3 | |  |
| Neck circumference (cm) | 40.2 ± 4.1 | |  |
| Max possible advancement (mm) | 10.4 ± 3.4 | |  |
| Final advancement (mm) | 9.2 ± 3.0 | |  |
| Mandibular advancement (%max.) | 88.5 ±14.8 | |  |
| AHI, total (events/hr) | 30.6 [24.4 – 43.5] | 11.3 [5.5 – 19.1] | <0.001 |
| *Percent reduction (ΔAHI)* | 67.4 [42.5 - 83.1] | |  |
| AHI, non-REM (events/hr) | 31.4 [21.5 – 48.6] | 7.3 [3.1 – 16.9] | <0.001 |
| AHI, REM (events/hr) | 48.8 [29.1 – 66.1] | 25.6 [6.1 – 45.3] | <0.001 |
| AHI, supinea (events/hr) | 52.9 [33.0 – 72.4] | 16.7 [8.7 – 38.4] | <0.001 |
| Arousal Index (events/hr) | 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 |
| Total sleep time (min) | 356 ± 65 | 363 ± 65 | 0.43 |
| REM sleep time (%TST) | 16.8 ± 6.2 | 17.6 ± 6.9 | 0.45 |
| Supine sleep timea (%TST) | 40.1 [26.3 – 69.3] | 37.0 [21.2 – 80.6] | 0.83 |

On average, participants were typical OSA patients, middle aged, predominantly obese with moderate-to-severe OSA. Continuous variables are presented as mean ± SD or median [25th - 75th centile]. BMI, body mass index; AHI, Apnea hypopnea index; TST, total sleep time. aData available in N=62.

## Table 2: Traits associated with oral appliance efficacy: multiple regression

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Beta | SEM | Beta Std. | p-value | Interpretation |
| Constant | 47.1 | 5.1 | 1.4 | <0.0001 |  |
| Pharyngeal traits |  |  |  |  |  |
| VPASSIVE | -0.771 | 0.325 | -0.47 | 0.02 | Not severe and not mild collapsibility → Success |
| VPASSIVE2 | -0.0293 | 0.0095 | -1.1 | 0.003 |
| Compensation2 | 0.0215 | 0.0067 | 0.94 | 0.002 | Higher compensation → Failure  particularly when arousal threshold is low or response to arousal is high |
| Compensation x Arousal threshold | 0.0486 | 0.0107 | 1.4 | <0.0001 |
| Compensation x Response to arousal | -0.0171 | 0.0064 | -0.41 | 0.009 |
| Non-pharyngeal traits |  |  |  |  |  |
| Loop gain | -112 | 41 | -0.37 | 0.008 | Higher loop gain → Failure  particularly when compensation or response to arousal are high |
| Loop gain x Compensation | -6.95 | 2.23 | -0.52 | 0.003 |
| Loop gain x 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 threshold2 | 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 x VPASSIVE | -0.0212 | 0.0115 | -0.50 | 0.069 |

Oral appliance efficacy is defined as the percentage reduction in apnea-hypopnea index with treatment compared to baseline (transformed, see Methods). The Table describes final results (12/20 terms) after backward stepwise elimination (P-to-remove=0.157) which began with five traits, their squares and all interaction terms. Note that 10/20 terms were significant with p<0.05, 9/20 with P<0.01. Traits were mean-subtracted before terms were generated and applied to the model (see below). *Beta Std.* 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 non-responder to a typical responder). Mean values of the endotypic traits before mean substraction: VPASSIVE = 79.0±20.8, Loop gain = 0.43±0.11, Compensation = -9.5+27.0%, Arousal threshold = 141.8±26.0%, Response to Arousal = 36.3±22.6%. A regression model cutoff of 60% (predicted reduction in AHI, untransformed) was used to define predicted responders and predicted non-responders (maximized sensitivity plus specificity). *SEM* = standard error of the mean. Overall R2=0.30, adjusted R2=0.19, p=0.003.

# Figures and Figure Legends

**Figure 1.** Key slices of the 5-trait multivariable model (Table 2) illustrating how combinations of traits may influence oral appliance efficacy. Each slice illustrates the 2-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 non-responders (<50% reduction in 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 non-responders). *Top and left*: A U-shaped relationship between collapsibility (Vpassive) and efficacy is evident. For example, in Top, the light-green shading indicating predicted responders are only seen in a mid range of “moderate” collapsibility, and at lower loop gain. Note that non-responders with high Vpassive (mild collapsibility) tend to have high loop gain, low arousal threshold, 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 Vpassive), 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 simplifed two-trait view (i.e. 2-trait prediction differed from the full model prediction).

**Figure 2.** Based on combined endotype traits, predicted responders (black), compared with predicted non-responders (gray)**,** exhibited a greater oral appliance efficacy indicated by a greater reduction in apnea-hypopnea index (untransformed) from baseline **(A)** and a lower residual apnea-hypopnea index on treatment **(B).** Error bars llustrate 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.



