ERJ Express. Published on August 23, 2018 as doi: 10.1183/13993003.00674-2018

EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

Identifying obstructive sleep apnea patients responsive to supplemental oxygen therapy

Scott A. Sands, Bradley A. Edwards, Philip I. Terrill, James P. Butler, Robert L. Owens, Luigi Taranto-Montemurro, Ali Azarbarzin, Melania Marques, Lauren B. Hess, Erik T. Smales, Camila M. de Melo, David P. White, Atul Malhotra, Andrew Wellman

Please cite this article as: Sands SA, Edwards BA, Terrill PI, *et al*. Identifying obstructive sleep apnea patients responsive to supplemental oxygen therapy. *Eur Respir J* 2018; in press (https://doi.org/10.1183/13993003.00674-2018).

Identifying Obstructive Sleep Apnea Patients Responsive to Supplemental Oxygen Therapy

Scott A Sands[1,2*], Bradley A Edwards[1,3,4], Philip I Terrill[5], James P Butler[1], Robert L Owens[1,6], Luigi Taranto-Montemurro[1], Ali Azarbarzin[1], Melania Marques[1], Lauren B Hess[1], Erik T Smales[1], Camila M de Melo[1], David P White[1], Atul Malhotra[1,6], Andrew Wellman[1]

[1]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital and Harvard Medical School, Boston, USA.
[2]Department of Allergy, Immunology and Respiratory Medicine and Central Clinical School, The Alfred and Monash University, Melbourne, Australia.
[3]Sleep and Circadian Medicine Laboratory, Department of Physiology, Monash University, Melbourne, VIC, Australia.
[4]School of Psychological Sciences and Monash Institute of Cognitive and Clinical Neurosciences, Monash University, Melbourne.
[5]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia.
[6]Division of Pulmonary, Critical Care and Sleep Medicine, University of California San Diego, La Jolla CA, USA.

*Corresponding author: Scott A. 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 6172780911 | F: +1 6177327337

Running title: Targeting Ventilatory Control in Obstructive Sleep Apnea

Word count: 3559/3000

Author Contributions. Conception and design: SS, BE, PT, AM, AW. Methods development: SS, PT, AA. Data collection: SS, BE, CM, ES. Analysis: SS, LH. All authors interpreted data, edited the manuscript for important intellectual content, and approved the final draft.

Take-Home Message
A subgroup of patients with obstructive sleep apnea who benefit from stabilizing ventilatory control with supplemental oxygen therapy can be recognized by estimating pathophysiological mechanisms from a routine diagnostic sleep study.

Abstract

Rationale: A possible precision-medicine approach to treating obstructive sleep apnea (OSA) involves targeting ventilatory instability (elevated loop gain) using supplemental inspired oxygen in selected patients. Here we test whether elevated loop gain and three key endophenotypic traits (collapsibility/compensation/arousability)—quantified using clinical polysomnography—can predict the effect of supplemental oxygen on OSA severity.

Methods: 36 patients (apnea-hypopnea index [AHI] >20 events/hr) completed two overnight polysomnographic studies (single-blinded randomized-controlled cross-over) on supplemental oxygen (40% inspired) versus sham (air). OSA traits were quantified from the air-night polysomnography (Terrill et al. ERJ 2015). Responders were defined by ≥50% reduction in AHI (supine non-REM). Secondary outcomes included blood pressure and self-reported sleep quality.

Results: 9/36 patients (25%) responded to supplemental oxygen (ΔAHI=72±5%). Elevated loop gain was not a significant univariate predictor of responders/non-responder status (primary analysis). In post-hoc analysis, a logistic regression model based on elevated loop gain and other traits (better collapsibility and compensation; cross-validated) had 83% accuracy (89% before cross-validation); predicted responders exhibited improved OSA severity (ΔAHI: 59±6% vs. 12±7% in predicted non-responders, p=0.0001) plus lowered morning blood pressure and "better" self-reported sleep.

Conclusions: Patients whose OSA responds to supplemental oxygen can be identified by measuring their endophenotypic traits using diagnostic polysomnography.

Keywords

ventilatory instability | personalized medicine | endophenotype | loop gain

Introduction

Around half of patients diagnosed with obstructive sleep apnea (OSA) are currently untreated or non-adherent to continuous positive airway pressure (CPAP) [1, 2]. Thus, novel approaches to OSA therapy are required. In the last few years, the concept of personalized OSA therapy has emerged, based on the idea that OSA interventions have maximal impact when they match patients' underlying pathophysiology or 'endophenotypes' [3-7]. At the core of this notion is the recognition that OSA emerges as the consequence of different endophenotypic traits in different individuals, namely 1.) increased pharyngeal collapsibility, 2.) reduced ventilatory control stability (elevated *loop gain*, i.e. an exaggerated ventilatory drive response to reduced airflow and attendant hypoxia/hypercapnia), 3.) reduced respiratory arousal threshold (a small rise in ventilatory drive terminates sleep), and 4.) reduced compensatory pharyngeal dilator muscle activation [8-11].

A major hurdle for clinical implementation of personalized medicine is that assessment of the traits causing OSA has been confined to research laboratories [9, 12]. To overcome this barrier, we recently developed an automated technique for estimating the four key traits [13-15] using routine clinical sleep studies (polysomnography).

Here we prospectively tested the predictive value of phenotyping using polysomnography in a study of supplemental inspired oxygen, a therapy which acts specifically to lower loop gain [16] and substantially improves OSA in a subgroup of patients [3]. We tested the primary hypothesis that elevated loop gain—measured from clinical polysomnography—predicts a preferential reduction in OSA severity with a single night of supplemental oxygen (40% inspired) versus sham (air) in a randomized single-blind crossover study (NCT01751971). We also assessed the predictive value of elevated loop gain in combination with the other three traits (post-hoc analysis).

Methods

Participants

Patients with a previous clinical diagnosis of OSA with AHI>20 events/hr were eligible to participate. Patients using respiratory stimulants or depressants (including opioids, benzodiazepines) were excluded, as were those with diagnoses of heart failure or lung diseases, with central rather than obstructive sleep apnea (majority central respiratory events), and women who were pregnant. Participants provided written informed consent and approval was granted by the Partners' Institutional Review Board.

We enrolled 47 participants; 8 did not exhibit OSA and 3 did not attend the second overnight, leaving 36 patients who completed the protocol (Supplemental Figure S1).

Procedure

Patients completed two overnight polysomnographic studies, one week apart (randomized order). Supplemental oxygen or medical air (sham) was delivered via venturi mask (40% inspired oxygen, equivalent to ~4 L/min via cannula [17]; see [3, 4]). The single-blind design enabled real-time monitoring of appropriate inspired oxygen levels (Vacumed, Ventura CA). Patients slept predominantly-supine to minimize position effects. Supine blood pressure was measured (Dinamap Pro 100v2, GE Medical Systems, Tampa FL, USA) during quiet wakefulness immediately preceding lights out (after ≥1 hour of rest during setup and ~10 min in the supine position) and again ~10 min after lights on with oxygen/sham removed. We assessed self-reported sleep quality at study completion ("better", "same" or "worse" vs. previous study; scores −1, 0, 1 respectively) and alertness (Stanford Sleepiness Scale) on each morning.

Polysomnography

Standard clinical polysomnographic instrumentation was used [18]. Airflow was assessed with nasal pressure and oronasal thermistor. We prioritized recording of high quality nasal pressure signals. Hypopneas were scored based on a 30% reduction in airflow *without* an oxygen desaturation or arousal criterion (supplemental oxygen would otherwise mask hypopneas).

Quantifying the pathophysiological traits using polysomnography

Sham night polysomnography was used to quantify the baseline OSA traits using an automated method [13-15]:

Chemical drive. Phenotypic traits were quantified by first estimating "ventilatory drive"—i.e. intended ventilation—using a chemoreflex feedback control model (gain, response time, delay, arousal response) fit to ventilation data [13]. Briefly, nasal pressure (square-root transformed) provided a ventilatory flow surrogate that was integrated to yield a breath-to-breath ventilation signal (uncalibrated tidal volume × respiratory rate, mean-normalized). 7-min windows containing ≥1 respiratory event were identified (non-REM), and

estimated ventilatory drive (model output) was best fit to the ventilation signal between obstructive events (when the airway is patent).

Loop gain. For each window, loop gain was calculated from the feedback model; the median was used to represent the night [13]. Two parameters were quantified: LG_1 (*a priori* predictor) is the ventilatory drive response to a 1 cycle/min reduction in ventilation and reflects "hypersensitivity" (e.g. increased chemosensitivity or reduced lung volume). LG_n quantifies "instability" and also includes circulatory delay effects (LG_n>1.0 yields periodic central apneas) [3, 19].

Arousal threshold. The arousal threshold was taken as the median estimated ventilatory drive preceding scored EEG arousals [15].

Upper airway physiology. Pharyngeal collapsibility ("Vpassive") was defined here as the ventilation at normal/eupneic ventilatory drive [9, 12] during sleep and is quantified using an overnight breath-by-breath summary plot of ventilation versus estimated ventilatory drive [14]. Pharyngeal muscle "compensation" is taken as the increase in ventilation from $V_{passive}$ to the value at the arousal threshold ("V_{active}"; compensation=$V_{active}-V_{passive}$). Greater Vpassive and compensation indicate better pharyngeal patency and were expected to predict successful oxygen therapy.

Definition of response to therapy

Patients were considered "responders" if their AHI was reduced by at least 50% with treatment versus sham (*a priori* criterion) and were otherwise "non-responders". Responses were considered "complete" if the AHI was also <15 events/hr on treatment [4], equivalent to a >67% reduction in AHI in our population. Non-responders were considered "borderline" if they exhibited a >33% reduction in AHI.

Assessment of predictive value

The *predictive value* of loop gain and other traits was assessed based on whether there was a significant difference in the reduction in AHI (percent baseline) between defined "predicted responders" versus "predicted non-responders" subgroups (e.g. high versus low loop gain). The pre-specified choice of loop gain parameter (LG_1) and a *cutoff* of 0.7 (i.e. threshold that defines high versus low LG_1) was based on prior findings [13]. For all post-hoc subgroup classification tests, including multivariable analysis (below), cutoff values were selected to maximize sensitivity plus specificity [20]. Predictive values were estimated using "leave-one-out" cross-validation, whereby that each subject's response was predicted using a modified version of the same model with the subject's own data held out.

Multivariable model analysis

We employed logistic regression using the phenotypic variables—loop gain, collapsibility, arousal threshold, muscle compensation—to define subgroups of "predicted responders" and "predicted non-responders"; a "quadratic" model that includes interaction terms and squared terms was chosen based on clear evidence of interactions. Terms were selected using backward elimination (p-to-remove=0.157) [21]. To raise statistical

power, we included “training data” from a previous study employing supplemental oxygen (N=20) to help build a more robust regression model [4]; predictive value was reported exclusively for patients in the current study (see Online Supplement).

Statistical analysis

Unpaired Student’s t-tests or Wilcoxon rank-sum tests were used to compare differences between groups (responders versus non-responders; predicted responders versus predicted non-responders). Standard errors for proportions (positive/negative predictive values, i.e. proportion of correct predictions given a positive/negative prediction) employed the normal approximation to the binomial distribution; p-values indicate differences from chance values. Significance was accepted at $p<0.05$.

Results

Baseline characteristics are detailed in Table 1. Supplemental oxygen lowered AHI by ~30% overall (Supplemental Table S1) confirmed by a ~25% reduction in arousal index. Responders (N=9) exhibited ~70% reduction in AHI, accompanied by a ~50% reduction in the frequency of arousals from sleep, a ~50% reduction in the time spent in light sleep (stage 1 non-REM), and exhibited a ~7 mmHg reduction in the overnight change in blood pressure (Figure 1). These changes were not observed in non-responders (N=27). 6/9 responders also had an AHI<15 events/hr on oxygen (i.e. complete responders). Overall, most patients felt they slept "better" on the oxygen night (better:same:worse on oxygen vs. air = 19:9:7). There was no effect on subjective alertness (change in Stanford Sleepiness Scale: +0.1±0.2 units on a scale, higher value represents reduced alertness).

Loop gain and responses to oxygen

Example measurements in one responder and one non-responder are shown in Figure 2. Contrary to our primary hypothesis, elevated baseline loop gain was not a significant univariate predictor of the response to treatment (% reduction in AHI with supplemental oxygen vs. sham), i.e. there was no difference in the response between patients with higher versus lower loop gain based on LG_1 (35.7±6.6% vs 25.2±9.0%, p=0.4; pre-specified cutoff = 0.7; Supplemental Figure S2). However, there was a strong trend towards a greater response in those with higher loop gain based on LG_n (Figure 3).

Other physiological traits and responses to oxygen

Reduced collapsibility (higher $V_{passive}$) and greater compensation were strong predictors of the response to oxygen, and there was a trend towards a greater response in patients with lower arousal threshold (Figure 3).

Multivariable model analysis

When traits were considered *in combination* (multivariable logistic regression), a higher loop gain increased the likelihood of being a responder, particularly in patients with better compensation; a poor compensation and poor collapsibility reduced the likelihood of being a responder (Table 2, Figure 4; see also Table S2, Figure S3 Online Supplement). The regression model exhibited excellent predictive value overall (ΔAHI in predicted responders vs predicted non-responders: 62±5% versus 10±7%, p<0.0001; positive predictive value [PPV] = 69±13% [p=0.0005], negative predictive value [NPV] = 100±0% [p<0.0001], accuracy = 89±5% [p<0.0001]); after cross-validation results remained strong (ΔAHI = 58±6% versus 12±7%, p=0.0001; PPV = 62±13% [p=0.007], NPV = 96±4% [p<0.0001], accuracy = 83±6% [p=0.0002]).

Secondary outcomes in "predicted responders"

Predicted responders (cross-validated results) exhibited improvements with oxygen in arousal index, blood pressure and subjective sleep quality (slept better:same:worse on oxygen vs. air) that were not observed in predicted non-responders (Figure 5). Stanford Sleepiness Scale was unchanged in both subgroups.

Predictive values of non-physiological variables

Predicted responder/non-responder status based on the pathophysiology remained significantly associated with responses to oxygen (p=0.002-0.003) after adjusting for common clinical factors (age, sex, body mass index, neck circumference, current CPAP use, and AHI off treatment; logistic regression with covariates added separately, see Supplemental Figure S4). These clinical factors were not significant predictors.

Effect of oxygen on the physiological traits

To understand whether oxygen therapy may have adversely-affected traits other than loop gain, particularly in non-responders. we also assessed the effect of oxygen on the traits causing sleep apnea (Table 3, see Supplemental Figure S5 for individual data). As expected [13, 16], oxygen lowered loop gain ("instability"; LG_n) through a reduction in ventilatory control "sensitivity" (LG_1) but had no influence on collapsibility ($V_{passive}$) or compensation. However, oxygen lowered our measure of the arousal threshold. Effects were similar between responder and non-responder subgroups.

Discussion

The current study demonstrated that quantifying the pathophysiological traits of OSA patients using diagnostic polysomnography can identify patients whose condition is likely to be treatable using supplemental oxygen, and can rule out non-responders. In contrast to our specific hypothesis, elevated loop gain alone was not a strong predictor of the response to oxygen. Rather, elevated loop gain in combination with greater pharyngeal patency (less-severe collapsibility and greater compensation) predicted improved OSA; the combined multivariable model accurately ruled out a positive response to treatment (95% certainty), and identified a subgroup of "predicted responders" who: 1.) had a 62% likelihood of halving the frequency of respiratory events and a 46% likelihood of adequately treating OSA (AHI<15 events/hr), and 2.) experienced a significant reduction in morning blood pressure (relative to evening values) and an improvement in self-reported sleep quality (slept "better" than on the night without treatment). Our study shows that measuring and combining key endophenotypic causes of sleep apnea—based on a routine sleep study—provides insight into the response to an intervention and opens the door for personalizing intervention based on underlying mechanisms.

Physiological insight

Elevated loop gain. We hypothesized that elevated loop gain would predict the response to lowering loop gain with supplemental oxygen, based on the notion that patients with higher loop gain have the greatest range for lowering loop gain and improving OSA. For example, many forms of central sleep apnea—a high loop gain condition—can be effectively treated with supplemental oxygen [22, 23]. In addition, our prior study (N=12) in OSA found that oxygen is more effective in patients with higher versus lower loop gain [3, 13]. Yet here we found that loop gain alone was insufficient to predict the response to supplemental oxygen, confirming more recent studies illustrating that loop gain (or chemosensitivity) by itself is not strongly associated (or even inversely associated [24]) with the reduction in OSA severity with treatments that lower loop gain (oxygen/acetazolamide) [4, 6, 25]. The current study demonstrates that successfully targeting an abnormal pathophysiological trait for OSA treatment requires not just knowledge of the trait itself but also the other determinants of OSA.

Greater pharyngeal patency. Importantly, our study demonstrates that more severe pharyngeal collapsibility and poor muscle compensation (as measured using polysomnography) rule-out a positive OSA response to supplemental oxygen. These findings are consistent with OSA pathophysiology; patients who have both a high loop gain and a reduced airway patency will continue to have a reduced patency once loop gain is lowered, promoting residual OSA regardless of how high loop gain was at baseline [4, 16].

Baseline OSA severity. Remarkably, in the nine responders (25% of patients), oxygen lowered the AHI from an average of 57 to 18 events/hr (~70% reduction), illustrating that supplemental oxygen can have a large impact in the right patients even in severe OSA. Responders were not identified by milder sleep apnea at

baseline, i.e. a lower apnea-hypopnea index. Thus, OSA severity is unlikely to contribute meaningfully to clinical decisions regarding the use of loop gain lowering interventions for OSA (Table 1; Supplemental Table S1, Figure S4).

Ruling-out versus ruling-in. Polysomnographic phenotyping accurately ruled out non-responders to supplemental oxygen (high negative predictive value), but was somewhat less accurate at ruling-in responders to treatment. Thus, there remains uncertainty that a patient with appropriate pathophysiology will respond favorably to treatment. A more certain outcome might require that the intervention also consistently succeed at improving the pathophysiology (lowering loop gain) across patients, but the magnitude of such improvements vary considerably [3, 4, 13, 25]. Regardless, by ruling out clear non-responders, our method potentially allows clinicians and investigators to limit future trials of supplemental oxygen (or other agents lowering loop gain) to the subgroup of OSA patients with the most favorable pathophysiology.

Effect of oxygen therapy on the traits. We found that supplemental oxygen lowers our estimate of loop gain and has no influence on our estimates of upper-airway physiology (collapsibility and compensation), consistent with previous data [16]; there was also no evidence of a differential effect on collapsibility or compensation in non-responders versus responders (Table 2, Figure S5) to support the notion that oxygen may worsen upper-airway traits in non-responders. Interestingly, we also observed that oxygen lowered the arousal threshold, i.e. a deleterious effect. The direction of influence, and the clearer effect in responders, indicates that this effect is not a plausible cause of the improvement in OSA or differences between responders/non-responders. Lowering of the arousal threshold may occur consequent to a habituation effect accompanying the amelioration of OSA[26]. The observation is also consistent with our previous finding that reducing inspired oxygen levels (hypoxia) act to increase the arousal threshold [16].

Additional outcomes. Supplemental oxygen was recently confirmed to have no impact on lowering blood pressure in unselected patients with OSA, in contrast to CPAP [27, 28]. On the surface, those findings may appear to contrast with ours. However, in our study, blood pressure lowering effects—in the form of reduced morning minus evening values of systolic and diastolic blood pressure—were only exhibited in the subgroup that responded to treatment. Thus, the available evidence suggests that oxygen administration may exclusively lower blood pressure in the subgroup of patients in whom it also improves OSA. Longer-term investigations are warranted.

Clinical implications

A large proportion of OSA patients are intolerant of continuous positive airway pressure, and supplemental oxygen is currently used as a salvage therapy for OSA primarily to maintain nocturnal oxygenation. In unselected patients, improvements in oxygenation are consistent but improvements in OSA are modest (~10-30% reduction in AHI) [3, 6, 27-31]; in some individuals, oxygen may even increase OSA severity (Figures 2-5). However, in patients with a specific set of phenotypic traits, supplemental oxygen can have effects that

are clinically-relevant (>50% reduction in AHI; Figures 4 and 5). In addition, a recent trial indicated that adherence to supplemental oxygen, while imperfect, is superior to CPAP [27], suggesting that supplemental oxygen may be well tolerated in these individuals.

Our approach to identifying likely responders was to quantify the pathophysiological traits causing OSA using clinical sleep studies and a technique that requires no invasive equipment or positive airway pressure manipulation and is fully automated. Thus, our approach can be readily implemented in the clinical setting.

The specific four-trait regression model (Figure 4) was developed based on the available data (post-hoc) and tested using a leave-one-out cross-validation approach, which makes efficient use of limited data. We emphasize that our specific model needs to be confirmed via prospective testing, preferably over a longer treatment period, to confirm clinical utility. Since our findings are consistent with physiological principles, we believe that the likelihood of reproducibility is high.

Our study also provides proof of principle that it is possible to identify OSA responders to loop-gain-lowering therapies more generally. We hope that our results will also apply to the prediction of responses to other current and future OSA interventions targeting loop gain e.g. acetazolamide [25], H_2S inhibitors [32], and purinergic antagonists [33]. The general approach to combining non-invasive trait measurements may also help to predict responses to other therapies, such as pharyngeal surgery [34, 35]. Further investigations along these lines are needed.

Methodological considerations

In the current study, subjects were studied on one night of treatment with supplemental oxygen. Thus, possible additional improvements in OSA severity over time were not captured [26, 36]. However, available data indicate that oxygen has no major impact on OSA severity beyond the effects on the first night, and observed improvements in OSA severity are immediately reversed with treatment discontinuation [30]. Nonetheless, longer term studies are necessary to examine effects of oxygen on OSA symptoms and blood pressure in predicted responders. For purposes of scientific veracity we administered supplemental oxygen through a venturi mask at a moderate concentration (40% inspired, equivalent to ~4 L/min by nasal cannula), at a level used previously [3, 4], thereby 1.) avoiding interference between nasal cannula airflow measurement and oxygen/sham administration, 2.) avoiding a titration procedure which may have affected patient blinding and reduced the proportion of the night on optimal therapy, and 3.) minimizing dosage differences that might have confounded assessment of outcomes between subgroups.

We interpreted the results of our multiple-trait model (including high loop gain, less-severe collapsibility) as evidence that there is an identifiable patient subgroup that responds preferentially to oxygen therapy. We also consider the possibility that this phenotype could simply have a form of less-severe OSA (albeit previously unrecognized based on AHI or other polysomnographic measures), which might be easier to treat in general with *any* OSA intervention. While further studies are needed to demonstrate oxygen-therapy-

specificity, the available evidence refutes this notion: First, the predicted-responder subgroup exhibited selective improvements in blood pressure and self-reported sleep quality, demonstrating that the responder phenotype of OSA appears clinically-important i.e. is non-trivial. Second, higher loop gain measured using polysomnography has been shown to predict *non-responders* to pharyngeal surgery, illustrating that the high loop gain contribution to the oxygen-responder phenotype it is not a biomarker of easier-to-treat OSA [35]. Third, our recent data suggests that higher loop gain and less-severe collapsibility (the oxygen responder phenotype) predicts oral appliance *failure* [37], suggesting that oxygen responders are likely to be non-responders to other therapies; indeed high loop gain via CPAP manipulation also predicted oral appliance non-responders [5]. A study is underway to investigate this issue further (NCT03189173).

We note that the values for physiological traits obtained in the current study from routine polysomnography are surrogates and might differ from those obtained with invasive 'gold standard' measures used in specialized physiology laboratories. However, our measures compare favorably with gold standard values (correlation coefficients ~0.7 for each trait) [13-15].

The primary outcome variable was the AHI measured in supine non-REM sleep, with respiratory events scored using modified criteria that avoids reliance on the oxygen saturation signal (respiratory events require ≥30% reduction in airflow). Use of standard criteria for scoring respiratory events (≥30% reduction in flow with ≥3% desaturation or arousal), or inclusion of all sleep states and positions, yielded similar results (Table S1, Online Supplement).

Patients slept supine because this captures OSA pathophysiology at its worst; treatment of lateral OSA without mechanical intervention would have been a less formidable challenge. In a patient who sleeps entirely lateral, we see no reason that the same traits while lateral would not predict responses to oxygen therapy also in the lateral position. Additional testing is warranted.

Conclusions

For the first time we show that measuring the pathophysiologic variables causing OSA from a clinical sleep study (off treatment) can predict which patients are most suitable for supplemental oxygen therapy. A multivariable model incorporating increased loop gain with an improved airway (better collapsibility and compensation) accurately predicted responders to therapy: Predicted responders not only exhibited improvements in OSA severity, but also experienced improvements in blood pressure and perception of sleep quality. We consider that phenotyping in this manner will provide an avenue for personalizing interventions for patients who are intolerant of CPAP and may otherwise remain untreated.

References

1. Skinner T, McNeil L, Olaithe M, Eastwood P, Hillman D, Phang J, de Regt T, Bucks RS. Predicting uptake of continuous positive airway pressure (CPAP) therapy in obstructive sleep apnoea (OSA): a belief-based theoretical approach. *Sleep Breath* 2013: 17(4): 1229-1240.
2. Rotenberg BW, Murariu D, Pang KP. Trends in CPAP adherence over twenty years of data collection: a flattened curve. *Journal of otolaryngology - head & neck surgery = Le Journal d'oto-rhino-laryngologie et de chirurgie cervico-faciale* 2016: 45(1): 43.
3. Wellman A, Malhotra A, Jordan AS, Stevenson KE, Gautam S, White DP. Effect of oxygen in obstructive sleep apnea: role of loop gain. *Respir Physiol Neurobiol* 2008: 162(2): 144-151.
4. Edwards BA, Sands SA, Owens RL, Eckert DJ, Landry S, White DP, Malhotra A, Wellman A. 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.
5. Edwards BA, Andara C, Landry S, Sands SA, Joosten SA, Owens RL, White DP, Hamilton GS, Wellman A. 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.
6. Xie A, Teodorescu M, Pegelow DF, Teodorescu MC, Gong Y, Fedie JE, Dempsey JA. Effects of stabilizing or increasing respiratory motor outputs on obstructive sleep apnea. *J Appl Physiol* 2013: 115(1): 22-33.
7. Eckert DJ, Owens RL, Kehlmann GB, Wellman A, Rahangdale S, Yim-Yeh S, White DP, Malhotra A. 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(12): 505-514.
8. Younes M. Role of respiratory control mechanisms in the pathogenesis of obstructive sleep disorders. *J Appl Physiol* 2008: 105(5): 1389-1405.
9. Wellman A, Edwards BA, Sands SA, Owens RL, Nemati S, Butler J, Passaglia CL, Jackson AC, Malhotra A, White DP. A simplified method for determining phenotypic traits in patients with obstructive sleep apnea. *J Appl Physiol* 2013: 114(7): 911-922.
10. 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.
11. McGinley BM, Schwartz AR, Schneider H, Kirkness JP, Smith PL, Patil SP. Upper airway neuromuscular compensation during sleep is defective in obstructive sleep apnea. *J Appl Physiol (1985)* 2008: 105(1): 197-205.
12. Wellman A, Eckert DJ, Jordan AS, Edwards BA, Passaglia CL, Jackson AC, Gautam S, Owens RL, Malhotra A, White DP. A method for measuring and modeling the physiological traits causing obstructive sleep apnea. *J Appl Physiol* 2011: 110(6): 1627-1637.
13. Terrill PI, Edwards BA, Nemati S, Butler JP, Owens RL, Eckert DJ, White DP, Malhotra A, Wellman A, Sands SA. Quantifying the ventilatory control contribution to sleep apnoea using polysomnography. *Eur Respir J* 2015: 45(2): 408-418.
14. Sands SA, Edwards BA, Terrill PI, Taranto-Montemurro L, Azarbarzin A, Marques M, Hess LB, White DP, Wellman A. Phenotyping Pharyngeal Pathophysiology using Polysomnography in Patients with Obstructive Sleep Apnea. *Am J Respir Crit Care Med* 2018: In Press.
15. Sands SA, Terrill PI, Edwards BA, Taranto Montemurro L, Azarbarzin A, Marques M, de Melo CM, Loring SH, Butler JP, White DP, Wellman A. Quantifying the Arousal Threshold Using Polysomnography in Obstructive Sleep Apnea. *Sleep* 2018: 41(1).
16. Edwards BA, Sands SA, Owens RL, White DP, Genta PR, Butler JP, Malhotra A, Wellman A. Effects of hyperoxia and hypoxia on the physiological traits responsible for obstructive sleep apnoea. *J Physiol* 2014.
17. Wettstein RB, Shelledy DC, Peters JI. Delivered oxygen concentrations using low-flow and high-flow nasal cannulas. *Respiratory care* 2005: 50(5): 604-609.
18. Rechtschaffen A, Kales A. A Manual of Standardized Terminology, Techniques and Scoring Systems for Sleep Stages of Human Subjects. . UCLA Brain Information Service / Brain Research Institute, Los Angeles, CA, 1968.
19. Khoo MC, Kronauer RE, Strohl KP, Slutsky AS. Factors inducing periodic breathing in humans: a general model. *J Appl Physiol* 1982: 53(3): 644-659.
20. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950: 3(1): 32-35.
21. Heinze G, Dunkler D. Five myths about variable selection. *Transplant international : official journal of the European Society for Organ Transplantation* 2017: 30(1): 6-10.

22. Weintraub Z, Alvaro R, Kwiatkowski K, Cates D, Rigatto H. Effects of inhaled oxygen (up to 40%) on periodic breathing and apnea in preterm infants. *J Appl Physiol* 1992: 72(1): 116-120.
23. Javaheri S, Ahmed M, Parker TJ, Brown CR. Effects of nasal O2 on sleep-related disordered breathing in ambulatory patients with stable heart failure. *Sleep* 1999: 22(8): 1101-1106.
24. Wang D, Wong KK, Rowsell L, Don GW, Yee BJ, Grunstein RR. Predicting response to oxygen therapy in obstructive sleep apnoea patients using a 10-minute daytime test. *Eur Respir J* 2018: 51(1).
25. Edwards BA, Sands SA, Eckert DJ, White DP, Butler JP, Owens RL, Malhotra A, Wellman A. Acetazolamide improves loop gain but not the other physiological traits causing obstructive sleep apnoea. *J Physiol* 2012: 590(Pt 5): 1199-1211.
26. Loewen A, Ostrowski M, Laprairie J, Atkar R, Gnitecki J, Hanly P, Younes M. Determinants of ventilatory instability in obstructive sleep apnea: inherent or acquired? *Sleep* 2009: 32(10): 1355-1365.
27. Gottlieb DJ, Punjabi NM, Mehra R, Patel SR, Quan SF, Babineau DC, Tracy RP, Rueschman M, Blumenthal RS, Lewis EF, Bhatt DL, Redline S. CPAP versus oxygen in obstructive sleep apnea. *N Engl J Med* 2014: 370(24): 2276-2285.
28. Norman D, Loredo JS, Nelesen RA, Ancoli-Israel S, Mills PJ, Ziegler MG, Dimsdale JE. Effects of continuous positive airway pressure versus supplemental oxygen on 24-hour ambulatory blood pressure. *Hypertension* 2006: 47(5): 840-845.
29. Smith PL, Haponik EF, Bleecker ER. The effects of oxygen in patients with sleep apnea. *Am Rev Respir Dis* 1984: 130(6): 958-963.
30. Gold AR, Schwartz AR, Bleecker ER, Smith PL. The effect of chronic nocturnal oxygen administration upon sleep apnea. *Am Rev Respir Dis* 1986: 134(5): 925-929.
31. Loredo JS, Ancoli-Israel S, Kim EJ, Lim WJ, Dimsdale JE. Effect of continuous positive airway pressure versus supplemental oxygen on sleep quality in obstructive sleep apnea: a placebo-CPAP-controlled study. *Sleep* 2006: 29(4): 564-571.
32. Del Rio R, Marcus NJ, Schultz HD. Inhibition of hydrogen sulfide restores normal breathing stability and improves autonomic control during experimental heart failure. *J Appl Physiol (1985)* 2013: 114(9): 1141-1150.
33. Pijacka W, Moraes DJ, Ratcliffe LE, Nightingale AK, Hart EC, da Silva MP, Machado BH, McBryde FD, Abdala AP, Ford AP, Paton JF. Purinergic receptors in the carotid body as a new drug target for controlling hypertension. *Nature medicine* 2016: 22(10): 1151-1159.
34. Li Y, Ye J, Han D, Cao X, Ding X, Zhang Y, Xu W, Orr J, Jen R, Sands S, Malhotra A, Owens R. Physiology-Based Modeling May Predict Surgical Treatment Outcome for Obstructive Sleep Apnea. *J Clin Sleep Med* 2017: 13(9): 1029-1037.
35. Joosten SA, Leong P, Landry SA, Sands SA, Terrill PI, Mann D, Turton A, Rangaswamy J, Andara C, Burgess G, Mansfield D, Hamilton GS, Edwards BA. Loop Gain Predicts the Response to Upper Airway Surgery in Patients With Obstructive Sleep Apnea. *Sleep* 2017: 40(7).
36. Salloum A, Rowley JA, Mateika JH, Chowdhuri S, Omran Q, Badr MS. Increased propensity for central apnea in patients with obstructive sleep apnea: effect of nasal continuous positive airway pressure. *Am J Respir Crit Care Med* 2010: 181(2): 189-193.
37. Bamagoos AA, Cistulli PA, Sutherland K, Edwards BA, Eckert DJ, Hess LB, Wellman A, Sands SA. Phenotyping Using Polysomnography to Select Obstructive Sleep Apnoea Patients for Mandible Advancement Device Therapy [Abstract]. *Journal of Sleep Research* 2017: 26 (Suppl. 1): 5-33.

Acknowledgements

This work was supported by the American Heart Association (15SDG25890059 [S.A.S.]), the NIH (R01HL102321, R01HL128658, P01HL094307, and P01HL095491 [A.W. and S.A.S.]), the National Health and Medical Research Council of Australia (1053201 [S.A.S.], 1035115 [B.A.E.], and 1064163 [P.I.T., S.A.S., A.W., and B.A.E.]), the R. G. Menzies Foundation (S.A.S.), the American Thoracic Society Foundation (S.A.S.), and the Heart Foundation of Australia (Future Leader Fellowship 101167 [B.A.E.]). A.M. is PI on NIH R01HL085188, K24HL132105, T32HL134632 and Co-I on R21HL121794, R01HL119201, R01HL081823. This work was also supported by Harvard Catalyst (National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH Award UL1TR001102).

Tables

Table 1. Patient characteristics

Characteristic	All patients (N=36)	Responders* (N=9)	Non-Responders (N=27)	p-value¶
Demographics				
Age (years)	55±2	53±4	55±2	0.7
Sex (M:F)	26:10	6:3	20:7	0.7
Race (Black:White:Asian:Other)	9:25:0:1	5:3:0:0	4:22:0:1	0.015¶
Body mass index (kg/m^2)	31.1±0.7	32.3±1.2	30.6±0.8	0.3
Neck circumference (cm)	40.6±0.5	40.2±1.0	40.7±0.7	0.7
Systolic blood pressure† (mmHg)	136.8±2.4	134.9±3.7	137.5±2.9	0.6
Diastolic blood pressure† (mmHg)	80.7±1.9	79.7±3.8	81.1±2.2	0.8
Current use of anti-hypertensive medication, N (%)	12 (33)	1 (11)	11 (41)	0.2
Currently treated (CPAP:oral appliance:untreated)	12:2:22	1:0:8	11:2:14	0.06
Polysomnography#				
Apnea-hypopnea index, AHI (events/h)	57.9±22.1	56.6±7.7	58.3±4.3	0.8
Central events (% respiratory events)	4.7±2.0	8.5±5.7	3.4±1.8	0.3
Hypopneas (% respiratory events)	47.3±5.4	61.9±11.3	42.5±6.0	0.12
Arousals (% respiratory events)	88.8±4.9	79.1±5.2	92.1±6.2	0.3
Nadir oxygen saturation (%)	87.1±4.8	89.2±1.5	86.4±0.9	0.13
Stage 1 sleep (% total sleep time)	25.9±22.0	22.3±5.1	27.1±4.6	0.6

Values are mean±S.E.M. *Responders are defined by a ≥50% reduction in apnea-hypopnea index. ¶Student's t-tests were used for continuous variables and Fisher exact tests were used for categorical variables (including black versus not black, treated versus not treated). †Morning, supine, off treatment (sham night). #Polysomnography refers to the sham night; respiratory event and oxygenation data reflect supine non-REM sleep.

Table 2: Logistic regression model for predicting responses to oxygen therapy

Variable	β	SEM	odds ratio*	p	Interpretation
Constant	−1.97	1.02		0.01	
Loop gain	15.41	7.40	3.7	0.038	Higher loop gain→success
$V_{passive}$	5.27	3.71	4.8	0.15	Reduced collapsibility→success
Compensation	15.09	6.62	45.5	0.023	Greater compensation→success
$V_{passive}$ × Compensation	−58.53	29.97	0.11	0.036	Poor collapsibility & poor compensation→failure
Loop gain × Compensation	−80.34	34.16	0.17	0.019	Low loop gain & poor compensation→failure
Arousal threshold × Compensation	−86.43	29.53	0.012	0.003	Low arousal threshold & higher compensation→success

The Table describes the final regression results (6 terms) after backward stepwise elimination (p-to-remove=0.157) which began with four traits, their squares, and all interaction terms (full quadratic model, N terms = 14). SEM = standard error of the mean. *Odds ratio describes the increase in likelihood of being a responder per SD increase in each term. Traits were mean-subtracted before application to the regression model: mean Vpassive*=62.8%, mean loop gain [LG_n]=0.42, mean arousal threshold* = 157.6%, mean compensation = 6.1%. To promote normality, Vpassive and arousal threshold values were square-root transformed around 100% using $y=1+(x-1)^{0.5}$ and $y=1-(1-x)^{0.5}$ respectively (n.b. x=1 describes 100%). Patients were considered a "predicted responder" if Y = −1.97 + 15.41[Loop gain] + 5.27[$V_{passive}$] + 15.09[Compensation] – 58.53[$V_{passive}$ × Compensation] – 80.34[Loop gain × Compensation] – 86.43[Arousal threshold × Compensation] > −0.66 (use of this equation requires transformed, mean-subtracted traits). A simplified two-trait model is provided in the Supplement (Table S2, Figure S3).

Table 3: Effect of oxygen therapy on the physiological traits

Phenotypic Trait	Overall Effects (N=36)				Responders (N=9) versus Non-responders (N=26€)		
	Sham	Oxygen Therapy	Median Change	‡P-value	Median Change within Responders	Median Change within Non-responders	†P-value
Loop gain, LG_n (Instability)	0.41 [0.38 to 0.49]	0.38 [0.33 to 0.45]	−0.05 [−0.11 to 0.01]	0.026	−0.10# [−0.21 to −0.1]	−0.03 [−0.10 to 0.03]	0.16
Loop gain, LG_1 (Sensitivity)	0.63 [0.56 to 0.75]	0.46 [0.40 to 0.58]	−0.18 [−0.25 to −0.04]	<0.00001	−0.20* [−0.40 to −0.10]	−0.17*** [−0.24 to −0.03]	0.4
Delay (s)	10.1 [8.9 to 11.4]	12.8 [10.0 to 15.4]	+2.0 [0.9 to 4.4]	<0.00001	+1.6* [0.6 to 2.3]	+2.4*** [0.9 to 5.3]	0.3
Arousal threshold (%)	131 [116 to 151]	118 [109 to 150]	−9.8 [−20 to 2]	0.038	−9.8* [−21.6 to −2.2]	−9.5 [−17.4 to 10.7]	0.3
$V_{passive}$ (%) (¶Collapsibility)	91.5 [75.0 to 94.2]	87.3 [69.7 to 94.8]	0.0 [−11.2 to 4.0]	0.7	+0.6 [−0.7 to 3.8]	−2.1 [−13.4 to 4.0]	0.6
Compensation (%)	0 [-11.8 to 6.8]	−1.9 [-22.5 to 5.8]	−3.2 [−13.1 to 11.9]	0.5	−4.7 [−11.5 to 1.7]	−1.6 [−16.8 to 12.8]	0.9

Values are median [interquartile range]. ¶Higher values of $V_{passive}$ indicate less-severe collapsibility. ‡Wilcoxon Signed-Rank test. †Mann-Whitney U test. #P<0.1, *P<0.05, **P<0.01, ***P<0.001 Wilcoxon Signed-Rank test within responder/non-responder subgroups. €1 non-responder provided no data; no 7-min windows of non-REM sleep with no longer than 30-s of continuous wake were available for analysis.

Figure Legends

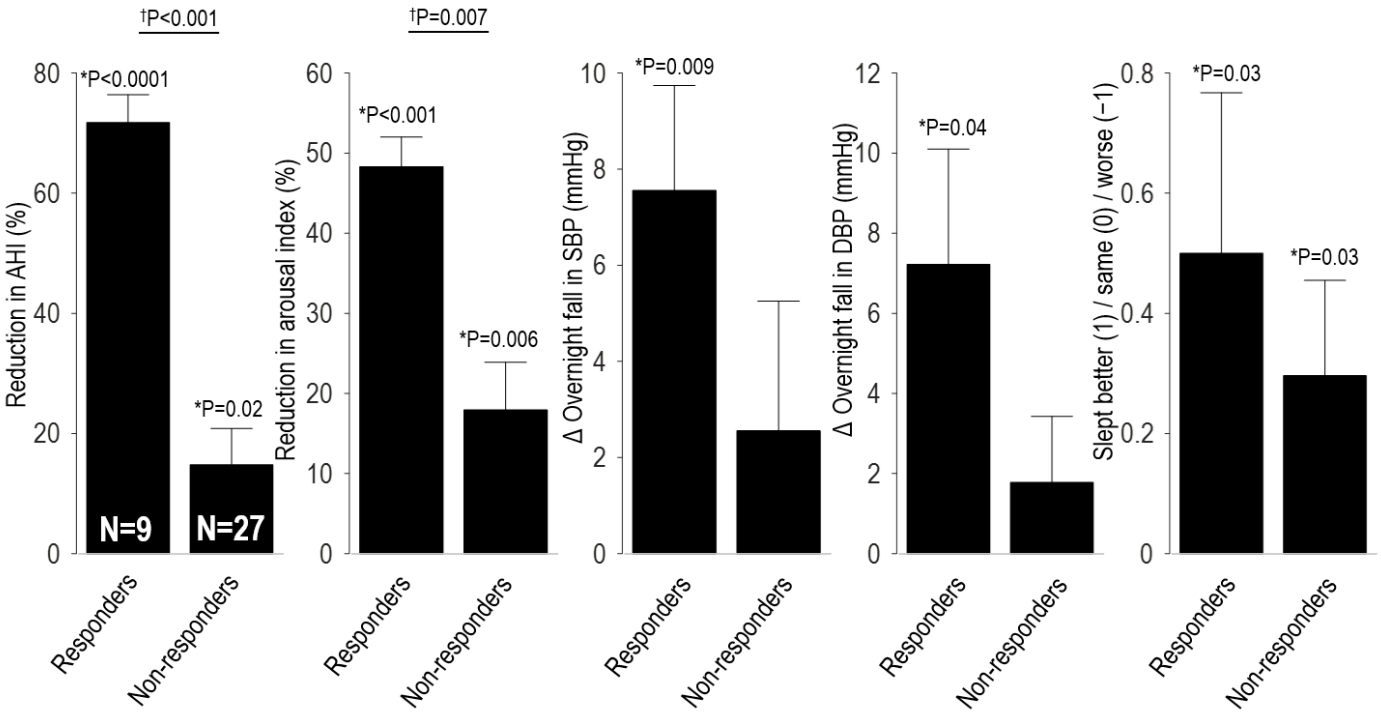
Figure 1. Effect of supplemental oxygen on primary and secondary outcomes in responders (N=9) and non-responders (N=27). In responders, improvements were observed in the apnea hypopnea index (AHI) by definition. In addition, responders exhibited improvements in the frequency of arousals from sleep (arousal index) as well as in blood pressure (change from evening to morning) and subjective sleep quality. There was no effect on the Stanford Sleepiness Scale (subjective morning alertness, not shown). *oxygen versus sham. †responders versus non-responders. SBP = systolic blood pressure. DBP = diastolic blood pressure. Error bars indicate SEM.

Figure 2. Example endophenotype data off treatment are shown for a responder (top, sham AHI=44.9, treatment AHI=9.7 events/hr) and a nonresponder (bottom, sham AHI=42.8, treatment AHI=45.2 events/hr). **Left:** Illustrative traces of sleep apnea and model estimation of ventilatory drive. Note that events are self-similar within a subject. In the responder, changes in ventilation track estimated ventilatory drive during obstructive events. By contrast, in the non-responder ventilation falls as ventilatory drive rises. **Right:** Summary plots of ventilatory versus ventilatory drive during sleep (black line: median, shading: IQR). The responder has a higher loop gain (LG_n), a lower ventilatory drive preceding arousal (arousal threshold) and less-severe collapsibility as inferred from the higher level of ventilation at normal ventilatory drive ($V_{passive}$). EEG = Electroencephalogram. Flow = square-root transformed nasal pressure. Th. and Ab. denote thoracic and abdominal excursions (piezoelectric respiratory belts); note signals are out of phase (paradox) during events in both subjects consistent with airflow obstruction. Ventilation and ventilatory drive are expressed as a proportion of the mean ventilation during the window ("eupnea"). Estimated ventilatory drive (green line, left) is shown partitioned into chemical drive (chemoreflex or "loop gain" contribution, black line) and the ventilatory response to arousal (arousal contribution, green minus black line).

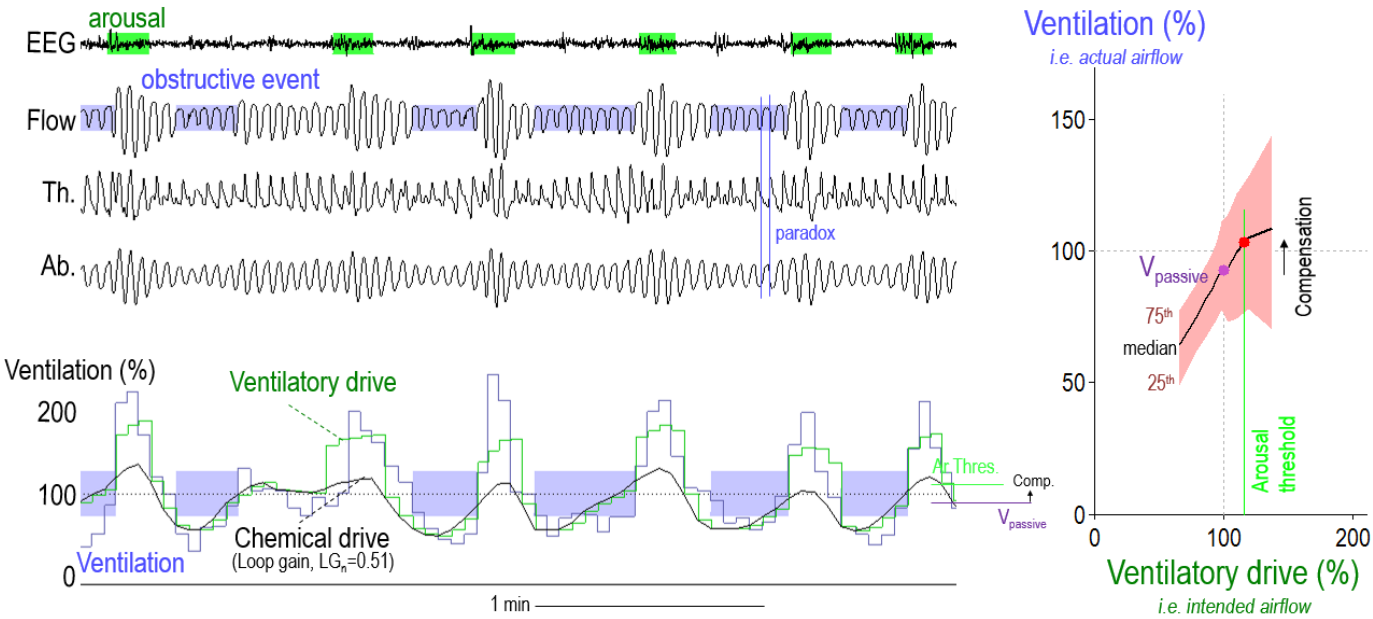
Figure 3. Predictive value of the endophenotypic traits causing OSA. Shading illustrates region of predicted responders and definition of high versus low for each trait subgroup. Bars illustrate the reduction in apnea-hypopnea index (AHI) with treatment in the high vs. low subgroups (mean±S.E.M., patients were assigned to subgroups using cross-validation). Loop gain LG_n indicates ventilatory instability i.e. the predisposition to spontaneous periodic breathing. Collapsibility, compensation and arousal threshold data are presented as a proportion of eupneic levels. See text for details. Note the y-axis scale is compressed below zero to faciliate visual interpretation of values above zero. Each trait had significant negative predictive value: Loop gain LG_n (reduction in AHI: 37.6±6.9% vs 14.0±11.5%; positive predictive value [PPV] = 35±10% [p=0.3], negative predictive value [NPV] = 92±7% [p=0.02]); Vpassive (49.3±6.9% vs 10.9±8.1; PPV = 47±12% [p=0.07], NPV = 95±5% [p<0.001]); Compensation (53.0±8.9% vs 18.5±7.2, PPV = 55±15% [p=0.049], NPV = 88±6% [p=0.046]); Arousal threshold (39.7±12.8% vs 22.2±6.0%; PPV = 50±13% [p=0.06], NPV = 91±6% [p=0.009]).

Figure 4. Multivariable analysis of the OSA traits. **(A-C)** 2-trait "slices" of the 4-trait regression model illustrate how the traits causing sleep apnea combine to predict responses to supplemental oxygen. Dots are individual patients (circles are patients from current study, squares are patients from Edwards et al [4]). Shading illustrates the regions of "predicted responders" (green) and "predicted non-responders" (red). Each 2-trait slice represents model predictions at constant values of the other 2 traits; data points that are far enough away from the slice such that the slice prediction does not match the overall model prediction (irrespective of correct/incorrect) are shown in light grey. **(D)** The continuous relationship between the reduction in AHI with oxygen and the regression model prediction is shown (Probability = $1/[1+e^{-Y}]$; see Table 2). Note the y-axis scale is compressed below zero to faciliate visual interpretation of values above zero.

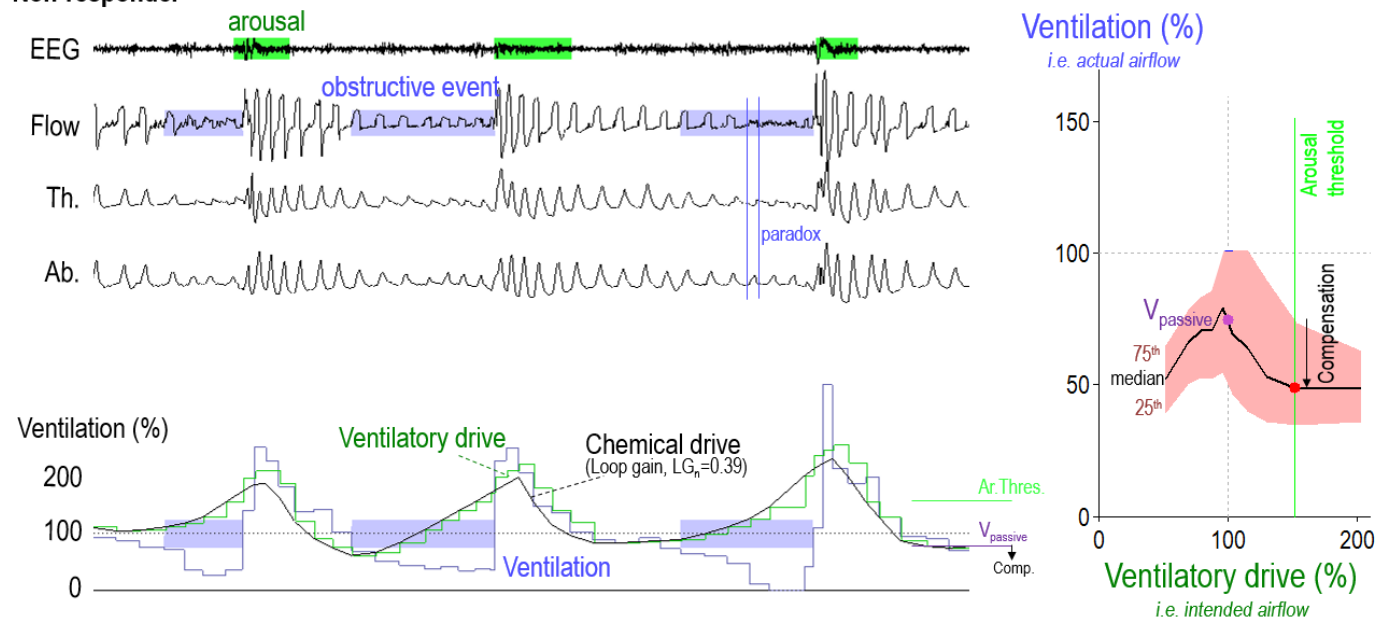
Figure 5. Effect of supplemental oxygen on primary and secondary outcomes in patients with suitable pathophysiology "predicted responders" (N=13) and patients with unsuitable pathophysiology "predicted non-responders" (N=23) defined based on endophenotypic traits (logistic regression, cross-validated; Table 2, Figure 4). In predicted responders, treatment led to an improvement in OSA severity (reduction in AHI); in contrast to Figure 1, differences between predicted responders/non-responders are not "by definition" because subgroups were assigned using only data from the other subjects (i.e. cross-validation). Predicted responders also exhibited improvements in the frequency of arousals from sleep, blood pressure (evening minus morning levels) and subjective sleep quality. There was no effect on the Stanford Sleepiness Scale (subjective morning alertness, not shown) in either subgroup. SBP = systolic blood pressure. DBP = diastolic blood pressure. Compare results with Figure 1. *oxygen versus sham. †responders versus non-responders.

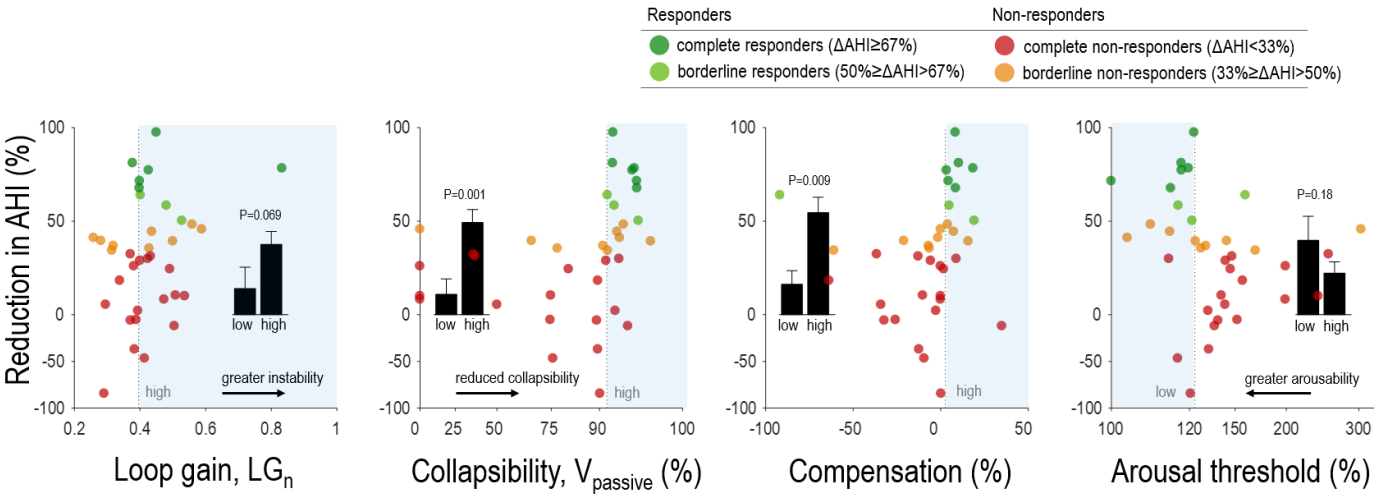

†P<0.001
†P=0.007
Reduction in AHI (%)
*P<0.0001
*P=0.02
N=9
N=27
Reduction in arousal index (%)
*P<0.001
*P=0.006
Δ Overnight fall in SBP (mmHg)
*P=0.009
Δ Overnight fall in DBP (mmHg)
*P=0.04
Slept better (1) / same (0) / worse (−1)
*P=0.03
*P=0.03
Responders
Non-responders

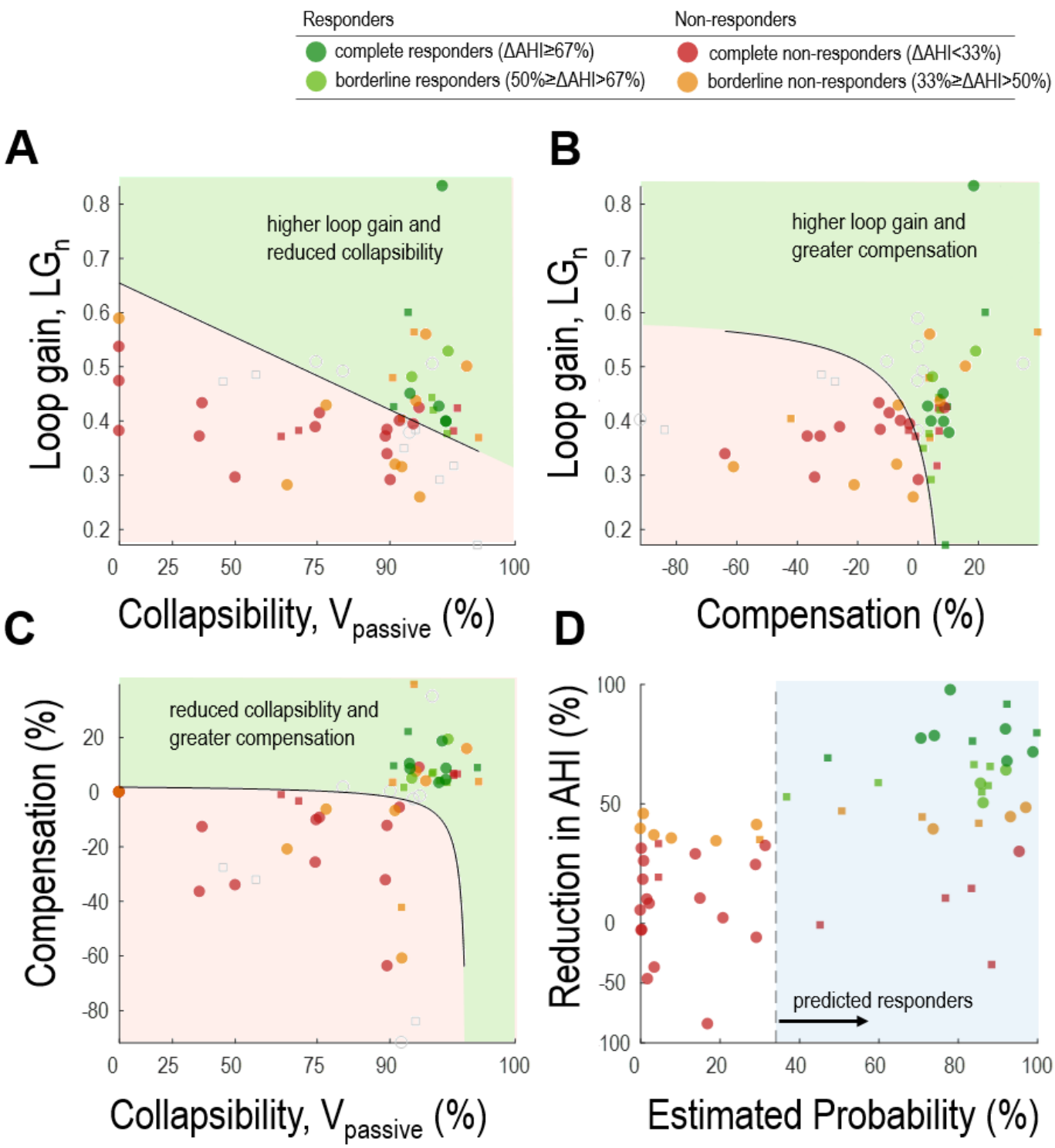
Responder

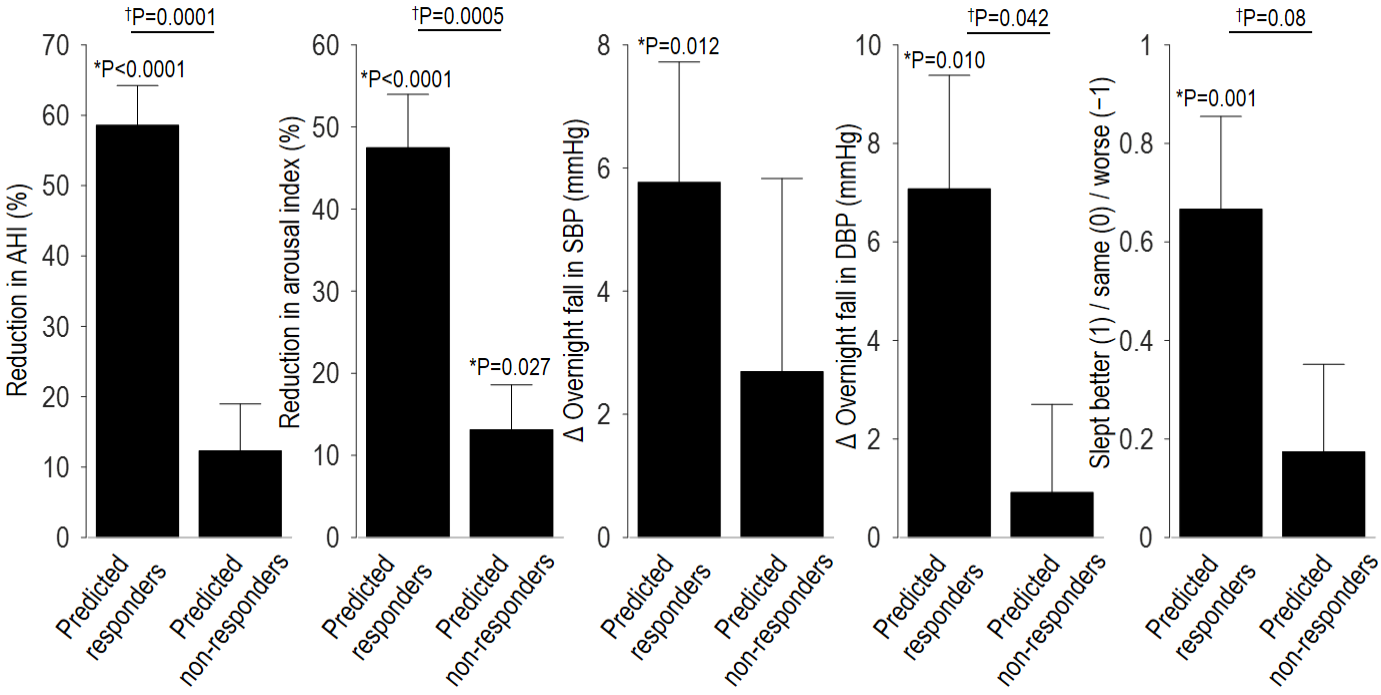


Non-responder




Responders
Non-responders
complete responders (ΔAHI≥67%)
borderline responders (50%≥ΔAHI>67%)
complete non-responders (ΔAHI<33%)
borderline non-responders (33%≥ΔAHI>50%)
Reduction in AHI (%)
P=0.069
low high
high
greater instability
Loop gain, LG$_n$
P=0.001
low high
reduced collapsibility
high
Collapsibility, V$_{passive}$ (%)
P=0.009
low high
high
Compensation (%)
P=0.18
low high
low
greater arousability
Arousal threshold (%)


Responders
Non-responders
complete responders (ΔAHI≥67%)
borderline responders (50%≥ΔAHI>67%)
complete non-responders (ΔAHI<33%)
borderline non-responders (33%≥ΔAHI>50%)
A
higher loop gain and
reduced collapsibility
Loop gain, LGn
Collapsibility, Vpassive (%)
B
higher loop gain and
greater compensation
Loop gain, LGn
Compensation (%)
C
reduced collapsiblity and
greater compensation
Compensation (%)
Collapsibility, Vpassive (%)
D
Reduction in AHI (%)
predicted responders
Estimated Probability (%)


†P=0.0001
*P<0.0001
Reduction in AHI (%)
70
60
50
40
30
20
10
0
Predicted responders
Predicted non-responders
†P=0.0005
*P<0.0001
*P=0.027
Reduction in arousal index (%)
60
50
40
30
20
10
0
Predicted responders
Predicted non-responders
*P=0.012
Δ Overnight fall in SBP (mmHg)
8
6
4
2
0
Predicted responders
Predicted non-responders
†P=0.042
*P=0.010
Δ Overnight fall in DBP (mmHg)
10
8
6
4
2
0
Predicted responders
Predicted non-responders
†P=0.08
*P=0.001
Slept better (1) / same (0) / worse (−1)
1
0.8
0.6
0.4
0.2
0
Predicted responders
Predicted non-responders

Online Data Supplement

Supplemental Methods

Participants

Detailed characteristics of participants are described in Table S1.

A minimum AHI of 20 events/hr was chosen to minimize the possibility that a clinically-important response to treatment (50% reduction in AHI) could occur by chance due to night-to-night variability (SD approximately 9 events/hr [S1]).

Power analysis

The study was powered (alpha = 0.05, power ~80%) to find a 1.0 SD difference in the response (by 33±33% percent reduction in AHI) between high loop gain and low loop gain subgroups (high: LG_1>0.7, i.e. "hypersensitivity") based on a prevalence of 1:2 (high:low). Ultimately the observed difference between groups was just +10.5±37.6% [mean±SD] (95%CI: -16 to 37%), a 0.28 SD difference; i.e. the best estimate of the difference was small and there was no more than a 37% greater reduction in patients with higher loop gain (using LG_1).

Percent change in AHI was chosen over absolute change in AHI because the percent change is typically least-strongly correlated with the baseline (sham) AHI (here: r=0.07 [p=0.7] versus r=0.4 [p=0.016] for absolute reduction). If we had used absolute reduction in AHI as the outcome variable, the observed difference would have been borderline non-significant (p=0.051), with a group difference of 13.3±18.9 events/hr [mean±SD] (95%CI: 0.1 to 26.6 events/hr). However, this difference can be explained by an increased baseline AHI in the high LG_1 group; after adjusting for baseline AHI the difference between groups became +6.0 events/hr (95%CI: -9.2 to 21.2 events/hr).

Power for multivariable analysis. Given the absence of appropriate existing data for a formal power analysis, we estimated that approximately (10×M)+10=50 subjects would be necessary (56 were used) to build a prediction model that would use at least M=4 terms. Robustness was assessed based on the loss in predictive value via cross-validation. We emphasize that the primary goal of the multivariable analysis was not to show that each trait contributes significantly to responses (there was no minimum detectable odds ratio). Rather, the objective was to define two subgroups that would have significantly different responses (after cross-validation); since these subgroups need to be powered to detect a difference in response, the power considerations are the same as for the initial analysis, i.e. 36 patients would provide ~80% power to detect a difference in the reduction in AHI by 33±33% (1 SD). Ultimately the difference was 46.3±30.3%

[mean±SD] (95%CI: 24.8 to 67.7 %), or 1.5 SD, which was largely unchanged after adjusting for baseline AHI (46.8%; 95%CI: 26.0 to 67.6%). The difference in absolute reduction in AHI between subgroups was also significant: 23.2±16.4 events/hr [mean±SD] (95%CI: 11.6 to 34.7 events/hr), or 1.4 SD; adjusting for baseline AHI had a minimal effect (24.2 events/hr; 95%CI: 14.6 to 33.8 events/hr).

Procedure

Studies were performed a week apart to facilitate between-study consistency of work and lifestyle factors that might contribute to sleepiness or blood pressure levels (e.g. exercise, diet, caffeine use). Participants were asked to keep routine medication use consistent between studies.

Medications for hypertension, when applicable, were administered at home on the morning prior to the overnight study, and then were not taken until after morning blood pressure measurements were made.

At arrival (~7pm), seated blood pressure measurements were made that served to familiarize participants with the measurement experience, reducing the chance of possible "first measurement" effects influencing the evening blood pressure values.

After study completion, patients were asked if they knew which night was oxygen and which was sham: 26% guessed correctly, 20% guessed incorrectly, and 54% were unsure (signed rank test $P=0.7$; correct=1, unsure=0, incorrect=−1) indicating that subjects were effectively blinded.

Ventilatory control tests (dynamic inspired CO_2) were also performed before and after sleep on both nights [S2]; data are not provided here to focus on polysomnographic predictors.

Polysomnographic setup

Care was taken to ensure high quality nasal pressure signals were recorded: a cannula without evidence of mechanical damping effects was selected [prongs 3.5 mm diameter] (Hudson RCI standard "over the ear" cannula, Teleflex, Morrisville NC). Cannulas were secured to the face with tape to minimize displacement (Tegaderm, 3M, Maplewood MN); signal amplification was DC coupled to preserve the baseline (Validyne, Northridge CA) and unfiltered signals were exported for analysis.

Hypopneas were scored based on a 30% reduction in airflow, avoiding the desaturation criterion given the use of supplemental oxygen.

An epiglottic pressure catheter (Millar Instruments, Houston TX) was used to adjudicate central versus obstructive hypopneas to confirm obstructive pathophysiology.

EEG arousals were scored using standard criteria (≥3-s change in EEG frequencies θ, α, β). All patients analyzed also had AHI>20 by standard hypopnea criteria (3% desaturation or arousal) [S3]. At baseline, events not associated with desaturation or arousal made up just 7.7±8.6% (mean±S.D.) of the scored events.

Quantifying the pathophysiological traits using polysomnography

Eupneic ventilation during OSA is inferred from the mean ventilation for each window of data on the basis that mean PCO_2 is not greatly deranged during this time. This assumption did not adversely affect chemical drive and loop gain measurements in our model simulations [S4]. Eupneic ventilation on CPAP also compares closely with the mean value of ventilation during sleep in patients with OSA [S5].

To construct each phenotypic summary plot of ventilation versus ventilatory drive during sleep, the following process was automated:

1. Values for ventilation and ventilatory drive were tabulated for each breath that appeared during windows of non-REM sleep [S4].
2. Breaths were also labelled based on whether or not a scored EEG arousal was present within the breath (from start inspiration to end expiration). Breaths within an arousal or ≤2 breaths after an arousal ended (after sleep onset) were excluded from analysis to minimize the possibility of including data influenced by wakefulness in the assessment of behavior during sleep.
3. Ventilatory drive data were sorted and divided into 10 groups or bins (deciles). For each decile, the median ventilation was measured and plotted against the median ventilatory drive for each decile.
4. Linear interpolation was used between deciles to find a) the value of ventilation at eupneic ventilatory drive ($V_{passive}$), and b) the value of ventilation at the arousal threshold (called V_{active}); compensation is given by V_{active} minus $V_{passive}$.

Definition of predictive model

The term "model" here is used to indicate a classifier plus the necessary coefficients/cutoffs for predicting responders/non-responders: Univariable models consist of a cutoff alone (threshold). Multivariable models comprise a set of selected features (phenotypic variables), a set of coefficients, as well as a cutoff. In all cases (univariable and multivariable), we sought to maximize sensitivity and specificity [S6]. Also in all cases, we employed leave-one-out cross validation to provide generalizable measures of performance.

Assessment of predictive value

Cross-validation. When assessing the performance or predictive value of a model (defined above) that has been developed (trained) on available data, it is best practice to use unseen data for model validation (testing) to prevent over-estimation of the predictive value for future applications. With the modest sample size available in our study (i.e. N=9 responders), use of a fully-separate dataset for development and validation was considered inefficient use of available data. Rather, we used a common procedure called (leave-one-out) "cross-validation". This procedure was used throughout the study for univariable and multivariable analyses (except for LG_1, the *a priori* primary predictor).

The procedure first involves developing a predictive model using all patients. To test the performance of this model, the entire process of developing the model was repeated but using all subjects except one who was left out. This "modified" predictive model was then tested on the unseen individual who was left out; we recorded the outcome of the prediction (true positive, false negative, etc.). This leave-one-out process was then repeated N times (here, N=36). A new modified model was developed each time a new individual was left out. Each time the model changes slightly, but we can be certain of the independence of the development and validation data. Of note, the actual model presented is that which is based on all subjects, since this is the best model we can present for future use based on all available data.

Multivariate model analysis

We chose logistic regression as a simple "machine learning" tool that is easily interpreted. Quadratic model terms were used given the observation of interactions between variables, i.e. a simpler linear model was not sufficient to explain responses. The process was designed to be simple and transparent. In brief, the process for identifying the model was as follows:

1. Terms are initially included in the model: all variables (N=4), their squares (N=4) and interaction terms (N=6). The number of terms starts with M=14.
2. A logistic regression model was fit to the data using M terms.
3. The term with the highest p-value (Wald test) was removed (if p>0.157) [S7-9], and Step 2 was repeated for the remaining terms.
4. Once no further terms were removed, a logistic regression model cutoff was selected to maximize sensitivity plus specificity (receiver operating characteristic analysis) [S10].

Additional considerations. Because two highly-correlated measures of loop gain were available (LG_1, LG_n), we tested the model performance with LG_1 and LG_n separately; LG_n was consistently the most predictive of these two variables and therefore chosen over LG_1. Vpassive was forced into the model because of (1) expert knowledge that collapsibility should contribute to responses [S11], and (2) its removal varied the model coefficients (betas) for other key traits (loop gain and compensation) considerably (by >25%). Model weights were used to balance the influence of patients per subgroup. To estimate the performance of this model when applied to unseen data, we repeated the above procedure using leave-one-out cross-validation (described above).

Including additional published data to build a robust multiple logistic regression model. Data from a previous study [S12] were used to help build the multiple regression model, but we did not seek to test outcomes in the additional individuals. Hence, during cross-validation, we used N=55 (20+36-1) patients to develop a regression model to predict the outcome for each of the 36 patients in the current study. There were, however, some differences in study design between the current and previous one: Edwards et al used the same inspired oxygen concentration as the current study but in fact tested the combination of oxygen and 3 mg eszopiclone versus sham/placebo: However, we argue that it is likely that eszopiclone had relatively little impact on the AHI in that study. Recent data [S13] illustrated that a similar dose of zopiclone (i.e. 3.75 mg of eszopiclone, plus 3.75 mg of its inactive stereoisomer) had no impact on AHI overall and none of the 8 patients with AHI>20 events/hr exhibited more than a 20% reduction in AHI with this treatment.

Sensitivity analysis also proved that the additional data from Edwards et al. were useful in building a robust multivariable regression model. Without these additional data, (1) the quadratic model was underspecified and could not be used, (2) a linear model identified all four traits as contributors but no parameter was significant suggesting findings may not be robust (indeed poor compensation tended to predict a positive outcome, which is likely to be erroneous), (3) the linear model performance was similar before cross-

validation but after cross-validation was slightly weaker (ΔAHI = 53±8% versus 15±7%, p=0.002; PPV = 54±14% [p=0.04], NPV = 91±6% [p<0.006], accuracy = 78±6% [p=0.003]) indicating reduced robustness compared with the inclusion of the additional data. However, findings relating to improved secondary outcomes in the predicted responder subgroup were all upheld.

Statistical analysis

Statistical analyses were performed using MATLAB (Statistics and Machine Learning Toolbox, Mathworks, Natick MA, USA).

$V_{passive}$ and arousal threshold data failed normality tests; their skewness were therefore minimized using square root transforms centered around the value of 100% (see manuscript for equations).

Adjustments were not made for multiple secondary outcomes; all outcomes assessed were presented regardless of significance. Exploratory outcomes that were significantly improved (e.g. percentage time in stage 1 non-REM sleep; Table S1) were not emphasized.

Use of clinical variables. A variety of clinical variables are available from which one might potentially build a separate predictive model that does not require the use our endophenotype traits. While we consider that such an effort would be highly-valuable, we caution that the use of endophenotype traits has a distinct advantage: there is a highly-plausible mechanistic basis for the association with the response to treatment. The use of patient characteristics that have little-to-no mechanistic basis is challenging statistically (spurious associations are expected when using a large number of variables in a relatively small dataset) and thereby requires a far greater number of patients. There is also the concern that any change in population characteristics (e.g. age, race) would likely require recalibration of the predictive model.

Supplemental References

S1. 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.

S2. Sands SA, Mebrate Y, Edwards BA, Nemati S, Manisty CH, Desai AS, Wellman A, Willson K, Francis DP, Butler JP, Malhotra A. Resonance as the Mechanism of Daytime Periodic Breathing in Patients with Heart Failure. *Am J Respir Crit Care Med* 2016.

S3. Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, Marcus CL, Mehra R, Parthasarathy S, Quan SF, Redline S, Strohl KP, Davidson Ward SL, Tangredi MM, American Academy of Sleep M. 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.

S4. Terrill PI, Edwards BA, Nemati S, Butler JP, Owens RL, Eckert DJ, White DP, Malhotra A, Wellman A, Sands SA. Quantifying the ventilatory control contribution to sleep apnoea using polysomnography. *Eur Respir J* 2015: 45(2): 408-418.

S5. Sands SA, Terrill PI, Edwards BA, Taranto Montemurro L, Azarbarzin A, Marques M, de Melo CM, Loring SH, Butler JP, White DP, Wellman A. Quantifying the Arousal Threshold Using Polysomnography in Obstructive Sleep Apnea. *Sleep* 2018: 41(1).

S6. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950: 3(1): 32-35.

S7. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. *Source code for biology and medicine* 2008: 3: 17.

S8. Heinze G, Dunkler D. Five myths about variable selection. *Transplant international : official journal of the European Society for Organ Transplantation* 2017: 30(1): 6-10.

S9. 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.

S10. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics* 2006: 7: 91.

S11. 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.

S12. Edwards BA, Sands SA, Owens RL, Eckert DJ, Landry S, White DP, Malhotra A, Wellman A. 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.

S13. Carter SG, Berger MS, Carberry JC, Bilston LE, Butler JE, Tong BK, Martins RT, Fisher LP, McKenzie DK, Grunstein RR, Eckert DJ. Zopiclone Increases the Arousal Threshold without Impairing Genioglossus Activity in Obstructive Sleep Apnea. *Sleep* 2016: 39(4): 757-766.

Supplemental Table

Table S1. Characteristics and Impact of Treatment

Characteristic	All patients (N=36)		Responders* (N=9)		Non-Responders (N=27)		Predicted Responders** (N=13)		Pred. Non-Responders (N=23)	
	Sham	Oxygen	Sham	Oxygen	Sham	Oxygen	Sham	Oxygen	Sham	Oxygen
Demographics										
Age (years)	55±2		53±4		55±2		53±3		55±3	
Sex (M:F)	26:10		6:3		20:7		8:5		18:5	
Race (black:white:asian:other)	9:25:0:1		5:3:0:0		4:22:0:1 ¶ ∆		7:6:0:0		2:19:0:1 ¶¶ ∆	
Body mass index (kg/m^2)	31.1±0.7		32.3±1.2		30.6±0.8		31.6±1.0		30.7±0.9	
Neck circumference (cm)	40.6±0.5		40.2±1.0		40.7±0.7		39.9±0.8		41.0±0.7	
Current treatment (CPAP:oral appliance:untreated)	12:2:22		1:0:8		11:2:14		4:0:9		8:2:13	
Medications										
Anti-hypertensives, N (%)	12 (33)		1 (11)		11 (41)		3 (21)		9 (41)	
Proton pump inhibitors, N (%)	5 (14)		1 (11)		4 (15)		1 (7)		4 (18)	
Statins	4 (11)		1 (11)		3 (11)		1 (7)		3 (14)	
Antidepressants/anti-anxiety	4 (11)		1 (11)		3 (11)		1 (7)		3 (14)	
Aspirin	3 (8)		0 (0)		3 (11)		1 (7)		2 (9)	
Levothyroxine	3 (8)		0 (0)		3 (11)		0 (0)		3 (14)	
Zolpidem	1 (3)		0 (0)		1 (4)		0 (0)		1 (5)	
Metformin	1 (3)		0 (0)		1 (4)		0 (0)		1 (5)	
Polysomnography										
Time in bed (min)	421±8	422±11	416±16	446±20	423±10	414±12	421±13	424±17	421±11	421±14
Apnea-hypopnea index† (events/hr)	57.9±3.7	40.5±3.8	56.6±7.7	17.6±4.6	58.3±4.3	48.1±3.9	56.1±5.7	23.9±4.0	58.9±4.9	49.8±4.6
Effect of oxygen (%)	−29.0±6.2 ###		−71.8±4.6 ¶¶¶ ###		−14.8±4.9 #		−58.6±5.6 ¶¶¶ ###		−12.3±7.2	
Arousal index† (events/hr)	50.3±3.7	35.9±3.4	46.0±8.1	23.1±3.9	51.7±4.2	40.1±4.0	47.1±6.0	22.5±3.3	52.0±4.7	43.4±4.3
Effect of oxygen (%)	−25.5±5.0 ###		−48.3±3.7 ¶¶ ###		−17.9±6.1 ##		−47.5±6.5 ¶¶¶ ###		−13.1±5.5 #	
Nadir oxygen saturation (%Hb)	87.1±0.8	97.1±0.4	89.2±1.5	97.9±0.5	86.4±0.9	96.9±0.6	88.2±1.4	97.5±0.6	86.5±1.0	96.9±0.6
Effect of oxygen (%Hb)	10.0±0.8 ###		8.7±1.5 ###		10.4±0.9 ###		9.2±1.0 ###		10.4±1.0 ###	
Stage 1 sleep (% total sleep time)	25.9±3.7	23.8±3.7	22.3±5.1	12.4±4.3	27.1±4.6	27.7±4.5	20.1±4.5	9.6±2.4	29.2±5.1	31.9±4.9
Effect of oxygen (%total sleep time)	−0.3[−8.1 to 3.7]		−10.5[−14.5 to −1.3] ¶		+0.8[−4.8 to 10.6]		−7.2[−14.5 to −0.1] ¶¶ #		1.1[−1.9 to 13.8]	
Additional outcomes										
ΔSystolic blood pressure‡ (mmHg)	+3.0±1.9	−0.8±1.4	+3.4±2.7	−4.1±1.8	+2.9±2.3	+0.3±1.7	+3.2±3.1	−2.5±2.2	+2.9±2.4	+0.2±1.8
Effect of oxygen (mmHg)	−3.8±2.1		−7.6±2.2 ##		−2.6±2.8		−5.8±2.0 #		−2.7±3.1	
ΔDiastolic blood pressure‡ (mmHg)	+4.1±1.5	+0.9±0.9	+6.6±2.8	−0.6±1.3	+3.2±1.8	+1.4±1.1	+6.4±2.5	−0.7±1.2	+2.7±1.9	+1.8±1.2
Effect of oxygen (mmHg)	−3.1±1.5 #		−7.2±2.9 #		−1.8±1.6		−7.1±2.3 ¶ ##		−0.9±1.8	
Slept Better:Same:Worse on oxygen ††	19:9:7 #		5:2:1		14:7:6		9:2:1 ¶ ###		10:7:6	
Alertness, Stanford Sleepiness Scale £	2.0±0.2	2.1±0.2	2.3±0.6	2.3±0.5	2.0±0.2	2.1±0.2	2.2±0.4	2.4±0.4	1.9±0.2	2.0±0.2
Effect of oxygen (points)	0.1±0.2		0.0±0.5		0.1±0.2		0.2±0.4		0.0±0.2	
OSA severity, alternate										
AHI standard scoring, supine non-REM (events/hr)	54.2±3.9	34.8±3.7	49.8±7.6	14.7±4.2	55.6±4.6	41.4±3.9	49.8±5.8	19.6±3.8	56.6±5.1	43.3±4.5
Effect of oxygen (%)	−36.2±5.7 ###		−73.9±4.8 ¶¶¶ ###		−23.6±5.7 ###		−62.8±5.3 ¶¶¶ ###		−21.0±6.6 ##	
AHI standard scoring, all states/positions	52.5±3.7	33.5±3.3	44.7±5.6	16.1±3.5	55.1±4.5	39.3±3.6	46.1±4.8	20.9±3.5	56.1±5.0	40.6±4.1
Effect of oxygen (%)	−37.5±4.4 ###		−66.1±5.5 ¶¶¶ ###		−28.0±4.3 ###		−56.8±4.9 ¶¶¶ ###		−26.7±5.2 ###	

Values are mean±S.E.M. or median[interquartile range]. *Responders are defined by a ≥50% reduction in apnea-hypopnea index. **Predicted responders are based on the cross-validated logistic regression model analysis. †Reported during non-REM supine

sleep. ‡Morning minus evening values are taken to reflect sleep apnea burden (supine). For 'Polysomnography' and 'Additional outcomes', statistical comparisons are shown for the "Effect of oxygen'. ¶P<0.05, ¶¶P<0.01, ¶¶¶P<0.001 responders versus non-responders. #P<0.05, ##P<0.01, ###P<0.001 oxygen vs sham. ΔFisher exact test (Black vs not Black). ††Not collected in one individual (responder) due to >1 month between studies (rescheduling difficulties); statistical differences were compared using ranks: Better=+1, Same=0, Worse=−1. £Taken >30 mins after lights on. Medication use was unchanged prior to each overnight study and there were no statistically differences between subgroups; antihypertensives included hydrochlorothiazide, lisinopril, losartan, labetolol, atenolol, amplodipine, verapamil, doxasozin; antidepressants and anti-anxiety medications included selective serotonin reuptake inhibitors and aripiprazole. "Standard scoring" denotes the definition of hypopneas based on ≥3% oxygen desaturation or arousal.

Table S2: Two-trait simplified logistic regression model for predicting responses to oxygen therapy

Variable	β	SEM	odds ratio*	p	Interpretation
Constant	−0.23	0.41		0.6	
Loop gain	9.73	4.89	2.7	0.046	Higher loop gain→success
$V_{passive}$	7.24	2.54	6.0	0.004	Reduced collapsibility→success
Loop gain × $V_{passive}$	-32.1	6.62	0.43	0.14	Low loop gain and greater collapsibility→failure

The Table describes the results (3 terms) after backward stepwise elimination (p-to-remove=0.157) which began with the two key traits (loop gain [LG_n], $V_{passive}$), an interaction term (included but not significant), and two squared terms (excluded since p>0.157). SEM = standard error of the mean. *Odds ratio describes the increase in likelihood of being a responder per SD increase in each term. Traits were mean-subtracted before application to the regression model: mean Vpassive*=62.8%, mean loop gain [LG_n]=0.42. To promote normality, Vpassive values were square-root transformed around 100% using $y=1+(x-1)^{0.5}$ (n.b. x=1 describes 100%). Patients were considered a "predicted responder" here if $Y = -0.23 + 9.73[\text{Loop gain}] + 7.24[V_{passive}] - 32.1[\text{Loop gain} \times V_{passive}] > 0.25$ (use of this equation requires transformed, mean-subtracted traits). The model included data from Edwards et al. [S12] such that N=56 (36+20). Predictive value (cross-validated) for patients in the current study (N=36): (ΔAHI = 53±7% in predicted responders versus 10±7% in predicted non-responders [p=0.0002]; PPV = 56±12% [p=0.01], NPV = 100±0% [p<0.0001], accuracy = 81±7% [p<0.0001]).

Supplemental Figures

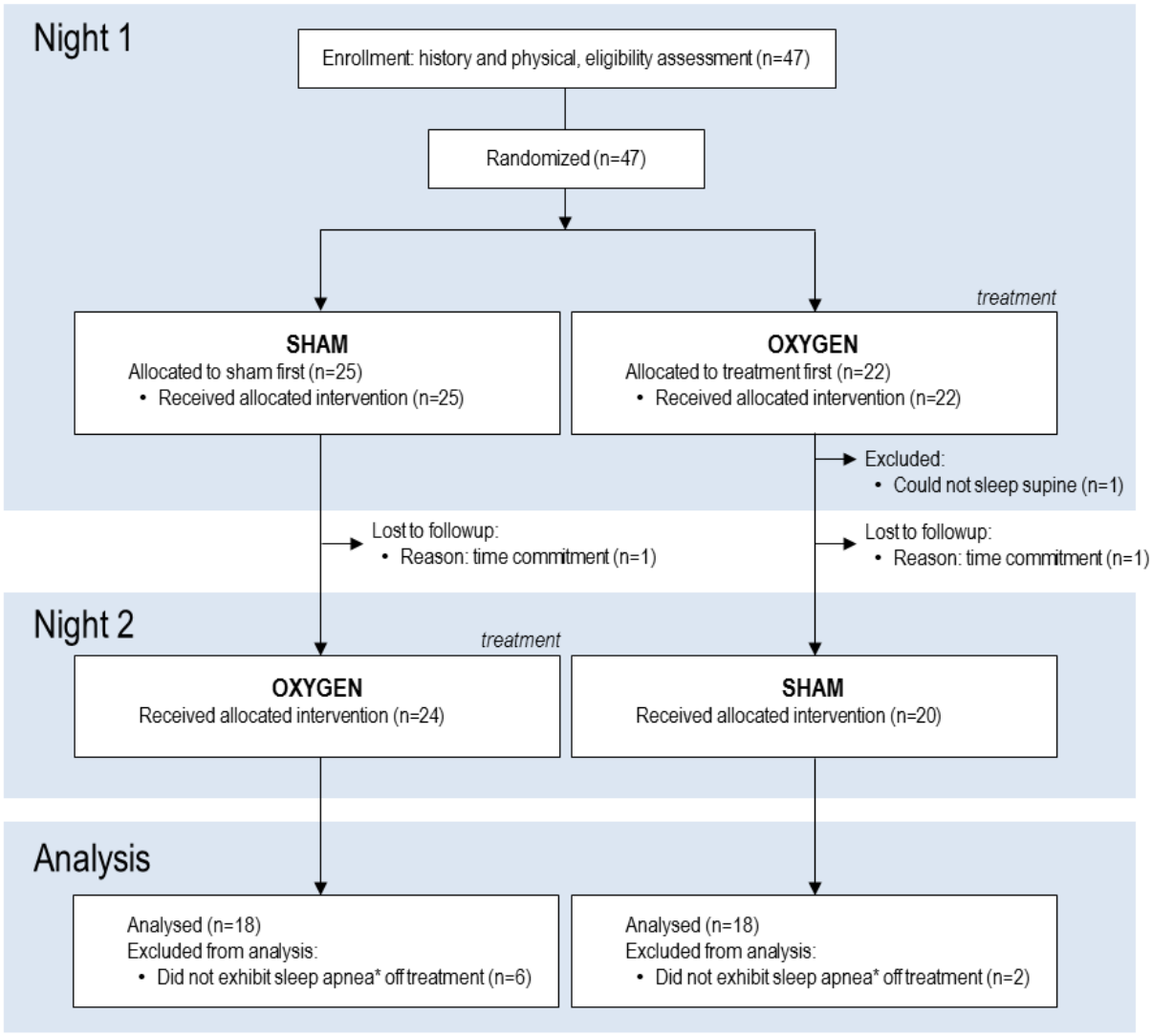


Figure S1. Study flow diagram. 47 patients with diagnosed OSA were randomized to either the sham first or treatment first arms. Randomization was performed using a computer random number generator in blocks of 2. As patients were excluded, new patients filled their slots to ensure equal group sizes for analysed data. Overall, 44 patients completed the study, but 8 patients did not have OSA on their sham study night (*criterion: non-REM AHI>20 events/hour) and therefore could not contribute data for analysis. By design, analysis was *per protocol* rather than *intention to treat*; sham night polysomnograms provided baseline data to measure phenotypic traits for categorizing patients into subgroups as well as for assessing the change in OSA severity (apnea-hypopnea index, AHI) with treatment. Of note, the goal was not to assess the effect of oxygen on OSA in unselected patients *per se*; rather it was to assess the relative reduction in AHI between phenotypic subgroups.

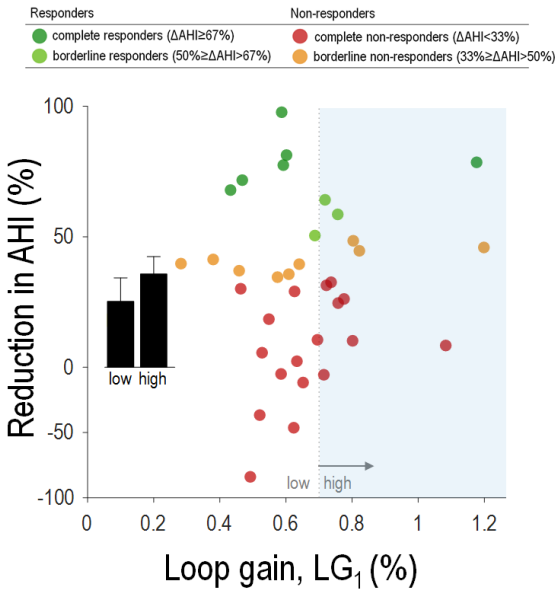


Figure S2. Contrary to our primary hypothesis, patients with high versus low loop gain based on LG_1 (pre-specified cutoff = 0.7, shading illustrates "high") did not show a significantly greater response to supplemental oxygen (reduction in apnea-hypopnea index AHI on treatment versus sham). Bars illustrate the reduction in AHI with treatment in the high vs. low subgroups. LG_1 is the magnitude of the chemoreflex ventilatory drive response to a 1 cycle/min swing in ventilation.

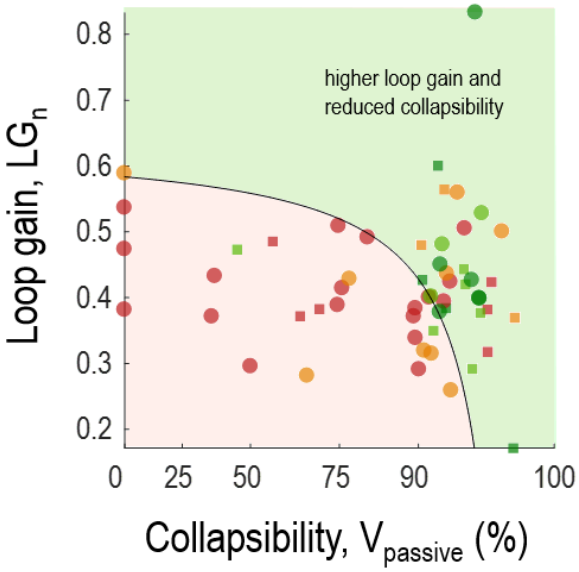


Figure S3. Two-trait simplified model confirming that loop gain and collapsibility ($V_{passive}$) can be combined to predict responses to oxygen therapy. Dots are individual patients (circles are patients from current study, squares are patients from Edwards et al [S12]); colors are consistent with figures in the main manuscript. Shading illustrates the regions of "predicted responders" (green) and "predicted non-responders" (red). See Table S2 for the equation for the logistic regression line.

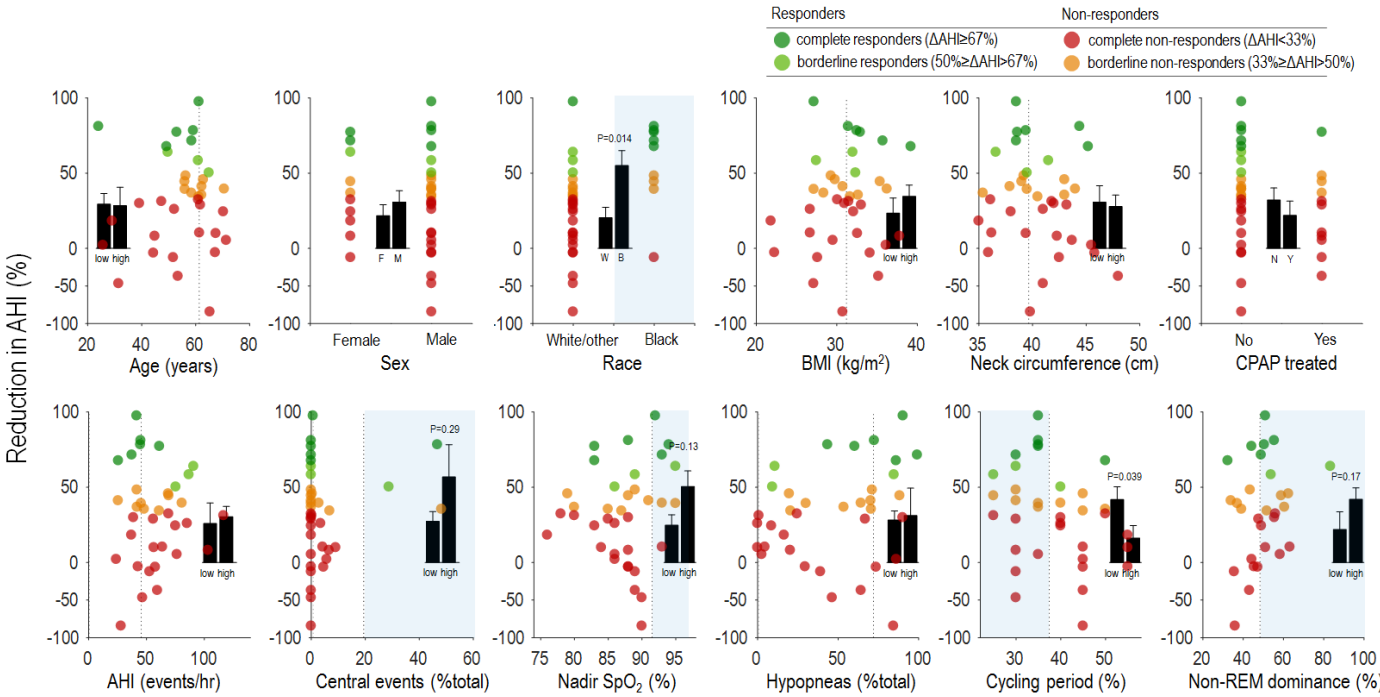


Figure S4. Clinical and other polysomnographic factors and the response to supplemental oxygen. Dashed vertical lines illustrate the optimal cutoffs. Bars illustrate the reduction in apnea-hypopnea index (AHI) with treatment in the subgroups. P-values>0.3 are not shown. There were no very strong predictors of the response to supplemental oxygen. Notably, however, black race significantly predicted a stronger response to treatment, which has not been reported previously. In addition, a faster cycling period (most common time from the end of one respiratory event to the end of the next, i.e. *mode*) was also a significant predictor. Non-significant trends were observed for a greater proportion of central events, a higher nadir oxygen saturation (SpO_2), and a greater non-REM dominance of OSA ($AHI_{non\text{-}REM}/[AHI_{non\text{-}REM}+AHI_{REM}]$; 0 = REM exclusive OSA, 100% = non-REM exclusive OSA, 50% = same OSA severity in non-REM and REM). BMI = body mass index, CPAP = continuous positive airway pressure, REM = rapid eye movement sleep. Note that p-values presented are not adjusted for multiple comparisons and variables were not proposed *a priori* as putative predictors.

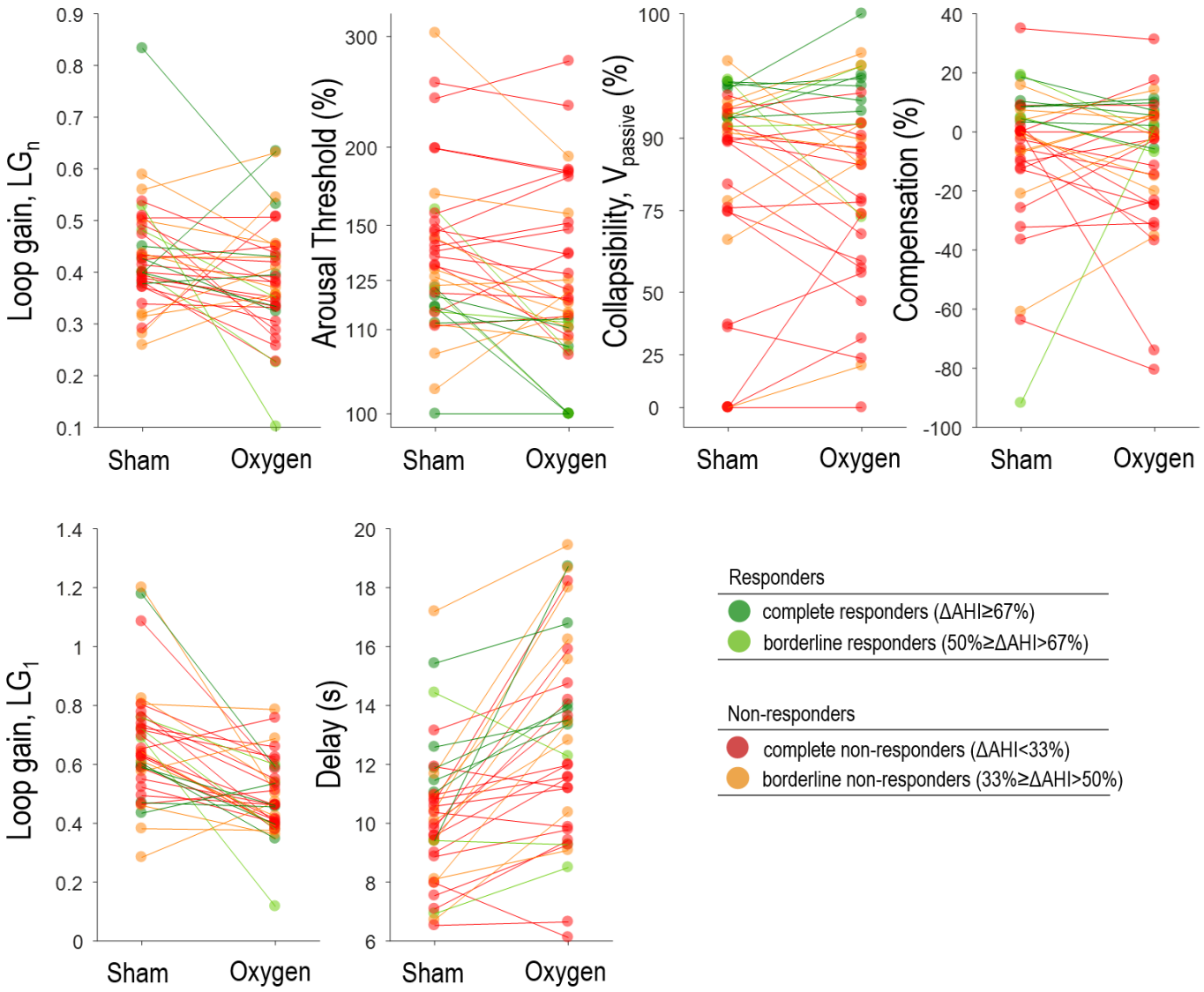


Figure S5. Effect of oxygen therapy on the physiological traits. Summary data are shown in Table 3. *Top:* The four traits causing sleep apnea are shown on sham and on oxygen therapy. Loop gain (LG_n, *instability*) was reduced, consequent to a reduction in feedback sensitivity (LG_1, *Bottom*), and was counteracted somewhat by an increase in estimated delay (*Bottom*). Arousal threshold was also slightly reduced with oxygen, possibly a direct physiological effect of oxygen, but could potentially be consequent to the improvement in sleep apnea severity. There was no evidence of a change in collapsibility or compensation with intervention. 35 patients contributed to these data; 1 individual had insufficient data on oxygen therapy for analysis.