pii: jc-18-00090 http://dx.doi.org/10.5664/jcsm.7484

JCSM
Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Awake Multimodal Phenotyping for Prediction of Oral Appliance Treatment Outcome

Kate Sutherland, PhD[1,2,3]; Andrew S.L. Chan, MD, PhD[1,2]; Joachim Ngiam, BDS(Hons), MSD, MPhil[1,2]; Oyku Dalci, PhD[4]; M. Ali Darendeliler, PhD[4]; Peter A. Cistulli, MD, PhD[1,2,3]

[1]Centre for Sleep Health and Research, Department of Respiratory and Sleep Medicine, Royal North Shore Hospital, Northern Sydney Local Health District, Sydney, Australia; [2]Sydney Medical School, University of Sydney, Sydney, Australia; [3]Charles Perkins Centre, University of Sydney, Sydney, Australia; [4]Discipline of Orthodontics, Faculty of Dentistry, University of Sydney, Sydney Dental Hospital, Sydney Local Health District, Sydney, Australia

Study Objectives: An oral appliance (OA) is a validated treatment for obstructive sleep apnea (OSA). However, therapeutic response is not certain in any individual and is a clinical barrier to implementing this form of therapy. Therefore, accurate and clinically applicable prediction methods are needed. The goal of this study was to derive prediction models based on multiple awake assessments capturing different aspects of the pharyngeal response to mandibular advancement. We hypothesized that a multimodal model would provide robust prediction.

Methods: Patients with OSA (apnea-hypopnea index [AHI] > 10 events/h) were recruited for treatment with a customized OA (n = 142, 59% male). Participants underwent facial photography (craniofacial structure), spirometry (mid-inspiratory flow at 50% vital capacity [MIF_{50}] and mid-expiratory flow at 50% vital capacity [MEF_{50}] and the ratio MEF_{50}/MIF_{50}) and nasopharyngoscopy (velopharyngeal collapse with Mueller maneuver and mandibular advancement). Treatment response was defined by 3 criteria: (1) AHI < 5 events/h plus ≥ 50% reduction, (2) AHI < 10 events/h plus ≥ 50% reduction, (3) ≥ 50% AHI reduction. Multivariable regression models were used to assess predictive utility of phenotypic assessments compared to clinical characteristics alone (age, sex, obesity, baseline AHI).

Results: Craniofacial structure and flow-volume loops predicted treatment response. Accuracy of the prediction models (area under the receiver operating characteristic curve) for each criterion were 0.90 (criterion 1), 0.79 (criterion 2), and 0.78 (criterion 3). However, these prediction models including phenotypic assessments did not provide a statistically significant improvement over clinical predictors only.

Conclusions: Multimodal awake phenotyping does not enhance OA treatment outcome prediction. These office-based, awake assessments have limited utility for robust clinical prediction models. Future work should focus on sleep-related assessments.

Commentary: A commentary on this article appears in this issue on page 1837.

Clinical Trial Registration: Registry: Australian New Zealand Clinical Trials Registry, Title: Multimodal phenotyping for the prediction of oral appliance treatment outcome in obstructive sleep apnoea, Identifier: ACTRN12611000409976, URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=336663

Keywords: obstructive sleep apnea, oral appliance therapy, mandibular advancement, treatment outcome prediction, sex

Citation: Sutherland K, Chan AS, 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.

BRIEF SUMMARY

Current Knowledge/Study Rationale: Predicting treatment response to oral appliance therapy in patients with sleep apnea has been an ongoing challenge and robust prediction methods are needed. A variety of simple awake prediction tests have been proposed but not successfully validated. This study aimed to combine multiple simple phenotypic assessments into a single prediction model.

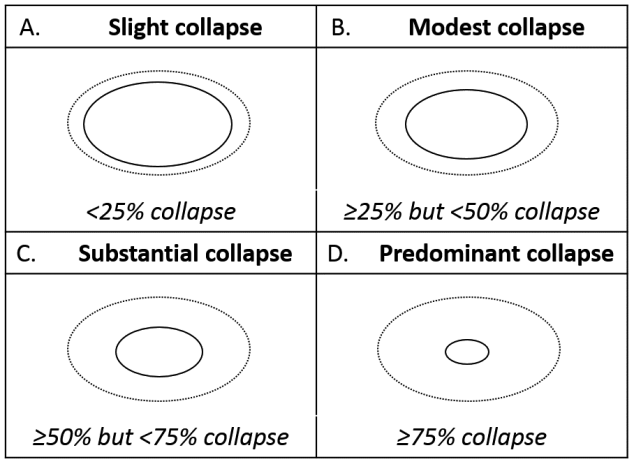
Study Impact: Our derived prediction models suggest that combined awake phenotyping tests do not improve prediction compared to clinical characteristics alone. This work suggests that robust prediction of oral appliance treatment response may need to include information from sleep assessments.

INTRODUCTION

Obstructive sleep apnea (OSA) is a chronic sleep disorder and is increasing in prevalence.[1] OSA is associated with a range of sequelae, including daytime symptoms and cardiometabolic and neurocognitive dysfunction.[2,3] Hence, OSA requires lifelong treatment adherence to prevent morbidity. First-line therapy, positive airway pressure (PAP), is effective but long-term adherence is often suboptimal.[4] An oral appliance (OA) is an alternative therapy that holds the mandible in a protruded position in order to enlarge the airway. An OA is generally used as a second-line treatment because of individual variations in efficacy.[5] Approximately one-third of patients will show minimal improvement in OSA with OA therapy.[6] Inability to predict which patients will not respond to OA therapy is a clinical barrier and reliable treatment prediction methods are needed.

A range of prediction methods for OA treatment outcome have been proposed and have been recently reviewed.[7] Some

Figure 1—Nasopharyngoscopic evaluation of the velopharyngeal response to the Mueller maneuver during mandibular advancement.



The velopharyngeal response was graded as one of four levels of collapse: **(A)** slight (< 25% collapse), **(B)** modest (25% to < 50% collapse), **(C)** substantial (50% to < 75% collapse), **(D)** predominant (≥ 75% collapse). The level of collapse was assigned after the operator was satisfied that three consistent observations were obtained.

prediction models show promise; however, prospective validation is necessary for clinical use. Our group has previously derived prediction models using awake-based assessments including craniofacial characteristics,[8,9] flow-volume loops (spirometry),[10] and observation of the pharyngeal response to mandibular advancement via nasopharyngoscopy.[11] However, in new patient samples predictive utility has degraded. For example, flow-volume loops were not predictive in a subsequent patient sample.[12] Heterogeneity in predictive performance may be explained by the small sample sizes used in derivation studies and the complexity of the pharyngeal response to mandibular advancement, which involves a range of structural and functional characteristics, complicated by sleep state changes. Prediction models derived from wakefulness-based assessments are attractive in terms of clinical applicability. We hypothesized that a range of awake phenotypic assessments with a previously demonstrated role in prediction and capturing different aspects of upper airway function would provide a more robust prediction model. The aim of this study was to use multimodal office-based phenotypic assessments to develop a clinically useful prediction model of OA treatment response.

METHODS

This study received ethical approval (Sydney Local Health District, Protocol No. X11-0134 & HREC/11/RPAH/192) and informed consent was obtained from all participants. This study was registered with the Australian New Zealand Clinical Trials Registry (ANZCTR Trial ACTRN12611000409976).

Participants

Participants were patients from a sleep clinic who were referred for evaluation of sleep-disordered breathing. Those with confirmed OSA (apnea-hypopnea index [AHI] > 10 events/h) by in-laboratory polysomnography and those who were willing to try OA therapy were recruited. Women were targeted for recruitment in an attempt to enrich the sample. Minimal exclusion criteria were set to ensure a generalizable sample. Exclusion criteria were limited to contraindications to OA therapy (eg, periodontal disease and insufficient teeth).

OA Therapy and Treatment Outcome Definitions

The OA used was a customized, two-piece appliance (SomnoDent, SomnoMed Australia) with previously established efficacy.[6] Patients initially received the device at 70% maximum protrusive level and incrementally titrated the device until maximum comfortable mandibular advancement limit was reached (average 14.6 ± 9.8 weeks). Maximal advancement was confirmed by the treating dentist/orthodontist. Treatment outcome was determined by in-laboratory polysomnography. For both the diagnostic and treatment study, apneas and hypopneas were scored using standard definitions. Specifically, hypopneas were defined as a ≥ 50% reduction of preceding airflow amplitude associated with either ≥ 3% oxygen desaturation or arousal. We used 3 criteria of treatment response[6]: criterion 1: treatment AHI < 5 events/h plus ≥ 50% reduction (complete response), criterion 2: treatment AHI < 10 events/h plus ≥ 50% reduction in AHI from baseline, criterion 3: ≥ 50% reduction in AHI from baseline (partial response).

Multimodal Phenotyping Assessments

Before commencing OA acclimatization, participants attended a research visit in which they underwent the following assessments.

Nasopharyngoscopy

Participants underwent a nasopharyngoscopy procedure, as previously described.[11] Prior to the procedure, patients were instructed to perform the Muller maneuver (maximal inspiration against a closed airway) until both patient and physician were confident in its performance. The pharyngeal response in the velopharyngeal region (behind the soft palate), both without and with maximum mandibular advancement, was recorded as one of four levels (**Figure 1**): < 25% (slight collapse), 25% to < 50% (modest collapse), 50% to < 75% (substantial collapse), ≥ 75% (predominant collapse).[13] Velopharyngeal collapse was recorded after three consistent observations.

Spirometry

Flow-volume loops were obtained in the erect position and were performed in accordance with the American Thoracic Society guidelines.[14] The highest values of three technically satisfactory performances were used. Variables of interest were: mid-inspiratory flow at 50% vital capacity (MIF_{50}) and mid-expiratory flow at 50% vital capacity (MEF_{50}) and the ratio MEF_{50}/MIF_{50}.[10]

Craniofacial Photography

Craniofacial assessment was performed by quantitative facial photography.[15,16] Briefly, a front and profile photograph was taken in natural head position with neutral facial expression. Photographs were analyzed using imaging software (ImageJ v1.51j8, National Institutes of Health, Bethesda, Maryland, United States) to obtain x and y coordinates of facial landmarks. Landmarks were used to calculate craniofacial angles and linear measurements. Craniofacial measurements retained for consideration in the multivariable model were those that showed a relationship with treatment response ($P < .1$, when controlled for influence of body mass index [BMI], height, age, and sex). These four craniofacial measurements are illustrated in **Figure 2**.

Statistical Analysis

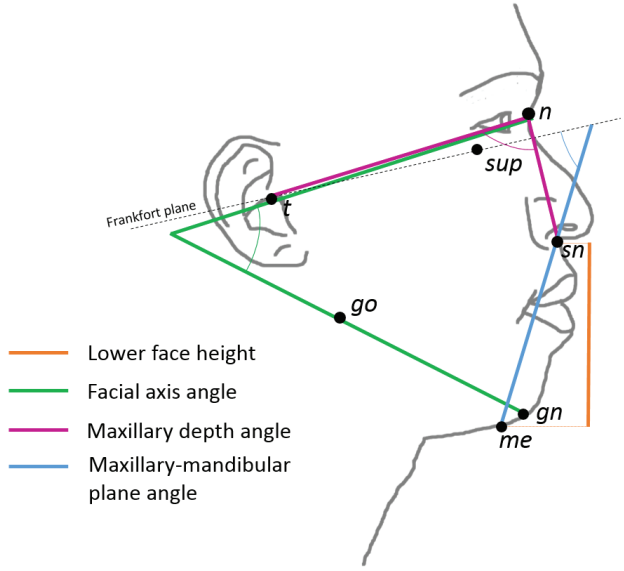
Statistical analysis was performed using SPSS software (version 24, IBM SPSS Statistics, Armonk, New York, United States). Treatment responders and nonresponders were compared by independent *t* test (continuous variables) and chi-square test (categorical variables). Interaction tests between sex and potential predictor variables were performed. This was done to confirm that there were no sex differences in the relationship between clinical/phenotypic variables and treatment response which would warrant separate models for males and females (data not shown). Prediction models for OA treatment response (logistic regression) were built in two steps. First, common clinical variables were entered. Second, phenotypic characteristics were considered (backward likelihood ratio method). Models were adjusted for height and ethnicity. Diagnostic statistics were calculated for each model (sensitivity, specificity, positive predictive value, negative predictive value). Receiver operating characteristic (ROC) curves were produced in SPSS using the predicted probabilities variable from the logistic regression models for each of the treatment response criteria. The areas under the ROC curves for the clinical-only and clinical plus phenotypic models were compared.[17] Statistical significance was accepted at $P < .05$.

RESULTS

Participant Characteristics

A total of 158 patients were recruited to the study. Of these, 149 (59% male) completed the protocol (**Table 1**). The reasons for withdrawal from the study (n = 16) were as follows: n = 1 withdrew due to unrelated illness, n = 1 withdrawn as scheduled for nasal surgery before final sleep study, n = 1 wanted to change to PAP therapy, n = 6 could not tolerate OA treatment, n = 5 were lost to follow-up, n = 2 completed final sleep study at submaximal mandibular protrusion. Participants were middle aged (56.3 ± 11.0 years), were overweight (BMI 29.5 ± 5.0 kg/m^2), reported Caucasian ethnicity (approximately 80%), and had moderate OSA. OA treatment reduced AHI by more than 50%, with a mean treatment AHI of 11.8 ± 13.2 events/h. In terms of treatment response, 32.5% were complete responders and 70.4% were partial responders. There was no difference in the final mandibular advancement level provided by the device between responders and nonresponders.

Figure 2—Illustration of facial measurements related to oral appliance treatment response.



Four craniofacial measurements derived from profile photographs differed between responders and nonresponders and were considered in prediction models. Lower face height (linear distance between landmarks *sn* and *me*), facial axis angle (angle formed between the line intersecting landmarks *t* and *n* and the line intersecting landmarks *go* and *gn*), maxillary depth angle (angle between landmarks *t*, *n*, and *sn*), and maxillary-mandibular plane angle (angle of the intersection of the Frankfort plane and line connecting through landmarks *me* and *sn*). Landmarks: *gn* = gnathion, *go* = gonion, *me* = menton, *n* = nasion, *sn* = subnasion, *sup* = infraorbital rim, *t* = tragion.

Clinical Characteristics and Treatment Response

Nonresponders tended to be older, although this was only statistically significant for criterion 1. Anthropometric measures (BMI, neck circumference, waist circumference) were on average higher in nonresponders (**Table 1**). Baseline AHI was higher in nonresponders, by response criteria 1 and 2. There was no sex difference in OA treatment response rates.

Multimodal Prediction Models

A base model was constructed that consisted of clinical predictors only (age, sex, BMI, neck circumference, waist circumference, and baseline AHI). The contribution of clinical variables to this base model are shown in **Table 2**. The clinical predictors varied by treatment response criteria. Baseline AHI was a significant predictor by all three criteria. For criteria 1 and 2, treatment response was associated with lower AHI. By criterion 3, treatment responders had a higher AHI, likely a function of the treatment response definition. By criteria 1 and 3, younger age was a predictor. BMI and neck circumference were borderline significant predictors ($P = .05$) in the criterion 1 model. Sex or waist circumference were not significant predictors.

Table 1—Sample characteristics.

	All	Criterion 1		Criterion 2		Criterion 3	
		Responder	Nonresponder	Responder	Nonresponder	Responder	Nonresponder
Clinical Data							
n (%)	142 (100)	46 (32.5)	96 (67.6)	75 (52.8)	67 (47.1)	100 (70.4)	42 (29.6)
Male sex, n (%)	84 (59.2)	31 (67.4)	53 (55.2)	45 (60)	39 (58.2)	58 (58.0)	26 (61.9)
Caucasian ethnicity, n (%)	112 (78.9)	33 (71.7)	79 (82.3)	54 (72.0)	58 (86.6)	76 (76.0)	36 (85.7)
Age (years)	56.3 ± 11.0	53.0 ± 11.1	57.8 ± 10.6*	55.4 ± 11.0	57.3 ± 10.9	55.3 ± 10.6	58.7 ± 11.5
BMI (kg/m^2)	29.5 ± 5.0	27.2 ± 3.9	30.6 ± 5.1***	28.0 ± 4.6	31.2 ± 4.9***	29.2 ± 5.2	30.3 ± 4.5
Neck circumference (cm)	39.9 ± 3.7	39.3 ± 3.2	40.1 ± 3.9	39.2 ± 4.9	40.6 ± 4.1*	39.5 ± 3.3	40.6 ± 4.3
Waist circumference (cm)	102.6 ± 12.2	98.4 ± 10.4	104.6 ± 12.6**	99.0 ± 10.5	106.5 ± 12.9***	101.3 ± 11.4	105.9 ± 13.5*
AHI (events/h)	28.7 ± 17.5	19.3 ± 9.8	33.2 ± 18.6***	23.6 ± 12.1	34.4 ± 20.8***	29.6 ± 17.5	26.6 ± 17.5
Severe OSA, n (%)	50 (35.2)	8 (17.4)	42 (43.8)**	21 (28.0)	29 (43.3)	40 (40.0)	10 (23.8)
Treatment Data							
Treatment AHI (events/h)	11.8 ± 13.2	4.3 ± 2.8	16.3 ± 13.9	4.3 ± 2.8	20.1 ± 15.1***	6.7 ± 5.7	23.4 ± 18.0***
%ΔAHI	−56.8 ± 41.9	−85.4 ± 10.7	−43.1 ± 44.4***	−80.5 ± 12.0	−30.3 ± 47.7***	−77.0 ± 12.9	−8.6 ± 47.6***
Final protrusion (mm)	8.7 ± 2.8	8.7 ± 2.9	8.7 ± 2.8	8.7 ± 2.7	8.7 ± 3.0	8.7 ± 2.8	8.7 ± 3.0
%maximum protrusion	81.2 ± 26.9	76.7 ± 24.6	83.1 ± 27.7	78.4 ± 24.0	84.2 ± 29.6	82.0 ± 28.5	79.6 ± 23.5

Data are presented as mean ± standard deviation or n (%) as indicated. Characteristics of the entire sample and treatment responders and nonresponders using the three criteria for response: (1) AHI < 5 events/h plus ≥ 50% reduction, (2) AHI < 10 events/h plus ≥ 50% reduction, (3) ≥ 50% AHI reduction. Asterisks indicate statistical significance: * = $P < .05$, ** = $P < .01$, *** = $P < .001$. AHI = apnea-hypopnea index, BMI = body mass index, OSA = obstructive sleep apnea.

Phenotypic variables considered for the multimodal model were: lower face height, maxillary depth angle, maxillary/mandibular plane angle, facial axis angle, MIF_{50}, MEF_{50}/MIF_{50}, level of velopharyngeal collapse with performance of Muller maneuver with mandibular advancement (< 25%, 25% to < 50%, 50% to < 75%, or ≥ 75%). For treatment response criterion 1, MEF_{50}/MIF_{50} and two craniofacial variables (lower face height, maxillary depth angle) contributed to the model. For criterion 2, two craniofacial variables (maxillary depth angle, maxillary/mandibular relationship angle) contributed to the model. For criterion 3, MEF_{50}/MIF_{50} and maxillary depth angle contributed to the model. Diagnostic accuracy of the multimodal prediction models is outlined in **Table 2**. The predictive accuracy of the models improved with addition of phenotypic variables compared to models using clinical data alone (defined by a higher area under the ROC curve). However, the improvement in predictive accuracy (area under the ROC curve) compared to the clinical model was not statistically significant (**Figure 3**).

Clinical Application of Oral Appliance Prediction Models

There was no significant improvement in the predictive accuracy of models including multimodal phenotypic variables compared to the clinical variable models. Therefore, there is no clinical justification for performing these additional assessments. The optimal logistic regression models (forward likelihood method) based on clinical variables alone are presented in **Table 3**. The formulas for predicted probability of OA treatment response for a given patient are provided to allow calculation of the probability of treatment response. The optimal cutoff points based on ROC curve analysis (selected for maximal sensitivity and specificity using Youden index) are presented for each. However, these presented models have not been cross-validated or tested in an independent sample. However, the presentation of these models is to give an indication of the clinical utility, or lack thereof, of models based on clinical variables.

DISCUSSION

This study describes a novel clinical prediction method for OA treatment response that uses multimodal tests applied during wakefulness. We additionally present prediction models by three commonly used criteria for OA treatment response. Our logistic regression models show good to fair predictive accuracy in terms of the area under the ROC curve. The awake phenotypic variables assessed in this study did contribute to the derived prediction models. However, there was not a significant improvement in predictive accuracy compared to a model based on clinical variables alone. Therefore, there is not a clinical justification for performing these additional awake-based tests to predict OA treatment response.

We included phenotypic assessments, which met two conditions: (1) they can be performed during wakefulness in an office setting, and (2) they have a previously demonstrated role in OA treatment outcome prediction. We included MIF_{50} and MEF_{50}/MIF_{50} variables from spirometry, which are surrogate markers of site of pharyngeal collapse.[10] The second assessment was awake nasopharyngoscopy, which has previously shown a large difference between responders and nonresponders in terms of percentage collapse of the velopharyngeal airway with performance of the Mueller maneuver (−80% versus +9%).[11] The third phenotype was craniofacial structure using facial photography. Our phenotypic data were predictive of OA

Table 2—Multimodal prediction models for oral appliance treatment response.

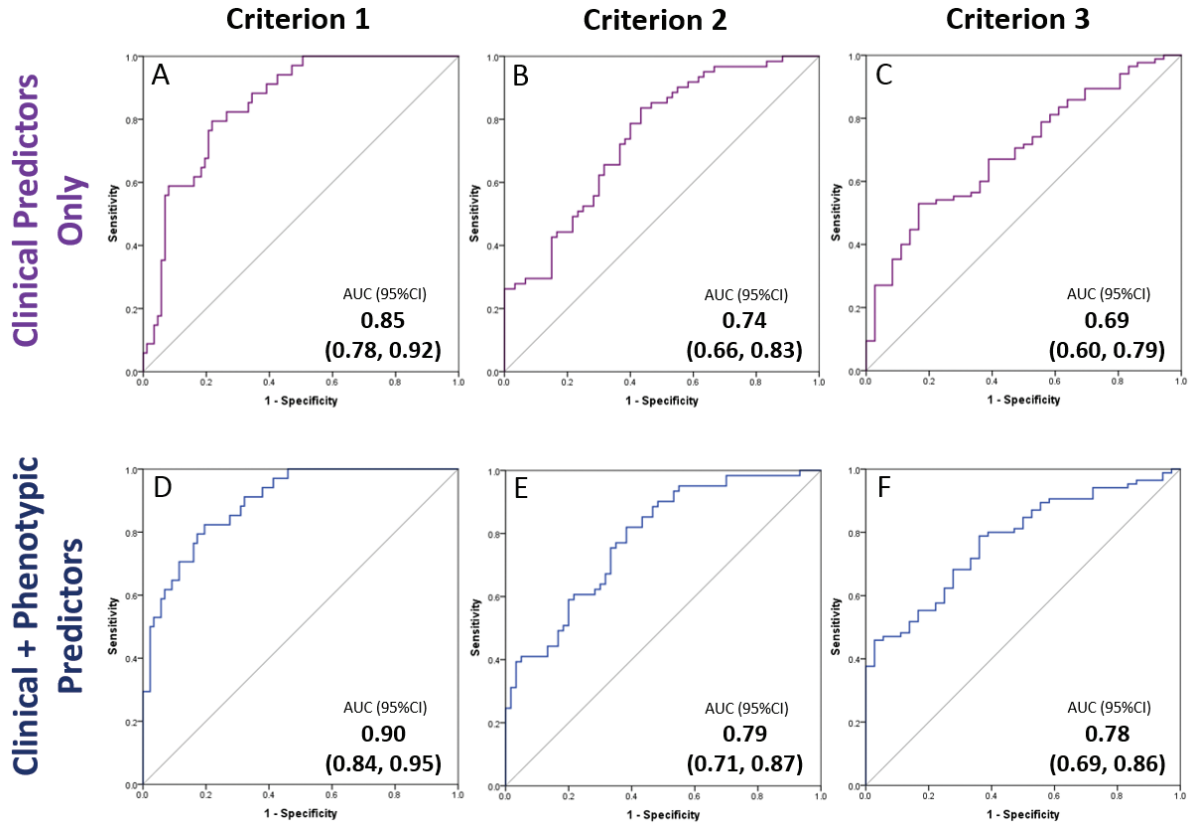
	Criterion 1			Criterion 2			Criterion 3		
	Variables	**Odds Ratio (95% CI)**	***P***	**Variables**	**Odds Ratio (95% CI)**	***P***	**Variables**	**Odds Ratio (95% CI)**	***P***
Clinical Predictors	Age	0.93 (0.87, 0.99)	.02	Age	0.99 (0.95, 1.03)	.63	Age	0.95 (0.91, 1.0)	.05
	Sex	0.90 (0.15, 5.39)	.91	Sex	0.67 (0.16, 2.65)	.55	Sex	0.82 (0.21, 3.25)	.77
	BMI	0.74 (0.55, 1.0)	.05	BMI	0.85 (0.69, 1.05)	.13	BMI	0.99 (0.80, 1.22)	.90
	Neck circumference	1.32 (1.0, 1.76)	.05	Neck circumference	1.11 (0.89, 1.34)	.35	Neck circumference	1.11 (0.89, 1.39)	.36
	Waist circumference	0.99 (0.86, 1.14)	.85	Waist circumference	0.99 (0.90, 1.08)	.76	Waist circumference	0.94 (0.86, 1.04)	.22
	AHI	0.90 (0.85, 0.96)	.01	AHI	0.97 (0.94, 1.00)	.04	AHI	1.05 (1.01, 1.09)	.01
Phenotypic Predictors	MEF_{50}/MIF_{50}	3.47 (1.00, 12.07)	.05	Maxillary depth angle	0.93 (0.84, 1.02)	.10	MEF_{50}/MIF_{50}	2.50 (0.82, 7.64)	.11
	Lower face height	4.25 (1.50, 12.05)	.01	Maxillary-mandibular relationship angle	0.94 (0.87, 1.01)	.09	Maxillary depth angle	0.89 (0.81, 0.98)	.02
	Maxillary depth angle	0.86 (0.77, 1.00)	.04						
Diagnostic Statistics	Correctly classified	81.8%		Correctly classified	70.2%		Correctly classified	72.7%	
	Sensitivity	64.7%		Sensitivity	73.8%		Sensitivity	91.8%	
	Specificity	88.5%		Specificity	66.7%		Specificity	27.8%	
	PPV	68.8%		PPV	69.2%		PPV	75.0%	
	NPV	86.5%		NPV	71.4%		NPV	58.8%	
	ROC AUC (95% CI)	0.90 (0.84, 0.95)		ROC AUC (95% CI)	0.79 (0.71, 0.87)		ROC AUC (95% CI)	0.78 (0.69, 0.86)	

Logistic regression was used to develop prediction models for treatment response by all three criteria: (1) AHI < 5 events/h plus ≥ 50% reduction, (2) AHI < 10 events/h plus ≥ 50% reduction, (3) ≥ 50% AHI reduction. The models were built by first entering clinical predictors and then phenotypic variables from nasopharyngoscopy, spirometry and craniofacial photography using the backward likelihood ratio method. Phenotypic variables entered into the model in this second stage are listed for each criterion. Models are adjusted for height and ethnicity. AHI = apnea-hypopnea index, AUC = area under the curve, BMI = body mass index, CI = confidence interval, MEF_{50} = mid-expiratory flow at 50% vital capacity, MIF_{50} = mid-inspiratory flow at 50% vital capacity, NPV = negative predictive value, PPV = positive predictive value, ROC = receiver operating characteristic.

treatment outcome; however, there was not a statistically significant difference in predictive accuracy compared to models based on clinical variables alone. Clinical characteristics (age, obesity measures, and baseline AHI) have a known association with OA treatment and are readily available. However, on their own these characteristics generally perform poorly in prediction of treatment outcome (less than 65% correctly classified).[6] A prediction model that requires additional effort in obtaining assessments must be able to improve classification of treatment response beyond clinical characteristics alone. The inclusion of awake assessments in our models did not strengthen predictive accuracy; therefore, we do not see a role for this combination of phenotypic assessments in robust prediction. The predictors in our models vary by treatment definition, suggesting these factors are not generalizable. Predictive accuracy was highest for models where the outcome was a complete response (AHI < 5 events/h, criterion 1). This most stringent definition of responders likely represents a less heterogeneous group in that they are able to resolve OSA by this form of treatment. With more liberal response definitions, particularly criterion 3 (≥ 50% AHI decrease), responders are heterogeneous in terms of their AHI level achieved through treatment.

Our results generally align with previous findings. A higher MEF_{50}/MIF_{50} ratio from flow-volume loops had predictive value in discriminating treatment responders, in concordance with the original prediction study.[10] Craniofacial structure has been associated with treatment response, primarily using lateral cephalometry. However, the exact craniofacial structures relating to response vary, due to factors such as small samples, different measurements, and variations in treatment response definitions.[18] In this study, we used a novel application of facial photographic phenotyping. We identified four

Figure 3—Comparison of ROC curves for oral appliance treatment outcome prediction models based on (1) clinical variables only and (2) clinical plus phenotypic assessment variables.



Separate prediction models were built for treatment response based on the three criteria: (1) AHI < 5 events/h plus ≥ 50% reduction, (2) AHI < 10 events/h plus ≥ 50% reduction, (3) ≥ 50% AHI reduction. The top row of ROC curves are models made from clinical variables only (for each treatment response definition criteria). The bottom row shows ROC curves for models with the addition of phenotypic awake assessments. There is an improvement in the AUC for all models from the top to the bottom row; however, this was not a statistically significant improvement ($P > .05$). AUC = area under the curve, CI = confidence interval, ROC = receiver operating characteristic.

Table 3—Clinical regression models for oral appliance treatment response.

	Criterion 1			Criterion 2			Criterion 3		
Predictor Variables	Variable	OR	*P*	Variable	OR	*P*	Variable	OR	*P*
	AHI	0.93	.001	AHI	0.97	.013	Waist circ.	0.97	.043
	Age	0.96	.041	BMI	0.89	.004			
	BMI	0.89	.014						
z	6.62 − [(0.072 × AHI) − (0.039 × Age) − (0.12 × BMI)]			4.46 − [(0.031 × AHI) − (0.12 × BMI)]			4.12 − (0.03 × waist circumference)		
Correctly Classified	73.6%			66.4%			71.4%		
Area Under ROC Curve	0.8			0.7			0.6		
Optimal Cutoff Point (Sensitivity, Specificity)	Predicted probability of 0.36 (80.4%, 70.8%)			Predicted probability of 0.61 (92.0%, 43.3%)			Waist circumference of 104.5 cm (52.4%, 65.3%)		

Because phenotypic tests did not provide a statistical improvement in prediction on top of clinical variables, we provide logistic regression models for treatment response based on clinical variables alone. Clinical predictors were included in the model using the forward likelihood ratio method for entering predictor variables into the model. The odds ratio indicates the change in odds of response for a one unit increase in the predictor (with the other predictors held constant). The predicted probability can be calculated for a given patient based on the model equation $(1 / (1 + e^{-z}))$. The value of z is given for each model. The performance of each model is indicated by the area under the ROC curve. ROC curves were used to define an optimal threshold value for classification of treatment responders and nonresponders and the associated sensitivity and specificity of prediction based on that cutoff is given. These models have not been cross-validated or tested in an independent sample. Criterion: (1) AHI < 5 events/h plus ≥ 50% reduction, (2) AHI < 10 events/h plus ≥ 50% reduction, (3) ≥ 50% AHI reduction. AHI = apnea-hypopnea index, BMI = body mass index, OR = odds ratio, ROC = receiver operating characteristic.

facial photographic measurements associated with treatment response using statistical analysis. Responders tended to have a longer lower face, increased facial axis angle, and reduced maxillary and mandibular position angles, suggestive of maxillary/mandibular retrusion. The results of this study suggest a relationship between facial surface measurements and OA treatment response and demonstrates that this simpler, nonradiographic technique may have utility in assessing craniofacial structure in relation to OSA treatment outcome. We have not found qualitative assessment of nasopharyngoscopy to be predictive.[13] However, a prospective study has found quantitative analysis of nasopharyngoscopic images in conjunction with baseline AHI to be predictive of OA outcome in Japanese patients with OSA.[19]

In this study we wished to explore the influence of sex on OA treatment response. We targeted females in recruitment in an attempt to achieve sex balance, ending up with a 60:40 split of men and women. Previous findings have suggested women are more likely to respond to OA treatment.[20] In our models, sex was not a significant predictor of treatment response and we have previously shown no difference in response rates in a larger dataset of women (n = 109).[6] Moreover, we did not find any sex differences in the relationship between treatment response and any of the clinical or phenotypic predictors. Nor did we identify a difference in treatment response rates between sexes.

We designed our study to include prediction tests based on specific criteria (ie, phenotypic assessments that would be simple to apply in wakefulness). However, our included assessments are by no means representative of all potential predictors. A Belgian study is underway to look at the predictive value of mandibular advancement on parameters from awake nasopharyngoscopy, drug-induced sleep endoscopy (DISE), and computational fluid dynamics derived pharyngeal resistance changes from patient-specific geometry.[21] Lower therapeutic PAP pressure requirement is associated with response, albeit with population-specific pressure thresholds.[22,23] Therapeutic PAP pressure appears to be a reasonable surrogate for pharyngeal collapsibility[24] and a highly collapsible pharyngeal airway is a poor substrate for OA efficacy.[25] There is increasing interest in nonanatomical pathophysiological traits that contribute to OSA.[26] These include characteristics such as respiratory dilator muscle insufficiency, premature arousal before dilatory muscles can take effect (low arousal threshold), and ventilatory overshoot to disturbance (high loop gain). Assessment of these traits in a small number of patients who have undergone treatment for OA identified higher loop gain as well as collapsibility as indicators of a smaller AHI reduction with OA therapy.[25] Overnight studies to measure nonanatomical traits such as loop gain are complex and restricted to research; however work is emerging to use algorithms to derive these characteristics from clinical polysomnography.[27] Our study does not incorporate sleep-based assessments such as DISE and nonanatomical pathophysiological contributors. It will be interesting to assess any additive role of these in prediction in future studies.

There is yet a robust, validated awake prediction method for OA treatment response[7] but there is an emergence of clinical sleep testing methodology for this purpose. Remote-controlled mandibular protrusion to monitor effects during sleep has been a concept in OA treatment prediction for some time, with studies with prototype protrusion devices published in 2002 and 2006.[28,29] Recent appearance of a commercial device for this purpose has allowed further investigation and demonstrated good predictive accuracy based on a prediction algorithm using respiratory events in supine REM sleep.[30] We have subsequently confirmed predictive accuracy of the method for discriminating responders and nonresponders, albeit with a higher test failure rate (> 20%) and requirement for maximal mandibular protrusive position to be reached during the study.[31] Despite these nuances, monitoring the effects of mandibular protrusion during sleep (whether by a temporary device, or remote controlled positioner) shows promise as a prediction method.

The current study includes a large sample to investigate multimodal assessments during wakefulness. We believe our study has the advantage of using a sample generalizable to our sleep clinic population through setting minimal inclusion criteria. We also enriched the sample with women to explore potential sex differences and have a range of OSA severity in the sample (35% severe OSA). Furthermore, we present the data using three commonly used criteria for treatment response. However, there are limitations to the study. Although we have a large sample for this type of study, multiple predictors would favor even larger samples. Our nasopharyngoscopic assessment did not produce quantitative analysis; however, we wanted to move toward a model that had a high possibility of clinical adoption. We have not provided a validation of the derived models, which is an essential step before adopting any prediction model into clinical practice. We have not validated the current models, as we do not deem the contribution of these phenotypic assessments robust enough to warrant this.

CONCLUSIONS

The ability to identify *a priori* which patients are likely to receive therapeutic benefit from OA treatment has been an ongoing challenge since the introduction of these devices for OSA treatment. A range of clinical characteristics and prediction tests ranging in complexity, in both sleep and wake states, have been proposed for this purpose. However, there are currently no adequately validated models. In this study, we combined multiple awake-based phenotypic assessments in an effort to improve prediction. Although we have produced models with reasonable diagnostic accuracies, we have not found that the addition of phenotypic assessments improves prediction beyond models using only clinical characteristics. Our work suggests the need to focus on sleep-related phenotypes in future studies.

ABBREVIATIONS

AHI, apnea-hypopnea index
BMI, body mass index

DISE, drug induced sleep endoscopy
MEF_{50}, mid-expiratory flow at 50% vital capacity
MIF_{50}, mid-inspiratory flow at 50% vital capacity
OA, oral appliance
OSA, obstructive sleep apnea
ROC, receiver operating characteristic

REFERENCES

1. Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Increased prevalence of sleep-disordered breathing in adults. *Am J Epidemiol.* 2013;177(9):1006–1014.
2. Drager LF, Polotsky VY, O'Donnell CP, Cravo SL, Lorenzi-Filho G, Machado BH. Translational approaches to understanding metabolic dysfunction and cardiovascular consequences of obstructive sleep apnea. *Am J Physiol Heart Circ Physiol.* 2015;309(7):H1101–H1111.
3. Leng Y, McEvoy CT, Allen IE, Yaffe K. Association of sleep-disordered breathing with cognitive function and risk of cognitive impairment: a systematic review and meta-analysis. *JAMA Neurol.* 2017;74(10):1237–1245.
4. Rotenberg BW, Murariu D, Pang KP. Trends in CPAP adherence over twenty years of data collection: a flattened curve. *J Otolaryngol Head Neck Surg.* 2016;45(1):43.
5. 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.
6. 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.
7. Okuno K, Pliska BT, Hamoda M, Lowe AA, Almeida FR. Prediction of oral appliance treatment outcomes in obstructive sleep apnea: a systematic review. *Sleep Med Rev.* 2016;30:25–33.
8. Mehta A, Qian J, Petocz P, Darendeliler MA, Cistulli PA. A randomized, controlled study of a mandibular advancement splint for obstructive sleep apnea. *Am J Respir Crit Care Med.* 2001;163(6):1457–1461.
9. Ng AT, Darendeliler MA, Petocz P, Cistulli PA. Cephalometry and prediction of oral appliance treatment outcome. *Sleep Breath.* 2012;16(1):47–58.
10. Zeng B, Ng AT, Darendeliler MA, Petocz P, Cistulli PA. Use of flow-volume curves to predict oral appliance treatment outcome in obstructive sleep apnea. *Am J Respir Crit Care Med.* 2007;175(7):726–730.
11. Chan AS, Lee RW, Srinivasan VK, Darendeliler MA, Grunstein RR, Cistulli PA. Nasopharyngoscopic evaluation of oral appliance therapy for obstructive sleep apnoea. *Eur Respir J.* 2010;35(4):836–842.
12. Chan AS, Lee RW, Srinivasan VK, Darendeliler MA, Cistulli PA. Use of flow-volume curves to predict oral appliance treatment outcome in obstructive sleep apnea: a prospective validation study. *Sleep Breath.* 2011;15(2):157–162.
13. Sutherland K, Chan ASL, Ngiam J, Darendeliler MA, Cistulli PA. Qualitative assessment of awake nasopharyngoscopy for prediction of oral appliance treatment response in obstructive sleep apnoea. *Sleep Breath.* 2018 Jan 23. doi: 10.1007/s11325-018-1624-8. [Epub ahead of print].
14. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med.* 1995;152(3):1107–1136.
15. Lee RW, Chan AS, Grunstein RR, Cistulli PA. Craniofacial phenotyping in obstructive sleep apnea--a novel quantitative photographic approach. *Sleep.* 2009;32(1):37–45.
16. Sutherland K, Lee RW, Petocz P, et al. Craniofacial phenotyping for prediction of obstructive sleep apnoea in a Chinese population. *Respirology.* 2016;21(6):1118–1125.
17. Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology.* 1982;143:29–36.
18. Alessandri-Bonetti G, Ippolito DR, Bartolucci ML, D'Anto V, Incerti-Parenti S. Cephalometric predictors of treatment outcome with mandibular advancement devices in adult patients with obstructive sleep apnea: a systematic review. *Korean J Orthod.* 2015;45(6):308–321.
19. Okuno K, Sasao Y, Nohara K, et al. Endoscopy evaluation to predict oral appliance outcomes in obstructive sleep apnoea. *Eur Respir J.* 2016;47(5):1410–1419.
20. Marklund M, Stenlund H, Franklin KA. Mandibular advancement devices in 630 men and women with obstructive sleep apnea and snoring: tolerability and predictors of treatment success. *Chest.* 2004;125(4):1270–1278.
21. Verbruggen AE, Vroegop AV, Dieltjens M, et al. Predicting therapeutic outcome of mandibular advancement device treatment in obstructive sleep apnoea (PROMAD): study design and baseline characteristics. *Journal of Dental Sleep Medicine.* 2017;4(3):119–138.
22. 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.
23. Tsuiki S, Kobayashi M, Namba K, et al. Optimal positive airway pressure predicts oral appliance response to sleep apnoea. *Eur Respir J.* 2010;35(5):1098–1105.
24. Landry SA, Joosten SA, Eckert DJ, et al. Therapeutic CPAP level predicts upper airway collapsibility in patients with obstructive sleep apnea. *Sleep.* 2017;40(6).
25. 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.
26. Eckert DJ. Phenotypic approaches to positional therapy for obstructive sleep apnoea. *Sleep Med Rev.* 2018;37:175–176.
27. 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.
28. Dort LC, Hadjuk E, Remmers JE. Mandibular advancement and obstructive sleep apnoea: a method for determining effective mandibular protrusion. *Eur Respir J.* 2006;27(5):1003–1009.
29. Petelle B, Vincent G, Gagnadoux F, Rakotonanahary D, Meyer B, Fleury B. One-night mandibular advancement titration for obstructive sleep apnea syndrome: a pilot study. *Am J Respir Crit Care Med.* 2002;165(8):1150–1153.
30. 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.
31. 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.

ACKNOWLEDGMENTS

The authors thank the trial coordinators involved in this study: Meghan Leigh, Peter Singh, and especially Amanda Davies for her role in study setup; Anel Blignaut for assistance with oral appliance treatment; the staff of the sleep investigation unit, Royal North Shore Hospital; and SomnoMed for provision of oral appliances for the study.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication February 18, 2018
Submitted in final revised form June 24, 2018
Accepted for publication July 26, 2018
Address correspondence to: Kate Sutherland, PhD, Department of Respiratory and Sleep Medicine, 8A, Acute Services Building, Royal North Shore Hospital, Reserve Road, St Leonards, NSW 2065, Australia; Tel: +61 2 9463 2924; Fax: +61 2 9463 2099; Email: kate.sutherland@sydney.edu.au

DISCLOSURE STATEMENT

All authors have seen and approved this manuscript. This study was funded by a National Health and Medical Research Council (NHMRC) project grant (GNT1024351), Australia. Pilot work was supported by a seed funding grant from the NHMRC Centre of Clinical Research Excellence, Centre for Integrated Research and Understanding of Sleep (CIRUS) and Dr. Sutherland was supported by a

CIRUS postdoctoral research fellowship. Dr. Cistulli holds an endowed Academic Chair position at the University of Sydney, funded by ResMed Inc. He has received research support from ResMed, SomnoMed, Zephyr Sleep Technologies. He has a pecuniary interest in SomnoMed, resulting from previous involvement in research and development. Oral appliances for this investigator study were provided by SomnoMed. The other authors report no conflicts of interest.