ORIGINAL RESEARCH

Simple and Unbiased OSA Prescreening: Introduction of a New Morphologic OSA Prediction Score

Naima Laharnar [1],*
Sebastian Herberger [1],*
Lisa-Kristin Prochnow [1]
Ning-Hung Chen [2]
Peter A Cistulli [3,4]
Allan I Pack [5]
Richard Schwab [5]
Brendan T Keenan [5]
Diego R Mazzotti [5,6]
Ingo Fietze [1,7]
Thomas Penzel [1,8]

[1]Department of Internal Medicine and Dermatology, Interdisciplinary Center of Sleep Medicine, Charité – Universitätsmedizin Berlin, Berlin, Germany; [2]Department of Pulmonary and Critical Care Medicine, Sleep Center, Chang Gung Memorial Hospital, Taipei, Taiwan; [3]Charles Perkins Centre, University of Sydney, Sydney, NSW, Australia; [4]Department of Respiratory Medicine, Royal North Shore Hospital, Sydney, NSW, Australia; [5]Department of Medicine/Division of Sleep Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [6]Department of Internal Medicine, Division of Medical Informatics, University of Kansas Medical Center, Kansas City, KS, USA; [7]The Federal State Autonomous Educational Institution of Higher Education I. M. Sechenov, First Moscow State Medical University of the Ministry of Health of the Russian Federation, Moscow, Russia; [8]Department of Biology, Saratov State University, Saratov, Russia

*These authors contributed equally to this work

Correspondence: Naima Laharnar
Department of Internal Medicine and Dermatology, Interdisciplinary Center of Sleep Medicine, Charité – Universitätsmedizin Berlin, Campus Charité Mitte, Charitéplatz 1, Berlin, 10117, Germany
Tel +49-30 450 513 120
Email naima.laharnar@charite.de

Purpose: An early prescreening in suspected obstructive sleep apnea (OSA) patients is desirable to expedite diagnosis and treatment. However, the accuracy and applicability of current prescreening tools is insufficient. We developed and tested an unbiased scoring system based solely on objective variables, which focuses on the diagnosis of severe OSA and exclusion of OSA.

Patients and Methods: The OSA prediction score was developed (n = 150) and validated (n = 50) within German sleep center patients that were recruited as part of the Sleep Apnea Global Interdisciplinary Consortium (SAGIC). Six objective variables that were easy to assess and highly correlated with the apnea–hypopnea index were chosen for the score, including some known OSA risk factors: body-mass index, neck circumference, waist circumference, tongue position, male gender, and age (for women only). To test the predictive ability of the score and identify score thresholds, the receiver-operating characteristics (ROC) and curve were calculated.

Results: A score ≥8 for predicting severe OSA resulted in an area under the ROC curve (ROC-AUC) of 90% (95% confidence interval: 84%, 95%), test accuracy of 82% (75%, 88%), sensitivity of 82% (65%, 93%), specificity of 82% (74%, 88%), and positive likelihood ratio of 4.55 (3.00, 6.90). A score ≤5 for predicting the absence of OSA resulted in a ROC-AUC of 89% (83%, 94%), test accuracy of 80% (73%, 86%), sensitivity of 72% (55%, 85%), specificity of 83% (75%, 89%), and positive likelihood ratio of 4.20 (2.66, 6.61). Performance characteristics were comparable in the small validation sample.

Conclusion: We introduced a novel prescreening tool combining easily obtainable objective measures with predictive power and high general applicability. The proposed tool successfully predicted severe OSA (important due to its high risk of cardiovascular disease) and the exclusion of OSA (rarely a feature of previous screening instruments, but important for better differential diagnosis and treatment).

Keywords: obstructive sleep apnea, diagnostic, sensitivity, specificity, screening

Introduction

Obstructive sleep apnea (OSA) is a chronic condition caused by recurrent partial (hypopnea) or complete (apnea) upper-airway collapse during sleep and presents with a wide range of symptoms, including snoring and daytime sleepiness.[1,2] The estimated global prevalence of moderate to severe OSA amounts to over 425 million adults.[3–5] OSA leads to significant performance impairment, increased risk for accidents and elevated risk for metabolic and cardiovascular disease, depression, dementia, and stroke. OSA is directly associated with increased overall morbidity and mortality.[6–14]

Diagnosis of OSA can be complex.[15] The most accurate way to diagnose OSA is the objective overnight recording of sleep and respiration, either in a sleep laboratory with polysomnography or at home with a portable polygraph system.

However, these methods are expensive, time-consuming, and often not readily available.[4,16] Less expensive and early pre-screening tools include patient-reported symptoms with questionnaires or prediction scores based on risk factors or a combination of these.[15] While questionnaires are practical, they rely on subjective symptom reporting, reaching very limited diagnostic accuracy and little differentiation of OSA severity.[17–20] OSA prediction scores often include some objective risk factors such as oral cavity, neck circumference, body-mass index (BMI), and craniofacial photography.[21] However, these scores were mainly evaluated within very specific clinical risk populations (eg, obese patients, pregnant women, children, etc.), did not differentiate OSA severity, and can be less practical and of limited statistical accuracy.[22–28]

Given the existing evidence, a recent review on OSA prescreening tools for the US Preventive Services Task Forces concluded that the accuracy and clinical utility of current screening tools is not sufficient.[15] Thus, to identify patients at risk in a timely and cost-efficient manner, a well-performing early screening step in the diagnostic process is needed. The early and objective risk prediction of OSA and the differentiation of severity is important in a clinical context. The exclusion of OSA is essential for effective differential diagnosis. Equally important is the detection of severe OSA due to its high risk of cardiovascular disease, morbidity and overall mortality.[29,30]

Our objective was to develop and test a simple OSA prediction score less prone to subjective biases as an early and effective screening tool with general clinical practicability. We aimed to include only simple to obtain objective features based on anthropometric measures and intraoral visibility, with a special focus on the accurate diagnosis of severe OSA, due to its clinical importance.

Patients and Methods
Participants and Data Collection Procedure

Our study used patient data that was collected as part of the Sleep Apnea Global Interdisciplinary Consortium (SAGIC). The Consortium is a collaborative, worldwide research project on sleep apnea in existence since 2011 with a major focus on personalized medicine approaches, craniofacial risk factors and the genetics of OSA. The participating sleep centers (University of Pennsylvania, Ohio State University, University of Sydney, University of Western Australia, University of Iceland, University of San Paulo, Korea University, Chang Gung Memorial Hospital, and the University Hospital Charité Berlin) collect data through questionnaires, standardized facial and intraoral photography, saliva samples, and polysomnography or polygraphy.

The morphological score was developed and validated with data from German SAGIC patients, using anthropometric measures and intraoral photography collected as part of SAGIC. The patients had to be 18 years or older with suspected OSA (eg, symptoms such as daytime sleepiness and non-restful sleep that were not explained by other factors). Exclusion criteria included pregnancy, history of psychiatric or neurological disorders (eg, stroke, epilepsy, or brain damage) or other relevant medical diagnoses (eg, requiring medications including beta-blocker, antihistamine, etc.) that may influence sleep, or a currently ongoing OSA therapy (eg, continuous positive airway pressure therapy). The patients underwent overnight sleep-laboratory polysomnography (SOMNOmedics GmbH, Randersacker, Germany) or portable polygraphy (Nox-T3, CareFusion Respiratory, CA, USA or Embletta, Natus Medical Incorporated, WI, USA), and completed a SAGIC questionnaire (including demographic questions, medication and medical history, etc). Anthropometric measurements (neck, waist, and hip), facial and intraoral photography were collected by the study team. Visual scoring of sleep and respiration were performed by certified sleep specialists according to AASM 2012 (American Academy of Sleep Medicine) criteria.[31]

Participation and patient recruitment for SAGIC was approved by the Institutional Ethics and Scientific Review Committee of the Charité – Universitätsmedizin Berlin, Germany (EA 1/245/12). Patients signed an informed consent form.

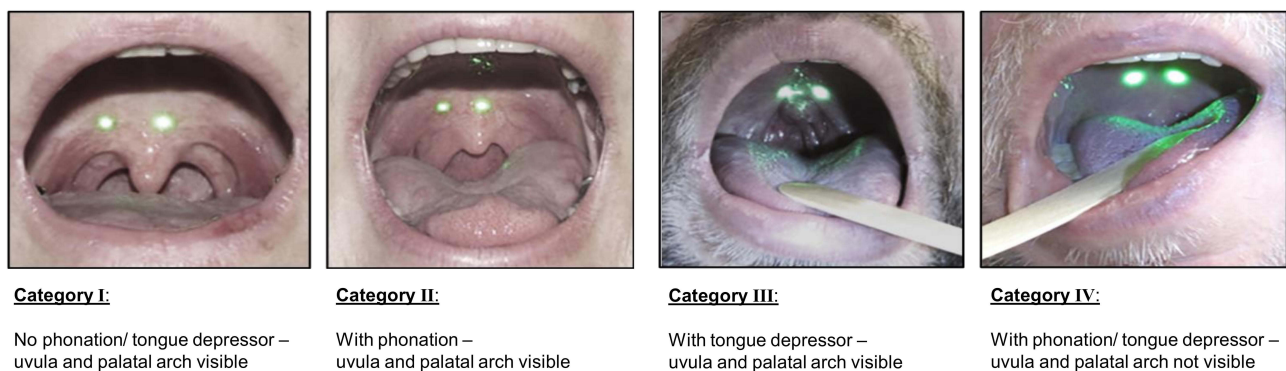## Intraoral Facial Photography and Modified Friedman Tongue Position

Classification of pharyngeal visibility was included in our scoring system based on intraoral photographs, although, in clinical practice classification may also be performed directly. The photographs were taken following a standardized SAGIC protocol and have been shown to provide useful anatomical data for OSA risk prediction.[32,33] In brief, a digital camera (Canon PowerShot SX 120 and 170 IS) and laser beam (single, class IIIa, 4.0–5.0 milliwatt green laser module) were connected and mounted on a monopod with tilt functionality

(Manfrotto 682B professional monopod with detachable legs). Patients were prepared (eg, hair tied back, glasses removed) and certain facial landmarks (eg, gonion profile, orbital rim, etc.) were marked for craniofacial pictures (detailed description in Lee et al 2009).[32] Patients were seated in correct posture and camera alignment (eg, camera approx. 40cm from subject). First, craniofacial pictures were taken (frontal, profile and neck extended photographs, described in more detail in Lee et al 2009).[32] Secondly, intra-oral pictures were taken using standardized instructions on how to open the mouth. The intraoral area was photographed in eight positions, four with the tongue in normal position and four with extended tongue, each first without and then with added tools such as phonation and tongue depressor (described in more detail in Schwab et al 2017).[33]

We used a modification of the Friedman Tongue Position (FTP) for an objective classification of the intraoral visibility. The FTP allows for an evaluation of the tongue's position in relation to pillars, uvula, soft and hard palate. As the FTP uses a neutral tongue position for visibility classification, we only analyzed the intraoral pictures with neutral tongue position. As our cohort consisted of OSA risk patients, we adapted the FTP categories by including phonation and tongue depression to allow for a better differentiation of the higher categories. Patients were categorized the following way (Figure 1): I. Visibility of the uvula and palatal arch in neutral tongue position without added tools, II. Visibility of the uvula and palates in neutral tongue position only with phonation, III. Visibility of the uvula and palates in neutral tongue position only with tongue depressor, IV. No visibility of the uvula and palatal arch with or without phonation or tongue depression.

Statistics

Statistical analysis was performed with SPSS (IBM SPSS Statistics, version 25). Non-parametric Spearman rank correlations were calculated between the AHI of the test cohort and several OSA risk factors to identify possible prediction variables for the score. We used non-parametric correlations as the AHI and some of the risk variables were not normally distributed. Significance level was set at $p = 0.01$. We weighted the identified score variables with 1–3 points based on previous studies and prediction ability.[24] To test the predictive ability of the score and identify score thresholds, the receiver-operating characteristics (ROC) and curve were calculated with a test cohort of n = 150 patients. The ROC curve is defined by sensitivity (true positive rate) on the y-axis and 1-specificity (false positive rate) on the x-axis. The area under the ROC curve (ROC-AUC) was used as a measure of model-predictive performance. The ROC-AUC is interpreted as the probability that a randomly selected subject with the outcome of interest (eg, case) has a larger model predicted probability than a randomly selected subject without the outcome (eg, control); thus, ROC-AUC values closer to 1.0 represent better performance.[34] The ROC curves were used to identify score thresholds to best predict patients with severe OSA (AHI $\geq$ 30) and patients with no OSA (AHI $\leq$ 5). Additionally, the positive and negative predictive values, the positive and negative likelihood ratios, and the test accuracy with a 95% confidence interval were calculated. These variables are based on the two-by-two outcome table of true positives, true negatives, false positives (Type I error), and false negatives (Type II error) (see Supplementary Table 1 for equations). The ROC analysis was repeated with an independent cohort of n = 50 patients to validate the results.

Category I:

No phonation/ tongue depressor – uvula and palatal arch visible

Category II:

With phonation – uvula and palatal arch visible

Category III:

With tongue depressor – uvula and palatal arch visible

Category IV:

With phonation/ tongue depressor – uvula and palatal arch not visible

Figure 1 Modified Friedman tongue position.

Results

Test and Validation Cohort

First, a total of 166 SAGIC patients were recruited to develop and test the scoring system; 16 patients were excluded from analysis due to missing or incomplete data. The remaining 150 patients made up the test cohort. Afterwards, another 50 patients were recruited as part of SAGIC and assigned to the validation cohort to validate the scoring system. The cohorts did not differ in general demographics (Table 1). Both cohorts consisted of slightly more men (test cohort: 60%; validation cohort: 56%), a mean age of almost 60 years (58±12 years; 60±12 years) and an overall mean BMI of 30±7 kg/m^2 (representing the overweight category). Both cohorts presented a similar range of AHI. The test cohort showed a range from 0.0 to 115.0 events/hour (mean±SD AHI: 19.7±22.0 events/hour), the validation cohort a range from 0.1 to 114.2 events/hour (mean±SD AHI: 28.2±24.5 events/hour).

Score Development

For the score, six objective and easy to assess variables that highly correlated with the AHI of the test cohort were included: Modified FTP (r = 0.40, p = 0.001), waist circumference (r = 0.66, p = 0.001), neck circumference (r = 0.56, p = 0.001), BMI (r = 0.53, p = 0.001), male sex (r = 0.34, p = 0.001) and age for women (r = 0.53, p = 0.001). As age was only marginally significant for the entire test cohort (r = 0.20, p=0.02) and not associated with AHI in men alone (r = 0.02, p = 0.860), we decided to use age only as a prediction variable for women. We used the same weighting system as Deflandre et al, assigning

Table 2 Variables of the Scoring System with Weighting Points

	1 Point	2 Points	3 Points
Modified Friedman Scale		II	III and IV
Waist circumference (cm)	> 95	> 110	> 125
Age (years)	Women ≥ 60	Women ≥ 65	Women ≥ 80
BMI (kg/m2)	> 28	> 39	> 41
Neck circumference (cm)	> 37	> 42	> 48
Sex	Male		

Notes: The score was inspired by the work of Deflandre et al who developed a score for anesthesiologists for the prescreening of OSA to decrease perioperative morbidity.[24] We used easier to access variables for a broader clinical applicability and modified Table 2 from Deflandre et al to reflect our variables and cut-off points. **Abbreviations:** BMI, body mass index; cm, centimeters.

each variable with up to three points, a higher point indicating a higher AHI and OSA risk, resulting in a score range from 0 to 15 points (Table 2).[24]

The score was applied to the test cohort to evaluate its overall predictive abilities for the detection of severe OSA (AHI ≥ 30) as well as the exclusion of OSA (AHI ≤ 5), and to identify the best possible score threshold for these categories (Supplementary Table 2).

For distinguishing patients with severe OSA (AHI ≥ 30, n = 34) and those without severe OSA (AHI < 30, including patients with an AHI below 5 and therefore no OSA at all, n = 116), the ROC-AUC was 0.90 (95% CI: 0.84, 0.95) with p < 0.001. Figure 2 displays the ROC curve with sensitivity and 1-specificity for different thresholds. The score threshold of 8 showed the best balance of predictive values with both, a high sensitivity of 82% (95% CI: 65%, 93%) and a high specificity of 82% (95% CI: 74%, 88%). As a higher score indicates a higher OSA risk, a score of 8 or more points was chosen for the prediction of patients with severe OSA and achieved
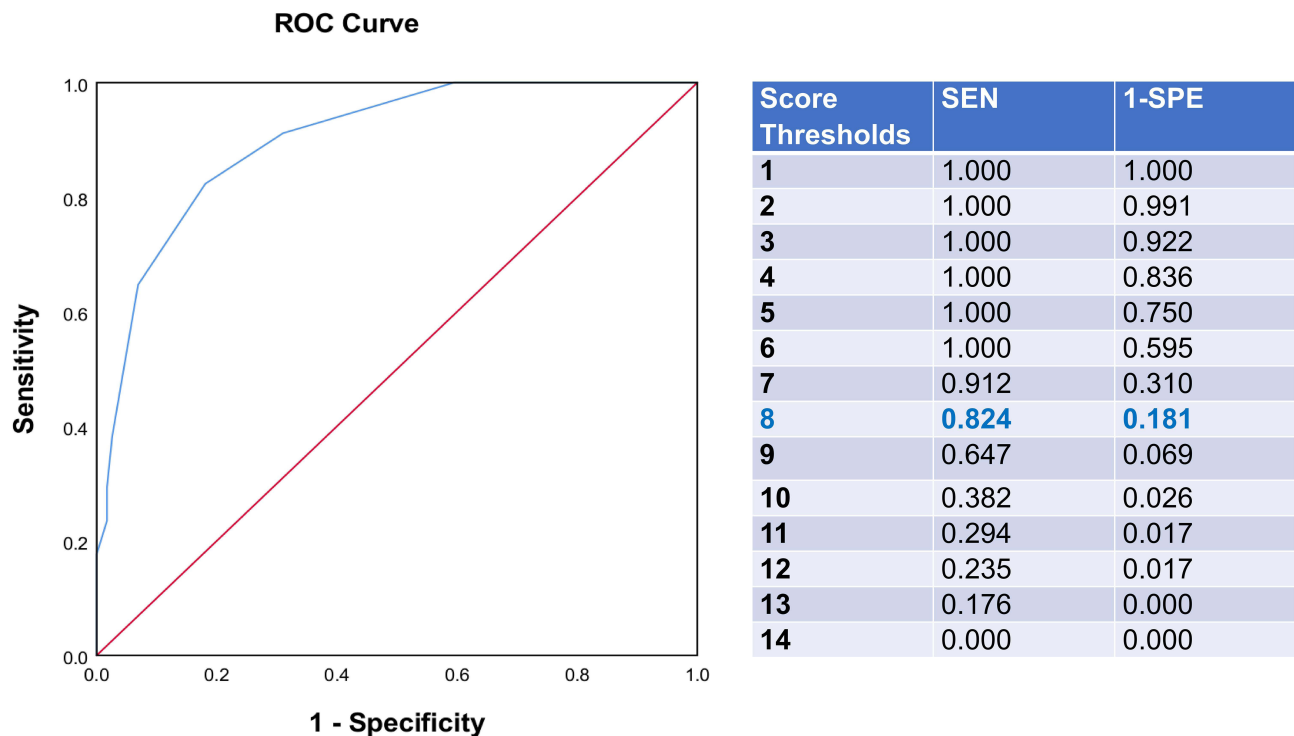
Table 1 Demographics of SAGIC Patients, Separated by Test and Validation Cohort

	Test Cohort n = 150	Validation Cohort n = 50	p
Men	90 (60%)	28 (56%)	0.618
Age (years)	57.5 ± 12.3	59.7 ± 12.4	0.267
BMI (kg/m^2)	30.2 ± 7.2	29.9 ± 6.7	0.783
Neck circumference (cm)	40.6 ± 4.9	40.6 ± 3.8	0.999
Waist circumference (cm)	107.8 ± 18.6	107.2 ± 13.9	0.847
Waist-to-Hip Ratio	0.9 ± 0.1	0.9 ± 0.1	0.862
Modified Friedman Scale	2.8 ± 0.9	2.8 ± 0.8	0.752
AHI (events/hour)	19.7 ± 22.0	28.2 ± 24.5	**0.021***
AHI ≤ 5	n = 39	n = 5	
5 < AHI < 30	n = 77	n = 27	
AHI ≥ 30	n = 34	n = 18	

Notes: Presented are numbers (%) and mean ± standard deviation for variables; chi-square for categorical variables or *t*-test for continuous variables was calculated with a significance level of p = 0.05. *Significant differences between the cohorts are highlighted (bold).
Abbreviations: BMI, body mass index; AHI, apnea–hypopnea index.

ROC Curve

Score Thresholds	SEN	1-SPE
1	1.000	1.000
2	1.000	0.991
3	1.000	0.922
4	1.000	0.836
5	1.000	0.750
6	1.000	0.595
7	0.912	0.310
8	**0.824**	**0.181**
9	0.647	0.069
10	0.382	0.026
11	0.294	0.017
12	0.235	0.017
13	0.176	0.000
14	0.000	0.000

Figure 2 Test cohort – receiver-operating curve and score thresholds for predicting severe obstructive sleep apnea (apnea–hypopnea index ≥30).
Notes: Left: ROC curve for score prediction of severe obstructive sleep apnea (apnea–hypopnea index ≥ 30) within the test cohort. The curve is defined by sensitivity (true positive rate) on the y-axis and 1-specificity (false positive rate) on the x-axis. Points above the diagonal of the plot represent good results, better than random. Right: Sensitivity and 1-specificity for different score thresholds. Highlighted (blue, bold): A score of 8 or above presents the best predictive values.
Abbreviations: ROC, receiver-operating characteristics; SEN, sensitivity; SPE, specificity.

a test accuracy of 82% (95% CI: 75%, 88%). Table 3 presents the statistical details including the positive and negative predictive values and the population independent positive and negative likelihood ratios.

Next (Figure 3, Table 3), we identified the best possible score threshold to distinguish patients without OSA (exclusion of OSA, AHI < 5, n = 39) and patients with OSA (AHI ≥ 5, n = 111). The ROC-AUC was 0.89 (95% CI: 0.83, 0.94) with $p < 0.001$. The score threshold of 5 showed the best predictive values with a sensitivity of 72% (95% CI: 55%, 85%) and a specificity of 83% (95% CI: 75%, 89%). As a lower score indicates a lower OSA risk, a score of 5 or less was chosen to predict patients with no OSA. The test accuracy for this score was 80% (95% CI: 73%, 86%).

Having defined the upper threshold score to identify patients with severe OSA (AHI ≥ 30; score ≥ 8) and the lower threshold to rule out OSA (AHI < 5; score ≤ 5), we verified the middle category for the diagnosis of mild-to-moderate OSA. Based on the two-by-two outcome table of true and false positives and negatives (Supplementary Table 2), we calculated sensitivity,

specificity, predictive values, and likelihood ratios (Table 3). In our test cohort, scores above 5 and below 8 predicted mild to severe OSA (5 ≤ AHI < 30) with a sensitivity of 49% (95% CI: 38%, 61%), a specificity of 78% (95% CI: 67%, 87%) and a test accuracy of 63% (95% CI: 55%, 71%).

Score Validation

We retested the scoring system with the validation cohort (n = 50, Table 1) by assessing the prediction power and the accuracy of the score thresholds. The ROC curve was significant for both, the prediction of severe OSA (ROC-AUC = 0.89 with 95% CI: 0.79, 0.90; $p < 0.001$) and exclusion of OSA (ROC-AUC=0.96 with 95% CI: 0.90, 1.00; $p = 0.001$). The thresholds were confirmed as best prediction cut-off values (Table 4, Supplementary Table 3, Supplementary Figures 1 and 2). The score threshold 8 presented a sensitivity of 83% (95% CI: 59%, 96%), specificity of 88% (95% CI: 71%, 97%), and a test accuracy of 86% (95% CI: 73%, 94%) for detecting severe OSA. The score threshold of 5 achieved a sensitivity of 100% (95% CI: 48%, 100%), a specificity

Table 3 Test Cohort (n = 150) – Receiver-Operating Characteristics for a Score ≥8 (Severe OSA), Score ≤5 (Exclusion of OSA) and Score Range 5–8 (Mild to Moderate OSA)

|	Score ≥ 8	Score ≤ 5	5 < Score < 8
ROC-AUC	0.90 (0.84–0.95)	0.89 (0.83–0.94)	n/a
SEN	0.82 (0.65–0.93)	0.72 (0.55–0.85)	0.49 (0.38–0.61)
SPE	0.82 (0.74–0.88)	0.83 (0.75–0.89)	0.78 (0.67–0.87)
PPV	0.57 (0.47–0.67)	0.60 (0.48–0.70)	0.70 (0.59–0.79)
NPV	0.94 (0.88–0.97)	0.89 (0.83–0.93)	0.59 (0.53–0.65)
LR+	4.55 (3.00–6.90)	4.20 (2.66–6.61)	2.26 (1.38–3.67)
LR -	0.22 (0.10–0.45)	0.34 (0.20–0.57)	0.65 (0.50–0.83)
Test accuracy	0.82 (0.75–0.88)	0.80 (0.73–0.86)	0.63 (0.55–0.71)

Notes: Presented are the means with confidence interval 95%. The score was applied in the test cohort. The variables were calculated for a score threshold ≥8 (separating patients with severe obstructive sleep apnea (apnea–hypopnea index ≥30) from patients without severe sleep apnea (apnea–hypopnea index < 30), for a score threshold ≤5 (separating patients without obstructive sleep apnea (exclusion of OSA) from patients with OSA (apnea–hypopnea index ≥ 5)), and for a score range 5–8 to identify patients with mild-to-moderate OSA (apnea–hypopnea index ≥ 5 and below 30)).
Abbreviations: OSA, obstructive sleep apnea; ROC-AUC, receiver-operating characteristic–area under curve; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio.

of 78% (95% CI: 63%, 89%) and a test accuracy of 80% (95% CI: 66%, 90%) for the exclusion of OSA. For the middle category (5 < score < 8), the score reached a sensitivity of 52% (95% CI: 32%, 71%), a specificity of 91% (95% CI: 72%, 99%) and a test accuracy of 70% (95% CI: 55%, 82%).

Discussion

This study was undertaken to develop and test an OSA prescreening tool which utilizes objective measurements and is easy to apply. The score is based exclusively on common anthropometric measures and intraoral visibility. It was developed as a morphological score and does not

ROC Curve
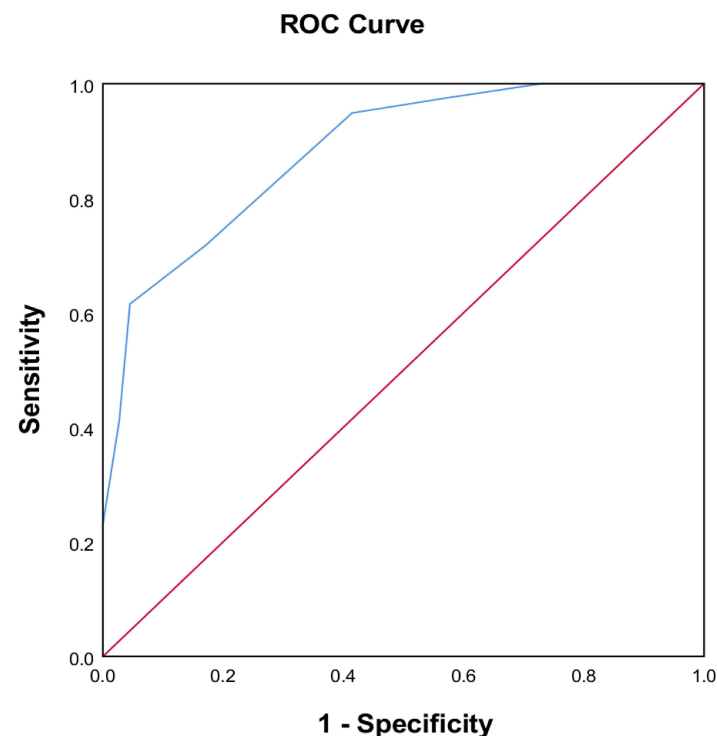
Score Thresholds	SEN	1-SPE
1	0.026	0.000
2	0.231	0.000
3	0.410	0.027
4	0.615	0.045
5	**0.718**	**0.171**
6	0.949	0.414
7	0.974	0.568
8	1.000	0.730
9	1.000	0.856
10	1.000	0.892
11	1.000	0.910
12	1.000	0.946
13	1.000	1.000
14	1.000	1.000

Figure 3 Test cohort – receiver-operating curve and score thresholds for predicting exclusion of obstructive sleep apnea (apnea–hypopnea index < 5).
Notes: Left: ROC curve for score prediction of exclusion of obstructive sleep apnea (apnea–hypopnea index <5) within the test cohort. The curve is defined by sensitivity (true positive rate) on the y-axis and 1-specificity (false positive rate) on the x-axis. Points above the diagonal of the plot represent good results, better than random. Right: Sensitivity and 1-specificity for the different score thresholds. Highlighted (blue, bold): A score of 5 or below presents the best predictive values.
Abbreviations: ROC, receiver-operating characteristics; SEN, sensitivity; SPE, specificity.

Table 4 Validation Cohort (n = 50) – Receiver-Operating Characteristics for a Score ≥8 (Severe OSA), Score ≤5 (Exclusion of OSA) and Score Range 5–8 (Mild to Moderate OSA)

	Score ≥ 8	Score ≤ 5	5 < Score < 8
ROC-AUC	0.89 (0.79–0.90)	0.96 (0.90–1.00)	n/a
SEN	0.83 (0.59–0.96)	1.00 (0.48–1.00)	0.52 (0.32–0.71)
SPE	0.88 (0.71–0.97)	0.78 (0.63–0.89)	0.91 (0.72–0.99)
PPV	0.79 (0.59–0.91)	0.33 (0.22–0.46)	0.88 (0.64–0.97)
NPV	0.90 (0.77–0.96)	1.00 (n/a)	0.62 (0.52–0.71)
LR+	6.90 (2.60–17.06)	3.00 (2.61–7.77)	5.97 (1.51–23.54)
LR -	0.20 (0.07–0.54)	0.00 (n/a)	0.5.3 (0.35–0.80)
Test accuracy	0.86 (0.73–0.94)	0.80 (0.66–0.90)	0.70 (0.55–0.82)

Notes: Presented are the means with confidence interval 95%. The score was applied in the validation cohort. The variables were calculated for a score threshold ≥8 (separating patients with severe obstructive sleep apnea (apnea–hypopnea index ≥30) from patients without severe sleep apnea (apnea–hypopnea index < 30), for a score threshold ≤5 (separating patients without obstructive sleep apnea (exclusion of OSA) from patients with OSA (apnea–hypopnea index ≥ 5)), and for a score range 5–8 to identify patients with mild-to-moderate OSA (apnea–hypopnea index ≥ 5 and below 30)).
Abbreviations: OSA, obstructive sleep apnea; ROC-AUC, receiver-operating characteristic–area under curve; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio.

contain subjective questions, thereby reducing bias related to subjective evaluation measurements.

It differentiates the prediction between severe OSA and the exclusion of OSA with high test accuracy, sensitivity, specificity, and general applicability. It represents an effective and valuable new addition to the clinical tools addressing the most relevant diagnostic questions in suspected OSA patients; however, it does not predict an OSA syndrome (OSAS).

The majority of OSA patients remain undiagnosed with estimates of unrecorded cases of at least 80–90%.[35,36] The past decades have shown that OSA is a heterogeneous disease, with different phenotypes and degrees of severity leading to different outcomes.[37,38] For the sleep physician, an early diagnostic tool is especially relevant to effectively prepare for further diagnosis or treatment. Especially the exclusion of OSA is important in establishing a correct differential diagnosis when presented with common symptoms like severe snoring, trouble staying asleep, or daytime fatigue. OSA can be excluded when the patient presents an AHI smaller than 5 events/hour and without specific symptoms. The prescreening is also especially relevant because an OSA patient without symptoms are often not recognized. This is of great importance for patients with severe OSA, as it may be associated with elevated cardiovascular risk.[29,30] If a patient presents specific symptoms, then prescreening with our score may only indicate the severity of the disease. A continuing diagnostic is necessary, independent of the AHI revealed during prescreening. So, if a patient during prescreening displays an AHI smaller than 5 but with clinically relevant symptoms, a sleep specialist or a physician trained in sleep medicine will initiate further diagnostical steps. With an AHI of 8 or larger but without symptoms, the prescreening is important to identify OSA patients otherwise overlooked and to initiate appropriate and often extremely necessary further treatment based on relevant comorbidity.

Currently, no simple prescreening system – not prone to subjective biases and applicable for wide clinical use without expensive and time-consuming assessment of overnight sleep – exists that could answer these questions with adequate accuracy.[15] The score presented in this study is a simple to use prescreening tool.

Previous scoring systems for predicting OSA have been limited. Next to the time-consuming and expensive systems of overnight sleep assessment (eg, polysomnography) or imaging of airway measurements, the less expensive and easily applied screening instruments include questionnaires, containing items that are subjective measures, potentially associated with inaccuracy.[4,39] Moreover, often the results do not differentiate the predicted outcome in a clinically meaningful way. The popular Epworth Sleepiness Scale (ESS), assessing daytime sleepiness with eight simple questions, only reaches a medium sensitivity of 53% and a medium specificity of 59% for the general diagnosis of OSA.[40,41] The Berlin Questionnaire predicts general OSA without differentiation with a high sensitivity of 93%, but at a cost of a very low specificity of 16%.[42] Another questionnaire, the STOP-Bang differentiates in predicting low, moderate, and severe OSA with a medium sensitivity of 77% and specificity of 45% for having a moderate to severe OSA.[43]

While other prediction scores are based on more objective OSA risk factors that are less prone to bias, they are often validated within very specific populations (eg, obese patients, pregnant women, children, etc.) and do not differentiate OSA severity and lack accuracy.[22–28] The Morphometric Equation, introduced by Kushida, Efron and Guilleminault in 1997, uses only known anatomical OSA risk-factors (oral cavity, neck circumference, BMI) to calculate a probability of OSA. Their score reached a high sensitivity of 98% and specificity of 100% for predicting the presence of OSA, but it was not possible to differentiate degrees of OSA severity. The application within certain OSA patients failed.[26,44,45] The P-SAP (perioperative sleep apnea prediction) score, developed and validated within a specific surgical population, includes several objective demographical, medical, and airway measures, as well as snoring as a symptom. The score is lacking either in sensitivity or specificity (lower scores with a sensitivity of 94% and specificity of 32%, higher scores with a sensitivity of 24% and specificity of 91%).[27] The NoSAS (neck, obesity, snoring, age, sex) score, validated within a sleep clinic, includes objective measures and the subjective symptom of snoring. While sensitivity of 94% was quite high for general OSA, specificity was lacking (30%).[23] Deflandre et al attempted to differentiate and predict the common severity levels of OSA with the DES-OSA prediction score.[24] It is based on the Mallampati Score, distance between thyroid and chin, neck circumference, BMI, gender. While the score was able to differentiate between mild, moderate, and severe OSA with a good combination of relatively high sensitivity (75–83%) and specificity (72–77%), it did not focus on the exclusion of OSA.

We developed a morphological score without subjective questions in order to reduce a possible subjective bias. It was specifically developed for clinical application within sleep medicine. However, its easy applicability may make it a useful tool for general clinical practice as well. The development of our scoring system was guided by previous scores, such as the DES-OSA.[24] The Mallampati Score and the distance between thyroid and chin used by Deflandre et al were replaced with other variables that were highly correlated with the AHI and easier to assess. In particular, the distance between thyroid and chin is difficult to measure with high-risk OSA patients and has been criticized as not being an adequate tool for airway assessment.[46] Instead of the Mallampati Score with protruded tongue position, we used a modified

version of the FTP with a neutral tongue position. The scale was modified to adapt for higher values of OSA patients, and therefore to achieve a more even distribution between the newly adapted four categories with our patient cohort: 7% fit into category I, 33% into category II, 33% into category III, and 26% into category IV. While we used the intraoral photographs obtained as part of SAGIC data collection, intraoral visibility can also be classified without photographs during routine clinical practice. We suggest that the modified FTP is added to the clinical routine when other OSA-related symptoms are seen and/or the patients score high on the more commonly used scores for pharyngeal visibility (eg, Mallampati Score).

Our score was developed on the one hand, to predict severe OSA due to its high risk of cardiovascular comorbidities. Here, with a score of 8 or higher, the accuracy of the test was above 80%. The score reached a high sensitivity and specificity in both the test cohort (82% sensitivity and 82% specificity) and validation cohort (83% sensitivity and 88% specificity). It also reached population-independent likelihood ratios in an acceptable range in both the test cohort (positive likelihood ratio 4.6, negative likelihood ratio 0.2) and validation cohort (positive likelihood ratio 6.9, negative likelihood ratio 0.2). A positive likelihood ratio above 4 and a negative likelihood ratio below 0.3 indicate good applicability of the scoring system with a general population.

On the other hand, the score was supposed to indicate the exclusion of OSA, which was important for an effective differential diagnosis and to avoid unnecessary and costly diagnostic procedures. Here, a "true positive" would mean recognizing a non-OSA patient correctly as not having OSA. A score of 5 or lower showed a test accuracy of 80% in both, the test and the validation cohort, indicating a good test efficacy. The score reached a high specificity of 83% (test cohort), identifying the "negatives" (OSA patients) correctly. The sensitivity of 72% in the test cohort, recognizing the "positives" (non-OSA patients), was not as high but still in an acceptable range. Within the validation cohort, the specificity decreased to 78% and the sensitivity increased to 100%; however, the validation cohort was relatively small, and results may not be as conclusive. Also, the positive likelihood ratios within the test cohort with over 4 indicated a strong general applicability. The score had a higher likelihood of recognizing non-OSA patients and correctly excluding them (true positives) than of diagnosing non-OSA patients with OSA (false negative, Supplementary Table 2).

Limitations

There are certain limitations to the development of our scoring system. This score was developed to focus on the detection of severe OSA and on the exclusion of OSA, an often overlooked feature in a prescreening instrument. The score was not intended to differentiate between the severity of OSA (medium or moderate OSA). It is a prescreening tool that will give first information on how to proceed and which further diagnostic steps to take. It is also noteworthy, that this is a score for screening of OSA, not OSAS. At this time, the prescreening did not focus on further clinical symptoms or cardiovascular symptomatic. This could be included in further studies. Also, we based the score on previous scores and accepted the predictive value of certain high-risk variables. However, we confirmed that those variables were highly correlated with the AHI. Another limitation was the selection bias based on using the SAGIC cohort with suspected OSA patients. OSA was only suspected and nonetheless, 25% of our test cohort and 10% of our validation cohort had an AHI below 5, presenting a complete variation of OSA severity categories within the cohorts. Also, the likelihood ratios, which are population-independent, had values in an acceptable range, indicating a good applicability of the scoring system with a general population and in general clinical practice. However, the validation cohort was quite small and certain results may not be as conclusive as necessary.

We recommend that the score be retested within a larger and more diverse sample, especially regarding the predictive capabilities of the exclusion of OSA due to some marginal values. At that point, it would also be interesting to compare AHI assessment with full overnight polysomnography versus portable polygraphy. Also, to confirm the score's applicability within general clinical practice, the sample should include patients not recruited from sleep clinics, such as patients seen in a primary care setting. The score may need to be adapted if applied in a younger population and regarding sex (eg, waist and neck circumference) and international application, as studies have demonstrated ethnic and sex variations with regard to anatomical risk factors and OSA severity.[47]

Conclusion

This score introduces a new early and easily applied prescreening tool for OSA. It combines objective and precise morphological measurements with predictive power, focusing on answering the most relevant clinical questions, especially regarding the identification of severe OSA patients independent of clinical symptoms. Also, without specific clinical symptoms, a relevant OSA can be excluded. We believe that it can be a valuable tool for widespread screening applicable in general clinical practice and may reduce unnecessary expensive testing in some scenarios.

Disclosure

Outside the submitted work, PAC has an appointment to an endowed academic Chair at the University of Sydney that was created from ResMed funding. He receives no personal fees and this relationship is managed by an Oversight Committee of the University. He has received research support from ResMed, SomnoMed, Zephyr Sleep Technologies, and Bayer. He is a consultant/adviser to Zephyr Sleep Technologies, Signifier Medical Technologies, SomnoMed, ResMed, and Bayer. He has a pecuniary interest in SomnoMed related to a previous role in R&D (2004). AIP is the John Miclot Professor of Medicine. Funds for this endowment are provided by the Phillips Respironics Foundation. RS reports grants from ResMed, Inspire, CryOSA; he is on the Advisory Board for eXciteOSA and has royalties from UptoDate and Merck Manual, outside the submitted work. IF has received research grants from Löwenstein, ResMed, Weinmann and Philips at the Charite University Hospital. TP reports grants from Cidelec, grants and personal fees from Löwenstein Medical, grants from Novartis, personal fees from Jazz Pharma, Bayer Healthcare, Cerebra, Philips, and Neuwirth, speaker fee from National Sleep Foundation, outside the submitted work; and shareholder of Advanced Sleep Research, The Siestagroup GmbH, Nukute. The authors report no other conflicts of interest in this work.

References

1. Rosenthal LD, Dolan DC. The Epworth sleepiness scale in the identification of obstructive sleep apnea. *J Nev Ment Dis.* 2008;196 (5):429–431. doi:10.1097/NMD.0b013e31816ff3bf

2. Rizzatti FG, Mazzotti DR, Mindel J, et al. Defining extreme phenotypes of obstructive sleep apnea across international sleep centers. *Chest*. 2020;pii:S0012-3692(20) 30680-2. doi:10.1016/j.chest.2020.03.055

3. Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687–698. doi:10.1016/S2213-2600(19)30198-5

4. Fietze I, Laharnar N, Obst A, et al. Prevalence and association analysis of obstructive sleep apnea with gender and age differences - results of SHIP-trend. *J Sleep Res*. 2019;28(5):e12770. doi:10.1111/jsr.12770

5. Senaratna CV, Perret JL, Lodge CJ, et al. Prevalence of obstructive sleep apnea in the general population: a systematic review. *Sleep Med Rev*. 2016;34:70–81. doi:10.1016/j.smrv.2016.07.002

6. Andrade AG, Bubu OM, Varga AW, Osorio RS. The relationship between obstructive sleep apnea and alzheimer's disease. *J Alzheimers Dis*. 2018;64(s1):S255–S270. doi:10.3233/JAD-179936

7. Heinzer R, Vat S, Marques-Vidal P, et al. Prevalence of sleep-disordered breathing in the general population: the HypnoLaus study. *Lancet Respir Med*. 2015;3(4):310–318. doi:10.1016/S2213-2600(15)00043-0

8. Kerner NA, Roose SP. Obstructive sleep apnea is linked to depression and cognitive impairment: evidence and potential mechanisms. *Am J Geriatr Psychiatry*. 2016;24(6):496–508. doi:10.1016/j.jagp.2016.01.134

9. Newman AB, Nieto FJ, Guidry U, et al. Relation of sleep-disordered breathing to cardiovascular disease risk factors. *Am J Epidemiol*. 2001;154(1):50–59. doi:10.1093/aje/154.1.50

10. Shahar E, Whitney CW, Redline S, et al. Sleep-disordered breathing and cardiovascular disease—cross-sectional results of the Sleep Heart Health Study. *Am J Respir Crit Care Med*. 2001;163(1):19–25. doi:10.1164/ajrccm.163.1.2001008

11. Schipf S, Werner A, Tamayo T, et al. Regional differences in the prevalence of known type 2 diabetes mellitus in 45–74 years old individuals: results from six population-based studies in Germany (DIAB-CORE Consortium). *Diabet Med*. 2012;29(7):e88–e95. doi:10.1111/j.1464-5491.2012.03578.x

12. Somers VK, White DP, Amin R, et al. Sleep apnea and cardiovascular disease: an American Heart Association/American College of Cardiology Foundation scientific statement from the American Heart Association Council for High Blood Pressure Research Professional Education Committee, Council on Clinical Cardiology, Stroke Council, and Council on Cardiovascular Nursing. *Circ*. 2008;118 (10):1080–1111.

13. Trost M, Machleit-Ebner A, Neddermann M, Young P. Assessment of ability to drive in obstructive sleep apnea patients using a driving simulator. *Somnologie*. 2019;23(1):36–42. doi:10.1007/s11818-019-0192-x

14. Young T, Finn L, Peppard PE, et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. *Sleep*. 2008;31(8):1071–1078.

15. Jonas DE, Amick HR, Feltner C. Screening for obstructive sleep apnea in adults: evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2017;317(4):415–433. doi:10.1001/jama.2016.19635

16. Flemons WW, Littner MR. Measuring agreement between diagnostic devices. *Chest*. 2003;124(4):1535–1542. doi:10.1378/chest.124.4.1535

17. Chiu HY, Chen PY, Chuang LP, et al. Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: a bivariate meta-analysis. *Sleep Med Rev*. 2017;36:57–70. doi:10.1016/j.smrv.2016.10.004

18. Li Z, Tang XD. [The accuracy of STOP-Bang questionnaire in the screening of patients with obstructive sleep apnea]. *Zhonghua Yi Xue Za Zhi*. 2020;100(14):1057–1062. Chinese. doi:10.3760/cma.j.cn112137-20191031-02365

19. Seneviratne U, Puvanendran K. Excessive daytime sleepiness in obstructive sleep apnea: prevalence, severity, predictors. *Sleep Med*. 2004;5(4):339–343. doi:10.1016/j.sleep.2004.01.021

20. Shahid A, Shen J, Shapiro CM. Measurements of sleepiness and fatigue. *J Psychosom Res*. 2010;69(1):81–89. doi:10.1016/j.jpsychores.2010.04.001

21. Lee RW, Petocz P, Prvan T, Chan AS, Grunstein RR, Cistulli Pa. Prediction of obstructive sleep apnea with craniofacial photographic analysis. *Sleep*. 2009;32(1):46–52.

22. Ahlin S, Manco M, Panunzi S, et al. A new sensitive and accurate model to predict moderate to severe obstructive sleep apnea in patients with obesity. *Med*. 2019;98(32):e16687. doi:10.1097/MD.00 00000000016687

23. Costa JC, Rebelo-Marques A, Machado JN, et al. Validation of NoSAS (neck, obesity, snoring, age, sex) score as a screening tool for obstructive sleep apnea: analysis in a sleep clinic. *Pulmonology*. 2019;25(5):263–270. doi:10.1016/j.pulmoe.2019.04.004

24. Deflandre E, Degey S, Brichant J-F, Poirrier R, Bonhomme V. Development and validation of a morphologic obstructive sleep apnea prediction score: the DES-OSA score. *Anesth Analg*. 2016;122(2):363–372. doi:10.1213/ANE.0000000000001089

25. Izci-Balserak B, Zhu B, Gurubhagavatula I, Keenan BT, Pien GW. A screening algorithm for obstructive sleep apnea in pregnancy. *Ann Am Thorac So*. 2019;16(10):1286–1294. doi:10.1513/AnnalsATS.201902-131OC

26. Kushida CA, Efron B, Guilleminault C. A predictive morphometric model for the obstructive sleep apnea syndrome. *Ann Intern Med*. 1997;127(8 Pt 1):581–587. doi:10.7326/0003-4819-127-8_Part_1-199710150-00001

27. Ramachandran SK, Kheterpal S, Consens F, et al. Derivation and validation of a simple perioperative sleep apnea prediction score. *Anesth Anal*. 2010;110(4):1007–1015. doi:10.1213/ANE.0b013e318 1d489b0

28. Soh HJ, Rowe K, Davey MJ, Horne RSC, Nixon GM. The OSA-5: development and validation of a brief questionnaire screening tool for obstructive sleep apnea in children. *Int J Pediatr Otorhinolaryngol*. 2018;113:62–66. doi:10.1016/j.ijporl.2018.07.029

29. Marshall NS, Wong KK, Cullen SR, Knuiman MW, Grunstein RR. Sleep apnea and 20-year follow-up for all-cause mortality, stroke, and cancer incidence and mortality in the Busselton Health Study cohort. *J Clin Sleep Med*. 2014;10(4):355–362. doi:10.5664/jcsm.3600

30. Wang X, Ouyang Y, Wang Z, Zhao G, Liu L, Bi Y. Obstructive sleep apnea and risk of cardiovascular disease and all-cause mortality: a meta-analysis of prospective cohort studies. *Int J Cardiol*. 2013;169 (3):207–214. doi:10.1016/j.ijcard.2013.08.088

31. Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med*. 2012;8(5):597–619. doi:10.5664/jcsm.2172

32. Lee RWW, Chan ASL, Grunstein RR, Cistulli PA. Craniofacial phenotyping in obstructive sleep apnea – a novel quantitative photographic approach. *Sleep*. 2009;32(1):37–45.

33. Schwab RJ, Leinwand SE, Bearn CB, et al. Digital morphometrics: a new upper airway phenotyping paradigm in OSA. *Chest*. 2017;152 (2):330–342. doi:10.1016/j.chest.2017.05.005

34. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29–36. doi:10.1148/radiology.143.1.7063747

35. 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. doi:10.1093/aje/kws342

36. Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep*. 1997;20(9):705–706. doi:10.1093/sleep/20.9.705

37. Saaresranta T, Hedner J, Bonsignore MR, et al. Clinical phenotypes and comorbidity in European sleep apnoea patients. *PLoS One*. 2016;11(10):e0163439. doi:10.1371/journal.pone.0163439

38. Keenan BT, Kim J, Singh B, et al. Recognizable clinical subtypes of obstructive sleep apnea across international sleep centers: a cluster analysis. *Sleep J.* 2018;41(3):1–14.

39. Chaudhry U, Cohen JR, Al-Samawi Y. Use of cone beam computed tomography imaging for airway measurement to predict obstructive sleep apnea. *Cranio.* 2020;12:1–7. doi:10.1080/08869634.2020.1765602

40. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep.* 1991;14(6):540–545. doi:10.1093/sleep/14.6.540

41. Kicinski P, Przybylska-Kuc SM, Tatara K, et al. Reliability of the Epworth Sleepiness Scale and the Berlin Questionnaire for screening obstructive sleep apnea syndrome in the context of the examination of candidates for drivers. *Med Pr.* 2016;67(6):721–728. doi:10.13075/mp.5893.00494

42. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. *Ann Intern Med.* 1999;131(7):485–491. doi:10.7326/0003-4819-131-7-199910050-00002

43. Orbea CAP, Lloyd RM, Faubion SS, Miller VM, Mara KC, Kapoor E. Predictive ability and reliability of the STOP-BANG questionnaire in screening for obstructive sleep apnea in midlife women. *Maturitas.* 2020;135:1–5. doi:10.1016/j.maturitas.2020.02.004

44. Sares MCM, Bittencourt LR, Zonato AL, Gregorio LC. Application of the Kushida morphometric model in patients with sleep-disordered breathing. *Braz J Otorhinolaryngol.* 2006;72(4):541–548. doi:10.1016/S1808-8694(15)31002-8

45. Islam S, Aleem F, Ormiston IW. Does the Kushida morphometric model predict outcomes following maxillomandibular advancement surgery for obstructive sleep apnoea? *J Craniomaxillofac Sur.* 2014;42(8):1675–1678. doi:10.1016/j.jcms.2014.05.012

46. Forozeshfard M, Razavi MR. Airway assessment: predictor of thyromental distance in the patients with chin augmentation surgery. *Br J Anaesth.* 2016;117(eLetters Supplement):0007–0912. doi:10.1093/bja/el_13354

47. Sutherland K, Keenan BT, Bittencourt L, et al. A global comparison of anatomic risk factors and their relationship to obstructive sleep apnea severity in clinical samples. *J Clin Sleep Med.* 2019;15(4):629–639. doi:10.5664/jcsm.7730