

SCIENTIFIC INVESTIGATIONS

A sleep apnea prediction model developed for African Americans: the Jackson Heart Sleep Study

Dayna A. Johnson, PhD, MPH, MS, MSW^{1,2}; Tamar Sofer, PhD^{2,3}; Na Guo, MS²; James Wilson, MD⁴; Susan Redline, MD, MPH^{2,3,5}

¹Department of Epidemiology, Emory University, Atlanta Georgia; ²Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, Massachusetts; ³Department of Sleep Medicine, Harvard Medical School, Cambridge, Massachusetts; ⁴Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi; ⁵Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts

Study Objectives: African Americans have a high prevalence of severe sleep apnea that is often undiagnosed. We developed a prediction model for sleep apnea and compared the predictive values of that model to other prediction models among African Americans in the Jackson Heart Sleep Study.

Methods: Participants in the Jackson Heart Sleep Study underwent a type 3 home sleep apnea study and completed standardized measurements and questionnaires. We identified 26 candidate predictors from 17 preselected measures capturing information on demographics, anthropometry, sleep, and comorbidities. To develop the optimal prediction model, we fit logistic regression models using all possible combinations of candidate predictors. We then implemented a series of steps: comparisons of equivalent models based on the C-statistics, bootstrap to evaluate the finite sample properties of the C-statistics between models, and fivefold cross-validation to prevent overfitting.

Results: Of 719 participants, 38% had moderate or severe sleep apnea, 34% were male, and 38% reported habitual snoring. The average age and body mass index were 63.2 (standard deviation:10.7) years and 32.2 (standard deviation: 7.0) kg/m². The final prediction model included age, sex, body mass index, neck circumference, depressive symptoms, snoring, restless sleep, and witnessed apneas. The final model has an equal sensitivity and specificity of 0.72 and better predictive properties than commonly used prediction models.

Conclusions: In comparing a prediction model developed for African Americans in the Jackson Heart Sleep Study to widely used screening tools, we found a model that included measures of demographics, anthropometry, depressive symptoms, and sleep patterns and symptoms better predicted sleep apnea.

Keywords: sleep apnea, prediction model, epidemiology, African Americans, Jackson Heart Study

Citation: Johnson DA, Sofer T, Guo N, Wilson J, Redline S. A sleep apnea prediction model developed for African Americans: the Jackson Heart Sleep Study. *J Clin Sleep Med.* 2020;16(7):1171–1178.

BRIEF SUMMARY

Current Knowledge/Study Rationale: With the growing recognition of the burden of sleep apnea, particularly among African Americans, it is important to develop screening tools for use. Also, the predictive ability of commonly used screening tools among African Americans is unknown.

Study Impact: Using a range of measures we were able to develop internally validated screening models, which generally had a better prediction performance than other well-known screening tools such as STOP-Bang, NoSAS (Neck, Obesity, Snoring, Age, Sex), and the Hispanic Community Health Study. The prediction models we developed have utility in both clinical settings and research.

INTRODUCTION

Obstructive sleep apnea (OSA) is a highly prevalent sleep disorder, estimated to occur in approximately 25 million American adults, and is increasing in prevalence.^{1,2} The prevalence is higher among males, individuals who are overweight/obese, and racial/ethnic minorities.^{1,3,4} OSA is a disorder in which recurrent periods of apneas and hypopneas occur, resulting in intermittent hypoxemia and sleep disruption.⁵ The disorder is often undiagnosed and untreated,^{6,7} particularly among African Americans.⁸ Epidemiologic studies have implicated OSA as associated with several chronic conditions including obesity, diabetes, hypertension, stroke, and depression.^{9–11} To potentially reduce its public health burden, it is important to screen individuals at increased risk for OSA, such as African Americans, who have a high prevalence but commonly undiagnosed and untreated.

Snoring is one of the most commonly recognized symptoms of sleep apnea.¹ It is estimated that approximately 25–50% of individuals who snore loudly have OSA. Information on self-reported snoring is more commonly available in large cohort studies and is more easily ascertained clinically than objectively measured sleep apnea. However, relying solely on self-reports of snoring is not adequate for OSA screening. Snoring is often underreported, and lack of bed partners to report snoring may introduce misclassification. Furthermore, population groups may variably report snoring information. For example, the Cleveland Family Study reported that among individuals with OSA, snoring was underreported in African American males, those without a bed partner, and those with less than a high school education.¹² Snoring may also be underreported among women, in whom symptoms of depressed mood and insomnia may co-occur with OSA.^{13,14} Given the high prevalence of

undiagnosed OSA among African American men and women, there is a clear need to develop a prediction model for OSA screening that is inclusive of several OSA symptoms beyond snoring to improve prediction of OSA in this population.

There are several validated screening tools for sleep apnea, such as the STOP-Bang questionnaire,¹⁵ Berlin,¹⁶ Neck, Obesity, Snoring, Age, Sex (NoSAS),¹⁷ OSA-50,¹⁸ and Multivariate Apnea Prediction model.¹⁹ These scales assess a variety of symptoms and clinical conditions including snoring, age, sex, witnessed choking/gasping or stopped breathing, obesity, sleepiness, high blood pressure, and others. Most of the scales were validated among a population that was mostly non-Hispanic white, which limits generalizability. More recently, a 4-item (age, sex, snoring, and body mass index [BMI]) prediction equation for sleep apnea was developed among a Hispanic population.²⁰ The ability to predict sleep apnea based on this prediction equation for African Americans is unknown. Given that African Americans have a high prevalence of obesity,²¹ measurement of body fat distribution, such as neck or waist circumference, may provide better prediction than BMI.

We developed a comprehensive prediction model for OSA among African Americans in the Jackson Heart Sleep Study (JHSS), an epidemiologic study of sleep disorders and risk factors in African Americans. To develop the model, we considered demographic, adiposity, sleep symptoms, and comorbidity data. We evaluated various combinations of the data to an objective measure of OSA derived from in-home sleep apnea testing. We also compared the predictive power of the new model to well-established screening models such as STOP-Bang questionnaire, NoSAS, and the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) model. We performed a rigorous assessment of model performances using bootstrap and cross-validation to prevent overfitting to the specific dataset.

METHODS

The Jackson Heart Study (JHS) is a longitudinal study of 5,306 African American adults 21–95 years of age from 3 counties in Jackson, Mississippi (Hinds, Madison, and Rankin). The details of the JHS design have been previously published.²² In brief, JHS was designed to prospectively assess the etiology of cardiovascular disease among African Americans. The baseline assessment was between September 2000 and March 2004, followed by 2 examinations. The current paper uses data from the JHSS, which was conducted between December 2012 and May 2016 after the third examination. Institutional review board approval was obtained from the University of Mississippi and Partners Research Committee, and written informed consent was obtained from all participants.

In brief, eligible participants in the JHSS were those who participated in the third JHS follow-up examination and other ancillary studies. Details regarding the recruitment into the JHSS has been previously published.⁸ Participants attended a clinic visit and underwent in-home sleep apnea testing, 1-week wrist actigraphy, fasting venipuncture, anthropometry, blood pressure and other vascular studies, and completed interviewer-administered sleep and health questionnaires.

Sleep Measures

Sleep apnea was measured with a validated type 3 home sleep apnea device (Embletta-Gold device, Embla, Broomfield, CO),^{23,24} recording nasal pressure (measuring airflow); thoracic and abdominal inductance plethysmography; finger pulse oximetry; body position; and electrocardiography. The respiratory event index (REI) was derived by as the sum of all apneas plus hypopneas associated with $\geq 3\%$ desaturation divided by the estimated sleep time. Sleep apnea was characterized by the standard REI categories of moderate or more severe OSA as $REI \geq 15$ compared with none/mild ($REI < 15$). In sensitivity analyses, the REI was calculated based on all apneas plus hypopneas associated with a $\geq 4\%$ desaturation.

From the self-administered sleep questionnaire, we collected information on participants' sleep patterns and symptoms. The sleep questionnaire was composed of the components based on the HCHS sleep questionnaire, STOP-bang, NoSAS, and the Epworth Sleepiness Scale. Frequency of snoring, witnessed breathing pauses, trouble falling asleep, and multiple awakenings at night were assessed on a 5-point Likert scale (0, <1, 1–2 times, 3–4 times, and ≥ 5 times per week), which were further dichotomized into binary variables (<3 and ≥ 3 times per week). "Don't know" was added as the third category of the final snoring variable for those who reported so regarding their snoring status. Restless or very restless sleep was dichotomized from self-reported quality of typical night's sleep (very sound or restful, sound to restful, average quality, restless, and very restless). Sleepiness was measured by the Epworth Sleepiness Scale, which has 8 questions asking participants to rate the likelihood of falling asleep under 8 scenarios on a 4-point Likert scale (0–3). The sum of 8-item scores (Epworth Sleepiness Scale score), ranging from 0 to 24, was dichotomized to indicate excessive daytime sleepiness (Epworth Sleepiness Scale score > 10).²¹ Weekly average sleep duration was calculated as the weighted average of self-reported sleep duration on weekdays and weekends. Participants who reported napping for at least 5 minutes once a week or more often were classified as having a napping habit.²⁰

Demographics

Participants reported date of birth and sex (male or female).

Anthropometry

Trained staff following a standardized protocol measured height, weight, and waist and neck circumference. BMI was calculated as $\text{weight}/\text{height}^2$ (kg/m^2).

Medical conditions

Depressive symptoms were assessed by the Center for Epidemiologic Studies of Depression scale. The Center for Epidemiologic Studies of Depression scale is a standardized, 20-item, self-reported instrument that measures the frequency of recently experienced depressive symptoms.²⁵ Participants who have a Center for Epidemiologic Studies of Depression total score ≥ 16 were classified as having high depressive symptoms. Seated blood pressure measurements were obtained using an Omron HEM907XL blood pressure monitor (OMRON IntelliSense Professional Digital Blood Pressure Monitor, Bannockburn, IL) after 5 minutes of rest. Three seated blood pressure readings

were taken 1 minute apart, and the last 2 were averaged. Hypertension was defined as a systolic blood pressure ≥ 130 mm Hg or a diastolic blood pressure ≥ 80 mm Hg, use of antihypertensive medications (self-report or identified from a medication inventory), or self-reported history of hypertension. Diabetes was defined as fasting glucose ≥ 126 mg/dL, use of antidiabetic medication, or self-reported diabetes diagnosis. History of heart diseases was defined as having any of the following: self-reported heart attack, heart bypass, stent procedure, or heart failure.

STOP-Bang score

STOP-Bang questionnaire is a validated screening tool for sleep apnea, including 8 dichotomous questions regarding loud snoring, frequent tiredness/fatigue/sleepiness, observed breathing pause/choking/gasping, high blood pressure, obesity, age, neck size, and sex.¹⁵ Similar questions were collected in JHSS. A STOP-Bang total score was calculated by adding up the number of positive endorsement of 8 modified criteria: (1) snoring as louder than talking or very loud that can be heard in adjacent rooms in the last month, (2) responding ≥ 3 times per week to any of the following questions: “Did you feel overly sleepy during the day?/In the last 4 weeks, how often have you felt tired or fatigued after your sleep?/During your waking time in the last 4 weeks, have you felt tired, fatigued or not up to par?,” (3) reporting at least once a week for witnessed breathing pauses, (4) meeting the aforementioned definition of hypertension, (5) BMI > 35 kg/m², (6) older than 50 years of age; (7) neck circumference ≥ 43 cm for male or ≥ 41 cm for female; and (8) self-identified as male. We evaluated the predictive properties of the STOP-Bang score under the widely used threshold (≥ 3) for moderate or high risk of OSA and the optimal cutoff score that maximized Youden’s index in our analytic sample.

NoSAS score

A group of researchers in Switzerland developed a simple clinical tool (NoSAS score) to screen people at risk for significant sleep-disordered breathing.¹⁷ Only neck circumference, overweight and obesity status, snoring, age, and sex were needed for calculating the NoSAS score: 4 points for neck size > 40 cm, 3 points for overweight (BMI = 25 to < 30 kg/m²), 5 points for obese (BMI ≥ 30 kg/m²), 2 points for snoring, 4 points for older age (> 55 years), and 2 points for being male. We assessed the prediction performance of the NoSAS score using the recommended cutoff score (≥ 8) in addition to a threshold identified in our sample to be optimal.

HCHS sleep apnea prediction model

Shah et al²⁰ developed a risk calculator for predicting sleep apnea in a cohort of Hispanic/Latino adults enrolled in the HCHS/SOL. We calculated the predicted probability using their prediction equations: $1/(1 + \exp[-(-10.2561 + 0.0655 \times \text{Age} + 0.1391 \times \text{BMI} + 0.7006 \times [1 \text{ if male, } 0 \text{ if female}] + 0.9481 \times [\text{if snoring} \geq 3 \text{ times a week, } 0 \text{ if otherwise}]^{20} + 0.1012 \times [1 \text{ if doesn't know snoring status, } 0 \text{ if otherwise}]])$). Its prediction performance was evaluated under the given cutoff probability (≥ 0.12) and the optimal threshold probability in our sample.

Statistical approach

Participants with complete data on sleep apnea and candidate predictors were included in the analytic sample. Frequency and

percentage of categorical variables and mean and standard deviation of continuous variables were generated among the overall sample and by sleep apnea status. We conducted χ^2 tests and Wilcoxon rank-sum tests to compare the distribution of categorical and continuous measures between participants with and without moderate-severe OSA as defined by the REI.

To identify nonlinear terms in the prediction model, we allowed for nonlinear terms (interactions, squared, and cubic terms) if they passed the following screening step. For age, BMI, and neck and weight circumferences, we conducted likelihood ratio tests to examine the significance of quadratic and cubic terms in improving the goodness of fit of unadjusted logistic regression models separately for each predictor, retaining variables with $P < .05$. To identify potential modification of each of these by sex, we tested interaction terms and retained these based on $P < .1$.

We constructed a base prediction model that included age, BMI, sex, and the two snoring variables described earlier. We considered additional models that include additional predictors. We fitted all possible models by adding different combinations of candidate predictors to the basic model while forcing in lower-order or main-effect terms if nonlinear or interaction terms were present (eg, if cubic age was included, linear and quadratic age was included as well). Area under the receiver operating characteristic curve, also called the C-statistic, was used to compare the prediction performance for all candidate models. We rounded the C-statistic to the third digit, assuming that greater precision is not meaningful.

To select the best prediction model, we took a series of steps, with the following goals: (1) find the best model for each possible complexity level (defined by degrees of freedom); (2) remove models for which simpler models have equivalent performance; and (3) select the final probability model as the one that has the best out-of-sample performance. In detail, we first grouped the models by degrees of freedom and the model(s) with (tied) maximum C-statistic within each group were carried over to the next step. Second, we generated 500 bootstrap datasets with a sample size of 719 participants that were randomly resampled from the analytic sample with replacement. We evaluated the C-statistic of the retained models in each bootstrap dataset to examine the variability of its prediction performance over 500 samples. Then, we started from the basic (least complex) model and compared it with the model with the next complexity level (one degree of freedom greater). To compare these models, we obtained the bootstrap distribution of differences in C-statistic. If the median difference was ≥ 0.01 , we kept the more complex model. Otherwise, we removed it from further considerations and took the simpler model for the next comparison. We repeated this procedure until all the models from the first step were evaluated. Third, we conducted fivefold cross-validation to evaluate the out-of-sample prediction performance of the models that were retained from the second step. We partitioned the analytic dataset evenly into 5 subsets, fitted each remaining model using 4 subsets as training datasets and validated them using the fifth subset as a testing dataset. Average C-statistic was calculated over 5 training and testing datasets. The model with the greatest average C-statistic among the testing dataset was chosen as the final prediction model.

Standard measures of prediction model accuracy are specificity, sensitivity, and Youden’s index. Sensitivity is the

Table 1—Selected study participant characteristics by sleep apnea severity (n = 719).

Characteristic	Overall (n = 719, 100%)	REI ≥ 15 (n = 272, 38%)	REI < 15 (n = 447, 62%)	P
Demographics				
Male, n (%)	242 (33.7%)	110 (40.4%)	132 (29.5%)	.003
Age (y), mean ± SD	63.2 ± 10.7	64.5 ± 10.3	62.5 ± 10.9	.029
Anthropometry				
BMI (kg/m ²), mean ± SD	32.2 ± 7.0	34.7 ± 7.3	30.6 ± 6.4	<.001
Waist circumference (cm), mean ± SD	106.2 ± 16.3	112.5 ± 15.7	102.4 ± 15.4	<.001
Neck circumference (cm), mean ± SD	38.6 ± 4.1	40.2 ± 4.3	37.7 ± 3.6	<.001
Sleep patterns				
Average sleep duration (h), mean ± SD	6.4 ± 1.5	6.5 ± 1.6	6.4 ± 1.4	.315
Naps (≥once a week), n (%)	409 (56.9%)	173 (63.6%)	236 (52.8%)	.005
Restless or very restless sleep, n (%)	127 (17.7%)	58 (21.3%)	69 (15.4%)	.045
Multiple awakenings at night, n (%)	266 (37.0%)	106 (39.0%)	160 (35.8%)	.392
Sleep quality symptoms				
Trouble falling asleep, n (%)	124 (17.3%)	42 (15.4%)	82 (18.3%)	.318
Sleep apnea symptoms				
Snoring, n (%)				<.001
≥3 times a week	273 (38.0%)	134 (49.3%)	139 (31.1%)	
<3 times a week	257 (35.7%)	87 (32.0%)	170 (38.0%)	
Don't know	189 (26.3%)	51 (18.7%)	138 (30.9%)	
Witnessed apneas (≥3 times a week), n (%)	32 (4.5%)	18 (6.6%)	14 (3.1%)	.028
Sleepiness (ESS > 10), n (%)	148 (20.6%)	61 (22.4%)	87 (19.5%)	.341
Medical conditions				
High depressive symptoms (CESD-20 ≥ 16), n (%)	131 (18.2%)	44 (16.2%)	87 (19.5%)	.268
Hypertension, n (%)	612 (85.1%)	241 (88.6%)	371 (83.0%)	.041
Diabetes, n (%)	197 (27.4%)	92 (33.8%)	105 (23.5%)	.003
History of heart diseases, n (%)	43 (6.0%)	19 (7.0%)	24 (5.4%)	.376

BMI = body mass index, CESD = Center for Epidemiologic Studies of Depression, ESS = Epworth Sleepiness Scale, REI = respiratory event index.

proportion of individuals classified as high risk of sleep apnea out of those with sleep apnea; specificity is the proportion of individuals classified as low risk of sleep apnea out of those without sleep apnea; and Youden's index is defined as the sum of sensitivity and specificity – 1. To compare the predictive properties of our models with other screening tools, 3 more logistic regression models were fit using the analytic dataset with continuous STOP-Bang score, NoSAS score, and HCHS sleep apnea prediction equation as a single predictor. For these three models and our retained models, we obtained the C-statistic, used the DeLong method (pROC packaged in R) to compute its 95% confidence interval, found the optimal cutoffs that maximized Youden's index, and report the sensitivity, specificity, and Youden's index for each model under the optimal and recommended cutoffs.

SAS version 9.4 (SAS Institute, Inc., Cary, NC) was used to generate descriptive statistics and screen for nonlinear terms. R 3.4.0 (R Foundation for Statistical Computing, Vienna, Austria) and R package pROC (version 1.10.0), caret (version 6.0-77), and PredictABEL (version 1.2-2) were applied for model development and evaluation procedures.

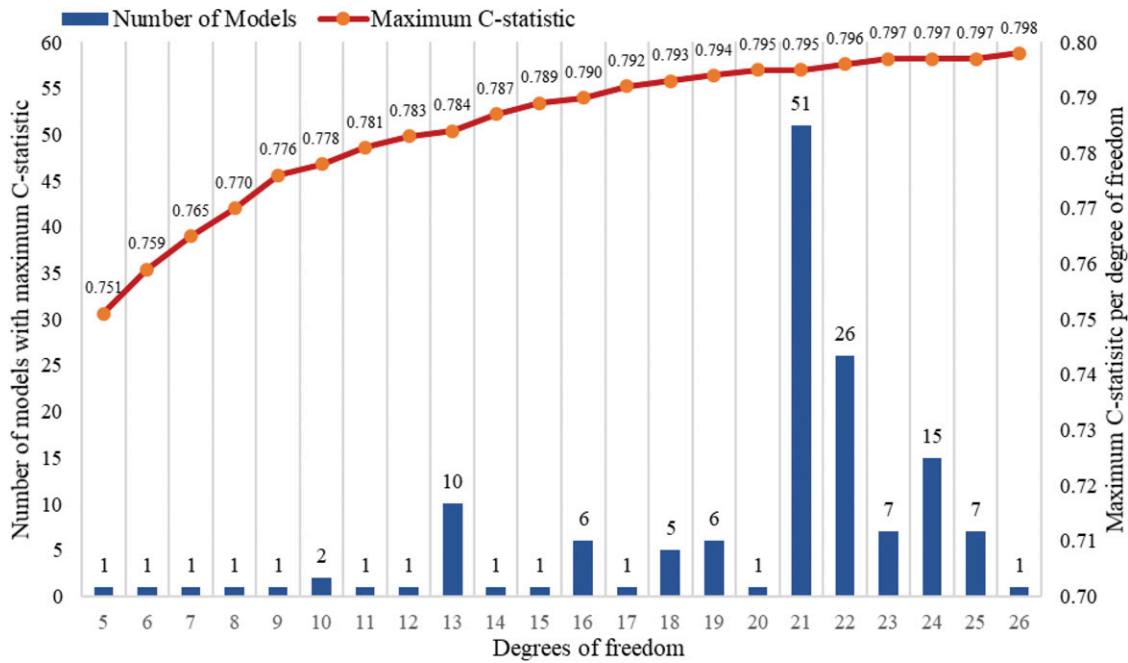
RESULTS

A total of 719 JHSS participants were included in the analyses, of whom, 38% have moderate or severe sleep apnea. The descriptive statistics of measures of demographics, anthropometry, sleep patterns and symptoms, and medical conditions among the total sample and by sleep apnea status are presented in **Table 1**. Overall, 34% were male and 38% were habitual snorers. The mean age and BMI were 63.2 (standard deviation: 10.7) years and 32.2 (standard deviation: 7.0) kg/m², respectively. On average, participants with sleep apnea were older, had larger BMI and neck and waist circumferences, and were more likely to be male, report napping, have restless or very restless sleep, snore habitually, have witnessed apneas, and have measured hypertension and diabetes.

Model development

Based on our predefined covariates, and following the criteria for inclusion of higher order terms (quadratic, cubic, and interactions) 26 variables were selected as candidate predictors, including age, age², age³, waist, waist², waist³, neck, neck², neck³, sex, BMI,

Figure 1—Number of models with maximum C-statistic per degrees of freedom.



X-axis represents the number of degrees of freedom in the model, which ranges from 5 to 26. Blue bars were used to mark the number of models with the same degrees of freedom and maximum C-statistic, whose values correspond to the Y-axis on the left side. The red line above the bars was connected by the orange dots showing the maximum C-statistic per degrees of freedom (Y-axis on the right side). For example, there are 51 models with 21 degrees of freedom that have a C-statistic of 0.795.

BMI², average sleep duration, snoring, napping habit, restless or very restless sleep, multiple awakenings at night, trouble falling asleep, witnessed apneas, sleepiness, high depressive symptoms, hypertension, diabetes, history of heart diseases, and a sex-interaction term with sleepiness. We constructed all logistic regression models that included the basic variables (age, sex, BMI, snoring) and all other combinations of predictors. These formed 22 groups by levels of complexity (degrees of freedom, range: 5–26). Of these, 147 models with (tied) maximum C-statistic per group were retained. **Figure 1** shows the distribution of tied model numbers and maximum C-statistic per group.

Model performance and cross-validation

In the second step, we removed 142 models because simpler models had equivalent performance based on the bootstrap analysis. Five models remained for the final cross-validation step (see **Table S1** for statistical models in the supplemental material). **Table 2** presents the C-statistics (95% confidence interval) of these five retained models that were evaluated in the complete dataset and by fivefold cross-validation. The model with the best cross-validation performance included age, BMI, male sex, snoring, restless or very restless sleep, BMI², neck size, neck size², high depressive symptom, witnessed apneas, age², and age³, and had an average C-statistic of .76 computed on the test datasets.

Comparison of proposed and existing prediction models

Table 3 provides various performance measures across the existing and the proposed prediction models, computed over

JHSS participants. The optimal computed thresholds for the HCHS, NoSAS, and STOP-Bang prediction models that maximized the sum of sensitivity and specificity were all larger than the recommended thresholds, which are ≥ 4 versus 3 for STOP-Bang score, ≥ 10 versus 8 for NoSAS score, and ≥ 0.35 versus 0.12 for HCHS prediction probability. **Table 3** shows the frequency and prevalence of positive and accurate predictions and predictive properties under optimal and recommended cutoffs for each prediction method. Under the recommended cutoffs, high sensitivity (0.79–0.96) and low specificity (0.28–0.45) were observed for the STOP-Bang score, NoSAS score, and HCHS prediction probability, resulting in a large proportion of false-positive findings and prediction accuracy less than 60%. Application of population-specific cutoffs improved the prediction accuracy by 10–17% for these 3 methods. Of all the retained models, our model 3 (age, BMI, male sex, snoring, restless or very restless sleep, neck size, age², and age³) under the optimal cutoff prediction probability (≥ 0.44) had the best prediction accuracy of 74% because of its high specificity (0.81). Our best prediction model 4 (age, BMI, male sex, snoring, restless or very restless sleep, BMI², neck size, neck size², high depressive symptom, witnessed apneas, age², and age³) had the second-best prediction accuracy of 72%, with equally acceptable sensitivity and specificity (0.72).

In secondary analyses, we repeated the analyses with 4% REI. There were fewer candidate models (**Figure S1**), with 4 final models (**Table S2**). Model 3, which included age, BMI, male sex, snoring, restless or very restless sleep, neck size, witnessed apneas, depressive symptoms and average sleep

Table 2—C-statistic of 5 models retained after Bootstrap in the final step of 5-fold cross-validation. C-statistics are provided for the full analytic data set and averaged in the 5-fold cross-validation.

Prediction Model	Number of Measures	Degrees of Freedom	Complete Dataset (C-Statistic and 95% CI)	Fivefold Cross-Validation (Average C-statistic) (Testing Datasets)
Model 1: Age, BMI, male, snoring	4	5	0.751 (0.715, 0.787)	0.740
Model 2: Model 1 + restless or very restless sleep + neck size	6	7	0.765 (0.730, 0.800)	0.752
Model 3: Model 2 + age ² + age ³	6	9	0.776 (0.741, 0.810)	0.762
Model 4: Model 3 + witnessed apneas + BMI ² + high depressive symptoms + neck size ²	8	13	0.784 (0.750, 0.818)	0.763
Model 5: Model 3 + witnessed apneas + sleepiness + high depressive symptoms + waist size + waist size ² + waist size ³ + neck size ² + neck size ³ + sleepiness × male	10	18	0.793 (0.759, 0.826)	0.761

C-statistics are provided for the full analytic data set and averaged in the fivefold cross-validation. BMI = body mass index, CI = confidence interval.

Table 3—Predictive properties by model and cutoff.

Method	C-Statistic (95% CI)	Cutoffs	High Risk (%)	Accuracy (%)	Sensitivity (95% CI)	Specificity (95% CI)	Youden's Index
STOP-Bang score (range: 0–8)	0.686 (0.647, 0.726)	Predefined (score ≥ 3)	64%	56%	0.79 (0.72, 0.84)	0.44 (0.35, 0.49)	.23
		Optimal (score ≥ 4)	37%	66%	0.56 (0.45, 0.61)	0.74 (0.63, 0.78)	.30
NoSAS score (range: 0–17)	0.697 (0.659, 0.735)	Predefined (score ≥ 8)	65%	58%	0.82 (0.74, 0.86)	0.45 (0.35, 0.50)	.27
		Optimal (score ≥ 10)	38%	66%	0.57 (0.48, 0.64)	0.73 (0.64, 0.78)	.30
HCHS prediction model (range: 0.018–0.990)	0.745 (0.709, 0.782)	Predefined (probability ≥ .12)	80%	53%	0.96 (0.90, 0.98)	0.28 (0.17, 0.32)	.24
		Optimal (probability ≥ .35)	35%	70%	0.58 (0.47, 0.64)	0.79 (0.69, 0.83)	.37
Model 1	0.751 (0.715, 0.787)	Optimal (probability ≥ .46)	33%	72%	0.56 (0.46, 0.62)	0.81 (0.70, 0.85)	.37
Model 2	0.765 (0.730, 0.800)	Optimal (probability ≥ .42)	39%	72%	0.64 (0.55, 0.70)	0.76 (0.66, 0.81)	.40
Model 3	0.776 (0.741, 0.810)	Optimal (probability ≥ .44)	36%	74%	0.64 (0.52, 0.69)	0.81 (0.68, 0.84)	.45
Model 4	0.784 (0.750, 0.818)	Optimal (probability ≥ .38)	45%	72%	0.72 (0.61, 0.77)	0.72 (0.60, 0.77)	.44
Model 5	0.793 (0.759, 0.826)	Optimal (probability ≥ .35)	47%	72%	0.76 (0.66, 0.81)	0.70 (0.57, 0.74)	.46

The true prevalence of moderate or sleep apnea in the complete dataset was 38%. CI = confidence interval, HCHS = Hispanic Community Health Study, NoSAS = Neck, Obesity, Snoring, Age, Sex.

duration had the best accuracy (74%) with the highest specificity (0.78) (**Table S3**). This model included three additional variables (eg, witnessed apneas, depressive symptoms, sleep duration) than the most accurate (74%) model for 3% REI. Our models for 4% REI, similar to 3% REI, had better predictive properties than the commonly used prediction models.

DISCUSSION

In a large community sample of African Americans with a high burden of moderate or severe OSA (38%), we developed and cross-validated a prediction model with better accuracy, sensitivity, and specificity compared with widely used screening

tools. Of the alternative models, the STOP-Bang score, NoSAS score, and HCHS prediction probability, the highest accuracy was seen for the HCHS prediction model when a threshold of ≥ 0.35 was used, resulting in a sensitivity of 0.58 and specificity of 0.79. This was comparable to our model 1 that similarly was based on age, sex, BMI, and self-reported habitual snoring. Notably, neither the STOP-Bang nor the NoSAS performed well in our population, with a maximum accuracy of 66% (for the NoSAS with a threshold of ≥ 4). After consideration of a large number of potential predictors and rigorous evaluation, we also identified 4 additional models that showed modest improvements in prediction over the minimal model. For example, by adding a self-reported item of restless sleep and measured neck size, the C-statistics for the model increased from .74 to .75 and

the sensitivity increased from 0.56 to 0.64. Further inclusion of age², age³, witnessed apneas, depressive symptoms, BMI², and neck size² improved the sensitivity of the prediction model from 0.56 to 0.72. These suggest that the HCHS prediction model or any of our models may be superior to the STOP-Bang or NoSAS in our population of older African Americans who are predominantly obese women. Moreover, especially for research purposes, incorporation of additional information, through measurement of neck and/or waist circumference and screening questions related to sleep quality, observed apneas, and mood, can improve both sensitivity and specificity, thus improving classification.

Sleep apnea is highly prevalent and largely undiagnosed,^{2,26} particularly among African Americans. Despite the high burden of sleep apnea among African Americans,⁸ prior studies on screening and diagnosis of sleep apnea has mainly occurred among non-Hispanic white populations. There is a clear need to test the predictive ability of commonly used screening tools among African Americans. The prediction models we developed have utility in both clinical settings and research (online calculators for each of our models are available at <https://osascreeen.org/>). We provide models that range from simple assessments to more detailed measurements. Use of the more comprehensive models may help reduce false negatives and false positives while collecting information relevant for general physical health (waist circumference) and mental health (depressive symptoms). Health care systems are beginning to incorporate standardized questionnaires into electronic health records, using the data from these for population health screening and monitoring. Our findings highlight several priority questions and measurements to consider as electronic health data capture is expanded and integrated into clinical practice.

In the community-based sample of 12,158 Hispanic/Latino adults, Shah et al²⁰ considered 17 candidate predictors that included demographics, snoring, BMI, waist and neck circumference, hypertension, diabetes, heart disease, and several poor sleep symptoms. Their final model included self-reported snoring, age, sex, and BMI, with a sensitivity of 0.77 and a specificity of 0.75. In reproducing this model, we found a lower sensitivity of 0.56 and higher specificity of 0.81. Our most sensitive and specific model (0.72 and 0.72) included more detailed measures (eg, depression, sleep symptoms, neck and waist circumference), which may be necessary to improve accuracy in older samples. In the paper by Shah et al,²⁰ the authors found no improvement with additional covariates. The different findings may be attributable to the difference in age of the populations (HCHS/SOL is a younger sample) and the cutoff value of apnea-hypopnea index (apnea-hypopnea index ≥ 5 events/h compared with REI ≥ 15 events/h for the current study). Also, the addition of neck or waist circumference may add more utility beyond BMI given the high prevalence of obesity in our sample. Our findings suggest that optimal thresholds for published equations may need to be re-evaluated in specific populations.

Prior studies have found the NoSAS screening tool to have a better predictive performance than the STOP-Bang and Berlin questionnaires.¹⁷ NoSAS includes neck circumference, overweight and obesity status, snoring, age, and sex. In assessing NoSAS in our sample, we found the accuracy to be 58%. In comparing our prediction model 2 to the NoSAS, we included

BMI as opposed to overweight and obesity status and added restless sleep and found a better accuracy score of 72%. The NoSAS score was developed in a sample from Lausanne, Switzerland, with a lower average BMI (26 kg/m²) and lower prevalence of hypertension and diabetes, 41% and 10%, respectively, than in the current study. The study characteristics may explain the difference in the prediction performance.

Commonly assessed screening tools often include history of hypertension, a common sleep apnea comorbidity.¹⁶ However, in a population with a high prevalence of hypertension (eg, African Americans), including hypertension in the prediction model may not improve the accuracy of the model. In testing combinations of variables to develop the prediction model, hypertension and diabetes did not predict sleep apnea in our sample.⁸

Although we used rigorous methods to develop our prediction model, there are limitations to our study. Our study is unique in its collection of in-home sleep apnea testing among a large population of African Americans; however, we did not conduct polysomnography, which is the gold standard for sleep apnea measurement. Thus, sleep apnea severity may be underreported in the current study. These models may perform differently among individuals who live alone because of differences in awareness of sleep-related symptoms (snoring, observed apneas)¹²; however, data on available bedpartners were not available to assess this. This is particularly relevant given that one of the model parameters is witnessed apneas. The prediction models differed based on commonly used metrics of OSA (3% REI vs 4% REI). Although 3% REI and 4% REI were highly correlated,⁸ 4% REI results in some under-recognition of OSA in women²⁷ and possibly in less obese individuals. We also used self-reported habitual snoring, which is prone to measurement error similar to any self-reported measure. We used rigorous assessments of model performances using bootstrap and cross-validation and assessed internal validity but not external validity. One of the strengths of our study is that our sample is comprised of African Americans; however, all participants are from Jackson, Mississippi, and are not representative of African Americans in the United States.

CONCLUSIONS

To our knowledge, JHHS is one of the largest African American samples with both self-reported and objective (in-home sleep apnea testing and actigraphy) sleep measurement. We previously reported a high prevalence of moderate or severe sleep apnea in our sample, 37% of which was largely undiagnosed.⁸ With the growing recognition of the burden of sleep apnea, it is important to develop screening tools for use. Using a range of measures, we were able to develop internally validated screening models, which generally had a better prediction performance than other well-known screening tools. With the availability of online calculators, prediction models can be made available to improve current screening methods. Moreover, identifying specific symptoms and risk factors associated with increased OSA can help clinicians prioritize screening items and identify relevant comorbidities. This improvement can help clinicians make more accurate referrals for sleep studies. Future studies should explore

the generalizability of this predictive model among a wider population of African Americans.

ABBREVIATIONS

BMI, body mass index
 CESD, Center for Epidemiologic Studies of Depression
 CI, confidence Interval
 ESS, Epworth Sleepiness Scale
 HCHS, Hispanic Community Health Study
 JHS, Jackson Heart Study
 JHSS, Jackson Heart Sleep Study
 NoSAS, Neck, Obesity, Snoring, Age, Sex
 OSA, obstructive sleep apnea
 REI, respiratory event index
 SD, standard deviation
 SOL, Study of Latinos

REFERENCES

- Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proc Am Thorac Soc.* 2008;5(2):136–143.
- 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.
- Friedman M, Bliznik D, Klein M, Duggal P, Somenek M, Joseph NJ. Comparison of the incidences of obstructive sleep apnea-hypopnea syndrome in African-Americans versus Caucasian-Americans. *Otolaryngol Head Neck Surg.* 2006;134(4):545–550.
- Redline S, Tishler PV, Hans MG, Tosteson TD, Strohl KP, Spry K. Racial differences in sleep-disordered breathing in African-Americans and Caucasians. *Am J Respir Crit Care Med.* 1997;155(1):186–192.
- Motamedi KK, McClary AC, Amedee RG. Obstructive sleep apnea: a growing problem. *Ochsner J.* 2009;9:149–153.
- Chen X, Wang R, Zee P, et al. Racial/ethnic differences in sleep disturbances: the Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep.* 2015;38:877–888.
- Olafiranye O, Akinboboye O, Mitchell JE, Ogedegbe G, Jean-Louis G. Obstructive sleep apnea and cardiovascular disease in blacks: a call to action from the Association of Black Cardiologists. *Am Heart J.* 2013;165(4):468–476.
- Johnson DA, Guo N, Rueschman M, Wang R, Wilson J, Redline S. Prevalence and correlates of obstructive sleep apnea among African Americans, the Jackson Heart Study. *Sleep.* 2018;41(10):zsy154.
- Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med.* 2002;165(9):1217–1239.
- Tishler PV, Larkin EK, Schluchter MD, Redline S. Incidence of sleep-disordered breathing in an urban adult population: the relative importance of risk factors in the development of sleep-disordered breathing. *JAMA.* 2003;289(17):2230–2237.
- Golbidi S, Badran M, Ayas N, Laher I. Cardiovascular consequences of sleep apnea. *Lung.* 2012;190(2):113–132.
- Kump K, Whalen C, Tishler PV, et al. Assessment of the validity and utility of a sleep-symptom questionnaire. *Am J Respir Crit Care Med.* 1994;150(3):735–741.
- Chervin RD. Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea. *Chest.* 2000;118(2):372–379.
- Pillar G, Lavie P. Psychiatric symptoms in sleep apnea syndrome: effects of gender and respiratory disturbance index. *Chest.* 1998;114(3):697–703.
- Nagappa M, Liao P, Wong J, et al. Validation of the STOP-Bang questionnaire as a screening tool for obstructive sleep apnea among different populations: a systematic review and meta-analysis. *PLoS One.* 2015;10(12):e0143697.
- 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.

- Marti-Soler H, Hirotsu C, Marques-Vidal P, et al. The NoSAS score for screening of sleep-disordered breathing: a derivation and validation study. *Lancet Respir Med.* 2016;4(9):742–748.
- Chai-Coetzer CL, Antic NA, Rowland LS, et al. A simplified model of screening questionnaire and home monitoring for obstructive sleep apnoea in primary care. *Thorax.* 2011;66(3):213–219.
- Maislin G, Pack AI, Kribbs NB, et al. A survey screen for prediction of apnea. *Sleep.* 1995;18(3):158–166.
- Shah N, Hanna DB, Teng Y, et al. Sex-specific prediction models for sleep apnea from the Hispanic community health study/study of Latinos. *Chest.* 2016;149(6):1409–1418.
- National Center for Health Statistics. *Health, United States, 2016: With Chartbook on Long-term Trends in Health.* Hyattsville, MD: Government Printing Office; 2017.
- Fuqua SR, Wyatt SB, Andrew ME, Sarpong DF, Henderson FR, Cunningham MF, Taylor HA Jr. Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. *Ethn Dis.* 2005;15:S618–S629.
- Oldenburg O, Lamp B, Horstkotte D. Cardiorespiratory screening for sleep-disordered breathing. *Eur Respir J.* 2006;28(5):1065–1067.
- Dingli K, Coleman EL, Vennelle M, Finch SP, Wraith PK, Mackay TW, Douglas NJ. Evaluation of a portable device for diagnosing the sleep apnoea/hypopnoea syndrome. *Eur Respir J.* 2003;21(2):253–259.
- Zimmerman M, Posternak MA, Chelminski I. Using a self-report depression scale to identify remission in depressed outpatients. *Am J Psychiatry.* 2004;161(10):1911–1913.
- Colten HR, Altevogt BM, eds. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem.* National Academies Press, Washington, DC; 2006.
- Won CH, Reid M, Sofer T, et al. Sex differences in obstructive sleep apnea phenotypes, the Multi-Ethnic Study of Atherosclerosis. *Sleep.* 2019;42:zsz274.

ACKNOWLEDGMENTS

The authors thank the staff and participants of the Jackson Heart Study.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication January 24, 2020

Submitted in final revised form March 24, 2020

Accepted for publication March 24, 2020

Address correspondence to: Dayna A. Johnson, PhD, MPH, MS, MSW, Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd. NE, Atlanta, GA 30322; Tel: (404) 727-8708; Email: dayna.johnson@emory.edu

DISCLOSURE STATEMENT

All authors have seen and approved this manuscript. Work for this study was performed at Emory University, Brigham and Women's Hospital, and the University of Mississippi Medical Center. This study was funded by National Heart, Lung, and Blood Institute Grants R01HL110068 and K01HL138211. Dr. Wilson is supported by National Institute of General Medical Sciences Grant U54GM115428. Dr. Redline was supported in part by NHLBI grant 5R35HL135818. The Jackson Heart Study is supported and conducted in collaboration with Jackson State University (HHSN2682018000131), Tougaloo College (HHSN2682018000141), the Mississippi State Department of Health (HHSN2682018000151 and HHSN26800001), and the University of Mississippi Medical Center (HHSN2682018000101, HHSN2682018000111, and HHSN2682018000121) contracts from the National Heart, Lung, and Blood Institute and the National Institute for Minority Health and Health Disparities. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the US Department of Health and Human Services. The authors report no conflicts of interest.