Original Research

Exhaled Nitric Oxide in Obstructive Sleep Apnea: Compartment-Specific Associations and Incremental Predictive Value

Tongshan Qiu[1,2†], Bokai Li[2†], Zicheng Lu[2†], Ran Zhang[3], Jingjing Huang[1,4], Xi Jin[1,4*], Rui Fang[1,4*], Huiping Luo[1,4*]

[1] *ENT Institute and Department of Otorhinolaryngology, Eye & ENT Hospital of Fudan University, Shanghai, China*

[2] *Shanghai Medical School, Fudan University, Shanghai, China*

[3] *Department of Otorhinolaryngology, Head and Neck Surgery, Aviation General Hospital of China Medical University, Beijing, China*

[4] *The Therapy Center of Sleep-disordered Breathing, Eye & ENT Hospital of Fudan University, Shanghai, China*

[*] *Corresponding authors. Xi Jin, E-mail: 15801895988@163.com; Rui Fang, E-mail: fangrui2000@139.com; Huiping Luo, E-mail: hluo13@fudan.edu.cn*

[†] *These authors contributed equally to this work.*

© The Author(s) 2026. This article is published by Higher Education Press at journal.hep.com.cn.
This is an open access article distributed under the Creative Commons Attribution License 4.0 (CC BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Airway inflammation is implicated in obstructive sleep apnea (OSA), but nitric oxide signals from different respiratory compartments may provide distinct information. We examined compartment-specific exhaled nitric oxide (eNO) profiles and whether eNO improved OSA prediction beyond routine clinical and anthropometric characteristics.

Methods: This prospective cross-sectional study included 965 participants who underwent polysomnography and measurement of fractional exhaled nitric oxide at 50 mL/s (FeNO50), alveolar nitric oxide concentration (CaNO), and fractional nasal nitric oxide at 10 mL/s (FnNO10). Base models were compared with FeNO50-augmented and complete-eNO-panel models using seven classification algorithms and repeated stratified nested cross-validation. Discrimination, calibration, class-specific performance, and clinical net benefit were evaluated for OSA and sleep-related hypoxemia.

Results: Of the 965 participants, 798 had OSA and 167 did not. FeNO50 was higher in participants with OSA (median, 21.00 [interquartile range, 16.00–29.00] vs. 18.00 [13.00–23.00] ppb; $p < 0.001$). CaNO showed a modest between-group difference, whereas FnNO10 did not differ between groups. FeNO50 was elevated across OSA severity categories without a clear dose–response gradient. It was not correlated with apnea-hypopnea index (AHI) and showed only a very weak inverse correlation with nadir oxygen saturation, while remaining independently associated with OSA after multivariable adjustment. Adding FeNO50 resulted in small, nonsignificant changes in area under the receiver operating characteristic curve (ROC-AUC) for OSA (ΔAUC, −0.009 to 0.009) and hypoxemia (ΔAUC, −0.009 to 0.002). The complete eNO-panel models achieved ROC-AUCs of 0.769–0.839 for OSA and 0.788–0.820 for hypoxemia.

Conclusions: FeNO50 captures an airway inflammatory phenotype associated with OSA but does not closely reflect polysomnographic severity. Its incremental predictive contribution beyond established clinical and anthropometric characteristics was limited, supporting its use as complementary phenotypic information rather than a stand-alone OSA screening test.

KEYWORDS

obstructive sleep apnea, fractional exhaled nitric oxide, airway inflammation, machine learning, incremental predictive value

Article history: Received Jun 5, 2026; Accepted Oct 12, 2026

1 Introduction

Obstructive sleep apnea (OSA) is characterized by recurrent episodes of partial or complete upper-airway collapse during sleep, resulting in intermittent nocturnal hypoxemia, hypercapnia, and sleep fragmentation. Common clinical features include snoring, witnessed apneas, and daytime sleepiness[1]. Beyond its impact on sleep quality, OSA is associated with cardiovascular, metabolic, and neurocognitive morbidity, making it a major public health concern[1]. An estimated 936 million adults aged 30–69 years worldwide have mild-to-severe OSA, including approximately 425 million with moderate-to-severe disease[2]. This burden highlights the need for accessible approaches to identify individuals who require objective sleep testing.

A growing body of evidence implicates airway inflammation in the pathophysiology of OSA[3–5]. Mechanical stress from recurrent upper-airway collapse and vibration, together with intermittent hypoxia–reoxygenation, may induce local oxidative stress and inflammatory signaling. These processes may promote mucosal edema and increase susceptibility to airway collapse, although the direction and clinical importance of this relationship remain incompletely defined. Assessment of airway inflammation may therefore provide phenotypic information complementary to conventional clinical evaluation.

Nitric oxide is produced throughout the respiratory tract, and its concentration in exhaled gas can be measured simply, reproducibly, and noninvasively[6,7]. Fractional exhaled nitric oxide at 50 mL/s (FeNO50) predominantly reflects nitric oxide from the conducting airways, whereas alveolar nitric oxide concentration (CaNO) estimates the distal or alveolar contribution and nasal nitric oxide reflects nasal and paranasal sources. FeNO is established primarily as a biomarker of type 2 airway inflammation in asthma and can support treatment decisions when interpreted alongside other clinical information[7,8]. Studies in OSA have generally reported higher FeNO, but findings for FeNO, CaNO, and nasal nitric oxide have varied across populations and measurement protocols[9,10].

Most previous studies have focused on group-level differences in exhaled nitric oxide (eNO) or its correlations with polysomnographic indices. Whether eNO provides incremental information for individual-level identification of OSA beyond routinely available clinical and anthropometric characteristics remains uncertain. We therefore characterized FeNO50, CaNO, and fractional nasal nitric oxide measured at 10 mL/s (FnNO10) in participants with polysomnographically defined OSA and evaluated whether FeNO50 or the complete eNO panel improved the prediction of OSA and sleep-related hypoxemia across seven established classification algorithms.

2 Materials and Methods

2.1 Study subjects

This prospective clinical study enrolled adult patients with OSA who were diagnosed in the Therapy Center of Sleep-disordered Breathing in the Eye & ENT Hospital of Fudan University and healthy controls. Data were collected from October 2021 to December 2024. Subjects with space-occupying lesions of the upper airway (such as nasal polyps), chronic obstructive pulmonary disease, and current respiratory tract infection were excluded (Figure 1).

Clinical history and Epworth Sleepiness Scale scores were collected by the same trained nurse. The clinical and anthropometric variables considered in the analyses included sex, age, hypertension, diabetes mellitus, heart disease, hyperlipidemia, weight, neck circumference, waist circumference, hip circumference, and body mass index.

2.2 Ethical consideration

This study was approved by the institutional review board of the Eye & ENT Hospital of Fudan University (Nos. 2021134 and 2025288), and all patients and control subjects gave written informed consent before participating in the study.

2.3 Polysomnography (PSG)

An overnight PSG test (Embla N7000, Embla, USA) was conducted. The details of sleep studies have been described in a previous study[11] and are briefly described as below. Each subject underwent a nocturnal polysomnography from 9:00 p.m. to 6:00 a.m. In the sleep ward, EEG, EOG, and chin EMG were recorded with surface electrodes. Airflow was monitored by the nasal pressure. Chest and abdominal movements were evaluated by respiratory inductive plethysmography. Continuous nocturnal pulse oxygen saturation ($LSaO_2$) was recorded by an oximeter connected to the finger. Electrocardiograms and sleep postures were also recorded. Sleep records were scored automatically by the software, and then manually re-scored by the sleep ward physician. The PSG reports were judged by the same registered polysomnographic technologist (RPSGT) according to the American Academy of Sleep Medicine (AASM) manual for the scoring of sleep and associated events version 2.4[12].

OSA was defined as apnea-hypopnea index (AHI) ⩾ 5/h

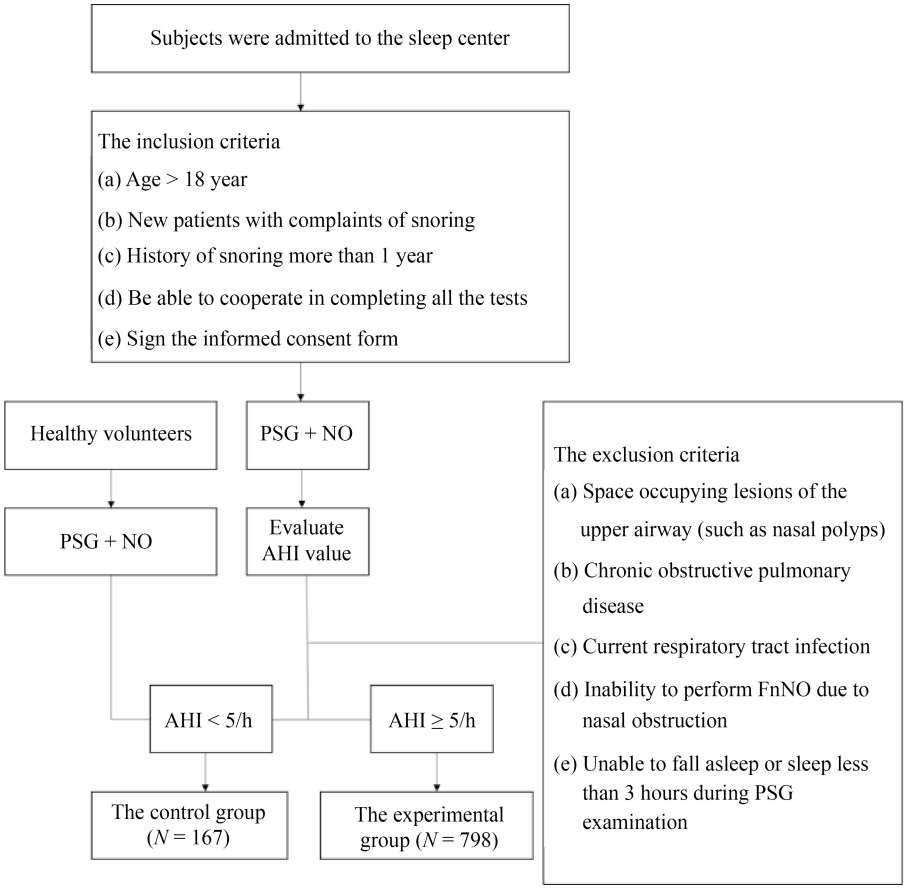


Fig. 1 Subject enrollment process. AHI, apnea–hypopnea index; FnNO, fractional nasal nitric oxide; NO, nitric oxide measurement; PSG, polysomnography.

and OSA was not considered when the AHI was < 5/h. AHI < 5/h, normal; 5–14.9/h, mild OSA; 15–29.9/h, moderate OSA; and AHI ≥ 30/h, severe OSA.

The secondary outcome was sleep-related hypoxemia, defined as $LSaO_2$ below 90% during polysomnography.

2.4 Exhaled nitric oxide measurement

The fractional nasal nitric oxide (FnNO), the fractional exhaled nitric oxide (FeNO), and the alveolar nitric oxide concentration (CaNO) were measured by the same technician after the PSG monitoring, using the nitric oxide analyzer (SV-Eno-03, Sunvou, China)[13]. All measurements were based on the European Respiratory Society (ERS) guidelines[14]. A 30 s self-calibration was started automatically before the test. When measuring FnNO, maintaining the expiratory pressure ≥ 10 cm H_2O was necessary in order to ensure the soft palate was closed. Before the test, subjects were asked to choose one nostril without congestion. When accepting the test, the subjects blocked one nostril by using the customized rubber plug and whistled. The equipment simultaneously drew from the plug at a constant speed of 10 mL/s for 10 s. When measuring FeNO, the expiratory pressure was required to reach at least 5 cm H_2O to ensure that the soft palate was closed in order to eliminate the effect of the nasal cavity. When doing the test, the subjects kept seated and covered the inspiratory filter closely to their mouth. After inhaling for about 2 to 3 s, they exhaled slowly and constantly with a flow rate of 50 mL/s for at least 10 s, following the instructions on the computer. When measuring the CaNO, after breathing at a flow rate of 200 mL/s, the machine automatically calculates the CaNO value. High-intensity exercises and smoking were not allowed before the test. The standard packaged gas provided by the manufacturer was used for manual calibration every 1 to 2 months. The calibration was also performed each time before the consumable was replaced.

2.5 Conventional statistical analysis

Continuous variables were summarized as mean (standard deviation) when approximately normally distributed and as median (interquartile range) when skewed. Categorical variables were reported as numbers and percentages. Between-group comparisons used the independent-samples *t* test for approximately normally distributed variables and the Mann–Whitney *U* test for skewed variables. Categorical variables were compared using the χ^2 test or Fisher exact test. Comparisons across OSA

severity groups used one-way analysis of variance with Bonferroni-adjusted post-hoc comparisons or the Kruskal–Wallis test, according to data distribution. Associations between skewed eNO measures and polysomnographic parameters were evaluated using Spearman rank correlation.

Multivariable logistic regression analysis was performed to determine independent factors associated with OSA. Clinically relevant variables were entered into the model, and adjusted odds ratios (ORs) with 95% confidence intervals (CIs) were calculated. All hypothesis tests were two-sided, with $p < 0.05$ considered statistically significant.

2.6 Machine-learning model development

Machine-learning models were developed for two binary outcomes: OSA, defined as an apnea–hypopnea index of 5 events/h or greater, and sleep-related hypoxemia, defined as $LSaO_2$ below 90% during polysomnography. To determine whether FeNO added predictive information beyond routinely available clinical characteristics, we compared models constructed using different predictor sets under the same validation framework.

The base predictor set included sex, age, hypertension, diabetes mellitus, heart disease, hyperlipidemia, weight, neck circumference, waist circumference, hip circumference, and body mass index. For the primary incremental-value analysis, fractional exhaled nitric oxide at 50 mL/s (FeNO50) was added to the base predictor set. In a secondary analysis, a complete eNO panel was evaluated by adding FeNO50, alveolar nitric oxide concentration (CaNO), and fractional nasal nitric oxide measured at a sampling flow of 10 mL/s (FnNO10) to the base predictors. The predictor sets were defined before model fitting and were applied consistently across all algorithms.

Seven classification algorithms were evaluated: logistic regression, random forest, balanced random forest, extreme gradient boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), histogram-based gradient boosting, and a stacking ensemble. Logistic regression served as an interpretable reference model, whereas the tree-based and ensemble algorithms were included to capture potentially nonlinear associations and interactions among clinical, anthropometric, and eNO variables.

Binary predictors were encoded as indicator variables, and continuous predictors were entered as numerical variables. The predictor sets were specified before model fitting, and no data-driven feature selection was performed. All data-dependent preprocessing procedures were incorporated into the modeling pipeline and estimated using training data only. Only records with complete predictor and outcome information were included; therefore, no imputation was required.

2.7 Internal validation and hyperparameter optimization

Model development and internal validation were performed using repeated stratified nested cross-validation; no separate hold-out test set was used. The outer validation procedure consisted of two repetitions of five-fold stratified cross-validation. Within each outer training fold, three-fold inner cross-validation was used for hyperparameter optimization, with mean ROC-AUC as the selection criterion. Stratification preserved the outcome distribution in each training and validation fold.

All preprocessing and hyperparameter optimization were confined to the corresponding training data. After the optimal hyperparameter combination had been selected in the inner cross-validation, the model was refitted using the complete outer training fold and evaluated in the held-out outer fold. Out-of-fold predicted probabilities were retained for performance estimation. Identical cross-validation partitions were used when comparing the base and FeNO-augmented predictor sets, allowing paired assessment of the incremental contribution of FeNO. A random seed of 2026 was used throughout the analysis.

For logistic regression, the inverse regularization parameter (C) was evaluated at 0.01, 0.1, 1.0, and 10.0. For random forest, the candidate numbers of trees were 300, 500, and 800; maximum tree depth was unrestricted or set to 5, 10, or 20; minimum samples required to split an internal node were 2, 5, and 10; and minimum samples required at a leaf node were 1, 2, and 4. For balanced random forest, the number of trees was evaluated at 300, 500, and 800, maximum tree depth was unrestricted or set to 5, 10, or 20, and the minimum number of samples required to split an internal node was 2, 5, or 10.

For XGBoost, the number of boosting iterations was evaluated at 200, 400, and 600; maximum tree depth at 3, 5, and 7; learning rate at 0.01, 0.05, and 0.10; and subsampling fraction at 0.70, 0.80, and 1.00. For LightGBM, the corresponding search space comprised 200, 400, and 600 boosting iterations; learning rates of 0.01, 0.05, and 0.10; 15, 31, and 63 leaves; and maximum tree depths of 5 and 10 or no depth restriction. For histogram-based gradient boosting, learning rates of 0.01, 0.05, and 0.10, maximum iteration numbers of 100, 200, and 400, and maximum leaf-node numbers of 15, 31, and 63 were evaluated. The stacking ensemble was fitted using its prespecified configuration without additional hyperparameter searching. For each algorithm, the hyperparameter combination yielding the highest mean inner-fold ROC-AUC was selected separately within each outer training fold.

2.8 Model performance and incremental predictive value

Model discrimination was assessed primarily using the

ROC-AUC. Given the imbalance between OSA and non-OSA participants, performance was additionally characterized using positive-class precision–recall AUC (PR-AUC), sensitivity, specificity, positive predictive value, negative predictive value, balanced accuracy, and positive-class F1 score. For the complete eNO-panel OSA models, non-OSA F1 and macro-averaged F1 scores were also reported. Overall accuracy was considered a secondary metric.

Calibration was evaluated using the Brier score, calibration intercept, calibration slope, and plots of observed against predicted probabilities. A calibration intercept of 0 and a calibration slope of 1 represented ideal calibration. Decision-curve analysis was used to quantify clinical net benefit over a range of threshold probabilities, with the model-based strategies compared with the strategies of referring all participants and referring none[15].

The primary incremental-value analysis compared the base predictor set with the same predictor set plus FeNO50. Because both models were evaluated in the same participants using identical cross-validation partitions, differences in ROC-AUC were assessed as paired comparisons using DeLong tests. Changes in ROC-AUC with 95% confidence intervals and Brier scores for both models were reported. The secondary analysis evaluated the predictive performance of models incorporating the complete eNO panel. For a screening-oriented assessment, specificity was also calculated at a target sensitivity of 90%.

Model performance estimates were derived from out-of-fold predictions. For the complete eNO-panel models, 95% confidence intervals for ROC-AUC and PR-AUC were obtained using nonparametric bootstrap resampling. Model performance was interpreted using discrimination, class-specific classification measures, calibration, and clinical net benefit.

2.9 Software

Statistical analyses were performed using SPSS 25.0 (IBM, Armonk, NY, USA) and GraphPad Prism 9.0 (GraphPad Software, San Diego, CA, USA). Machine-learning analyses were conducted using Python version 3.10.20. Model development and evaluation were implemented using scikit-learn version 1.7.2, XGBoost version 3.2.0, and LightGBM version 4.6.0. The random seed was fixed at 2026 to support reproducibility. Reporting followed the TRIPOD+AI recommendations for clinical prediction models developed using regression or machine-learning methods[16].

3 Results

3.1 Subjects

A total of 1027 subjects were voluntarily enrolled from October 2021 to December 2024. Complete data of 965 patients were collected. The characteristics of study participants are shown in Table 1. There were 798 subjects (age 41.77 ± 9.53, BMI [body mass index] 27.31 ± 3.89 kg/m^2, AHI [apnea-hypopnea index] 43.50 ± 25.53/h) with AHI ⩾ 5/h, who were diagnosed as OSA. There were 167 subjects (age 40.24 ± 9.43, BMI 23.96 ± 3.10 kg/m^2, AHI 2.70 ± 1.32/h) whose had an AHI < 5/h and were enrolled as the control group. No significant

Table 1 Characteristics of the patients with OSA and healthy controls.

		OSA group (n = 798)	Control group (n = 167)	t/χ^2	p value
Age (years)		41.77 ± 9.53	40.24 ± 9.43	−1.897	0.058
Sex (%)	Male	689 (86.3%)	106 (63.5%)	49.760	0.001*
	Female	109 (13.7%)	61 (36.5%)		
Hypertension (%)		124 (15.5%)	5 (3%)	18.766	0.001*
Hyperlipidemia (%)		18 (2.3%)	5 (3%)	0.338	0.561
Diabetes (%)		18 (2.3%)	2 (1.2%)	0.383	0.554
BMI (kg/m^2)		27.31 ± 3.89	23.96 ± 3.10	−10.472	0.001*
Waist-to-hip ratio		0.93 ± 0.06	0.89 ± 0.06	−8.666	0.001*
Neck circumference (cm)		38.89 ± 3.40	35.72 ± 4.15	−10.492	0.001*
Waist circumference (cm)		93.18 ± 9.65	84.71 ± 7.99	−10.610	0.001*
Hip circumference (cm)		99.60 ± 6.92	95.10 ± 5.88	−7.839	0.001*
$LSaO_2$ (%)		76.25 ± 10.91	90.58 ± 5.41	16.548	0.001*
$ASaO_2$ (%)		93.48 ± 3.24	95.76 ± 7.10	4.060	0.001*
AHI (events/h)		43.50 ± 25.53	2.70 ± 1.32	−20.640	0.001*

Note: *$p < 0.05$. OSA, obstructive sleep apnea; BMI, body mass index; $LSaO_2$, lowest saturation of oxygen; $ASaO_2$, average saturation of oxygen; AHI, apnea-hypopnea index.

difference was found in age between the two groups.

3.2 Exhaled nitric oxide (eNO)

FeNO levels were significantly higher in participants with OSA than in controls (21.00 [16.00–29.00] ppb vs. 18.00 [13.00–23.00] ppb; $Z = -4.417$; $p < 0.001$). CaNO levels were also modestly higher in the OSA group than in the control group (4.90 [3.00–7.20] ppb vs. 4.30 [2.78–7.93] ppb; $Z = -2.071$; $p = 0.039$). In contrast, FnNO levels did not differ significantly between participants with and without OSA (322.00 [232.00–440.00] ppb vs. 340.50 [240.25–433.75] ppb; $Z = 0.413$; $p = 0.680$) (Table 2).

Further, the FeNO in the mild, moderate, and severe OSA groups was all significantly higher than those in the control group (Figure 2a). However, there were no significant difference in CaNO or FnNO among different groups (Figure 2b and 2c; Table 3).

3.3 The correlation between eNO and PSG parameters

FeNO50 was not correlated with AHI (Spearman $\rho = 0.036$, $p = 0.264$) (Figure 3a). It was weakly correlated with FnNO10 ($\rho = 0.221$, $p < 0.001$) (Figure 3b) and showed a very weak inverse correlation with nadir peripheral oxygen saturation (SpO_2) ($\rho = -0.070$, $p = 0.030$) (Figure 3c).

3.4 Multivariable logistic regression analysis

In multivariable logistic regression, FeNO50 remained independently associated with OSA (adjusted OR per 1-ppb increase, 1.034; 95% CI, 1.013–1.056; $p = 0.001$) (Table 4). Hypertension, waist circumference, and body mass index were positively associated with OSA, whereas female sex and hip circumference were inversely associated with OSA. Neck circumference, FnNO10, and

Table 2 FnNO, FeNO, and CaNO in OSA and control groups.

	FnNO (ppb)	FeNO (ppb)	CaNO (ppb)
Control group	340.50 [240.25–433.75]	18.00 [13.00–23.00]	4.30 [2.78–7.93]
OSA group	322.00 [232.00–440.00]	21.00 [16.00–29.00]	4.90 [3.00–7.20]
Z	0.413	−4.417	−2.071
p value	0.680	0.001*	0.039*

Note: Data are presented as median [IQR]. *$p < 0.05$. FnNO, fractional nasal nitric oxide; FeNO, fractional exhaled nitric oxide; CaNO, alveolar nitric oxide concentration; OSA, obstructive sleep apnea; ppb, parts per billion.

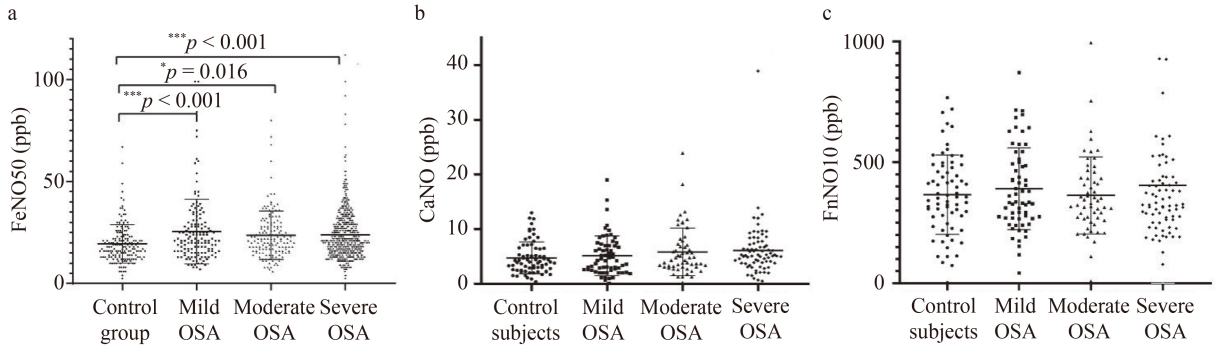


Fig. 2 Distribution of exhaled nitric oxide measures across OSA severity groups. (a) FeNO50; (b) CaNO; and (c) FnNO10 in participants without OSA and those with mild, moderate, or severe OSA. Asterisks indicate multiplicity-adjusted comparisons with the non-OSA group: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. CaNO, alveolar nitric oxide concentration; FeNO50, fractional exhaled nitric oxide at 50 mL/s; FnNO10, fractional nasal nitric oxide measured at 10 mL/s; OSA, obstructive sleep apnea.

Table 3 FnNO, FeNO and CaNO in mild OSA, moderate OSA, severe OSA and control group.

	FnNO (ppb)	FeNO (ppb)	CaNO (ppb)
Control group (*n* = 167)	340.50 [240.25–433.75]	18.00 [13.00–23.00]	4.30 [2.78–7.93]
Mild OSA (*n* = 142)	315.50 [236.75–440.25]	21.00 [15.75–29.00] *	4.70 [2.80–6.75]
Moderate OSA (*n* = 146)	331.00 [251.00–418.75]	21.00 [16.00–28.25] *	5.10 [3.00–7.70]
Severe OSA (*n* = 510)	319.00 [224.75–444.25]	21.00 [16.00–29.00] *	4.80 [3.00–7.13]
p value	0.932	0.001	0.619

Note: Data are presented as median [IQR]. *$p < 0.05$, compared with control group. FnNO, fractional nasal nitric oxide; FeNO, fractional exhaled nitric oxide; CaNO, alveolar nitric oxide concentration; OSA, obstructive sleep apnea.

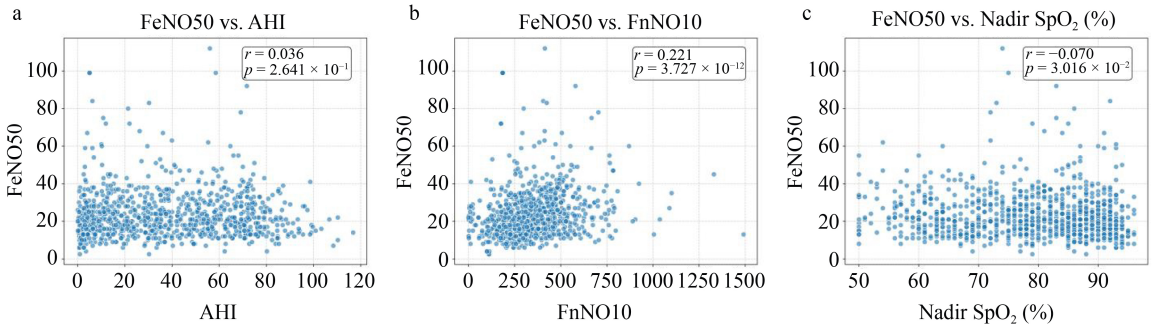


Fig. 3 Correlations of FeNO50 with polysomnographic and nitric oxide measures. (a) FeNO50 and AHI; (b) FeNO50 and FnNO10; and (c) FeNO50 and nadir SpO_2. Correlations were assessed using Spearman rank coefficients. AHI, apnea–hypopnea index; FeNO50, fractional exhaled nitric oxide at 50 mL/s; FnNO10, fractional nasal nitric oxide measured at 10 mL/s; SpO_2, peripheral oxygen saturation.

Table 4 Multivariate logistic regression analysis for OSA.

Items	B	Standard error	Wald	*p*	OR	OR (95% CI)
Sex	−0.550	0.271	4.118	0.042*	0.577	0.339–0.981
Hypertension (%)	1.458	0.494	8.725	0.003*	4.297	1.633–11.307
Neck circumference	0.055	0.044	1.586	0.208	1.056	0.970–1.151
Hip circumference	−0.071	0.026	7.329	0.007*	0.932	0.885–0.981
Waist circumference	0.056	0.018	9.213	0.002*	1.057	1.020–1.096
BMI	0.205	0.051	16.508	< 0.001*	1.228	1.112–1.356
FeNO	0.033	0.010	10.140	0.001*	1.034	1.013–1.056
FnNO	0.001	0.001	0.163	0.687	1.000	0.999–1.001
CaNO	0.004	0.009	0.158	0.691	1.004	0.986–1.022

Note: *$p < 0.05$. BMI, body mass index; CaNO, alveolar nitric oxide concentration; CI, confidence interval; FeNO, fractional exhaled nitric oxide; FnNO, fractional nasal nitric oxide; OR, odds ratio; OSA, obstructive sleep apnea; SE, standard error.

CaNO were not independently associated with OSA.

3.5 Primary analysis: incremental predictive value of FeNO50

The primary analysis evaluated whether FeNO50 provided incremental predictive information beyond routinely available clinical and anthropometric variables. Changes in ROC-AUC ranged from −0.009 to 0.009 across the seven algorithms, and none reached statistical significance (Table 5). The largest numerical increase was observed for XGBoost (ΔAUC = 0.009; 95% CI, −0.003 to 0.020; $p = 0.125$), whereas stacking showed the largest decrease (ΔAUC = −0.009; 95% CI, −0.023 to 0.005; $p = 0.217$).

Brier scores decreased slightly for all seven algorithms, and specificity at 90% sensitivity increased for six. Overall, adding FeNO50 did not consistently improve discrimination beyond the clinical and anthropometric predictors. Calibration and decision-curve analyses for

Table 5 Incremental predictive value of FeNO50 for OSA prediction across seven algorithms.

Model	Base ROC-AUC	FeNO50-augmented ROC-AUC	ΔAUC (95% CI)	DeLong *p* value	Base Brier score	FeNO50-augmented Brier score	Base specificity at 90% sensitivity	FeNO50-augmented specificity at 90% sensitivity
Logistic regression	0.800	0.808	0.008 (−0.006 to 0.021)	0.261	0.184	0.180	0.437	0.473
Random forest	0.812	0.808	−0.004 (−0.024 to 0.016)	0.664	0.139	0.137	0.521	0.551
Balanced random forest	0.824	0.833	0.008 (−0.003 to 0.019)	0.136	0.153	0.148	0.509	0.581
XGBoost	0.819	0.828	0.009 (−0.003 to 0.020)	0.125	0.114	0.109	0.527	0.551
LightGBM	0.816	0.820	0.003 (−0.010 to 0.017)	0.628	0.154	0.147	0.473	0.515
Histogram-based gradient boosting	0.812	0.816	0.005 (−0.012 to 0.022)	0.608	0.145	0.136	0.437	0.563
Stacking	0.775	0.766	−0.009 (−0.023 to 0.005)	0.217	0.198	0.196	0.329	0.281

Note: Differences in ROC-AUC were calculated using unrounded estimates. AUC, area under the receiver operating characteristic curve; CI, confidence interval; FeNO50, fractional exhaled nitric oxide at 50 mL/s; OSA, obstructive sleep apnea; ROC-AUC, area under the receiver operating characteristic curve.

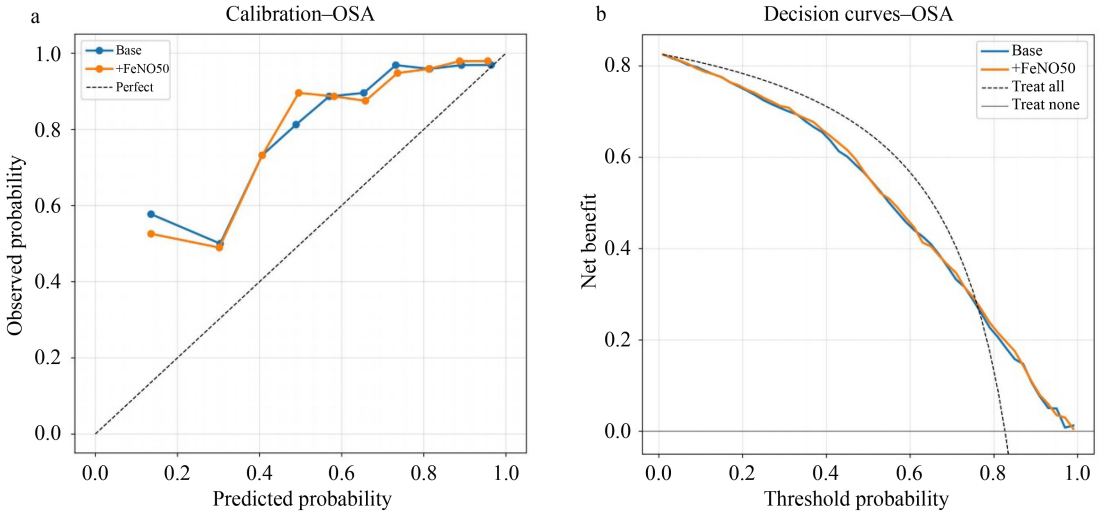


Fig. 4 Calibration and decision-curve analyses of the base and FeNO50-augmented models for OSA prediction. (a) Calibration curves comparing observed and predicted OSA probabilities. (b) Decision curves comparing net benefit across threshold probabilities. FeNO50, fractional exhaled nitric oxide at 50 mL/s; OSA, obstructive sleep apnea.

the base and FeNO50-augmented models are shown in Figure 4.

3.6 Secondary performance analysis of the complete eNO panel

In the secondary analysis, models incorporating FeNO50, CaNO, and FnNO10 achieved ROC-AUCs ranging from 0.769 to 0.839 (Table 6; Suppl. Figure 1A). Balanced random forest had the highest numerical ROC-AUC (0.839; 95% CI, 0.804–0.872), followed by random forest (0.833; 95% CI, 0.797–0.866) and XGBoost (0.829; 95% CI, 0.794–0.863). Balanced random forest also had the highest balanced accuracy (0.768), with a sensitivity of 0.812 and a specificity of 0.725. With OSA designated as the positive class, PR-AUCs ranged from 0.938 to 0.954, whereas F1 scores for non-OSA ranged from 0.416 to 0.559 (Suppl. Figure 1B; Tables 6 and 7).

Calibration performance varied across algorithms. Calibration intercepts ranged from −0.056 to 1.489, calibration slopes from 0.639 to 1.181, and Brier scores from 0.108 to 0.192 (Table 7). Calibration and decision curves comparing the base and complete eNO-panel models are shown in Suppl. Figure 1C and 1D. Compared descriptively with the corresponding FeNO50-augmented models, changes in ROC-AUC ranged from −0.013 to 0.025 and varied in direction across algorithms.

3.7 Secondary analysis of hypoxemia

Adding FeNO50 produced small changes in ROC-AUC across the seven algorithms, ranging from −0.009 to 0.002, and none of the differences reached statistical significance (Suppl. Table 1). Changes in Brier score and specificity at 90% sensitivity were inconsistent across algorithms. Calibration and decision curves for the base and FeNO50-augmented models are shown in Suppl. Figure 2.

Models incorporating the complete eNO panel achieved ROC-AUCs ranging from 0.788 to 0.820 (Table 8; Suppl. Figure 3A). Balanced random forest had the highest numerical ROC-AUC (0.820; 95% CI, 0.786–0.854),

Table 6 Discrimination and classification performance of the complete eNO-panel models for OSA.

Model	ROC-AUC (95% CI)	OSA PR-AUC (95% CI)	Accuracy	Balanced accuracy	Sensitivity	Specificity	PPV	NPV	F1-OSA
Logistic regression	0.795 (0.758–0.832)	0.941 (0.923–0.958)	0.722	0.733	0.717	0.749	0.932	0.356	0.810
Random forest	0.833 (0.797–0.866)	0.951 (0.934–0.966)	0.856	0.707	0.935	0.479	0.896	0.606	0.915
Balanced random forest	0.839 (0.804–0.872)	0.954 (0.938–0.968)	0.797	0.768	0.812	0.725	0.934	0.446	0.869
XGBoost	0.829 (0.794–0.863)	0.952 (0.936–0.967)	0.852	0.636	0.966	0.305	0.869	0.654	0.915
LightGBM	0.821 (0.785–0.856)	0.950 (0.933–0.965)	0.803	0.755	0.828	0.683	0.926	0.454	0.874
Histogram-based gradient boosting	0.814 (0.776–0.850)	0.941 (0.920–0.960)	0.822	0.755	0.857	0.653	0.922	0.489	0.888
Stacking	0.769 (0.733–0.806)	0.938 (0.921–0.954)	0.715	0.719	0.713	0.725	0.925	0.346	0.805

Note: CI, confidence interval; eNO, exhaled nitric oxide; F1-OSA, F1 score for the OSA class; NPV, negative predictive value; OSA, obstructive sleep apnea; PPV, positive predictive value; PR-AUC, area under the precision–recall curve; ROC-AUC, area under the receiver operating characteristic curve.

Table 7 Minority-class and calibration performance of the complete eNO-panel models for OSA.

Model	F1-non-OSA	Macro-F1	Brier score	Calibration intercept	Calibration slope
Logistic regression	0.483	0.646	0.184	1.489	0.799
Random forest	0.535	0.725	0.110	0.417	1.051
Balanced random forest	0.553	0.711	0.144	1.230	1.181
XGBoost	0.416	0.666	0.108	−0.056	0.968
LightGBM	0.545	0.710	0.140	1.113	0.896
Histogram-based gradient boosting	0.559	0.724	0.131	0.934	0.639
Stacking	0.468	0.637	0.192	1.489	0.937

Note: eNO, exhaled nitric oxide; F1-non-OSA, F1 score for the non-OSA class; Macro-F1, macro-averaged F1 score; OSA, obstructive sleep apnea.

Table 8 Performance of the complete eNO-panel models for predicting hypoxemia.

Model	ROC-AUC (95% CI)	PR-AUC (95% CI)	Accuracy	Balanced accuracy	Sensitivity	Specificity	PPV	NPV	F1-hypoxemia
Logistic regression	0.818 (0.788–0.849)	0.949 (0.937–0.961)	0.732	0.740	0.726	0.753	0.923	0.403	0.813
Random forest	0.816 (0.783–0.850)	0.943 (0.926–0.958)	0.807	0.733	0.855	0.611	0.900	0.509	0.877
Balanced random forest	0.820 (0.786–0.854)	0.944 (0.927–0.959)	0.773	0.757	0.783	0.732	0.922	0.453	0.847
XGBoost	0.807 (0.773–0.842)	0.940 (0.924–0.956)	0.819	0.623	0.946	0.300	0.846	0.576	0.893
LightGBM	0.808 (0.776–0.841)	0.943 (0.928–0.958)	0.748	0.732	0.759	0.705	0.913	0.417	0.829
Histogram-based gradient boosting	0.798 (0.764–0.833)	0.942 (0.927–0.956)	0.769	0.709	0.808	0.611	0.894	0.438	0.849
Stacking	0.788 (0.754–0.820)	0.938 (0.923–0.952)	0.683	0.729	0.653	0.805	0.932	0.363	0.768

Note: CI, confidence interval; eNO, exhaled nitric oxide; F1-hypoxemia, F1 score for the hypoxemia class; NPV, negative predictive value; PPV, positive predictive value; PR-AUC, precision–recall area under the curve; ROC-AUC, area under the receiver operating characteristic curve.

followed by logistic regression (0.818; 95% CI, 0.788–0.849) and random forest (0.816; 95% CI, 0.783–0.850). With hypoxemia designated as the positive class, PR-AUCs ranged from 0.938 to 0.949 (Suppl. Figure 3B). Calibration and decision curves comparing the base and complete eNO-panel models are shown in Suppl. Figure 3C and 3D. Compared descriptively with the corresponding FeNO50-augmented models, changes in ROC-AUC after the addition of CaNO and FnNO10 ranged from −0.007 to 0.004 and varied in direction across algorithms.

4 Discussion

4.1 Main findings summary

In this cross-sectional study of 965 participants, FeNO50 was higher in OSA, including mild disease, and remained independently associated with OSA after adjustment for clinical and anthropometric factors. FeNO50 changed little across OSA severity categories and was only weakly related to AHI and nadir oxygen saturation. CaNO showed a small overall difference between participants with and without OSA, whereas FnNO10 was similar between groups. This compartment-specific pattern suggests that eNO reflects an airway phenotype that is not proportional to conventional measures of OSA severity. Across seven algorithms, adding FeNO50 did not significantly improve discrimination for OSA or hypoxemia, and the complete eNO panel showed no consistent numerical advantage over the FeNO50-augmented models.

4.2 Compartment-specific eNO profiles in OSA

Repeated upper-airway closure may affect the respiratory tract through several related pathways. Snoring vibration and pharyngeal collapse produce mechanical stress at the mucosal surface, while intermittent hypoxia followed by reoxygenation promotes oxidative stress and inflammatory signaling. Airway studies in untreated OSA have demonstrated neutrophilic inflammation, increased interleukin-8 concentrations, and greater expression of inflammatory and oxidative-stress markers[3,4,17]. These processes provide a plausible basis for altered airway nitric oxide production. In the present study, FeNO50 was higher in participants with OSA than in those without OSA, and this difference was already apparent in mild disease. This finding is consistent with a meta-analysis of 16 studies showing higher FeNO in OSA and with longitudinal observations of lower FeNO after effective continuous positive airway pressure treatment[10,18]. The similar FeNO50 distributions across mild, moderate, and severe OSA indicate that the inflammatory signal and respiratory-event frequency capture different aspects of the disease.

The three nitric oxide measurements used in this study sample different components of respiratory NO exchange. FeNO50 is weighted toward nitric oxide originating from the conducting airways, although it is not anatomically confined to a single airway compartment. CaNO is an estimate of the distal or alveolar contribution to exhaled nitric oxide and is influenced by the assumptions used to separate airway and alveolar sources. FnNO10 primarily reflects nitric oxide within the nasal cavity and paranasal sinuses and is sensitive to sampling flow, velum closure, sinus ostial patency, and nasal ventilation[7,14]. FeNO50, CaNO, and FnNO10 should therefore be interpreted as related but non-interchangeable measures rather than as repeated measurements of a single inflammatory process.

This distinction helps to place the observed differences into context. FeNO50 showed the most consistent association with OSA, whereas the higher CaNO in the overall OSA group was modest and was not accompanied by a severity gradient. FnNO10 did not differ between participants with and without OSA. Previous CaNO findings have also been inconsistent: higher CaNO associated with nocturnal desaturation has been reported[19], whereas another study observed lower CaNO in OSA with partial recovery after continuous positive airway pressure treatment[20]; pooled analyses have not demonstrated a consistent overall difference[10]. Nasal nitric oxide has likewise been elevated in some OSA cohorts, including in association with nasal neutrophilia and increased inflammatory cytokines[9]. Variation in respiratory phenotype, coexisting nasal or lower-airway disease, measurement technique, sampling flow, and the timing of measurements relative to sleep may account for part of this heterogeneity. Our findings therefore do not exclude nasal or distal-airway inflammation in OSA; rather, they indicate that the measurable nitric oxide signal was not uniformly expressed across the nasal, conducting-airway, and alveolar compartments in this cohort.

4.3 FeNO as an associated biomarker and an incremental predictor

FeNO50 remained associated with OSA after adjustment for clinical and anthropometric factors, indicating that its between-group difference was not fully accounted for by the measured covariates. At the same time, FeNO50 was not correlated with AHI, showed only a very weak inverse correlation with nadir oxygen saturation, and did not increase across OSA severity categories. FeNO50 therefore characterizes an airway inflammatory feature associated with OSA rather than the frequency of respiratory events or the extent of nocturnal desaturation. Association and individual-level discrimination are distinct properties of a biomarker[21,22].

The prediction analyses examined whether FeNO50 contributed information beyond routine clinical assessment. The base models already captured major anatomical and metabolic features through age, sex, comorbidities, BMI, body weight, and measures of central and upper-body adiposity. Adding FeNO50 produced only small changes in discrimination, none of which reached statistical significance. The overlap in eNO distributions between participants with and without OSA may explain why a group-level difference translated into limited incremental discrimination. This is a familiar pattern in biomarker research, in which an independently associated marker may provide complementary biological information but little additional separation after established predictors are included[21–23].

4.4 Model performance, class imbalance, and hypoxemia prediction

No single algorithm performed best across all evaluation measures. For OSA prediction using the complete eNO panel, balanced random forest achieved the highest ROC-AUC and balanced accuracy, whereas random forest and XGBoost achieved higher overall accuracy through high sensitivity accompanied by lower specificity. For hypoxemia, balanced random forest, logistic regression, and random forest had similar ROC-AUCs. Algorithmic complexity therefore did not confer a uniform performance advantage, consistent with a systematic review showing no clear benefit of machine learning over logistic regression in low-bias clinical prediction studies[24]. Calibration also varied across algorithms, indicating that discrimination and probability estimation should be considered separately[25].

Performance estimates should also be interpreted in the context of the high prevalence of OSA in this referral cohort. High accuracy and OSA-class PR-AUC may coexist with limited recognition of non-OSA participants because precision–recall performance depends on the prevalence of the designated positive class[26]. In the complete eNO-panel models, positive predictive value (PPV) ranged from 0.869 to 0.934, whereas negative predictive value (NPV) ranged from 0.346 to 0.654 and non-OSA F1 ranged from 0.416 to 0.559. Specificity, NPV, non-OSA F1, and balanced accuracy therefore provided information that was not apparent from overall accuracy alone. A similar pattern has been reported in an otolaryngology sleep-clinic cohort, in which high sensitivity and positive predictive value were accompanied by lower specificity and negative predictive value[27].

The hypoxemia analysis yielded a similar interpretation. Adding FeNO50 provided no consistent improvement in the prediction of a nadir oxygen saturation below 90%. Nocturnal desaturation is determined not only by respiratory-event frequency but

also by event duration, baseline saturation, sleep stage, body position, obesity, and pulmonary oxygen reserve[28]. Measures incorporating the depth and duration of desaturation, such as hypoxic burden, also predict clinically important outcomes beyond the information conveyed by AHI alone[29]. FeNO50 reflects airway NO biology rather than ventilation or gas exchange, which may explain its weak association with nadir oxygen saturation and limited incremental contribution to hypoxemia prediction.

4.5 Clinical implications

FeNO measurement is rapid, noninvasive, and repeatable. In OSA assessment, its most relevant role may be to complement clinical phenotyping rather than replace polysomnography or home sleep apnea testing. This positioning is consistent with AASM guidance that objective sleep testing remains necessary for diagnosis[30].

FeNO50 may be particularly informative when it is already available during respiratory assessment, helping to identify patients in whom OSA coexists with an inflammatory airway signal. Interpretation should account for asthma, atopy, smoking, corticosteroid exposure, and other determinants of FeNO[8,14]. Its reported reduction after continuous positive airway pressure treatment also supports further study of eNO-defined phenotypes and longitudinal treatment responses[18].

4.6 Strengths and limitations

This study has three main strengths. First, data were collected prospectively, with OSA defined by attended polysomnography and eNO measured using standardized procedures. Second, FeNO50, CaNO, and FnNO10 were assessed simultaneously, allowing direct comparison of nitric oxide signals from different respiratory compartments. Third, incremental value was examined by comparing prespecified predictor sets under the same nested cross-validation framework across seven algorithms. Performance was evaluated beyond ROC-AUC by considering calibration, class-specific metrics, and clinical net benefit.

Three limitations should be acknowledged. First, this was a single-center referral cohort with a high prevalence of OSA, which may limit generalizability and inflate metrics influenced by disease prevalence. External validation in community and primary-care populations is therefore required. Second, the cross-sectional design and single eNO assessment preclude conclusions about temporality, causality, or treatment-related change. Longitudinal studies are needed to determine whether FeNO identifies a reproducible OSA inflammatory phenotype and whether it changes with effective therapy. Third, information on asthma, allergic rhinitis, chronic rhinosinusitis, smoking status, inhaled corticosteroid use, and other medications affecting FeNO was not systematically available. Residual confounding from these factors therefore cannot be excluded.

5 Conclusion

Multi-compartment eNO profiling identified a compartment-specific pattern in OSA, with FeNO50 showing the most consistent difference and remaining independently associated with OSA. Its incremental contribution beyond clinical and anthropometric information was limited, and the complete eNO panel showed no consistent numerical advantage over the FeNO50-augmented models across algorithms. FeNO50 may therefore be better positioned as a complementary marker of airway phenotype. Validation in diverse, well-phenotyped populations is warranted.

Acknowledgements

We sincerely thank Professor Wu Haitao for his valuable suggestions, guidance and selfless assistance to the younger generation. Many thanks to the residents and our nursing team for taking good care of the study participants. We also thank Wilson Chen and Jenny Lee for their encouragement and support in spirit. Much love to Dr. Luo's cats Man, Song and Wu for accompanying us all the way.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Ethical statement

This study was approved by the Institutional Review Board of the Eye & ENT Hospital of Fudan University (Nos. 2021134 and 2025288). All participants provided written informed consent before participation.

Funding information

This work was supported by the Shanghai Shenkang Hospital Development Center Emerging Frontier Technology Joint Research Project of Municipal Hospitals (No. SHDC12025112) to H.L.; the Shanghai Shenkang Hospital Development Center Shanghai Clinical Cohort Project (Reserve Clinical Cohort) (No. SHDC2025CCS044) to H.L.; the General Program of Shanghai Municipal Health Commission to H.L.; the Shanghai Key Laboratory of Genetic Editing and Cell Therapy for Rare Diseases (Nos. GECT-2025-B07 and GECT-2025-B08) to H.L. and Z.L.; and the Fudan University Undergraduate Academic Research Training

Program “AI+” (No. 250107) to H.L. and T.Q.

Data availability statement

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

AI statement

The authors used ChatGPT (version 5.6 Sol) as a language-editing tool to improve grammar, clarity, and readability. The AI tool was not used for data analysis, data generation, image processing, or scientific content creation. All outputs were carefully reviewed and verified by the authors, who take full responsibility for the final content.

Author contributions

Tongshan Qiu: Data collection and analysis, Writing–original draft; Bokai Li: Writing–review & editing; Zicheng Lu: Data collection and analysis; Ran Zhang: Data collection and analysis; Jing Jing Huang: polysomnography analysis, patient enrollment; Xi Jin: data collection, polysomnography analysis, supervision; Rui Fang: patient enrollment, supervision; Hui Ping Luo: Conceptualization, Project administration, patient enrollment, Data curation, Formal analysis, Writing–review & editing. The authors agree to be accountable for all aspects of the work.

Supplementary files

Supplementary material is available in the online version of this article at https://doi.org/10.15302/ENTD.2026.120001 and is accessible for authorized users.

References

1. Gottlieb DJ, Punjabi NM. Diagnosis and management of obstructive sleep apnea: a review. *JAMA*. 2020, 323(14): 1389–1400.
2. 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.
3. Kimoff RJ, Hamid Q, Divangahi M, et al. Increased upper airway cytokines and oxidative stress in severe obstructive sleep apnoea. *Eur Respir J*. 2011, 38(1): 89–97.
4. Devouassoux G, Lévy P, Rossini E, et al. Sleep apnea is associated with bronchial inflammation and continuous positive airway pressure-induced airway hyperresponsiveness. *J Allergy Clin Immunol*. 2007, 119(3): 597–603.
5. Carpagnano GE, Spanevello A, Sabato R, et al. Systemic and airway inflammation in sleep apnea and obesity: the role of ICAM-1 and IL-8. *Transl Res*. 2010, 155(1): 35–43.
6. American Thoracic Society, European Respiratory Society. ATS/ERS recommendations for standardized procedures for the online and offline measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide, 2005. *Am J Respir Crit Care Med*. 2005, 171(8): 912–930.
7. Högman M, Bowerman C, Chavez L, et al. ERS technical standard: Global Lung Function Initiative reference values for exhaled nitric oxide fraction $(F_{\mathrm{ENO}_{50}})$. *Eur Respir J*. 2024, 63(1): 2300370.
8. Dweik RA, Boggs PB, Erzurum SC, et al. An official ATS clinical practice guideline: interpretation of exhaled nitric oxide levels (FENO) for clinical applications. *Am J Respir Crit Care Med*. 2011, 184(5): 602–615.
9. Zhang D, Xiao Y, Luo J, et al. Measurement of fractional exhaled nitric oxide and nasal nitric oxide in male patients with obstructive sleep apnea. *Sleep Breath*. 2019, 23(3): 785–793.
10. Zhang D, Luo J, Qiao Y, Xiao Y, Huang R, Zhong X. Measurement of exhaled nitric oxide concentration in patients with obstructive sleep apnea: a meta-analysis. *Medicine*. 2017, 96(12): e6429.
11. Luo H, Tang X, Xiong Y, Meng L, Yi H, Yin S. Efficacy and mechanism of mandibular advancement devices for persistent sleep apnea after surgery: a prospective study. *J Otolaryngol Head Neck Surg*. 2016, 45(1): 56.
12. Berry RB, Brooks R, Gamaldo C, et al. AASM scoring manual updates for 2017 (Version 2.4). *J Clin Sleep Med*. 2017, 13(5): 665–666.
13. Lei W, Li F, Tang X, Bian S, Wang J, Huang J. The comparison of two exhaled nitric oxide analyzers: NIOX VERO and SUNVOU-CA2122. *J Breath Res*. 2021, 15(2): 026007.
14. Horváth I, Barnes PJ, Loukides S, et al. A European Respiratory Society technical standard: exhaled biomarkers in lung disease. *Eur Respir J*. 2017, 49(4): 1600965.
15. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006, 26(6): 565–574.
16. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024, 385: e078378.
17. Salerno FG, Carpagnano E, Guido P, et al. Airway inflammation in patients affected by obstructive sleep apnea syndrome. *Respir Med*. 2004, 98(1): 25–28.
18. Chua AP, Aboussouan LS, Minai OA, Paschke K, Laskowski D, Dweik RA. Long-term continuous positive airway pressure therapy normalizes high exhaled nitric oxide levels in obstructive sleep apnea. *J Clin Sleep Med*. 2013, 9(6): 529–535.
19. Hua-Huy T, Le-Dong NN, Duong-Quy S, Luchon L, Rouhani S, Dinh-Xuan AT. Increased alveolar nitric oxide concentration is related to nocturnal oxygen desaturation in obstructive sleep apnoea. *Nitric Oxide*. 2015, 45: 27–34.
20. Foresi A, Leone C, Olivieri D, Cremona G. Alveolar-derived

exhaled nitric oxide is reduced in obstructive sleep apnea syndrome. *Chest*. 2007, 132(3): 860–867.

21. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*. 2004, 159(9): 882–890.
22. Ware JH. The limitations of risk factors as prognostic tools. *N Engl J Med*. 2006, 355(25): 2615–2617.
23. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008, 27(2): 157–172.
24. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019, 110: 12–22.
25. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016, 74: 167–176.
26. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*. 2015, 10(3): e0118432.
27. Cho SW, Jung SJ, Shin JH, Won TB, Rhee CS, Kim JW. Evaluating prediction models of sleep apnea from smartphone-recorded sleep breathing sounds. *JAMA Otolaryngol Head Neck Surg*. 2022, 148(6): 515–521.
28. Peppard PE, Ward NR, Morrell MJ. The impact of obesity on oxygen desaturation during sleep-disordered breathing. *Am J Respir Crit Care Med*. 2009, 180(8): 788–793.
29. Azarbarzin A, Sands SA, Stone KL, et al. The hypoxic burden of sleep apnoea predicts cardiovascular disease-related mortality: the Osteoporotic Fractures in Men Study and the Sleep Heart Health Study. *Eur Heart J*. 2019, 40(14): 1149–1157.
30. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. *J Clin Sleep Med*. 2017, 13(3): 479–504.