Monaldi Archives for Chest Disease

eISSN 2532-5264 https://www.monaldi-archives.org/

Monaldi Arch Chest Dis 2026 [Online ahead of print]

To cite this Article:
Salik M, Saad M, Burkes WS. **Agreement between apnea-hypopnea index and hypoxic burden for severity classification of obstructive sleep apnea**. *Monaldi Arch Chest Dis* doi: 10.4081/monaldi.2026.4016

Submitted: 30-03-2026
Accepted: 8-07-2026

Agreement between apnea-hypopnea index and hypoxic burden for severity classification of obstructive sleep apnea

Muhammad Salik, Muhammad Saad, William S Burkes

Baptist Hospitals of Southeast Texas, Beaumont, Texas, United States

Correspondence: Muhammad Salik, Baptist Hospitals of Southeast Texas, Beaumont, Texas, United States. Tel.: +1 (409) 926-2800. E-mail: salikshoaib96@gmail.com

Contributions: Muhammad Salik: conceptualization, methodology, software, formal analysis, investigation, data curation, data collection, visualization, validation, writing original draft, and reviewing and editing. Muhammad Saad: data curation, data collection, validation, writing original draft, and reviewing and editing. William S Burkes: conceptualization, methodology, software, investigation, validation, writing original draft, and reviewing and editing.

Conflict of interest: the authors declare that there is no conflict of interest.

Ethics approval and consent to participate: ethical approval was obtained from the Ethical Review Committee of the Baptist Hospitals of Southeast Texas, Texas, United States (Protocol # BHSET23.035). The study was performed in accordance with the ethical standards outlined in the 1964 Declaration of Helsinki and its future amendments.

Informed consent: not applicable.

Patient consent for publication: not applicable.

Availability of data and materials: the datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of generative artificial intelligence and artificial intelligence-assisted technologies in the writing process: during the preparation of this work, the authors used ChatGPT solely for grammatical and language-editing purposes. The authors reviewed and edited the output and take full responsibility for the content of this publication.

Abstract

This study aimed to evaluate the agreement between apnea-hypopnea index (AHI)- and hypoxic burden (HB)-based obstructive sleep apnea (OSA) severity classifications. A retrospective single-center study of adults aged 18-79 years undergoing full-night in-laboratory polysomnography was conducted. OSA severity by AHI was categorized as normal (<5 events/hour), mild (5 to <15), moderate (15 to <30), or severe (≥30). HB severity was categorized as mild (<3.34), moderate (3.34 to 6.66), or severe (≥6.67). Agreement among patients with OSA (AHI ≥5) was assessed using Cohen's kappa. Ninety-five patients were included (mean age 58.4±18.3 years); 73 (76.8%) had OSA. Among these, HB severity exceeded AHI severity in 27.4%, was concordant in 61.6%, and was lower in 11.0%. Observed agreement was 30.1%, with slight concordance beyond chance (κ=0.168, $p<0.001$). Continuous AHI and HB were moderately correlated ($r=0.663$, $p<0.001$). Patients with HB greater than AHI severity demonstrated significantly lower nadir oxygen saturation ($p=0.011$). AHI- and HB-based OSA severity classifications demonstrate limited agreement. Incorporation of HB may provide a more physiologically informative assessment of OSA severity beyond event frequency alone.

Key words: apnea-hypopnea index; hypoxic burden; obstructive sleep apnea; severity classification.

Introduction

Obstructive sleep apnea (OSA) is a chronic sleep disorder characterized by recurrent upper airway obstruction during sleep [1] which has been shown to be associated with a number of other chronic diseases, for instance, excessive daytime sleepiness [2], neurocognitive dysfunction [3], and cardiovascular diseases [4]. The pathophysiological underpinning of OSA is recurring episodes of either complete (apnea) or partial (hypopnea) obstruction of upper airways, leading to repetitive events of oxygen desaturation and sleep disturbance [5,6].

For severity classification of OSA, the apnea-hypopnea index (AHI) is a well-established tool in sleep medicine [6]. Despite being a widely used clinical tool to predict complications associated with OSA, AHI has several limitations that limits its clinical utility. AHI reflects event frequency but does not account for desaturation depth, duration, or cumulative hypoxic exposure. It also does not incorporate sleep-stage distribution or positional dependence. Consequently, patients with similar AHI values may experience substantially different physiological burdens [7]. These limitations have prompted investigation into alternative metrics for assessing OSA severity [6].

Of late, the sleep apnea-targeted 'hypoxic burden (HB)' measure has been denoted as an ideal tool in stratification of high risk OSA individuals [8,9]. The HB is described as the 'sum of individual areas under the oxygen desaturation curve'. HB integrates both depth and duration of oxygen desaturation events, capturing cumulative hypoxic exposure beyond event frequency. Most importantly, the HB specific to OSA has been shown to be associated with adverse clinical outcomes in a number of recent research investigations [8,10]. A study by Azarbarzin *et al.* (2019) quantified OSA severity via HB and assessed its predictive ability for mortality. Contrasting with AHI, oxygen desaturation area under the curve events derived from HB efficiently predicted all-cause mortality and mortality associated with cardiovascular disease [8]. In a randomized controlled trial by Peker and colleagues (2025), they found that high HB was considerably associated with major cardiovascular and cerebrovascular adverse events compared to AHI of ≥30 events/hour in adults with moderate-to-severe OSA [10].

In the light of uncertainty towards AHI utility and growing scientific literature supporting use of HB for OSA severity classification, it is imperative to examine whether AHI as a standalone measure is sufficed or HB can act as a complimentary metric to facilitate identification of high-risk individuals with OSA. To delineate this objective, the present research study was aimed to assess whether or not there is agreement between AHI- and HB-based severity classification of OSA.

Materials and Methods

Ethical approval

Before initiating the study, an ethical approval was obtained from the Ethical Review Committee of the Baptist Hospitals of Southeast Texas, Texas, United States (Protocol # BHSET23.035). Due to retrospective nature of the research study, informed consent was waived by the Ethical Review Committee. The study was performed in accordance with the declaration of Helsinki 1975, and patients' confidentiality was maintained.

Study design, setting, and patients

This was a retrospective single-center observational study conducted at an accredited sleep laboratory in Texas. Patients (aged 18-79) who underwent full-night, in-laboratory diagnostic polysomnography between August 1, 2025 and August 31, 2025 were eligible for inclusion. Patients were excluded if they had prior exposure to positive airway pressure therapy, underwent split-night studies, had incomplete or inadequate overnight oximetry data, or demonstrated a predominance of central sleep apnea. Since the study was designed to compare conventional AHI-based OSA severity with a HB-based severity metric, clinical outcomes were not evaluated.

Polysomnography and sleep metrics

All polysomnographic studies were performed and scored according to standard clinical practice. According to the American Academy of Sleep Medicine (AASM) Clinical Practice Guideline [11], severity of OSA was graded based on the AHI. Currently, AHI is the widely accepted yardstick for severity assessment of OSA owing to its scoring simplicity and

objective measure of apnoea and hypopnea events [12]. The AHI was defined as the total number of apnoea and hypopnoea events per hour of sleep, and the AHI severity was categorized using standard thresholds as per AASM [11]: normal (<5 events/hour), mild (5 to <15), moderate (15 to <30), and severe (≥30). Oxygenation metrics derived from polysomnography reports included average oxygen saturation, lowest oxygen saturation (nadir SpO_2), and the percentage of total time in bed spent below predefined oxygen saturation thresholds (<50%).

Derivation of hypoxic burden (HB) and severity classification

Due to absence of event-level oxygen desaturation curves, HB was calculated through a derived composite proxy which incorporated cumulative hypoxemia severity, duration, and depth. A Weighted Hypoxemia Index (WHI) was first calculated using cumulative time spent below graded oxygen saturation thresholds. To account for progressive physiological burden with deeper hypoxemia, increasingly greater weights were applied to lower oxygen saturation bands. Discrete saturation bands between adjacent thresholds were used to avoid double-counting cumulative hypoxemia. The WHI was then normalized within the cohort using percentile ranking and scaled to a 0 to 10 severity score, where higher values represented greater HB. Two additional components were also integrated: (1) oxygen desaturation depth, defined as the magnitude by which the nadir SpO_2 fell below 90%, and (2) respiratory event frequency, represented by AHI. Each component was similarly transformed into percentile-based 0 to 10 sub scores. The final HB score from 0 to 10 was calculated as a weighted composite emphasizing cumulative hypoxemia while preserving contributions from desaturation depth and event frequency. Relative weighting prioritized cumulative hypoxemia while retaining proportional contributions from desaturation depth and event frequency. Higher scores indicated greater HB. For categorical analyses, HB severity was stratified a priori into three groups using fixed cut points on the 0 to 10 scale: mild (<3.34), moderate (3.34 to 6.66), and severe (≥6.67). These thresholds were selected to approximate tertile distribution within the cohort for comparative purposes and are not validated clinical cut points.

Statistical analysis

Data management and statistical analyses were conducted using Stata version 17.0 (Statistical Software. College Station, TX: StataCorp LLC). Continuous variables were summarized using means with standard deviations, while categorical variables were summarized as counts and percentages. Distribution of AHI- and HB-based OSA severity was assessed using cross-tabulations and Fisher's Exact test *p* value was reported. The correlation between continuous AHI and HB scores was assessed using Pearson correlation coefficient (*r*). Concordance beyond chance was quantified using Cohen's kappa (κ), restricted to patients with OSA (AHI ≥5 events/hour). Discordance between classification methods was examined by categorizing patients into three groups: HB severity greater than AHI severity, equal severity, or HB severity lower than AHI severity. Comparisons for lowest SpO_2 across discordance groups were performed using the Kruskal-Wallis test and a two-sided *p*-value <0.05 was considered statistically significant.

Results

A total of 95 patients met inclusion criteria and were included in the present research study. Among these, 73 (76.8%) patients met the criteria for OSA based on AHI ≥5 events/hour and were included in agreement and discordance analyses. The remaining patients had normal AHI values and were retained for descriptive assessment of HB distribution.

Distribution of AHI- and HB-based OSA severity

Table 1 presents the cross-classification of AHI-based and HB-based severity among study participants. A significant association was observed between the two classifications (Fisher's Exact *p* <0.001), indicating that the distribution of HB severity differed significantly across AHI severity categories. Among individuals classified as severe by AHI, the majority were also classified as severe by HB (23/29, 76.7%), suggesting concordance at the highest severity level. In contrast, agreement was lower in the lower and intermediate severity categories. For example, among those with mild AHI, HB severity was most frequently classified as moderate (14/28, 43.8%), with only 12 patients (36.4%) classified as mild by both measures. Notably, participants with normal AHI demonstrated substantial

heterogeneity in HB, with nearly half classified as mild HB (16/22, 48.5%) and others classified as moderate or severe HB.

Correlation between continuous AHI and HB measures

The scatter plot demonstrated a positive correlation between AHI and HB measures; higher AHI values tend to be associated with higher HB. However, the relationship was moderate and highly variable (Pearson correlation coefficient (r) = 0.663, p <0.001). At low AHI values (approximately 0 to 25), HB values ranged from near 0 to 8. At high AHI values, HB clusters near the upper range (8 to 10). Severe AHI almost always corresponds to high HB, but not vice versa (Figure 1).

Agreement between AHI- and HB-based OSA severity (AHI≥5)

Among patients with OSA, observed agreement between AHI- and HB-based severity classifications was 30.1%. Expected agreement by chance was 17.0%, yielding slight concordance beyond chance (Cohen's $\kappa = 0.168$, $p < 0.001$).

Discordant phenotype analysis (AHI≥5)

Patients were stratified into three groups based on discordance between severity classifications. The HB-based OSA severity was higher than AHI in 20 (27.4%) patients with OSA, equal in 45 (61.6%), and lower in 8 (11.0%), demonstrating frequent discordance between classification methods. Figure 2 presents the comparison of the lowest oxygen saturation (SpO_2 %) across three groups defined by the relationship between HB and AHI: HB > AHI, HB = AHI, and HB < AHI. The boxplot shows that the median lowest SpO_2 is lowest in the HB > AHI group and highest in the HB < AHI group. The interquartile range is widest for HB = AHI, suggesting greater variability in oxygen desaturation in this group. Overall, the figure suggests that patients whose HB exceeds AHI tend to experience more severe oxygen desaturation, whereas those with HB less than AHI maintain higher minimum SpO_2 levels (p=0.011).

Discussion

The AHI remains the standard metric for OSA severity classification [12,13]. However, it is now a well-established fact that AHI measure has a number of limitations which restricts its ability to capture clinically meaningful severity of OSA and determine its physiological repercussions, especially the extent of repetitive nocturnal hypoxemia. In this regard, HB is considered a potential measure to effectively determine OSA severity as it accounts for duration and depth of OSA-associated episodes of hypoxemia, in addition to frequency of respiratory events [6,9,12,13]. In the present research study, we aimed to assess the agreement between AHI and HB measure for OSA severity classification in adult individuals who underwent in-laboratory polysomnography studies. Classification and discordance analysis was performed that compared conventional AHI-based severity with a HB-based severity (derived as a 0 to 10 weighted composite score combining a percentile-normalized WHI) measure. Notably, HB-based severity was significantly higher than that of AHI-based severity in 27.4% of patients with OSA, concordant in 61.6%, and lower in 11.0%. Overall, agreement between AHI- and HB-based severity classifications was poor, with an observed agreement of 30.1% and a low kappa value ($\kappa = 0.168$, $p < 0.001$). Nocturnal oxygen saturation was lower in patients in whom HB exceeded AHI severity, highlighting greater physiologic aberration despite comparable event frequency. These findings suggest that reliance on AHI alone may inadequately capture physiological hypoxic burden.

Our findings demonstrate poor agreement between AHI-based and HB-based severity classification of OSA, suggesting that the two measures account for diverse dimensions of OSA severity. The AHI essentially captures the frequency of respiratory events. On the other hand, HB measures the overall duration and depth of oxygen desaturation episodes which may be ideal to constitute physiological stress in patients with OSA [6,9,14,15]. It is therefore patients who have similar AHI values may encounter significantly different hypoxic indices, resulting in conflicting OSA severity classification. The findings of the present research study highlight that dependence on AHI alone may result in underestimation of OSA severity and clinically relevant hypoxemia-associated burden may go unnoticed.

The pathophysiological mechanisms implicated in complications associated with OSA are more closely reflected in HB, for instance, oxidative stress and systemic inflammation [16]. The poor agreement demonstrated in this research study signifies that a significant fraction of patients categorized as mild or moderate by AHI may still be exposed to considerable hypoxic stress. In contrast, some patients with higher AHI values may have a relatively lower burden of cumulative hypoxia, potentially suggesting a dissimilar clinical risk profile. This disagreement supports the notion that HB may offer a complementary information to AHI for a more detailed examination of OSA severity.

Our research study did not assess the outcomes associated with OSA severity, for example, excessive daytime sleepiness, neurocognitive dysfunction, and major cardiovascular and cerebrovascular events. However, the findings of our research study are in line with previous scientific literature documenting poor concordance between frequency-based metrics and hypoxemia-based metrics in OSA severity assessment [9]. Earlier research investigations have shown that HB demonstrates stronger association with all-cause mortality and cardiovascular-associated mortality in comparison with AHI [8,10], strengthening the clinical pertinence of hypoxemia-based measures. Azarbarzin *et al*. (2019) conducted a study in which OSA severity was quantified using the HB and its ability to predict mortality was evaluated. Unlike the AHI, oxygen desaturation events measured as the area under the curve from HB effectively predicted both all-cause mortality and cardiovascular-related mortality [8]. Similarly, in a randomized controlled trial, Peker *et al*. (2025) demonstrated that a high HB was significantly associated with major cardiovascular and cerebrovascular adverse events, showing a stronger predictive value than an AHI ≥30 events/hour in adults with moderate-to-severe OSA [10]. The observed disagreement in severity categories in our study are consistent with reports showing that AHI is poorly correlated with downstream clinical repercussions in some patient cohorts [17]. Recently, Hui *et al*. (2025) demonstrated significant prediction ability of HB for cardiovascular disease associated mortality; however, no correlation was observed for AHI [17]. Altogether, these observations support the emerging standpoint that AHI and HB should not be considered substitutable metrics.

The lack of agreement between AHI- and HB-based classification of OSA severity implicates potential limitations of using AHI alone for clinical decision-making and risk stratification.

Therefore, integration of HB into routine clinical practice of sleep study reporting may aid in identification of patients at higher physiological risk who might otherwise be categorized as lower severity by AHI. This method could perhaps support more personalized assessment of OSA severity and possibly influence treatment prioritization and follow-up plans. Overall, our findings suggest that a combined framework using both AHI and HB may offer more clinically meaningful classification of OSA severity.

Strengths, limitations, and future directions

This study has several limitations. Firstly, it was a single-center research study which restricts the generalizability of findings to other populations and patients who underwent sleep assessment study in other sleep laboratories. Secondly, the present study had a relatively small sample size of 95 individuals which could possibly reduce the statistical power. Thirdly, the study solely focused on patients who underwent sleep assessment; therefore, there is a possibility of selection bias due to non-accountability of individuals with suspected sleep disordered breathing. Another limitation is that analysis involved single-night sleep assessment and therefore night-to-night fluctuation in AHI and HB could not be evaluated. Also, the derived HB score incorporated AHI as one of three weighted components alongside cumulative hypoxemia and desaturation depth, introducing partial construct overlap between the two compared metrics. Consequently, the classification comparison between AHI and HB should be interpreted as an examination of how a hypoxemia-weighted composite reclassifies patients relative to AHI, rather than a validation of two fully independent severity systems. Finally, we did not assess the outcomes associated with OSA severity such as excessive daytime sleepiness and cardiovascular events and therefore we were unable to assess the prognostic utility of HB-based OSA severity classification.

Nevertheless, the present research study is strengthened by below mentioned considerations. This study compared conventional AHI-based OSA severity with HB-based OSA severity classification in the same cohort. Furthermore, the study utilized real-world clinical data derived from routine practice in sleep laboratory. Another strength of the current investigation is that objective measurements were obtained from polysomnography

examinations which ensures the standardized indices of respiratory events and oxygen desaturations. Furthermore, the assessment of agreement between two clinically relevant OSA severity measures is directly pertinent to risk determination and clinical decision-making in sleep medicine.

Herein, we would also take the opportunity to outline future directions based on the limitations of the present research study. Firstly, multiple center research studies with large sample size must be conducted to further validate the complimentary clinical value of HB towards AHI-based OSA severity classification. More studies are needed to evaluate the clinical outcomes such as major cardiovascular and cerebrovascular events using the HB measure in OSA patients. It is equally important to assess the treatment responses to OSA by comparing the predictive ability of both AHI and HB measures. It is also advisable to include population with diverse characteristics such as sex distribution, different age brackets, and comorbidity status to account for OSA heterogeneity. Finally, an appropriate and standardized threshold for HB measure must be developed and validated for routine assessment in sleep laboratories. Prospective validation in larger multi-center cohorts is required before clinical implementation of HB-based severity classification.

Conclusions

In this single-center observational study, AHI- and HB-based severity classifications demonstrated only slight agreement. These findings suggest that reliance on AHI alone may incompletely characterize the physiological consequences of obstructive sleep apnea. Incorporating HB into routine polysomnographic assessment may enhance risk stratification by capturing cumulative hypoxic exposure beyond event frequency.

References

1. 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;7:687-98.

2. Tang Y, Li D, Yang M, et al. Prevalence of excessive daytime sleepiness (EDS) and its association with quality of life in patients with obstructive sleep apnea (OSA): data from

a sleep-center in Shenzhen, a single-center cross-sectional study. J Thorac Dis 2024;16:8216-29.

3. Quan SF, Chan CS, Dement WC, et al. The association between obstructive sleep apnea and neurocognitive performance--the Apnea Positive Pressure Long-term Efficacy Study (APPLES). Sleep 2011;34:303-14B.

4. Yeghiazarians Y, Jneid H, Tietjens JR, et al. Obstructive Sleep apnea and cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2021;144:e56-67.

5. Abbasi A, Gupta SS, Sabharwal N, et al. A comprehensive review of obstructive sleep apnea. Sleep Sci 2021;14:142-54.

6. Parekh A. Hypoxic burden - definitions, pathophysiological concepts, methods of evaluation, and clinical relevance. Curr Opin Pulm Med 2024;30:600-6.

7. Borsini E, Nogueira F, Nigro C. Apnea-hypopnea index in sleep studies and the risk of over-simplification. Sleep Sci 2018;11:45-8.

8. 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;40:1149-57.

9. Martinez-Garcia MA, Sánchez-de-la-Torre M, White DP, Azarbarzin A. Hypoxic burden in obstructive sleep apnea: present and future. Arch Bronconeumol 2023;59:36-43.

10. Peker Y, Celik Y, Zinchuk A, et al. Association of hypoxic burden with cardiovascular events: a risk stratification analysis of the randomized intervention with CPAP in coronary artery disease and sleep apnea cohort. Chest 2025;168:1481-93.

11. 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;13:479-504.

12. Soori R, Baikunje N, D'sa I, et al. Pitfalls of AHI system of severity grading in obstructive sleep apnoea. Sleep Sci 2022;15:285-8.

13. Malhotra A, Ayappa I, Ayas N, et al. Metrics of sleep apnea severity: beyond the apnea-hypopnea index. Sleep 2021;44:zsab030.

14. Redline S, Azarbarzin A, Peker Y. Obstructive sleep apnoea heterogeneity and cardiovascular disease. Nat Rev Cardiol 2023;20:560-73.
15. Coso C, Solano-Pérez E, Romero-Peralta S, et al. The hypoxic burden, clinical implication of a new biomarker in the cardiovascular management of sleep apnea patients: a systematic review. Rev Cardiovasc Med 2024;25:172.
16. Pau MC, Zinellu A, Mangoni AA, et al. Evaluation of inflammation and oxidative stress markers in patients with obstructive sleep apnea (OSA). J Clin Med 2023;12:3935.
17. Hui X, Cao W, Xu Z, et al. Hypoxic indices for obstructive sleep apnoea severity and cardiovascular disease risk prediction: a comparison and application in a community population. Respirology 2024;29:825-34.

Table 1. Comparison of apnea-hypopnea index severity versus hypoxic burden severity measures.

	HB classification			
AHI Classification	**Mild**	**Moderate**	**Severe**	**Row total**
Mild	12 (36.4%)	14 (43.8%)	2 (6.7%)	28 (29.5%)
Moderate	2 (6.1%)	10 (31.3%)	4 (13.3%)	16 (16.8%)
Normal	16 (48.5%)	5 (15.6%)	1 (3.3%)	22 (23.2%)
Severe	3 (9.1%)	3 (9.4%)	23 (76.7%)	29 (30.5%)
Total	33 (100.0%)	32 (100.0%)	30 (100.0%)	95 (100.0%)

AHI, Apnea-Hypopnea Index; HB, Hypoxic Burden.

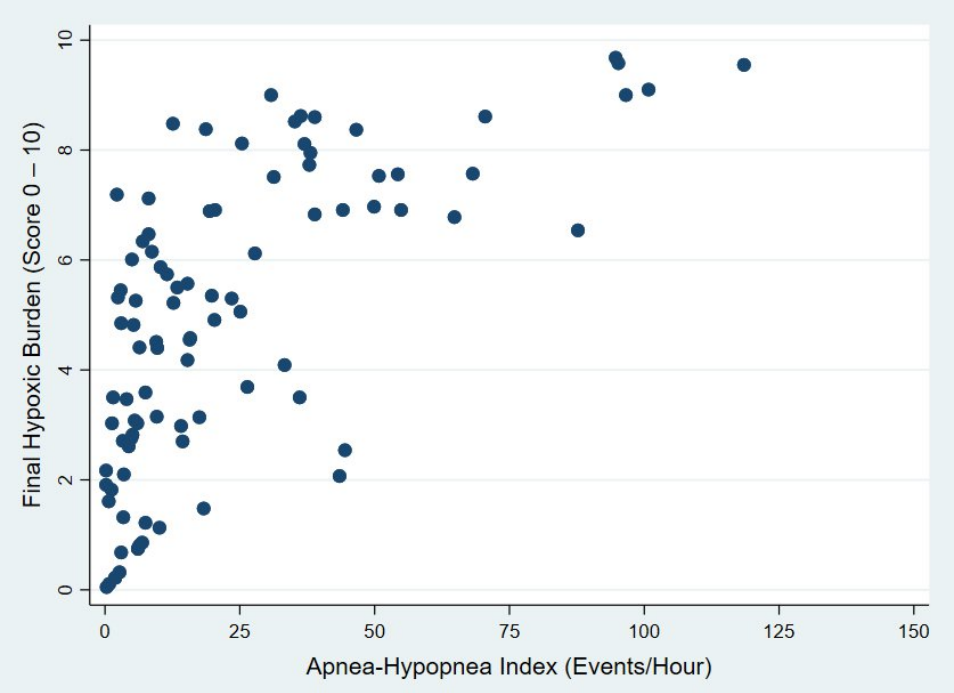

Figure 1. Correlation between apnea-hypopnea index severity *vs.* hypoxic burden severity measures.

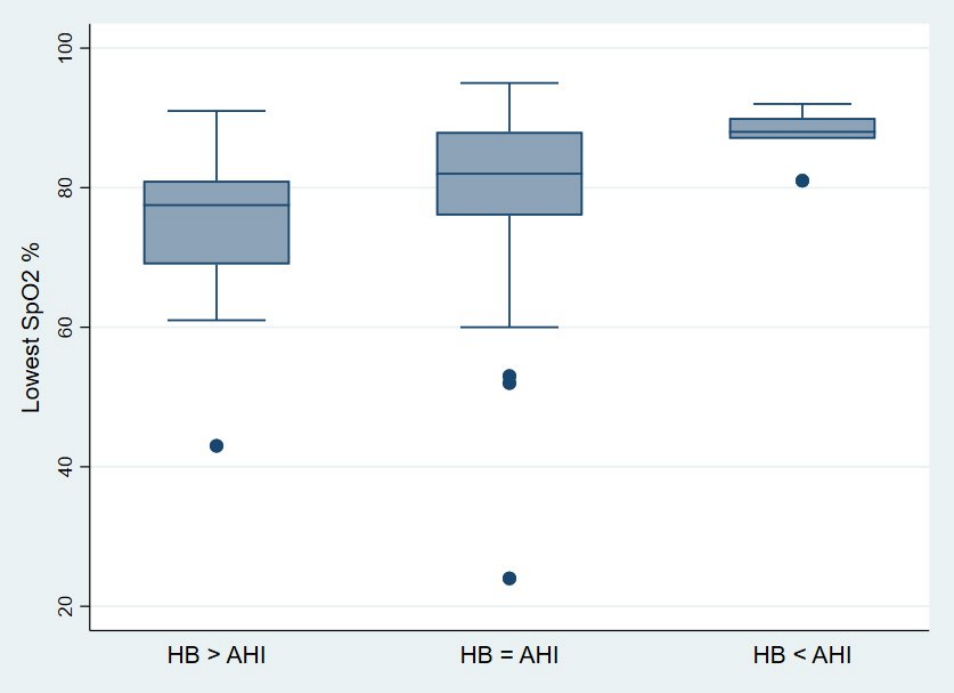

Figure 2. Comparison of the lowest oxygen saturation (SpO_2 %) across three discordant groups (HB > AHI, HB = AHI, and HB < AHI).