Evaluating the Impact of Multi-Night Home Sleep Apnea Testing for Obstructive Sleep Apnea Diagnosis

Yoav Nygate, MSc*, Jan Wodnicki, BSc*, Sam Rusk, BSc*, Chris Fernandez, MSc*, Dennis Hwang, MD**, Nathaniel F. Watson, MD, MSc***
* EnsoData Research, Ensodata, Madison, WI, USA **Kaiser Permanente, Southern California, USA
** Department of Neurology, University of Washington School of Medicine, Seattle, WA

Introduction

Obstructive sleep apnea (OSA) is a prevalent sleep-related breathing disorder affecting a substantial portion of adults in the general population.

Despite growing awareness, a large proportion of individuals with OSA – even those with moderate to severe disease – remain undiagnosed and untreated.

This diagnostic gap is due in part to limited access to sleep testing and the shortcomings of current diagnostic approaches.

The gold-standard for OSA diagnosis is an overnight polysomnography (PSG) performed in a sleep lab. However, many patients face long wait times for in-lab studies, and sleep laboratories have limited capacity, contributing to underdiagnosis.

Home sleep apnea testing (HSAT) has emerged as an accessible alternative, overcoming some limitations of in-lab PSG. Yet, whether conducted in-lab or at home, a single-night study may not capture the full picture of a patient's sleep-disordered breathing.

OSA severity is typically quantified by the apnea-hypopnea index (AHI), which can fluctuate from night to night due to variations in sleep depth, body position, and other factors.

It is well established that the "first night effect" (sleep disruption due to unfamiliar environment) and normal night-to-night variability can alter sleep quality and respiratory event frequency.

Advances in digital health technology provide new opportunities to improve OSA diagnostics. Modern FDA-cleared wearable devices can monitor sleep and breathing over multiple nights in the patient's home.

Multi-night home testing with such wearables could capture night-to-night variability in OSA severity more reliably than a single-night snapshot.

By aggregating data from several nights, clinicians may obtain a more comprehensive assessment of a patient's condition, reducing the risk of missed diagnoses or underestimation of disease severity.

The present study was designed to evaluate the impact of multi-night HSAT on OSA diagnosis accuracy.

Methodology

We conducted a retrospective observational study of patients who underwent a HSAT study in 2024-2025 using a wearable PPG-based device.

The data was collected in a "real-world" clinical context, reflecting typical patients presenting for OSA evaluation.

A total of **14,060** adult patients were included. All patients had at least three nights of home sleep recordings available.

To ensure data quality, we included only nights with a minimum of 4 hours of analyzable PPG signals and at least 1 hour of recorded sleep.

The FDA-cleared HSAT system used is a PPG-based wearable device coupled with an automated scoring algorithm that produces an AHI for each night.

The overall severity of OSA for a given night was classified into standard clinical categories based on AHI: No OSA (AHI < 5 events/hour), Mild (AHI 5 to 15), Moderate (AHI 15 to 30), or Severe (AHI ≥ 30 events/hour).

The following analyses were conducted:

1. **Variability Assessment:** Per-subject night-to-night variability was quantified using coefficient of variation (CV; Standard Deviation AHI / mean AHI).
2. **Confidence Interval Precision:** To determine the relationship between number of testing nights and diagnostic precision, we calculated the 95% CI half-width for the mean AHI across all available nights for each patient. Mean CI widths were aggregated across subjects by night count.
3. **Agreement and Bias Analysis:** Bland–Altman analysis was used to compare Night 1 AHI against each subject's multi-night mean. Bias and limits of agreement (LOA) were calculated to assess the reliability of single-night testing.
4. **Severity Stability Evaluation:** OSA severity transitions were tracked by comparing severity on Night 1 with that across subsequent nights. We calculated the percentage of subjects who would have been missed, experienced an increase, or remained stable in their diagnosis.

Results

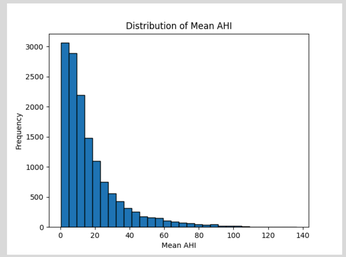

Figure 1. AHI distribution. Mean AHI was calculated across all available nights for each subject. The mean AHI was used as the "true" AHI for all subject going forward in the following analyses. Overall, the mean AHI of this distribution was 17.115 with a standard deviation of 17.443.

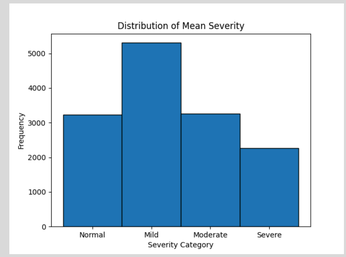

Figure 2. OSA severity distribution. The mean severity was determined based on the mean AHI calculated across all available nights for each subject. The mean severity was used as the "true" OSA severity for all subject going forward in the following analyses. 23% had no sleep apnea, 38% had mild sleep apnea, 23% had moderate sleep apnea, and 16% had severe sleep apnea.

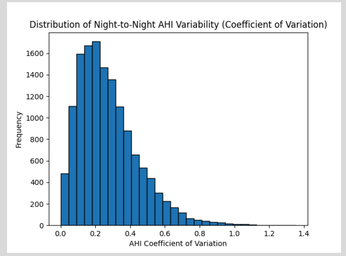

Figure 3. Night-to-Night AHI variability. Variability was assessed for the entire patient cohort utilizing the coefficient of variation (CV). CV is calculated by normalizing the multi-night standard deviation of the AHI by the multi-night mean AHI of each subject.

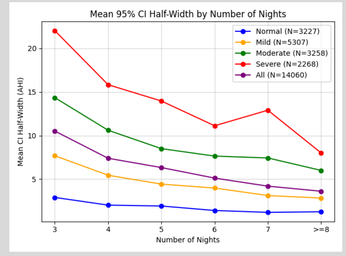

Figure 4. Mean 95% CI half-width by number of nights. This plot shows how the mean 95% CI half-width for AHI declines as more nights are included, increasing the precision for AHI determination. CIs were calculated using the t-distibution and confirmed precision plateaus after ~7 nights.

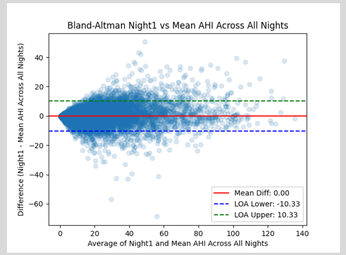

Figure 5. Bland-Altman analysis. A Bland–Altman plot showing differences between Night 1 AHI and the mean across all nights. Mean bias and 95% limits of agreement are overlaid to highlight under- or overestimation trends from relying on a single-night diagnosis.

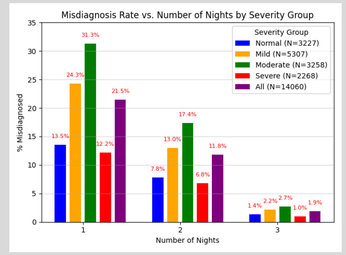

Figure 6. Misdiagnosis rate by number of nights. This chart shows the percentage of patients who would be misdiagnosed based on 1, 2, or 3 nights compared to the mean OSA severity determined from all available nights. Misdiagnosis dropped significantly with each additional night, increasing the diagnosis certainty.

Conclusions

This study provides compelling evidence that multi-night HSAT substantially improves diagnostic accuracy compared to traditional single-night assessments.

Our analysis shows that while single-night testing yields a misclassification rate of 12%–21.5%, extending to just three nights reduces this rate to approximately 1%–2%.

Moreover, as additional nights are included, the diagnostic precision increases. Testing over seven or more nights narrows the CI of the mean AHI to within ±4-8 events/hour, enhancing certainty around the diagnosis.

Based on these findings, a minimum of three nights would produce a more reliable diagnostic classification.

For cases where greater AHI precision is clinically relevant—such as borderline severity or treatment planning—seven nights of data could increases the robustness of the AHI determination.

Multi-night testing captures intermittent or fluctuating OSA patterns that single-night studies may miss.

This ensures treatment decisions are grounded in a more comprehensive assessment—whether based on maximum observed severity or a stabilized mean—rather than potentially anomalous nightly values.

For patients near diagnostic thresholds, additional nights can resolve ambiguity in OSA diagnosis, which has direct implications for treatment eligibility and urgency.

Furthermore, multi-night testing reduces the need for repeat studies. A multi-night approach provides greater confidence from the outset, streamlining the care pathway and reducing patient burden.

In summary, multi-night HSAT enhances diagnostic fidelity, informs more accurate and individualized treatment decisions, and improves overall efficiency in the management of OSA.