Making Type III HSAT Recordings Conclusive: Large-Scale Validation of an FDA-cleared AI AHI Classifier from HSAT Signals

Ágústson JS[1], Hugosdóttir R[1], Finnsson E[1], Sigmarsdóttir ÞB[1], McPhee M[1], Jonsson SÆ[1], Arnardóttir E[1], Erlingsson E[1], Árnadóttir G[1], Montazeri K[1], Wilcox D[1], Riney H[2]

1. *Nox Research, Nox medical ehf, Reykjavík, Iceland*
2. *Nox Research, Nox Health llc., Alpharetta, GA, United States.*

Presented at SLEEP 2026, Baltimore, Maryland, USA

thora@noxmedical.com

Introduction

Type III home sleep apnea tests (HSAT) lack electroencephalography (EEG) for sleep and arousal scoring. Currently, a respiratory event index based on monitoring time, rather than an apnea-hypopnea index (AHI), is commonly reported from HSATs. This underestimates disease severity and leaves HSAT studies inconclusive, necessitating in-lab polysomnography (PSG). We evaluated whether DeepRESP v2.0 (K252330), an FDA 510(k) cleared medical device, yields conclusive AHI-based severity classification aligned with attended PSG from HSAT signals commonly acquired in patients' homes.

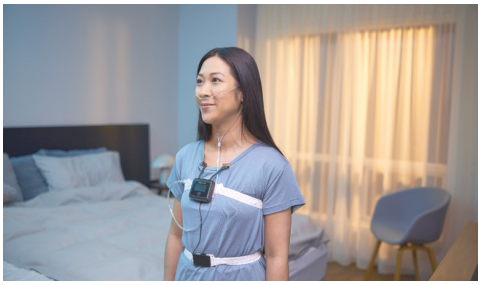


Figure 1. Type III home sleep apnea tests (HSAT), see Figure 1 A, lack the EEG signals commonly recorded in Type I/II polysomnography recordings, see Figure 1 B, used for arousal and sleep scoring.

Methods

We conducted a retrospective, multicenter validation using 5,771 Type I/II polysomnograms (PSGs(from accredited US sleep programs, manually scored by registered technologists and board-certified sleep physicians. To our knowledge, this dataset is substantially larger than the cohorts typically used in FDA 510(k) validation studies of HSAT devices. HSAT channels (nasal pressure, thoracoabdominal inductance belts, oximetry, body position) were fed to the device to score wake/NREM/REM, arousals, and respiratory events, and to compute AHI. Reference AHI came from full-montage manual scoring. Positive percentage (PPA), negative percentage (NPA), and overall percentage agreement (OPA) with manual AHI categories (≥5, ≥15, ≥30 events/hour) were estimated with 95% confidence intervals. Subgroups summarized heterogeneity in age, sex, BMI, severity, and race/ethnicity.

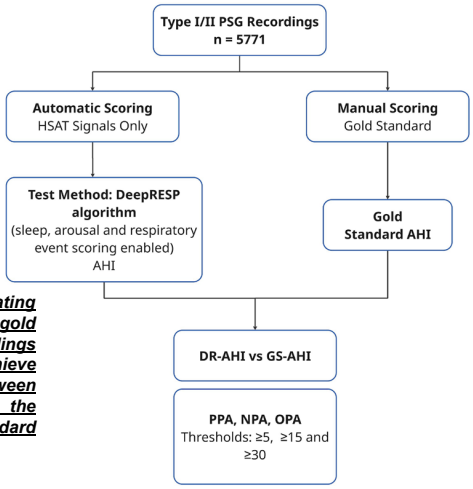


Figure 2: A flow diagram illustrating how the original gold standard PSG recordings were utilised to achieve comparable results between the test method and the current gold standard method.
GS: Gold Standard

Results

The cohort was heterogeneous: adults aged 18 to >65 years; 35% were female; 58% had BMI ≥30 kg/m²; AHI ranged from normal to severe; and recordings included PSGs from both in-laboratory and home settings. For AHI classification, agreement with manual scoring was high across thresholds: for AHI ≥5, PPA 91.0%, NPA 78.0%, OPA 90.6%; for AHI ≥15, PPA 78.1%, NPA 93.9%, OPA 81.7%; and for AHI ≥30, PPA 69.9%, NPA 97.5%, OPA 85.0%. High agreement was observed across demographic subgroups (age, sex, BMI, and race/ethnicity), supporting consistent performance across a diverse adult population.

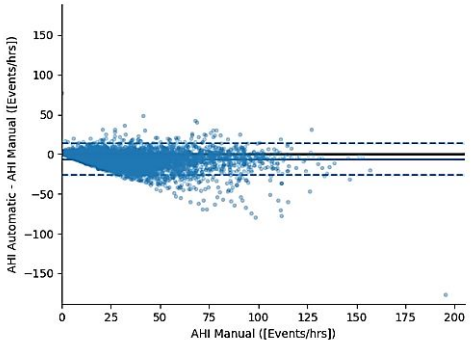
Table 1: PPA, NPA and OPA of DeepRESP scoring based on HSAT signals, vs. Gold Standard PSG scoring, with regards to classifying AHI ≥ 5, AHI ≥ 15, and AHI ≥ 30

	PPA % [95%CI]	NPA % [95%CI]	OPA % [95%CI]
	AHI ≥ 5		
DeepRESP HSAT Scoring	91.0 [90.2, 91.7]	78.0 [71.9, 83.9]	90.6 [89.9, 91.4]
	AHI ≥ 15		
DeepRESP HSAT Scoring	78.1 [76.8, 79.2]	93.9 [92.4, 95.1]	81.7 [80.7, 82.7]
	AHI ≥ 30		
DeepRESP HSAT Scoring	69.9 [68.2, 71.6]	97.5 [96.9, 98.0]	85.0 [84.0, 85.9]

Table 1: Subgroup analysis results, PPA, NPA and OPA of DeepRESP scoring based on HSAT signals, vs. Gold Standard PSG scoring, with regards to classifying AHI ≥ 5, AHI ≥ 15, and AHI ≥ 30

	PPA % [95%CI]	NPA % [95%CI]	OPA % [95%CI]
	AHI ≥ 5		
Female	82.0 [80.2, 83.8]	80.5[72.7, 88.2]	82.0 [80.2, 83.8]
Male	95.7 [95.0, 96.3]	73.9 [61.6, 84.0]	95.7 [95.0, 96.3]
African American	97.0 [93.9, 99.3]	100 [100, 100]	97.0 [93.9, 99.3]
Caucasian	94.8 [92.7, 96,8]	87.5 [50, 100]	94.8 [92.7, 96,8]
	AHI ≥ 15		
Female	60.0 [57.3, 62.8]	96.7 [95.4, 97.9]	73.1 [71.2, 75.2]
Male	85.5 [84.3, 86.7]	90.6 [88.3, 92.8]	86.3 [85.2, 87.4]
African American	90.8 [85.0, 95.4]	90.0 [76.7, 100.0]	90.4 [85.2, 94.8]
Caucasian	87.5 [84.1, 90.7]	89.9 [82.1, 96.1]	87.9 [84.8, 90.8]

Figure 3. Bland-Altman for AHI, with mean bias and LOA, from DeepRESP v2.0 HSAT configuration compared to manual scoring



Conclusions

AHI derived from the device closely aligns with conventional AHI from attended polysomnography, supporting conclusive severity classification from HSAT and potentially reducing inconclusive HSAT studies and repeat in-lab testing, thereby streamlining the diagnostic pathway and earlier linkage to therapy.