Epoch-level Agreement of Manual Sleep Stage and Arousal Scoring using a Reduced Montage versus Conventional Polysomnography

M.E. McPhee Christensen[1], K. Montazeri[1], S.Æ. Jónsson[1], E. Arnardóttir[1], R. Hugosdóttir[1], Þ.B. Sigmarsdóttir[1], S. Helgason[1], H. Riney[2], F. Kramer[3], and J.S. Ágústsson[1]

1. Nox Research, Nox medical ehf, Reykjavík, Iceland
2. Nox Research, Nox Health llc., Alpharetta, GA, United States.
3. Medical Devices & eHealth, R&D Pharmaceuticals, BAYER AG, Wuppertal, Germany

Presented at SLEEP 2024, Houston, Texas, USA

thora@noxmedical.com

Introduction

Conventional polysomnography (PSG) is costly, time-consuming, and can be uncomfortable for patients. Home sleep apnea testing (HSAT) devices offer a simpler and more convenient alternative, but they typically lack the necessary electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG) signals which are used during manual sleep stage and arousal scoring. Here we introduce a reduced frontal EEG/EOG/EMG, montage (Nox SAS, Fig. 1) that can be self-applied following instruction from a healthcare professional, and used at home, while still providing the necessary signals to allow for sleep stage and arousal scoring.

The aim of this study was to compare epoch-level agreement of manual sleep stage and arousal scoring using the reduced frontal montage versus conventional PSG montage.

Figure 1: Image of the reduced frontal montage Nox SAS solution connected to a Nox A1s Recorder

Methods

Retrospective data was obtained from 102 adults with obstructive sleep apnea, who underwent up to 3 in-lab sleep studies each, with a dual hook-up (NCT04713826). Sleep studies were conducted simultaneously with 2 Nox A1s recorders: one hooked up as full PSG according to the AASM Scoring Manual, and the other recording only from the reduced frontal montage (including 5 double electrodes placed over E2/E4, E1/E3, AF3/AF7, AF4/AF8 and an AFz/ground pair, Fig. 2).

Figure 2: Reduced frontal montage electrode placements (Nox SAS Solution)

Studies were manually scored by a Registered Polysomnographic Technologist using either the reduced or full EEG montage while blinded to the other montage. All other signals were shared. Mean sensitivity, specificity, and accuracy were calculated for epoch-level agreement in each sleep stage and for arousal presence.
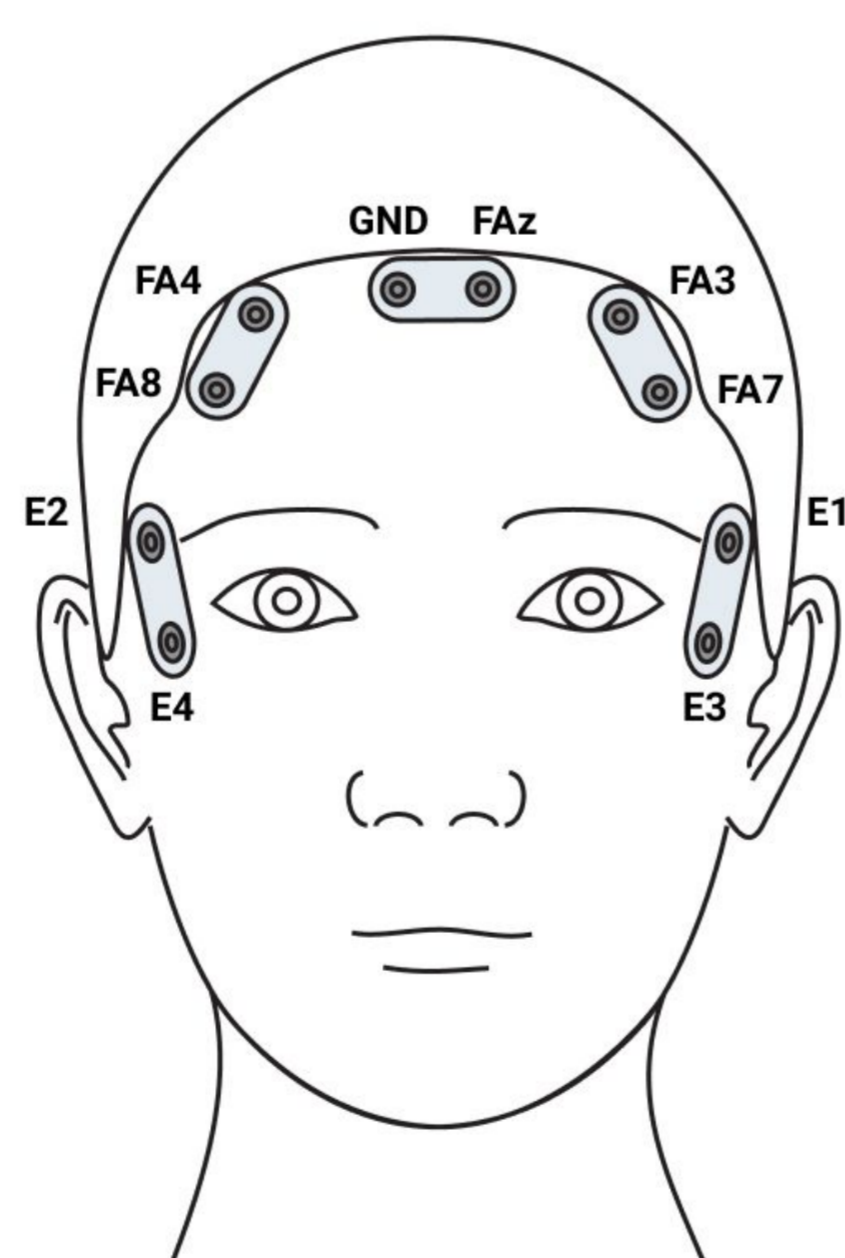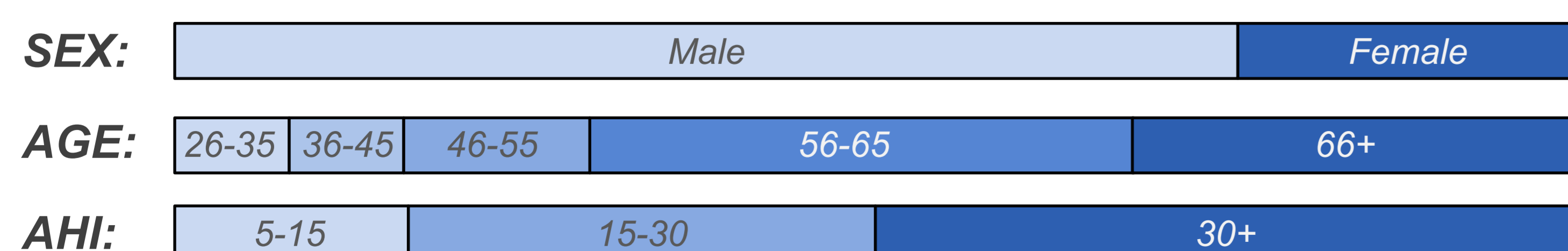
Study Population

Figure 3: Overview of sample characteristics shown as proportions per sex, age and apnea-hypopnea index (AHI) group (N = 102 patients and N = 244 studies)

Results

A total of 244 recordings were obtained from the study participants. An overview of sample characteristics is shown in Fig. 3. The sensitivity, specificity, and accuracy for each sleep stage, and for the presence of arousals, is shown in Table 1 below. The highest levels of agreement overall were observed for wake, REM and N3 sleep stages, and lowest for N1 and arousals which are notoriously difficult to score consistently.[1,2]

Table 1: Sensitivity, specificity and overall accuracy of epoch-level agreement in manual scoring from the reduced frontal montage compared to conventional PSG

Class	Sensitivity [95%CI]	Specificity [95%CI]	Accuracy [95%CI]
Sleep Stage Classification			
Wake	0.84 [0.82, 0.86]	0.95 [0.94, 0.95]	0.92 [0.91, 0.93]
REM	0.78 [0.75, 0.80]	0.99 [0.99, 0.99]	0.96 [0.96, 0.96]
N1	0.48 [0.46, 0.50]	0.94 [0.94, 0.95]	0.88 [0.87, 0.89]
N2	0.76 [0.75, 0.78]	0.87 [0.86, 0.88]	0.82 [0.81, 0.83]
N3	0.94 [0.92, 0.95]	0.92 [0.91, 0.93]	0.92 [0.91, 0.93]
Arousal Scoring			
Arousal	0.69 [0.67, 0.70]	0.89 [0.88, 0.89]	0.85 [0.85, 0.86]

Examination of confusion matrices (Table 2) reveals that misclassification most often occurs between adjacent sleep stages, for example with N1 typically misclassified as either N2 sleep or wake.

Table 2: Confusion matrices showing epoch-level agreement of manual scoring from the reduced montage with PSG for each sleep stage and arousal presence

		Reduced Frontal Montage (% of Epochs)					
	Sleep stage	WAKE	REM	N1	N2	N3	Total (N)
Conventional PSG	WAKE	**83.71%**	1.76%	6.74%	7.36%	0.43%	51,126
	REM	10.26%	**77.89%**	6.12%	5.62%	0.10%	27,207
	N1	15.14%	2.96%	**47.88%**	33.00%	1.02%	29,664
	N2	1.22%	0.57%	5.36%	**76.38%**	16.47%	93,910
	N3	0.06%	0.00%	0.04%	6.15%	**93.75%**	10,936
	Arousal presence	Arousal		No Arousal			Total (N)
	Arousal	**68.98%**		31.02%			51,126
	No Arousal	11.31%		**88.69%**			27,207

Cohen's kappa, which adjusts overall accuracy for data distribution, was calculated to compare to prior work (Table 3) and shows that the present results are consistent with the typical levels of interscorer reliability seen within manual scoring of reduced montage or PSG studies.

Table 3: Cohen's kappa for overall accuracy of manual scoring on reduced frontal montage versus conventional PSG in the present study compared to prior work

	Reduced frontal montage vs PSG	Meta-analysis of PSG scoring reliability[1]	Interscorer reliability on Nox SAS alone[3]	AASM interscorer reliability on PSG[4]
WAKE	0.81 [0.80, 0.83]	0.70 [0.63, 0.77]	0.76	0.82
REM	0.83 [0.81, 0.84]	0.69 [0.58, 0.81]	0.86	0.90
N1	0.49 [0.48, 0.51]	0.24 [0.15, 0.33]	0.41	0.52
N2	0.73 [0.72, 0.75]	0.57 [0.54, 0.60]	0.67	0.72
N3	0.53 [0.50, 0.57]	0.57 [0.42, 0.71]	0.74	0.53

Conclusions

The reduced frontal montage showed a high level of agreement with conventional PSG for manual scoring of sleep stages and arousals, consistent with levels of interscorer agreement reported for these tasks in prior work. The reduced montage may therefore be a suitable alternative to PSG for measuring sleep stages and arousals, and may be able to perform these measurements in a real-world home sleep environment. Accurate sleep stage and arousal scoring allows for more accurate estimates of downstream clinical parameters (such as AHI) than HSAT, though further studies are needed, especially in a home setting, to evaluate the accuracy of such parameters and determine the full potential of this reduced montage.

References:
1. Lee YJ, Lee JY, Cho JH, Choi JH. Interrater reliability of sleep stage scoring: a meta-analysis. *Journal of Clinical Sleep Medicine.* 2022;18(1):193-202. doi:10.5664/jcsm.9538; **2.** Pitkänen H, Nikkonen S, Rissanen M, et al. Multi-centre arousal scoring agreement in the Sleep Revolution. *Journal of Sleep Research.* Published online December 26, 2023:e14127. doi:10.1111/jsr.14127; **3.** Nikkonen S, Somaskandhan P, Korkalainen H, et al. Multicentre sleep-stage scoring agreement in the Sleep Revolution project. *Journal of Sleep Research.* 2024;33(1):e13956. doi:10.1111/jsr.13956; **4.** Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine Inter-scorer Reliability Program: Sleep Stage Scoring. *Journal of Clinical Sleep Medicine.* 2013;09(01):81-87. doi:10.5664/jcsm.2350