Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A novel, wearable, in-ear EEG technology to assess sleep and daytime sleepiness

Jonathan Berent

jb@nextsense.io

NextSense Inc.

Prabhjyot Saini
Emory University School of Medicine

Andrew X. Stewart
NextSense Inc.

Devansh Bheda
Infocusp, Inc.

Konstantin Borodin
NextSense Inc.

Akshay Paul
NextSense Inc.

Brian Tracey
Takeda Development Center Americas, Inc.

Caroline Dong
Emory University School of Medicine

Dmitri Volfson
Takeda Development Center Americas, Inc.

Derek L. Buhl
Takeda Development Center Americas, Inc.

David B. Rye
Emory University School of Medicine

Research Article

Keywords: In-ear EEG, Narcolepsy (NT1, NT2), Sleep Wearables, PSG, Excessive Daytime Sleepiness (EDS), Central Disorders of Hypersomnolence, Idiopathic Hypersomnia (IH), Maintenance of Wakefulness Test (MWT)

Posted Date: March 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9162683/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. J.B., A.X.S., K.B., and A.P. are employees of NextSense Inc. B.T., D.V., and D.L.B. were employees of Takeda Development Center Americas, Inc. at the time of the study. P.S., C.D., and D.B.R. declare no competing interests.

Version of Record: A version of this preprint was published at Bioelectronic Medicine on May 22nd, 2026. See the published version at https://doi.org/10.1186/s42234-026-00207-x.

A novel, wearable, in-ear EEG technology to assess sleep and daytime sleepiness

Jonathan Berent*.[1], Prabhjyot Saini[2], Andrew X. Stewart[1], Devansh Bheda[4], Konstantin Borodin[1], Akshay Paul[1], Brian Tracey[3], Caroline Dong[2], Dmitri Volfson[3], Derek L. Buhl†[3], and David B. Rye[2]

[1]NextSense Inc., Mountain View, California, USA
[2]Emory Sleep Center, Emory University School of Medicine, Atlanta, Georgia, USA
[3]Takeda Development Center Americas, Inc., Cambridge, Massachusetts, USA
[4]Infocusp, Inc.

Corresponding author

Jonathan Berent
Address: 703 N Shoreline Blvd, Mountain View CA 94043 **Email:** jb@nextsense.io

Abstract

In-ear EEG has emerged as a promising alternative to traditional sleep studies, offering greater comfort and convenience. Here we share a novel in-ear EEG system, including validation of sleep scoring accuracy in comparing both in-ear EEG and full scalp EEG polysomnography (PSG). We observed a high degree of agreement between in-ear EEG and full-scalp EEG PSG (Cohen's Kappa of **0.76**). For the primary MWT endpoint of sleep onset latency (SOL), at the per-subject level ($n = 24$)—averaging across trials as in standard clinical practice—agreement was good (ICC = 0.71, $r = 0.75$, MAD = 5.1 min). Among the 37 of 126 trials where both devices detected sleep, agreement was strong (ICC = 0.82, MAD = 1.9 min), with excellent agreement in healthy controls (ICC = 0.95, MAD = 1.2 min). Overall trial-level agreement across all 126 trials was moderate (ICC = 0.55), reflecting 22 discordant trials in which scalp EEG detected sleep but in-ear EEG did not—predominantly brief, subtle N1 transitions in CDH participants.

For overnight sleep architecture ($n = 16$ healthy controls), total sleep time ($r = 0.94$, ICC = 0.85), sleep efficiency ($r = 0.94$), and wake after sleep onset ($r = 0.93$) showed strong agreement, with small systematic biases consistent with reduced N1 detection sensitivity. This supports the ability of in-ear EEG data to detect different sleep stages, monitor the effects of medication on sleep, and assess daytime sleepiness. This paper validates a novel in-ear EEG device designed for enhanced comfort, user-friendly operation, and clinical-grade signal integrity.

*Corresponding author: Jonathan Berent. Address: 703 N Shoreline Blvd, Mountain View CA 94043. Email: jb@nextsense.io

†At the time of the study.

Keywords

In-ear EEG, Narcolepsy (NT1, NT2), Sleep Wearables, PSG, Excessive Daytime Sleepiness (EDS), Central Disorders of Hypersomnolence, Idiopathic Hypersomnia (IH), Maintenance of Wakefulness Test (MWT)

1 Introduction

Sleep is a fundamental biological process essential for the survival and wellbeing of mammalians including humans (Tobler 2011) that is tightly controlled and regulated (Achermann et al. 2017). Sleep plays multiple critical roles in various physiological and cognitive functions (Leong and Chee 2023) including learning (Stickgold et al. 2001), development (Mander, Winer, and Walker 2017; Paruthi et al. 2016), memory consolidation (Stickgold et al. 2001; Diekelmann and Born 2010), cardiovascular health, emotional regulation (Goldstein and Walker 2014), metabolic homeostasis (Xie et al. 2013), immune system regulation (Besedovsky, Lange, and Haack 2019), rhythm regulation (Saper, Scammell, and Lu 2005), intracellular trafficking (Cirelli, Gutierrez, and Tononi 2004), and generally our broader physical and psychological health and homeostasis (Buysse 2014).

Understanding sleep requires the ability to measure accurately, efficiently, effectively and longitudinally. The primary method for this in the clinic is through polysomnography (PSG), a costly and burdensome sleep study conducted in a specialized facility. PSG employs surface electrodes to monitor neural activity (EEG), eye movements (EOG), muscle activity (EMG), heart rhythms (ECG), and breathing functions during sleep (Vaughn and Giallanza 2008). However, the gold standard PSG setups are complex, time consuming (Malhotra et al. 2013), expensive (Flores et al. 2016), restrictive to in-laboratory settings only, may worsen sleep quality (Byun et al. 2019; Newell et al. 2012), and can be subject to diagnostic misclassification due to symptom variability across nights (Punjabi et al. 2020).

Wearable EEG technologies have made brain-based sleep measurement more accessible and precise, allowing for integration into longitudinal studies. Ear-centered electroencephalography, first introduced in 2011 (Looney, Kidmose, and Mandic 2011) and subsequently evaluated and validated against scalp EEG (Kidmose et al. 2013; Looney, Kidmose, and Mandic 2012; Mikkelsen, Villadsen, et al. 2019), has evolved considerably with multiple form factors including around-the-ear (Bleichner and Debener 2017), behind-the-ear (Sterr et al. 2018), customized earpieces (Mikkelsen, Kappel, et al. 2015; Kappel et al. 2019) and generic earpieces (Goverdovsky et al. 2016). Here, we present and validate a novel in-ear EEG system designed to (i) achieve high-quality, reliable biosignal acquisition for sleep monitoring; (ii) maximize comfort and practicality for sleep and daytime testing; and (iii) assess performance in both healthy controls and participants with central disorders of hypersomnolence (CDH). The hardware of this NextSense in-ear EEG system has been demonstrated in epilepsy and seizure prediction EEG work (Joyner et al. 2024), and here we evaluate this same in-ear EEG system for sleep state classification.

2 Methods

2.1 Study Design and Setting

This study evaluated the feasibility and validity of an in-ear EEG system (NextSense Inc., Mountain View, CA) for assessing sleep architecture and daytime sleepiness relative to conventional in-laboratory polysomnography (PSG) and American Academy of Sleep Medicine (AASM) scoring

standards. Recordings were conducted at the Emory Clinic Sleep Laboratory / Emory University Sleep Laboratory (Brain Health Center; 12 Executive Park Drive NE, Atlanta, GA 30329).

Briefly, the healthy control arm followed a single-visit design comprising an overnight, sleep-restricted PSG followed by daytime Maintenance of Wakefulness Testing (MWT). A separate patient arm enrolled participants with central disorders of hypersomnolence (CDH) for daytime MWT testing in a randomized crossover design (medication vs. medication-holiday) as described below.

Table 1: Participant characteristics by study.

Variable	**Category**	**Arm 1 (n=16)**	**Arm 2 (n=8)**	**p-value**
Demographics				
Age (y)	–	32.00 ± 1.72	42.14 ± 6.18	0.284
BMI	–	23.65 ± 1.21	26.80 ± 1.77	0.105
Initial Sleep Assessment				
ESS score	–	5.62 ± 0.66	15.00 ± 1.39	5.7×10^{-7}
Owl–Lark total	–	55.81 ± 2.81	51.75 ± 3.10	0.602
BDI total	–	1.25 ± 0.36	1.38 ± 0.65	1.000
Sex	Female	8 (50%)	2 (25%)	0.707
Sex	Male	8 (50%)	6 (75%)	—

2.2 Ethics, Oversight, and Consent

All participants provided written informed consent prior to participation. The study was conducted under Emory University Institutional Review Board approval (Emory IRB version 00-00-2019) and registered on ClinicalTrials.gov (healthy controls: NCT05114616; CDH cohort: NCT05066009). Study procedures were monitored by the principal investigator and trained sleep laboratory personnel. Adverse events were not anticipated; any events were documented and reported per institutional requirements.

2.3 Participants

2.3.1 Healthy Controls

Sixteen healthy adult volunteers (ages 18–60 years) residing within ∼20 miles of the Emory Sleep Center were recruited via flyers and local advertisements. Inclusion criteria included: normal BMI (18.5–28.0 kg/m^2), regular sleep-wake habits (7–8.5 hours nightly sleep with consistent bed/wake times within 1–2 hours, and typical bedtime between 9:30–11:30 p.m.), and no routine use of sleep aids or wake-promoting medications.

Exclusion criteria included: history of diagnosed or suspected sleep disorder (e.g., sleep-related breathing disorders, circadian rhythm sleep-wake disorders, central disorders of hypersomnolence, parasomnias, sleep-related movement disorders), regular caffeine consumption >200 mg/day, major psychiatric disorder or suicidal ideation, substance dependence within the past 6 months (including nicotine, marijuana, alcohol, or other substances likely to affect sleep), attention deficit/hyperactivity disorder, recent shift work (past 6 months), or inability to safely tolerate earbuds due to recent injury, skin breakdown, or infection.

2.3.2 Central Disorders of Hypersomnolence (CDH) Cohort

Participants with CDH (narcolepsy type 1 [NT1], narcolepsy type 2 [NT2], or idiopathic hypersomnia [IH]) were recruited from the Emory Sleep Center clinical database of patients who had provided consent to be contacted for research. Inclusion criteria included: age 18–60 years; prior clinical diagnosis of NT1, NT2, or IH; and use of daily wake-promoting medication(s) for $\geq$3 months with documented symptomatic benefit. Exclusion criteria included: suspected or diagnosed comorbid sleep-related breathing disorder, parasomnia, or other sleep-related movement disorder; unstable psychiatric disorder or active substance abuse; recent malignancy (active or remission $<$ 2 years) or active infection at screening; inability to safely tolerate earbuds; uncorrected near visual acuity worse than 20/50; or other investigator-determined conditions likely to affect safety or outcomes.

2.4 Recruitment and Screening

Healthy controls were screened by research staff for eligibility using medical history, medication review, and protocol inclusion/exclusion criteria. Participants meeting eligibility criteria completed an enrollment visit and were familiarized with the sleep laboratory environment. Prior to the visit, anatomical scanning was performed at least 3 weeks prior to the first in-clinic visit.

2.5 Devices and Instrumentation

2.5.1 Gold-Standard Polysomnography

Conventional PSG was performed using a AASM-compliant laboratory system (Philips ALICE 6 PSG) with EEG, EOG, chin EMG, leg EMG, ECG, and standard respiratory channels as clinically indicated. Sleep staging was performed on 30-second epochs per AASM scoring rules.

2.5.2 NextSense In-Ear EEG System

The NextSense in-ear EEG system consisted of custom-fit, EEG-enabled earbuds (EEGBuds) with dry-contact electrodes positioned within the external auditory canals. Custom earmolds were produced from high-resolution ear geometry scans using a ring/laser scanning approach (reported accuracy $\sim$0.1 mm) and fabricated using a 3D-printing workflow to ensure comfort and stable electrode-skin contact. The in-ear EEG system captured EEG and motion signals; in this validation study, in-ear EEG data were recorded concurrently with PSG during overnight sleep and daytime MWT sessions.

2.6 Study Procedures

Healthy controls underwent sleep-restricted overnight PSG followed by a 4-session MWT the following day. Participants with CDH completed two MWT days in a randomized crossover design (medication vs. medication-holiday) (Figure 1).

2.6.1 Overnight Sleep Recording (Healthy Controls)

Healthy participants underwent an overnight PSG in the accredited Emory sleep laboratory while wearing the in-ear EEG earbuds. To increase homeostatic sleep drive during the subsequent daytime MWT sessions, sleep was restricted by extending wakefulness approximately four hours beyond the participant's habitual bedtime, resulting in a sleep opportunity of approximately 3–4 hours. In the finalized study workflow, laboratory lights-out and lights-on were scheduled to restrict time in bed to 4 hours.

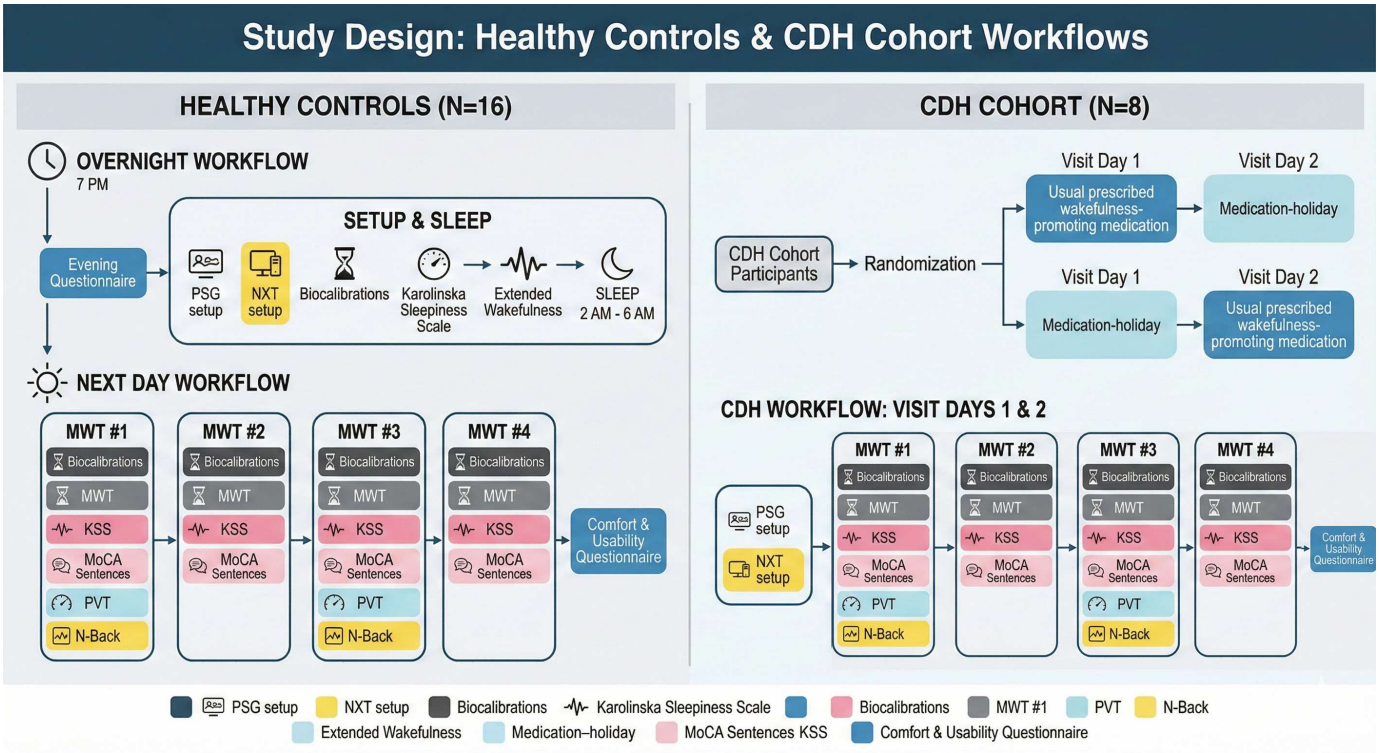

Figure 1: **Study Design** - workflow of sessions for healthy controls and CDH cohort. Controls - participants are subjected to light extended wakefulness before a reduced sleep session, while we record the whole session, and record 4x Maintenance of Wakefulness Tests (MWT) the following day, while presumably experiencing more-than-typical sleep pressure. CDH Cohort - sleep pressure is already higher in this population. There is a randomization of order - drug day or drug holiday on day 1, and the second day is the other condition. There are 4 full MWT sessions for each day.

2.6.2 Daytime Maintenance of Wakefulness Test (Healthy Controls and CDH)

The MWT consisted of four daytime trials scheduled at standardized times of 10:00 a.m., 12:00 p.m., 2:00 p.m., and 4:00 p.m. Each trial lasted up to 40 minutes. Participants were seated in a comfortable chair in a dark, quiet room and instructed to remain awake/vigilant. Standard PSG and in-ear EEG recordings were acquired concurrently during each MWT.

Sleep onset latency (SOL) was defined from "lights off" to sleep onset based on AASM criteria applied to 30-second epochs. Consistent with standard laboratory practice, sleep onset during MWT was scored as either: (i) the first epoch of any sleep stage other than N1 (N2, N3, or REM), or (ii) the first of three consecutive epochs of stage N1. SOL was computed from a 40-minute window beginning at the documented lights-off time, excluding pre-trial calibration epochs. Trials in which neither device detected sleep were assigned a SOL equal to the full trimmed trial duration (up to 40 min). Both in-ear EEG and scalp EEG SOL were derived from the consensus scorer (epoch-by-epoch majority vote of three technologists, with a pre-specified tiebreaker; see Section 2.10) to ensure methodological consistency.

Trials were terminated in real time by the monitoring technologist after scalp EEG showed three consecutive epochs of N1 sleep or one epoch of any other sleep stage, per standard MWT monitoring procedures. Because both devices recorded concurrently, the in-ear EEG annotation

files covered the same duration as the scalp files. The primary SOL analysis includes all 126 MWT trials. In trials where only the scalp EEG detected sleep (and in-ear EEG did not), the in-ear EEG SOL equals the full trial duration; these trials are analyzed separately to characterize the detection discrepancy (Section 3.3).

2.6.3 Patient (CDH) Crossover Procedures and Randomization

CDH participants completed two separate MWT test days (four-trial MWT each day) under a randomized crossover design: (i) medication day (usual prescribed wake-promoting medication[s]) and (ii) medication-holiday day (withholding usual wake-promoting medication[s]). The order of conditions was randomized 1:1 (Group A vs. Group B). Randomization was conducted using the 'Sealed Envelope' web service, with blocked random permuted blocks to balance assignment. Medication administration on study days was recorded by a study coordinator and kept blinded from scorers and analysts as applicable. The second test day was 11.5 ± 5 days after the first.

2.7 Behavioral, Cognitive, and Subjective Assessments

Prior to and following each MWT trial, participants completed:

- **Karolinska Sleepiness Scale (KSS):** a 10-point validated subjective sleepiness scale.
- **Psychomotor Vigilance Task (PVT):** a 10-minute reaction time test of sustained attention and vigilance.
- **N-back task:** a computerized working memory and executive function task (0-back through higher-load conditions as configured).

Between MWT trials, participants performed standardized read-aloud tasks (two sentences from the Montreal Cognitive Assessment [MoCA]) with audio recorded for exploratory biomarkers of alertness.

2.8 Data Acquisition, Synchronization, and Processing

Ear-EEG and PSG streams were time-aligned for analysis. Where applicable, synchronization between systems used a laboratory-compatible marker/sync channel. Depending on the output of an electrically-isolated RNG, the NextSense in-ear EEG device randomly writes a dedicated 'Optosync' channel to high / 100 µV, and simultaneously writes the next 5 values as high via a EEG-scale output channel, before returning to 0 µV for a RNG-determined random wait. This output can be simultaneously streamed and recorded by a clinical EEG or PSG system. With the signals from this Optosync signal recorded via both modalities and all data combined in one biosignal EDF file, any deviation in the timing of these two Optosync channels can be readily identified and corrected. Raw data were exported as EDF-format files for preprocessing and analysis, mostly using in-house Python tools (https://github.com/nextsense-io).

In-ear EEG preprocessing included resampling to match PSG sampling rates (e.g., 250 Hz to 200 Hz), high-pass filtering (e.g., 0.1 Hz), and notch filtering (e.g., 60 Hz) as specified in the study pipeline. PSG channels used standard channel-specific filtering (e.g., EEG 0.3 Hz to 35 Hz, EMG 10 Hz to 95 Hz, EOG high-pass 0.3 Hz) for scoring and analysis.

Spectral power features were computed across canonical frequency bands (delta, 0.5 Hz to 4 Hz; theta, 4 Hz to 8 Hz; alpha, 8 Hz to 13 Hz; beta, 13 Hz to 30 Hz) using FFT-based methods. For daytime sleepiness analyses, the theta/(alpha + beta) ratio was quantified using a moving-window

approach. Micro-sleeps were defined as EEG slowing arising from wakefulness lasting 3–15 seconds and were scored/derived in blinded fashion.

2.9 Manual Sleep Scoring

Three independent registered polysomnographic technologists (RPSGT certified) manually scored PSG and in-ear EEG recordings according to AASM criteria. Scorers were blinded to participant identity and clinical condition, and (for the CDH crossover) to medication assignment. Prior to scoring in-ear EEG, technologists completed an in-ear EEG training and certification procedure using practice data and were required to show competence prior to scoring study data. The certification was repeated at the study midpoint as part of quality control.

For both MWT and overnight sleep sessions, recordings were staged in 30-second epochs as:
0 - Wake (W), 1 - N1, 2 - N2, 3 - N3, 5 - REM
A - Arousal (brief awakening)
- - Empty / unknown

Each of our three expert sleep techs scored annotations alone.

2.10 Consensus Scoring

For both PSG and in-ear EEG, a consensus hypnogram was generated using epoch-by-epoch majority vote across the three technologists. If all three scorers assigned different stages for a given epoch (three-way tie), the stage from the scorer whose labels most frequently appeared in the majority consensus was used as the tiebreaker. The most consensus-aligned scorer was determined independently for each system and condition: LY for PSG MWT, TBR for in-ear EEG MWT, and AM for both PSG and in-ear EEG overnight sleep. Epochs marked as movement time or unscorable by any scorer were excluded from epoch-level agreement analyses.

2.11 Outcome Measures

The primary efficacy endpoint for daytime testing was MWT sleep onset latency (SOL). Secondary endpoints included epoch-level agreement metrics (confusion matrices, F1 scores, Cohen's kappa), inter-rater reliability (Fleiss' kappa), and derived sleep architecture variables including TST, SE, WASO, stage durations and latencies, and REM-related metrics. This manuscript reports the secondary endpoints of sleep staging agreement and overnight sleep architecture, as well as the primary MWT endpoint of sleep onset latency (SOL).

2.12 Analyses

Agreement between in-ear EEG and PSG was evaluated by:

(1) Epoch-level agreement using Cohen's kappa (pairwise) and Fleiss' kappa (multi-rater), with agreement also summarized as percent agreement.

(2) Quantitative agreement of derived sleep parameters using intraclass correlation coefficients (ICC) and absolute differences; and

(3) Visual inspection of EEG trace data, and hypnograms from multiple raters

2.13 Design of the NextSense In-Ear EEG System

The NextSense In-Ear EEG System device consists of a 3D-printed enclosure housing the electronics including an analog-to-digital converter (ADC), micro-SD card reader, lithium-ion battery, opto-sync mechanism for synchronizing with clinical systems, and additional microprocessors and components for communication and related functions (Figure 2).

Two custom-molded earbuds are connected to the enclosure via a micro-HDMI cable. Each earbud has been 3D printed with biocompatible resin from a 3D scan of the subject's ear using an ear scanning workflow and modeling software. Earbuds are coated with a proprietary conductive polymer on regions corresponding to the canal and cymba of each subject. Silver rivets are threaded through the earbud to bridge the conductivity between the earbud and the housing module (Waters et al. 2024). Proper insertion requires little-to-no preparation, other than avoiding an excess of wax. For EEG validation testing, the recording box was connected to the PSG headbox via an "Optosync" channel outputting a 20-sample, 250 Hz signal for post-processing alignment and validation.

The earbud electrodes were fitted and placed as described in previous seizure-prediction work using this in-ear EEG system (Joyner et al. 2024).



Figure 2: **NextSense In-Ear EEG device** - Wired headset with 2 electrodes for each ear (larger being ELW and ERW in left and right ear canal; smaller electrode contact in ELC and ERC - left and right cymba conchae) with micro-HDMI physical connector to the NextSense Xenon in-ear EEG recording device, with recording start / trigger button, micro-SD slot, and optional output of 'Optosync' timing synchronization channel over USB-C physical connector.

3 Results

3.1 In-Ear EEG and Full Scalp EEG-PSG Sleep Stage Agreement

With three qualified independent sleep scoring experts following AASM guidelines, we find broad agreement in reported sleep staging. The in-ear EEG sleep labels agreed with the full scalp EEG consensus at 80.9% agreement (5-class sleep labels, $\kappa = 0.76$).

We can use inter-rater agreement on full-scalp data compared to consensus (also on full-scalp) to contextualize variation of sleep label consistency. Here, the average agreement between a sleep

annotation label and the 'consensus' label was 94.8 % ($\kappa = 0.86$).

The distribution of label confusions can be shown in a label confusion matrix - see Figure 3. Agreement between in-ear EEG and scalp EEG was highest for Wake (93%), followed by N2 sleep (82%), REM (78%), and N3 sleep (67%). In contrast, stage N1 showed substantially lower agreement (53%), reflecting the known difficulty of reliably identifying transitional sleep.

Table 2: Sleep staging agreement: epoch-level metrics

	n	Cohen's κ		F1 Score	
		Ear vs. Scalp	Scalp vs. Scalp	Ear vs. Scalp	Scalp vs. Scalp
All Sessions	142	0.761	0.857	0.910	0.970
By Session Type					
MWT	126	0.767	0.854	0.921	0.972
Sleep	16	0.805	0.918	0.817	0.948
By Subject Group					
Healthy Controls	80	0.781	0.889	—	—
CDH	62	0.755	0.819	—	—

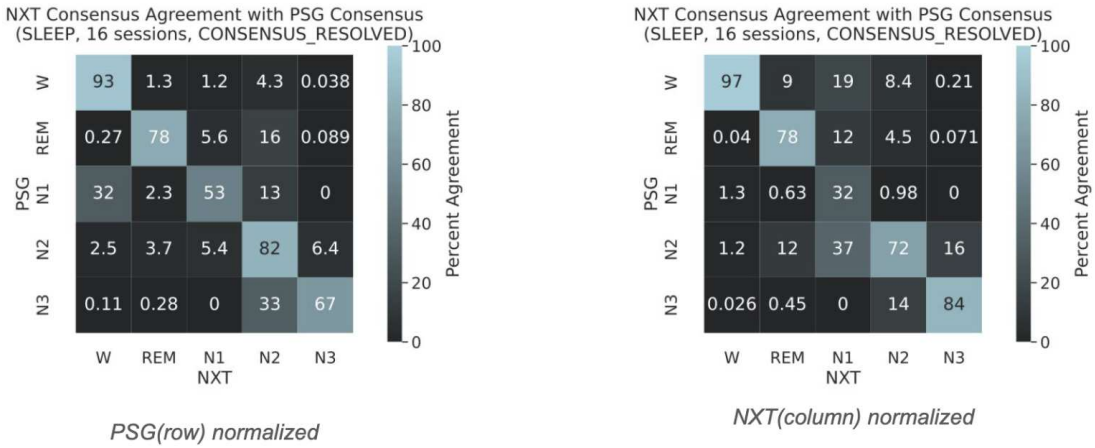

Figure 3: **Epoch-level confusion matrices for in-ear EEG vs. scalp EEG consensus** (16 overnight sleep sessions). Left: PSG (row) normalized, showing sensitivity per scalp-scored stage. Right: NXT (column) normalized, showing positive predictive value per ear-EEG-scored stage.

3.2 Sleep-Wake distribution

For both in-ear EEG and full scalp EEG-PSG, we can score the fraction of each session judged as wake or sleep. This can be additionally split by session type (MWT or sleep-night) and by each of our 3 expert sleep annotators. As shown in Figure 4, all three of our sleep tech expert annotators

find similar mode and median Wake Fraction across the sessions, from both in-ear EEG or scalp EEG.

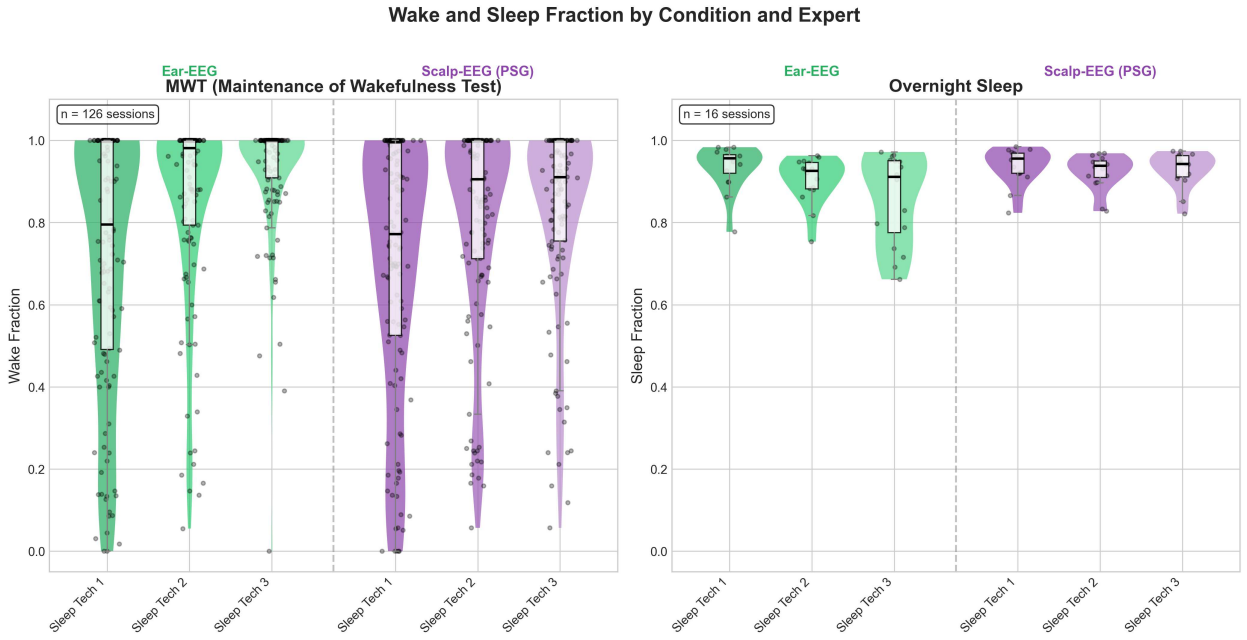

Figure 4: **Wake and Sleep Fraction distribution** -
Left - MWT - Among 126 MWT trials with complete individual-scorer annotations (40-minute windows from lights-off), the wake fraction distribution reflects a mix of trials where participants remained awake throughout (wake fraction = 1.0) and trials where sleep was detected, often early in the session. Because MWT trials were terminated upon sleep detection (mean SOL $\approx$ 8 min in sleep-detected trials), wake fractions span a wide range. Both green (in-ear EEG) and purple (scalp EEG) data sources show similar distributions across all three scorers.
Right - Overnight sleep - For the 16 overnight sleep sessions (healthy controls), the right panel shows the Sleep Fraction (proportion of epochs scored as any sleep stage). All three scorers show high agreement, with mean Sleep Fraction of 0.90 for in-ear EEG sources (green) and 0.93 for scalp-EEG sources (purple), consistent with the trimmed recording windows capturing the main sleep period.

3.3 Maintenance of Wakefulness Test: Sleep Onset Latency

Of the 126 MWT trials recorded across both arms (64 healthy, 62 CDH), sleep was detected by both devices in 37 trials, by scalp EEG only in 22 trials, by in-ear EEG only in 7 trials, and by neither device in 60 trials. A breakdown of trial disposition is provided in Table 4.

3.3.1 Overall Agreement

Across all 126 MWT trials, mean SOL was 29.6 ± 15.1 min for in-ear EEG and 24.9 ± 16.8 min for scalp EEG. The mean absolute difference (MAD) was 7.6 minutes, with a mean bias of +4.7 min ($p = 0.0005$). The Pearson correlation was $r = 0.58$ and intraclass correlation coefficient was ICC(2,1) = 0.55 (95% limits of agreement: -24.3 to $+33.6$ min; Table 3, Figures 5 and 6). The

Table 3: MWT Sleep Onset Latency: Ear-EEG vs. Scalp EEG

Group	n	SOL (min)		MAD (min)	LoA (min)	r	ICC(2,1)
		Ear-EEG	Scalp				
All trials							
Overall	126	29.6 ± 15.1	24.9 ± 16.8	7.6	[-24.3, 33.6]	0.58	0.55
Healthy	64	29.3 ± 15.1	27.5 ± 16.0	5.3	[-21.7, 25.5]	0.70	0.70
CDH	62	29.9 ± 15.2	22.3 ± 17.4	10.1	[-25.3, 40.4]	0.48	0.43
Per-subject	24	29.4 ± 10.6	25.7 ± 12.1	5.1	[-12.3, 19.7]	0.75	0.71
Both devices detected sleep							
Overall	37	9.2 ± 8.0	8.1 ± 7.3	1.9	[-7.9, 10.1]	0.82	0.82
Healthy	19	9.5 ± 8.4	8.9 ± 8.3	1.2	[-4.5, 5.6]	0.95	0.95
CDH	18	9.0 ± 7.8	7.4 ± 6.3	2.7	[-10.3, 13.6]	0.64	0.62

Table 4: Trial disposition for MWT sleep onset latency analysis

Category	Healthy	CDH
Total MWT trials	64	62
Neither device detected sleep	34	26
Both devices detected sleep	19	18
Scalp only detected sleep	7	15
Ear-EEG only detected sleep	4	3

moderate overall agreement reflects a substantial proportion of discordant trials in which one device detected sleep and the other did not (see below). Among the 37 trials where both devices detected sleep, agreement was good (ICC = 0.82, $r = 0.82$, MAD = 1.9 min), indicating that when sleep is detected by both systems, the SOL values are closely concordant.

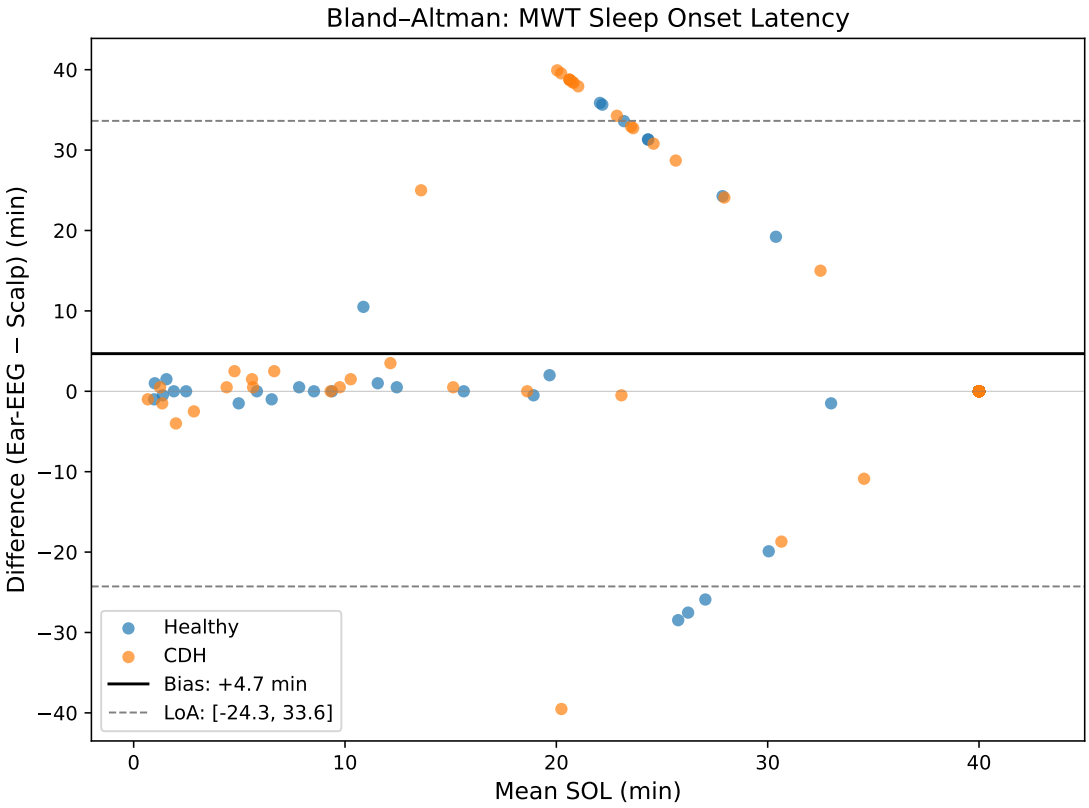

Figure 5: **Bland–Altman plot for MWT Sleep Onset Latency.** All 126 MWT trials. Blue: healthy controls. Orange: CDH. Black line: mean bias (+4.7 min). Dashed gray lines: ±1.96 SD limits of agreement [−24.3, +33.6 min].

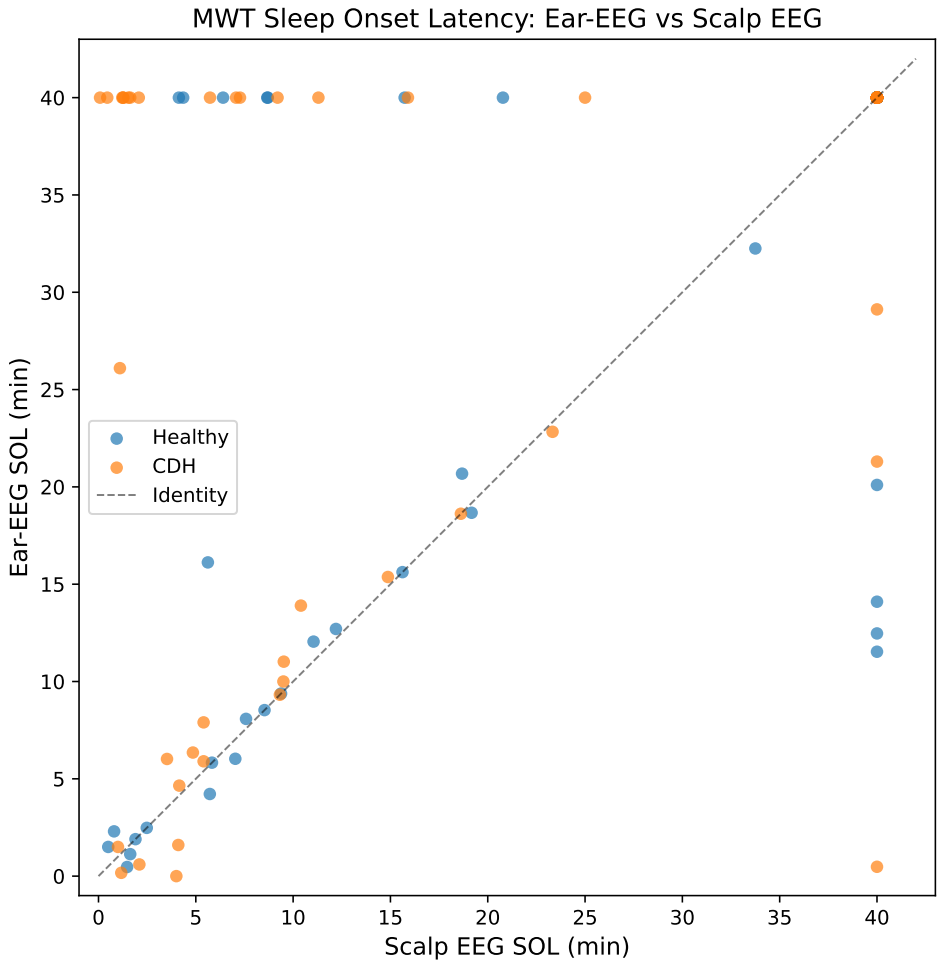

Figure 6: **Correlation of MWT Sleep Onset Latency** between in-ear EEG and scalp EEG. All 126 MWT trials ($r = 0.58$, ICC $= 0.55$). Dashed line: identity.

3.3.2 Healthy Controls

In the healthy control group ($n = 64$ trials from 16 subjects), mean SOL was 29.3 ± 15.1 min for in-ear EEG and 27.5 ± 16.0 min for scalp, with a non-significant bias of $+1.9$ min ($p = 0.22$), indicating no systematic difference between devices. Among the 19 healthy trials where both devices detected sleep, agreement was excellent (ICC $= 0.95$, MAD $= 1.2$ min). The overall ICC of 0.70 and wider limits of agreement (-21.7 to $+25.5$ min) reflect 7 discordant trials in which scalp EEG detected sleep but in-ear EEG did not.

3.3.3 CDH Cohort

In the CDH cohort ($n = 62$ trials from 8 subjects across 2 sessions each), mean SOL was 29.9 ± 15.2 min for in-ear EEG and 22.3 ± 17.4 min for scalp ($r = 0.48$, ICC $= 0.43$). A significant positive bias of $+7.6$ min ($p = 0.0007$) was observed, with a mean absolute difference of 10.1 minutes. The 95% limits of agreement were -25.3 to $+40.4$ min, wider than in healthy controls. This was driven primarily by 15 trials in which the scalp EEG detected sleep but the in-ear EEG did not (see

below). Among the 18 CDH trials where both devices detected sleep, agreement was moderate (ICC = 0.62, MAD = 2.7 min).

3.3.4 Per-Subject Mean SOL

In standard clinical practice, the MWT result for a patient is the average SOL across four trials. At the per-subject level ($n = 24$ subjects), averaging across all available trials, agreement was good ($r = 0.75$, ICC(2,1) = 0.71, MAD = 5.1 min), indicating that the in-ear EEG system captures between-subject differences in daytime sleepiness despite trial-level discordances.

3.3.5 Discordant Trials: Scalp-Only Sleep Detection

In 22 of 66 trials where at least one device detected sleep (7 healthy, 15 CDH), the consensus scorer identified sleep onset on scalp EEG but scored the entire in-ear EEG recording as wake. These discordant trials were concentrated in a subset of CDH participants: 4 subjects (204, 210, 211, 212) accounted for all 15 CDH discordant trials. Among the individual NXT scorers (AM, LY, TBR), one scorer (AM) detected sleep in some of these trials while the other two did not, suggesting that the ear-EEG signal contained ambiguous features at sleep onset that did not meet scoring criteria for most raters. The scalp-detected sleep in these trials was predominantly brief and subtle, consistent with the known difficulty of detecting transitional N1 sleep from ear-canal electrodes.

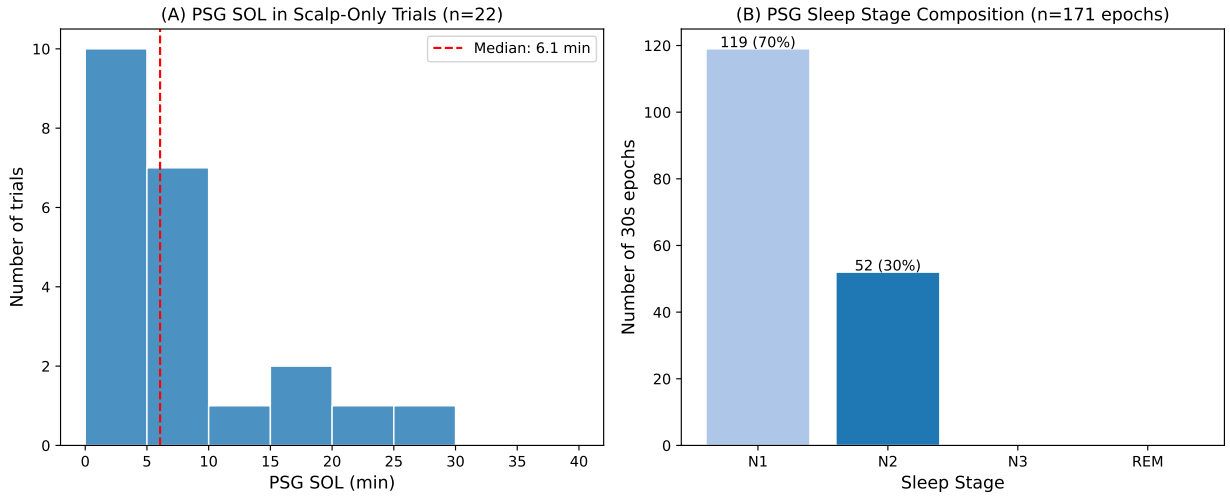

Figure 7: **Distribution and sleep-stage composition of scalp-only discordant MWT trials.** (A) Histogram of PSG sleep onset latency (SOL) for trials with PSG-only sleep detection ($n = 22$). (B) PSG sleep-stage composition across all discordant trials, showing that the missed sleep events were dominated by N1 epochs, consistent with the difficulty of detecting subtle transitional sleep using ear-canal electrodes. Counts represent 30-second PSG sleep epochs aggregated across all discordant trials.

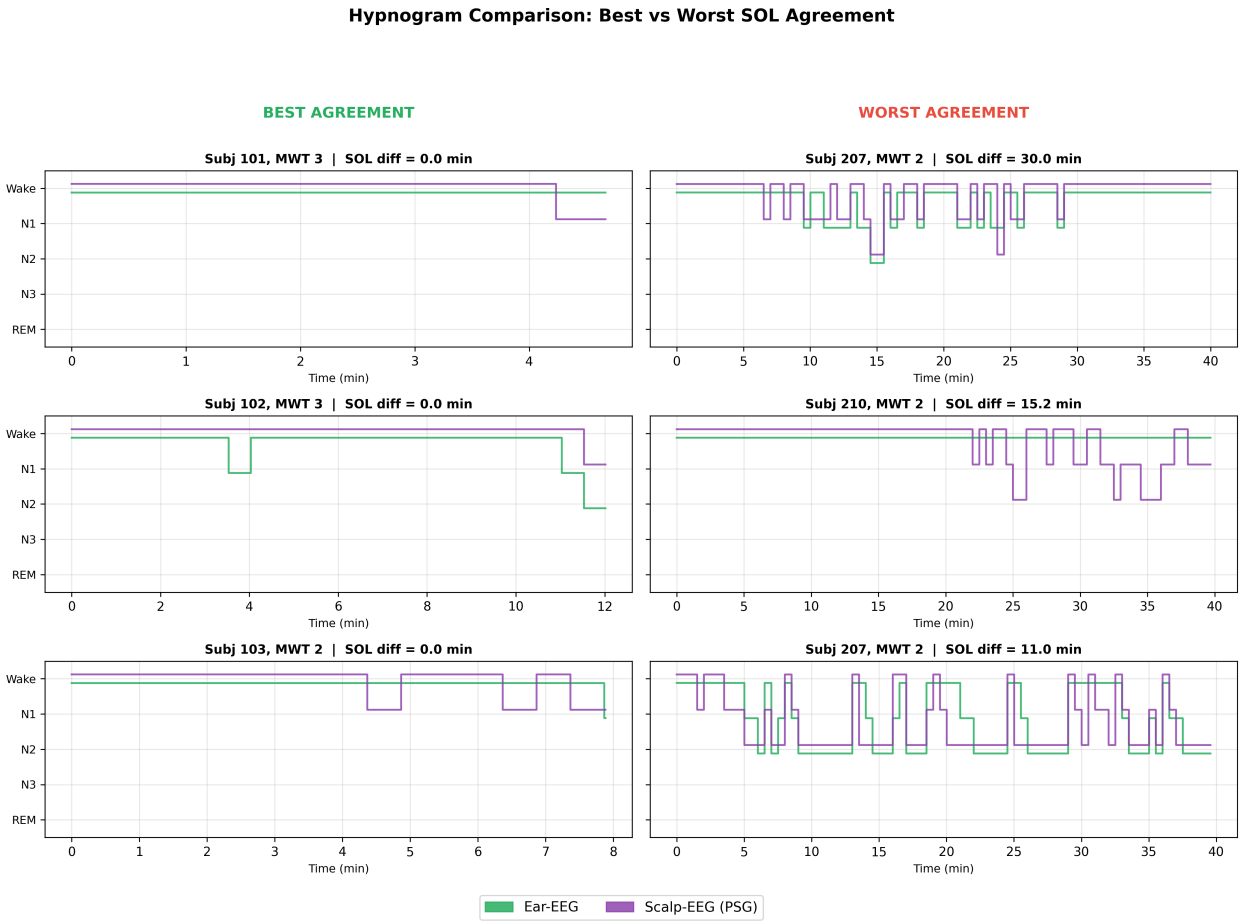

Figure 8: **Representative hypnogram comparisons for best and worst SOL agreement trials.** Left column: three trials with perfect agreement (SOL difference = 0.0 min). Right column: three trials with the largest SOL discrepancies (up to 30 min). Green: in-ear EEG (consensus scorer). Purple: scalp EEG (consensus scorer). In the worst-agreement cases, the scalp EEG shows clear transitions into N1/N2 sleep that are not detected by the in-ear EEG.

3.4 Overnight Sleep Architecture

Among the 16 healthy control subjects who completed overnight sleep-restricted PSG with concurrent in-ear EEG, sleep architecture metrics showed strong agreement for summary variables and moderate agreement for individual stage durations (Table 5).

Table 5: Overnight sleep architecture: Ear-EEG vs. Scalp EEG ($n = 16$ healthy controls)

Variable	Ear-EEG	Scalp	Bias	MAD	r	ICC(2,1)
SOL (min)	5.0 ± 4.1	4.9 ± 3.0	$+0.1$	1.6	0.76	0.74
TST (min)	218.0 ± 12.8	222.8 ± 10.3	-4.8^{**}	5.2	0.94	0.85
SE (%)	91.1 ± 5.3	93.1 ± 4.2	-2.0^{**}	2.2	0.94	0.85
WASO (min)	16.7 ± 10.8	12.0 ± 8.9	$+4.8^{**}$	5.2	0.93	0.82
N1 (min)	11.4 ± 4.3	8.8 ± 3.6	$+2.6^{*}$	3.8	0.46	0.39
N2 (min)	126.4 ± 22.0	117.2 ± 25.1	$+9.2$	21.7	0.42	0.40
N3 (min)	47.2 ± 23.8	58.7 ± 29.7	-11.5	19.6	0.64	0.59
REM (min)	33.0 ± 14.6	38.1 ± 16.2	-5.1	10.9	0.45	0.44

Values are mean ± SD. Bias = Ear-EEG − Scalp. MAD = mean absolute difference. $^{*}p < 0.05$; $^{**}p < 0.01$ (paired t-test). SOL = sleep onset latency; TST = total sleep time; SE = sleep efficiency; WASO = wake after sleep onset.

3.4.1 Summary Metrics

Sleep onset latency showed no systematic difference between devices (bias +0.1 min, $p = 0.89$, ICC = 0.74, MAD = 1.6 min). Total sleep time was highly correlated ($r = 0.94$, ICC = 0.85), with a small but significant bias of −4.8 min ($p = 0.001$), indicating that in-ear EEG scored slightly less total sleep. Sleep efficiency followed a similar pattern ($r = 0.94$, ICC = 0.85, bias −2.0%). Wake after sleep onset was also well-correlated ($r = 0.93$, ICC = 0.82), with in-ear EEG detecting approximately 5 minutes more wake ($p < 0.001$). The consistent direction of these biases—slightly less sleep, slightly more wake on in-ear EEG—is compatible with the reduced N1 detection sensitivity observed in epoch-level analyses (Section 3.1), whereby borderline sleep epochs are more often scored as wake on in-ear EEG.

3.4.2 Stage Durations

Individual sleep stage durations showed moderate agreement. N3 (slow-wave sleep) was the best-correlated stage ($r = 0.64$, ICC = 0.59), with in-ear EEG underestimating N3 by 11.5 minutes on average, though this did not reach significance ($p = 0.07$). N2 duration showed wide variability between devices (ICC = 0.40), likely reflecting epoch-by-epoch disagreements on N1/N2 boundaries. N1 duration showed a significant positive bias on in-ear EEG (+2.6 min, $p = 0.03$), consistent with ambiguous transitional epochs being scored as N1 rather than wake on in-ear EEG. REM duration showed moderate agreement (ICC = 0.44), with in-ear EEG underestimating by 5.1 minutes.

These stage-level discrepancies are consistent with the epoch-level confusion patterns reported in Section 3.1, where N1 was the most frequently misclassified stage. Importantly, the stage-level disagreements partially cancel at the summary level, yielding strong agreement for clinically relevant metrics (TST, SE, WASO).

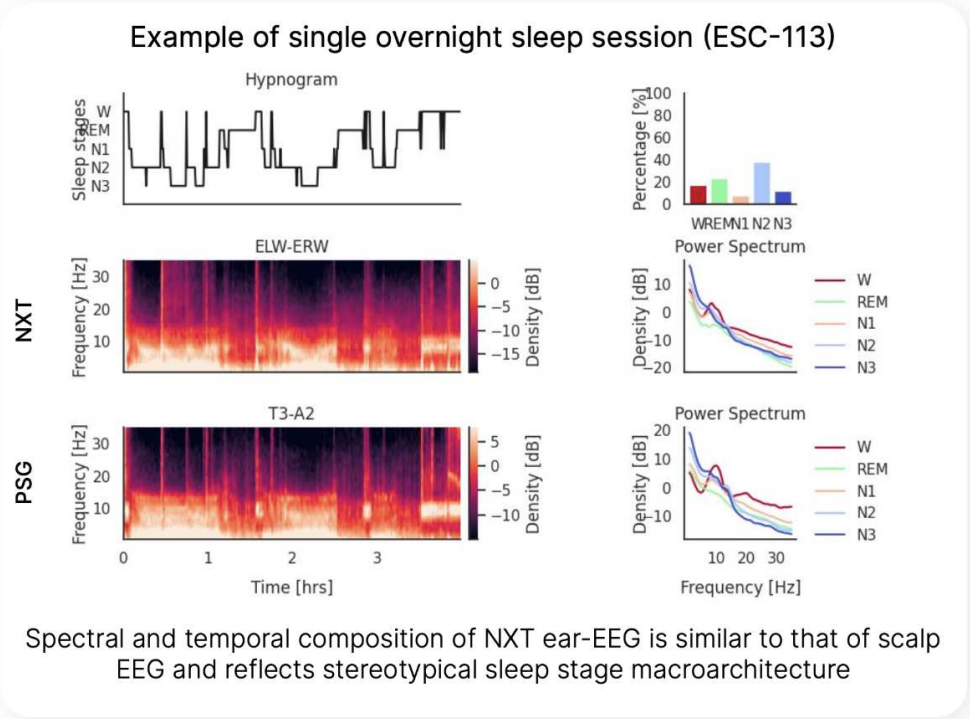

Figure 9: **NextSense in-ear EEG captures stereotypical sleep macroarchitecture.** Example overnight sleep session (ESC-113) showing hypnogram, spectrogram, and power spectral density from both NextSense in-ear EEG (channel ELW-ERW) and scalp PSG (channel T3-A2). The spectral and temporal composition of in-ear EEG is similar to that of scalp EEG and reflects stereotypical sleep stage macroarchitecture.

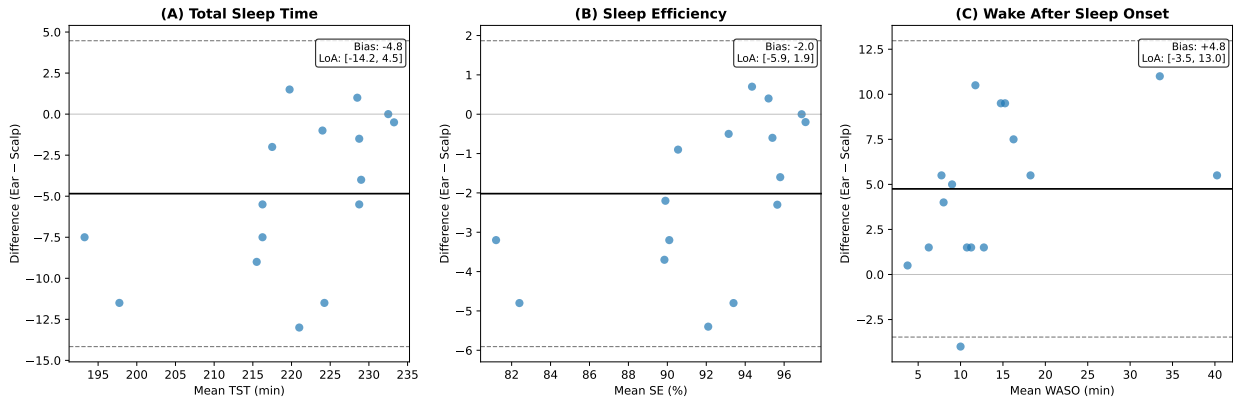

Figure 10: **Bland–Altman plots for overnight sleep architecture metrics (in-ear EEG vs. scalp EEG).** (A) Total sleep time (TST). (B) Sleep efficiency (SE). (C) Wake after sleep onset (WASO). Each point represents one of 16 healthy control subjects. Black line: mean bias. Dashed gray lines: ±1.96 SD limits of agreement.

3.5 Crosshead EEG Configuration Maximizes Dynamic Range and Sleep Feature Detection

To select an in-ear EEG derivation signal for manual and automated sleep staging, we compared an interaural (cross-head) bipolar channel to an ipsilateral (single-channel) bipolar channel. The crosshead configuration demonstrated superior sensitivity, a wider dynamic range, and substantially greater delta-band power. In preliminary overnight recording, both configurations exhibited the canonical $1/f$ power spectral profile and characteristic peaks across typical sleep frequency bands. However, the crosshead setup spanned a broader dynamic range, enhancing its capacity to discriminate between sleep stages. (Figure 11).

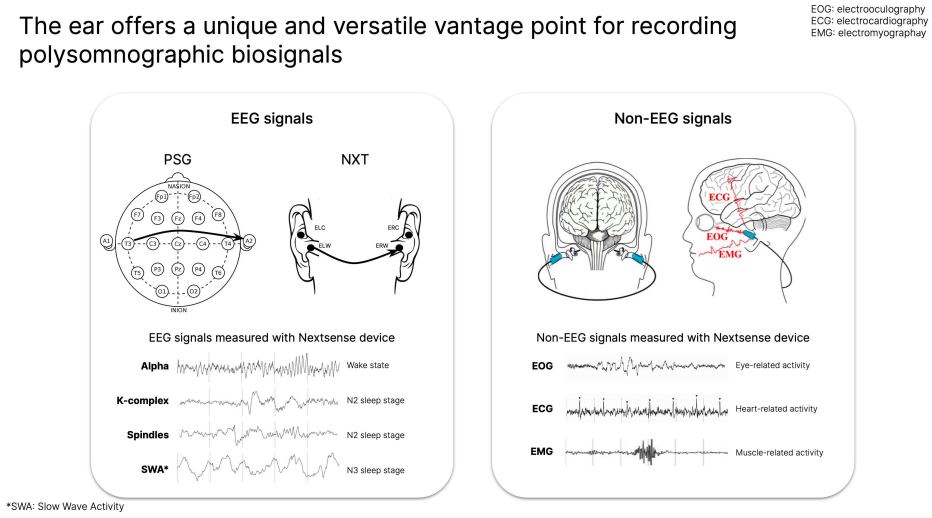

Figure 11: In-ear EEG signals from NextSense device

3.6 In-Ear EEG Signal Capture and Validation

Signal quality analyses were conducted prior to study launch and during the validation study, quantifying electrode impedance (IMP) alongside summary artifact/quality metrics including drift margin and line-noise contamination. A "settling effect" was observed wherein the conductive polymer interface improved over time. We identified a moderate relationship between electrode impedance and drift margin ($R^2 = 0.5725$), and wear-duration-dependent correlations between impedance and signal quality metrics (Figure 12).

4 Discussion

4.1 Advancements in Wearable EEG for Sleep Monitoring

Effective EEG-based sleep measurement requires maximizing signal quality. Key design considerations include: (1) inter-electrode distance for signal strength; (2) high-resolution digital acquisition; (3) inclusion of EOG and EMG features where possible; (4) electrode materials that maintain reliable skin contact; and (5) a form factor that minimizes sleep disturbance. The NextSense in-ear EEG system meets these requirements while enabling unobtrusive, longitudinal measurement.

The use of 4-channel in-ear EEG allowed high sleep scoring accuracy ($\kappa = 0.76$ overall, $\kappa = 0.80$ for overnight sleep), comparing relatively favorably to 20+ channel scalp EEG scorer-to-consensus

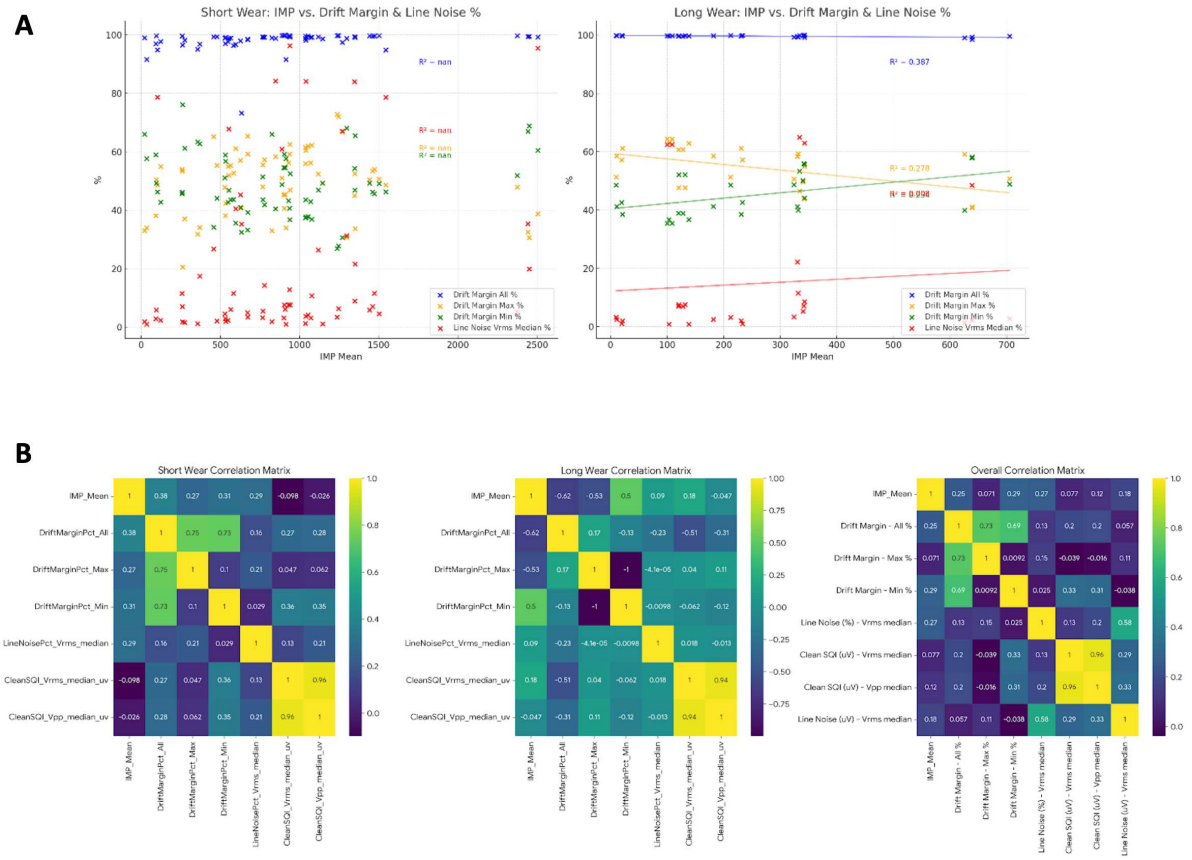

Figure 12: **Impact of electrode impedance on signal quality metrics.** (A) Scatter plots with linear regression fits illustrating the relationship between mean impedance (Ω) and signal quality metrics (Drift Margin %, Line Noise %) for Short Wear and Long Wear conditions. (B) Correlation matrices displaying Pearson coefficients (r) between impedance and signal quality indices.

agreement ($\kappa = 0.86$ overall, $\kappa = 0.92$ for overnight sleep). While inter-rater agreement was modestly lower for in-ear EEG compared to scalp EEG, it remained within acceptable clinical ranges, and the smaller form factor and capacity for unsupervised home recording may make in-ear EEG preferable for many clinical and research applications.

4.2 Comparison with Prior In-Ear EEG Validation Studies

These results compare favorably to prior in-ear EEG sleep staging validation work. Mikkelsen et al. (2019) reported $\kappa = 0.73$ for whole-night sleep staging using dry-contact in-ear EEG electrodes in a single-ear configuration (Mikkelsen, Villadsen, et al. 2019). The present study achieves $\kappa = 0.76$ overall and $\kappa = 0.80$ for overnight sleep sessions specifically, representing an improvement attributable to the cross-head (interaural) bipolar configuration, which provides greater dynamic range and delta-band power than single-ear derivations (Section 3.5). Sterr et al. (2018) demonstrated proof-of-concept behind-the-ear sleep staging with cEEGrid electrodes (Sterr et al. 2018), though with a smaller sample and fewer sleep stages evaluated. The present study is, to our knowledge, the largest in-ear EEG sleep staging validation to date in terms of total scored sessions ($n = 142$; 126 MWT trials + 16 overnight sleep sessions) across both healthy and clinical populations.

4.3 N1 Sleep Stage Agreement

Agreement for stage N1 was notably lower than for other stages (53%), a pattern consistent with the well-established difficulty of N1 scoring even among expert human scorers using full-scalp PSG. N1 is a transitional state with subtle EEG features (attenuation of alpha rhythm, emergence of low-amplitude mixed-frequency activity) that lacks the distinctive hallmarks of other stages such as K-complexes (N2), high-amplitude slow waves (N3), or rapid eye movements (REM). Inter-rater agreement for N1 is typically the lowest among all sleep stages in conventional PSG studies as well. The reduced signal amplitude available from ear-canal electrodes relative to scalp montages may further attenuate the already subtle N1 markers. Automated scoring algorithms trained specifically on ear-EEG data may help improve N1 detection in future work.

4.4 MWT Sleep Onset Latency Agreement

To our knowledge, this is the first validation of ear-EEG against scalp EEG for MWT sleep onset latency. The only prior ear-EEG sleep-latency validation used the Multiple Sleep Latency Test (MSLT) in healthy adults only: Alqurashi et al. 2018 reported $\kappa = 0.71$–0.79 with latency differences of <3 minutes in 84–94% of naps, but did not include clinical populations or report ICC. The present study extends this work to the MWT paradigm and includes both healthy controls and CDH participants.

The moderate overall trial-level agreement (Table 3) warrants careful interpretation. Nearly half of all trials resulted in neither device detecting sleep—an expected outcome of the MWT, in which participants actively resist sleep onset. When both devices detected sleep, SOL agreement was good to excellent, indicating that the in-ear EEG system reliably quantifies sleep onset timing once sleep is established. At the per-subject level, which mirrors standard clinical practice (averaging four trials per patient), agreement was good (ICC = 0.71), suggesting the device captures clinically meaningful between-subject differences in daytime sleepiness despite trial-level discordances.

It is also important to consider these results against the gold standard's own reproducibility. Tankere et al. 2024 showed that even two expert scorers using full-scalp EEG disagree on MWT

outcome in up to 32.6% of patients ($\kappa = 0.54$ for a 33-minute threshold, 0.27 for 19 minutes). The ear-versus-scalp agreement reported here falls within that range of inherent MWT scorer variability.

The discordant trials—predominantly scalp-detected sleep missed by the in-ear EEG—represent a genuine detection limitation rather than a protocol artifact. The scalp-detected sleep in these trials was predominantly brief and subtle, often consisting of scattered N1 epochs that did not reliably meet AASM criteria on the ear-EEG signal. Among the three in-ear EEG scorers, one detected sleep in some of these trials while the other two did not, confirming that the signal contained ambiguous features near the threshold of detectability. These discordant trials were concentrated in a subset of CDH participants (4 of 8 subjects accounted for all 15 CDH-discordant trials), suggesting that patient-specific factors such as electrode fit, signal quality, or the predominance of subtle N1-only sleep episodes may contribute. The in-ear EEG's reduced sensitivity to alpha rhythm attenuation—the primary marker of the wake-to-N1 transition—likely explains the lower detection rate for these events (see Section 4.3).

4.5 Overnight Sleep Architecture Agreement

The small, consistent biases observed in summary sleep architecture metrics—approximately 5 minutes less TST and 5 minutes more WASO on in-ear EEG—follow directly from the N1 detection limitation discussed above. Borderline sleep/wake transitions are more often scored as wake on in-ear EEG, reducing TST and inflating WASO, while leaving SOL essentially unaffected (bias +0.1 min). Crucially, these biases are predictable in direction and small in magnitude, making them amenable to correction in longitudinal applications.

Individual stage durations showed more variability, particularly at the N1/N2 boundary—a discrimination challenge even among expert scorers using full-scalp PSG. However, these stage-level disagreements partially cancel at the summary level, yielding clinically useful TST and SE estimates. For applications where night-to-night trends matter more than absolute values—medication titration, treatment monitoring, longitudinal research—the strong correlations ($r > 0.93$) may be more relevant than the modest absolute biases.

4.6 Clinical Implications

The ability to reliably score sleep stages from a comfortable, in-ear EEG device has significant implications for both clinical practice and research. For patients with central disorders of hypersomnolence—including narcolepsy and idiopathic hypersomnia—longitudinal monitoring of sleep architecture and daytime alertness currently requires repeated in-laboratory PSG, which is costly, disruptive, and limited to single-night snapshots. An in-ear EEG system capable of clinical-grade sleep staging in the home environment could enable multi-night assessment, capture night-to-night variability, and provide objective treatment monitoring over weeks or months. The crossover design in the CDH cohort (medication vs. medication-holiday) demonstrates that the in-ear EEG system can be deployed in a pharmacological intervention context, a prerequisite for its use in clinical trials and treatment optimization.

4.7 Constraints, Limitations, and Future Work

- **In-ear EEG is not exactly another scalp channel.** In-ear EEG signals are not natively identical to scalp EEG and are not directly interpretable without some minimal training. However, standardized scorer training and automated scoring approaches may mitigate these limitations.

- **High scalp inter-rater agreement.** In some large datasets, like the SHHS (Silva et al. 2009), the inter-rater agreement of $\kappa \geq 0.80$ is listed as very high agreement (Whitney et al. 1998). In the current study, we observe $\kappa = 0.76$ for ear-EEG versus scalp consensus and $\kappa = 0.86$ for scalp scorer-to-consensus agreement ($\kappa = 0.92$ for overnight sleep sessions specifically). Why might scalp self-agreement be so high? Part of this might be thought to be consensus (necessarily containing overlapping information). Another factor is the sleep session design, with built-in light-on time for mild sleep deprivation. With this, participants are kept awake for longer than they might in a typical SHHS sleep session. In some support of this, if we tightly clip the sleep sessions around the minutes of agreed sleep, agreement goes down, but not below 0.80.
- **Blinding concerns vs. sleep annotation process.** While we aimed to have sleep techs score the sleep session blindly (with respect to healthy/sleep disorder, order of session, drug treatment vs drug holiday, etc.), it was clear to our sleep techs which session was in-ear EEG vs scalp EEG, as we preserved the channel names in the sleep annotation software. This necessitates that we reveal channel identity, and so in-ear EEG vs scalp EEG file identity. Nonetheless, their task in each condition remains identical: to apply the most appropriate sleep class label given the 30s of biosignals before them.
- **Multimodal sleep detection.** The largest source of disagreement in the MWT analysis was 22 trials in which scalp EEG detected sleep but in-ear EEG did not, predominantly in the CDH cohort (15 of 22). The in-ear EEG device used in this study also contains an inertial measurement unit (IMU) that was not utilized for sleep staging. Future work could incorporate IMU-derived motion and head-position features into a multimodal sleep detection pipeline, potentially improving sensitivity to sleep onset in cases where the EEG signal alone is ambiguous.

Conflict of Interest

J.B., A.X.S., K.B., and A.P. are employees of NextSense Inc. B.T., D.V., and D.L.B. were employees of Takeda Development Center Americas, Inc. at the time of the study. P.S., C.D., and D.B.R. declare no competing interests.

Funding

This study was supported by NextSense Inc. and Takeda Development Center Americas, Inc.

Acknowledgements

The authors thank the expert sleep techs (LY, AM, TBR) for their patience and diligence. Thanks to Adrien Specht of NextSense and Stanford for data parsing help.

References

Achermann, P. et al. (2017). "Sleep homeostasis and models of sleep regulation." In: *Principles and Practice of Sleep Medicine*.

Alqurashi, Yousef D. et al. (2018). "A novel in-ear sensor to determine sleep latency during the Multiple Sleep Latency Test in healthy adults with and without sleep restriction". In: *Nature and Science of Sleep* 10, pp. 385–396. DOI: 10.2147/NSS.S171528.

Besedovsky, L., T. Lange, and M. Haack (2019). "The Sleep-Immune Crosstalk in Health and Disease." In: *Physiological Reviews* 99.3, pp. 1325–1380. DOI: 10.1152/physrev.00010.2018.

Bleichner, Martin G. and Stefan Debener (2017). "Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG". In: *Frontiers in Human Neuroscience* 11, p. 163. DOI: 10.3389/fnhum.2017.00163.

Buysse, D. J. (2014). "Sleep health: can we define it? Does it matter?" In: *Sleep* 37.1, pp. 9–17. DOI: 10.5665/sleep.3298.

Byun, Jong-Ho et al. (2019). "The first night effect during polysomnography, and patients' estimates of sleep quality". In: *Psychiatry Research* 274, pp. 27–29. DOI: 10.1016/j.psychres.2019.02.011.

Cirelli, C., C. M. Gutierrez, and G. Tononi (2004). "Extensive and Divergent Effects of Sleep and Wakefulness on Brain Gene Expression." In: *Neuron* 41, pp. 35–43. DOI: 10.1016/S0896-6273(03)00814-6.

Diekelmann, S. and J. Born (2010). "The memory function of sleep." In: *Nature Reviews Neuroscience* 11, pp. 114–126. DOI: 10.1038/nrn2762.

Flores, Natalia M. et al. (2016). "The Humanistic and Economic Burden of Narcolepsy". In: *Journal of Clinical Sleep Medicine* 12.3, pp. 401–407. DOI: 10.5664/jcsm.5594.

Goldstein, A. N. and M. P. Walker (2014). "The Role of Sleep in Emotional Brain Function." In: *Annual Review of Clinical Psychology* 10.1, pp. 679–708. DOI: 10.1146/annurev-clinpsy-032813-153716.

Goverdovsky, Valentin et al. (2016). "In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring". In: *IEEE Sensors Journal* 16.1, pp. 271–277. DOI: 10.1109/JSEN.2015.2469097.

Joyner, McGregor et al. (2024). "Using a standalone ear-EEG device for focal-onset seizure detection". In: *Bioelectronic Medicine* 10.1. Open Access, p. 4. DOI: 10.1186/s42234-023-00135-0. URL: https://doi.org/10.1186/s42234-023-00135-0.

Kappel, Søren L. et al. (2019). "Dry-contact electrode ear-EEG". In: *IEEE Transactions on Biomedical Engineering* 66.1, pp. 150–158. DOI: 10.1109/TBME.2018.2850218.

Kidmose, Preben et al. (2013). "A study of evoked potentials from ear-EEG". In: *IEEE Transactions on Biomedical Engineering* 60.10, pp. 2824–2830. DOI: 10.1109/TBME.2013.2267696.

Leong, R. L. and M. W. Chee (2023). "Understanding the need for sleep to improve cognition." In: *Annual Review of Psychology* 74.1, pp. 27–57. DOI: 10.1146/annurev-psych-032620-034127.

Looney, Damien, Preben Kidmose, and Danilo P. Mandic (2011). "Ear-EEG: User-centered and wearable brain monitoring". In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6882–6885. DOI: 10.1109/IEMBS.2011.6091709.

— (2012). "The in-the-ear recording concept: user-centered and wearable brain monitoring". In: *IEEE Pulse* 3.6, pp. 32–42. DOI: 10.1109/MPUL.2012.2201656.

Malhotra, Atul et al. (2013). "Performance of an automated polysomnography scoring system versus computer-assisted manual scoring". In: *Sleep* 36.4, pp. 573–582. DOI: 10.5665/sleep.2548.

Mander, B. A., J. R. Winer, and M. P. Walker (2017). "Sleep and Human Aging." In: *Neuron* 94, pp. 19–36. DOI: 10.1016/j.neuron.2017.02.004.

Mikkelsen, Kaare B., Søren L. Kappel, et al. (2015). "EEG recorded from the ear: characterizing the ear-EEG method". In: *Frontiers in Neuroscience* 9, p. 438. DOI: 10.3389/fnins.2015.00438.

Mikkelsen, Kaare B., Daniel B. Villadsen, et al. (2019). "Accurate whole-night sleep monitoring with dry-contact ear-EEG". In: *Scientific Reports* 9, p. 16824. DOI: 10.1038/s41598-019-53355-6.

Newell, Johan et al. (2012). "Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples". In: *Psychiatry Research* 200.2-3, pp. 795–801. DOI: 10.1016/j.psychres.2012.07.045.

Paruthi, S. et al. (2016). "Recommended Amount of Sleep for Pediatric Populations: A Consensus Statement of the American Academy of Sleep Medicine." In: *Journal of Clinical Sleep Medicine* 12.6, pp. 785–786. DOI: 10.5664/jcsm.5866.

Punjabi, Naresh M. et al. (2020). "Variability and Misclassification of Sleep Apnea Severity Based on Multi-Night Testing". In: *CHEST* 158.1, pp. 365–373. DOI: 10.1016/j.chest.2020.01.039.

Saper, C. B., T. E. Scammell, and J. Lu (2005). "Hypothalamic regulation of sleep and circadian rhythms." In: *Nature* 437, pp. 1257–1263. DOI: https://doi.org/10.1038/nature04284.

Silva, Graciela E. et al. (Aug. 2009). "Longitudinal Evaluation of Sleep-Disordered Breathing and Sleep Symptoms with Change in Quality of Life: The Sleep Heart Health Study (SHHS)". In: *Sleep* 32.8, pp. 1049–1057. ISSN: 0161-8105. DOI: 10.1093/sleep/32.8.1049. eprint: https://academic.oup.com/sleep/article-pdf/32/8/1049/13664948/sleep-32-8-1049.pdf. URL: https://doi.org/10.1093/sleep/32.8.1049.

Sterr, Alexander et al. (2018). "Sleep EEG derived from behind-the-ear electrodes (cEEGrid) compared to standard polysomnography: a proof of concept study". In: *Frontiers in Human Neuroscience* 12, p. 452. DOI: 10.3389/fnhum.2018.00452.

Stickgold, R. et al. (2001). "Sleep, learning, and dreams: off-line memory reprocessing." In: *Science* 294, pp. 1052–1057. DOI: 10.1126/science.1063530.

Tankere, Paul et al. (2024). "Revisiting the maintenance of wakefulness test: from intra-/inter-scorer agreement to normative values in patients treated for obstructive sleep apnea". In: *Journal of Sleep Research*. DOI: 10.1111/jsr.13961.

Tobler, I. (2011). "in *Principles and practice of sleep medicine*". In: *Principles and Practice of Sleep Medicine*. Elsevier, pp. 112–125.

Vaughn, B. V. and P. Giallanza (2008). "Technical review of polysomnography." In: *Chest* 134, pp. 1310–1319. DOI: https://doi.org/10.1378/chest.08-0812.

Waters, Samuel et al. (2024). "Domain Adaptation Using Large Scale Databases for Sleep Staging from a Novel In-Ear Sensor". In: *ResearchGate preprint*. DOI: 10.13140/RG.2.2.25639.76648.

Whitney, Coralyn W. et al. (Oct. 1998). "Reliability of Scoring Respiratory Disturbance Indices and Sleep Staging". In: *Sleep* 21.7, pp. 749–757. ISSN: 0161-8105. DOI: 10.1093/sleep/21.7.749. eprint: https://academic.oup.com/sleep/article-pdf/21/7/749/13661121/sleep-21-7-749.pdf. URL: https://doi.org/10.1093/sleep/21.7.749.

Xie, L. et al. (2013). "Sleep Drives Metabolite Clearance from the Adult Brain." In: *Science* 342.6156, pp. 373–377. DOI: 10.1126/science.1241224.

A Appendix

A.1 EEG and PSG Data Channel Details



Figure 13: NextSense in-ear EEG device

A.2 Scorer Certification Procedure

Prior to scoring study data, each sleep technologist completed a standardized certification procedure for in-ear EEG interpretation. A canonical set of previously scored in-ear EEG recordings was curated by an in-house sleep medicine expert (J. Dwyer, MD, PhD). Technologists scored these recordings in blinded, randomized fashion without access to the reference labels. To pass certification, each scorer was required to achieve $\geq$80% epoch-level agreement with the canonical labels. The certification procedure was repeated at the study midpoint as a quality-control measure to ensure scoring consistency throughout the study.

A.3 Design Changes

A.3.1 Ellcie Healthy Glasses

The original feasibility protocol included a commercially available eyeglass-based wearable (Ellcie Healthy Glasses) designed to capture eye movement and other physiologic signals during daytime testing. Due to discontinued manufacturer support and device inoperability during the study period, these data were not included in the analyses and are not reported in this manuscript.

Table 6: Polysomnography (PSG) data channels recorded using NextSense In-Ear EEG and Philips Respironics ALICE 6

Category	**Channels**	**Units**	**Notes**
EEG (Scalp) n=19	Fp1–A2, Fp2–A1; F3–A2, F4–A1; F7–A2, F8–A1; C3–A2, C4–A1, Cz–A2; P3–A2, P4–A1, Pz–A1; T3–A2, T4–A1, T5–A2; T6–A1, O1–A2, O2–A1	μV	Scalp EEG was recorded using the 10-20 system referenced to contralateral mastiod
EEG (Ear) n=5	ERW–ELC, ELW–ELC, ELW–ERW, ELW–ERC, ERW–ERC	μV	In-ear EEG derivations using ear-concha (EW) and ear-canal (EC)
EOG n=2	LOC–A2, ROC–A2	μV	Electrooculography for eye movement detection
EMG n=5	Chin, Leg 1, Leg 2; Aux1, Aux2	μV	Submental chin and bilateral leg
ECG n=2	ECG I, ECG II	μV	Cardiac activity
Respiration n=5	Flow (Thermistor), Flow (Pressure), BreathRate Aux1; Effort THO, Effort ABD	a.u. BrPM	Airflow breaths per min resp effort
Snore n=2	Mic, Pressure	a.u.	Snoring detected via microphone and pressure transducer
Oximetry n=3	SpO_2, PulseRate, Pleth	%, bpm, a.u.	Peripheral oxygen saturation and pulse-derived signals
PAP Therapy n=3	xPAP CPAP, IPAP, EPAP	cmH_2O	Delivered positive airway pressures
Motion / Position n=3	ACC_X, ACC_Y, ACC_Z	a.u.	Three-axis accelerometry
Annotations *n*=4	EDF Annotations: System annotations 3 AASM Sleep Techs	– –	Events, system events three independent qualified sleep techs using AASM scoring guidelines