SLEEP, 2026, **49**, 1–11
doi:10.1093/sleep/zsaf353
Advance Access publication 11 November 2025
Original Article

Original Article

The Insomnia EEG Score: a new tool for the classification of people with poor sleep

Jay Pathmanathan[1,*], David Little[1], Kolia Sadeghi[1], Tobias Di Marco[2], David Kleinschmidt[1], Jayne Nerrie[1], Nurkurniati Tjiptarto[1], Alex Arslan[1], Jeffrey Hubbard[2], Antonio Olivieri[2], Corey B. Puryear[1], M. Brandon Westover[1] and Jacob Donoghue[1]

[1]Beacon Biosignals Inc., Boston, MA 02210, United States and [2]Idorsia Pharmaceuticals Ltd., Hegenheimermattweg 91, 4123 Allschwil, Switzerland
*Corresponding author. Jay Pathmanathan, 22 Boston Wharf Rd., Suite 41, Boston, MA 02210, United States. Email: jay.pathmanathan@beacon.bio.

Abstract

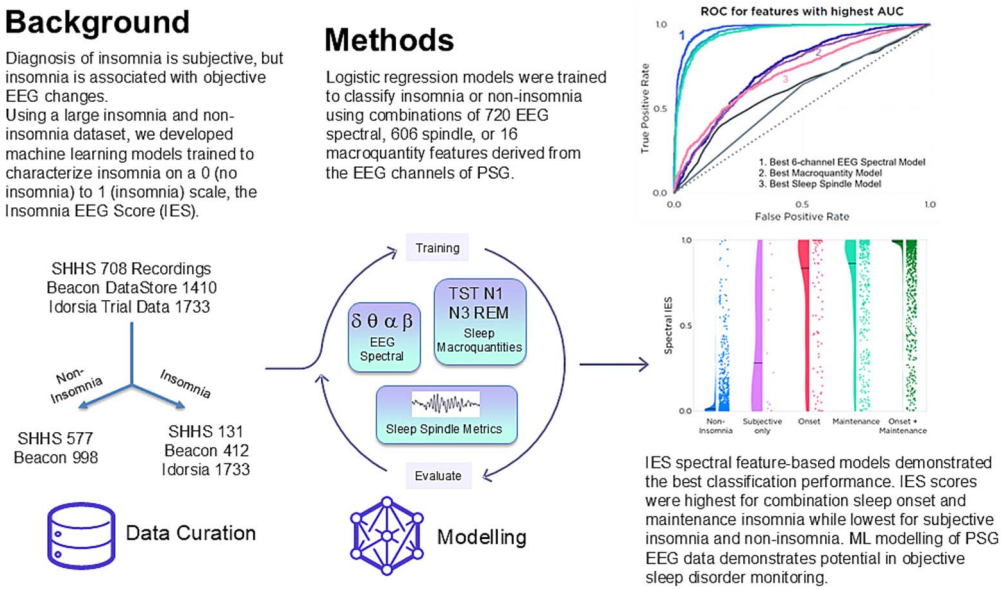
Quantitative features could help objectively identify and grade insomnia severity, though there is currently no pathophysiological biomarker of insomnia. In this study, we used the largest cohort of individuals with and without insomnia to date to train a model capable of distinguishing people with insomnia from those without insomnia. We identified 720 spectral, 606 spindle, and 16 macro-architecture features from electroencephalogram (EEG) channels of the polysomnography and examined their ability to classify people with insomnia and insomnia subtypes (subjective, maintenance, onset, and combined maintenance and onset), compared to individuals without insomnia. Consistent with prior work, these features poorly classified individuals when assessed independently. However, a linear combination of these features (the "Insomnia EEG Score" [IES]) was able to distinguish individuals with ($N = 2123$) and without ($N = 930$) insomnia, with spectral features providing the greatest discrimination (outperforming models based on hypnogram macroquantities or spindle features). These results demonstrate that insomnia is likely a quantifiable sleep disorder, albeit one requiring multiple objective measures (such as IES) to gauge disease subtype and, potentially, response to therapeutic interventions.

Clinical Trials

Study to Assess the Efficacy and Safety of ACT-541468 (Daridorexant) in Adult and Elderly Subjects with Insomnia Disorder, https://clinicaltrials.gov/study/NCT03545191, NCT03545191; Study to Assess the Efficacy and Safety of ACT-541468 (Daridorexant) in Adult and Elderly Subjects Suffering from Difficulties to Sleep, https://www.clinicaltrials.gov/study/NCT03575104, NCT03575104.

Key words: insomnia; electroencephalography; polysomnography; machine learning; biomarker

Graphical Abstract



Submitted: 13 February, 2025; **Revised:** 21 September, 2025; **Accepted:** 07 October, 2025

Statement of Significance

Although quantitative measurements of sleep are not standard and generally not needed for the diagnosis or management of insomnia, there are cases where objectively quantifying aspects of this disease could be of clinical or research value. Here, we demonstrate the ability of a machine learning model utilizing sleep EEG signals to distinguish individuals with insomnia from individuals without a diagnosis of insomnia and provide a quantitative insomnia score (the "Insomnia EEG Score"). The model was trained on the largest-to-date insomnia dataset, incorporating data from many sites and thousands of individuals across a diverse age spectrum. This model provides the basis for an objective characterization of different insomnia subtypes, which could prove valuable in evaluating therapeutic responses longitudinally.

Introduction

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) defines insomnia disorder as a sleep–wake disorder involving inadequate sleep associated with impairments in daytime function occurring at least three nights per week for at least 3 months, indicating the condition's persistence and chronicity [1]. DSM-5 diagnostic criteria [2] cluster insomnia symptoms into three types: difficulty falling asleep, difficulty maintaining sleep, and early morning awakening. Both the diagnosis and subtyping of insomnia disorder are based mainly on subjective patient reports, requiring no objective sleep measurement for diagnosis or management. However, it is well known that an individual's perception of sleep duration and quality can conflict with objective sleep measurements such as total sleep time (TST), latency to persistent sleep (LPS), and wake after sleep onset (WASO), as measured with polysomnography (PSG) [3–5]. For this reason, it is not uncommon to use both subjective and objective sleep metrics to characterize disease during the development of novel therapeutics for sleep disorders [6]. Although PSG is not a mandatory requirement for the diagnosis or monitoring of insomnia disorder, its use can be particularly beneficial in specific cases. According to the updated guidelines for the treatment of insomnia from the European Sleep Research Society published in November 2023 [7], PSGs should be considered in order to exclude other sleep disorders (such as sleep apnea, restless leg syndrome, or others), for patients with insomnia that don't respond to treatment, in patients with insomnia that are at risk for tiredness/fatigue-related accidents, and in cases where there is a significant discrepancy between subjective and objective sleep findings (sleep misperception). Objective data therefore still constitute an important tool for the management of insomnia disorder.

Prior studies, focusing on the hypnogram (sleep and wake staging) [4, 8–12] and EEG power spectral analyses (examining the power content of traditional EEG bands) [13–18], have provided varied and sometimes conflicting findings. Single features derived from hypnograms often show poor sensitivity and specificity in classifying insomnia [19], as patients with this pathology typically exhibit only slight differences in objective sleep macrofeatures, such as time in specific sleep stages, TST, LPS, and WASO when compared to healthy sleepers [4, 12]. Failure of the PSG-based hypnogram to better characterize insomnia could be due, at least in part, to the heterogeneity of specific insomnia subtypes with differing underlying neuropathology [20, 21]. As such, it also highlights that in addition to sleep quantity, qualitative aspects play an important role, as highlighted in individuals with sleep misperception [5, 11]. Power spectral analyses, which often complement analyses of sleep macrofeatures, have identified more consistent findings than hypnogram-based sleep studies. Most commonly, spectral analyses demonstrate reduced delta power in wake and N1 and increased power in faster frequency bands (particularly in the alpha and beta bands) [22, 23]. While a recent large study [14] suggested that the elevation in beta power could be due to inclusion of individuals taking sedative-hypnotic benzodiazepines, the shift of background frequencies to relatively faster rhythms could suggest that individuals with insomnia (or at least a large subset thereof) suffer from a state of neurophysiological "hyperarousal" [16, 17, 22, 24], causing difficulties in initiating and/or maintaining sleep throughout the night.

While these studies suggest that EEG-based spectral features during sleep could be indicative of insomnia, there is no currently accepted single EEG-derived measure. In fact, it is unlikely that a single EEG feature could provide such a measure, given the heterogeneity of the disorder and the low magnitude of EEG feature differences in individuals with and without insomnia. Assuming that brain activity in sleep is relevant to insomnia, it is more likely that a combination of measurable features could characterize disease. If so, large datasets coupled with machine learning (ML) analytics incorporating many EEG-based variables could provide substantially better insight [21]. Such ML models could advance the characterization of insomnia subtypes and evaluation of the effects of therapeutic interventions.

The present study leveraged the largest cohort assembled to date, with >2200 individuals with insomnia and >1400 individuals without insomnia, to develop a multivariate EEG-based neurophysiological biomarker of insomnia. The intent of this metric (the "Insomnia EEG Score," IES) was to leverage a data-driven ML approach to integrate a large number of EEG-derived features from many individuals to distinguish individuals with insomnia from individuals without insomnia.

Materials and Methods

Study cohort and subgroup selection

Data analyzed in this study came from four sources: two clinical trials (Idorsia Phase 3, studies 301 and 302, clinicaltrials.gov: NCT03545191 and NCT03575104) that included 1839 pre-dose screening overnight PSG recordings from 1839 individuals who went on to participate in the first *single-blind placebo* PSG; the Sleep Heart Health Study (SHHS) [25, 26], which included 6495 recordings from 4778 individuals; and the Beacon Clinico-PSG Database (Beacon), which included 7671 PSGs from 6441 individuals from a single-site US academic medical center. PSG recording characteristics differed across the sites, reflecting real-world data. SHHS used only two central channel EEG leads, while recordings from the Beacon dataset and the Idorsia trials utilized six EEG electrodes (F3/F4, C3/C4, and O1/O2), as described in Di Marco et al. [23] PSG duration was for the duration of natural sleep and was at least 6 h for Beacon dataset PSGs (recorded in clinical settings), at least 8 h for Idorsia clinical trial PSGs (per study protocol), and at least 4 h of recorded data for SHHS (per the study protocol, https://jhuccs1.us/shhs/details/design.htm).

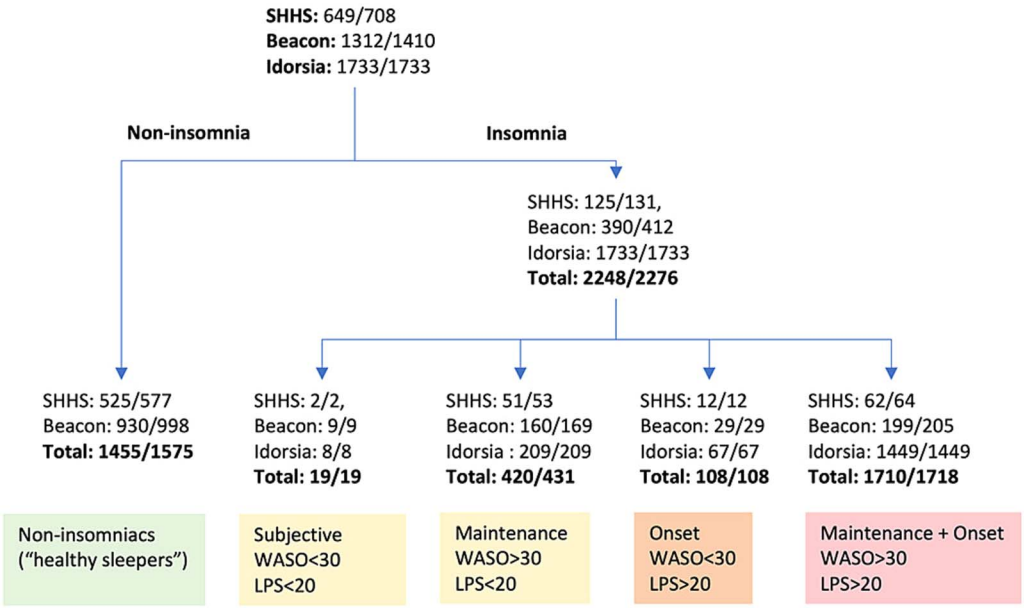


Figure 1. Diagram with number of individuals and recordings in each group and subgroup, broken down by dataset (number of individuals in the numerator; number of available recordings in the denominator). Individuals with insomnia were required to have sleep efficiency <0.85, a subjective complaint of insomnia, to not be taking medications affecting insomnia, and no lifestyle circumstances that poor sleep could be attributed to. Individuals without insomnia were required to have sleep efficiency ≥0.85, no subjective complaint of insomnia, and to not be taking certain medications. Individuals were excluded from all groups if they had certain medical conditions or if they had an Apnea Hypopnea Index >15.

The counts of individuals and PSGs are grand totals before exclusion/inclusion criteria were applied. All individuals were at least 18 years old. Exclusion criteria were chosen to reduce confounds from conditions other than insomnia. Specifically, we excluded individuals who had significant psychiatric, neurological, medical, or sleep conditions (e.g. sleep apnea, restless legs syndrome) other than insomnia (see Figure 1 and Supplementary Table S2), or who were taking medications known to impact sleep (see Supplementary Table S3). Included individuals were categorized as either not having insomnia or having insomnia. Note that some participants contributed more than 1 PSG to the cohort, as noted in Figure 1. These accounted for a small percentage of the total count and were mainly in the control cohort. To ensure individuals with repeat data did not bias the data, the best performing model was retrained on only the first PSG (discarding subsequent PSGs from any individual) and results were compared to the all-inclusive model. Individuals with insomnia were divided into four subgroups: subjective only, sleep onset insomnia, sleep maintenance insomnia, and both sleep onset and maintenance insomnia. Criteria used for these groupings are described below.

Individuals without insomnia

Individuals without insomnia were used as the comparison group and all individuals included came from Beacon and SHHS datasets. To be included in this group, individuals were required to not meet any of the exclusion criteria mentioned above and to have no subjective complaints of insomnia (see Supplemental Materials). However, absence of insomnia symptoms were not always documented, and an additional criteria was imposed that they have a sleep efficiency of ≥0.85, which is based on sleep efficiency values typical of healthy individuals in the age range of our data [27]. These criteria were designed to establish this group as a set of approximately "healthy" sleepers. Note that some of the data for this group originated from clinical sleep laboratories performing exams on referred patients, so we refer to this group as "non-insomnia" rather than healthy.

Individuals with insomnia

Individuals with insomnia were selected from Idorsia, Beacon, and SHHS sources (see Di Marco et al. 2024 [23] for full details). All individuals were required to have subjective sleep complaints or a formal insomnia diagnosis, not be taking medications for any sleep disturbance, and not report behaviors or other lifestyle factors that would confound an insomnia diagnosis (see Supplemental Materials for more details). Idorsia data were limited to the 1839 pre-dose screening PSGs of the individuals who went on to participate in the first *single-blind placebo* PSG, of which 1733 were associated with usable PSGs and requisite metadata. Across all three datasets, included individuals were required to have a recorded sleep efficiency of <0.85. All individuals in the insomnia group were then categorized into 4 insomnia subgroups based on WASO (the total number of minutes in bed spent awake after the first sleep epoch of persistent sleep), and LPS (the number of minutes from the beginning of time in bed until the first 10 minutes of uninterrupted sleep). The four subgroups were defined as follows:

1) Subjective insomnia—Subjective insomnia complaints; WASO <30 min, LPS <20 min.
2) Onset insomnia—Subjective insomnia complaints, WASO <30 min, LPS >20 min.
3) Maintenance insomnia—Subjective insomnia complaints, WASO >30 min, LPS < 20 min.
4) Onset + Maintenance insomnia—Subjective insomnia complaints, WASO >30 min, LPS > 20 min.

Note that because 59 (9 per cent) of SHHS and 93 (7 per cent) of Beacon individuals had more than one PSG recording (at most three PSGs, typically recorded years apart) each with different associated metadata, a small number of individuals had one PSG

in the "without insomnia" group and one in the "with insomnia" group, and a small number of individuals had PSGs with different insomnia subtypes. In the modeling and evaluations described below, each PSG was treated as one data point, ignoring diagnoses from other studies or the identity of the individuals who were being recorded. Note that subject-dependent PSGs (those coming from 1 participant) accounted for 1 per cent of the insomnia recordings and 8 per cent of the control (non-insomnia) recordings in total (see breakdown in Figure 1).

Artifact reduction and signal normalization

EEG signals were preprocessed as follows and as described in detail in our prior work [23]. All 3-s segments of EEG signals where at least half of the EEG channels had a root-mean square amplitude $\leq 1\ \mu V$ or $\geq 250\ \mu V$ were considered artifactual and removed. These thresholds were chosen to be well outside of the physiologic range of ~10 to 100 μV [28]. To reduce the influence of site-related differences, we utilize ratios of band powers instead of absolute band power values when analyzing spectral features of the EEG; see the next section for details. All PSG recordings (Beacon, SHHS, and Idorsia recordings) were re-sampled to 200 Hz and high-pass filtered (cutoff at 0.5 Hz). This high-pass filtering brought the distributions of relative powers across all sites within range of each other. Finally, we note that SHHS recordings included only the two central EEG leads (C3, C4), whereas recordings from other sites included all standard PSG leads (F3, F4, C3, C4, O1, O2). Therefore, only C3 and C4 are included in any analyses that included SHHS as a dataset, to enable fair comparisons.

Univariate analyses

For each PSG recording, we calculated a series of features to summarize the information captured across the entire night. We divided features into three broad types: (1) *spectral*: aggregate statistics of relative spectral powers in the delta, theta, alpha, beta bands; (2) *spindle*: characteristics derived from sleep spindles that were detected by an automated algorithm; (3) *macro-architecture*: quantities calculated from scored sleep stages over the course of the night (such as percentages of sleep time spent in a given sleep stage). Comparisons of each of these individual features between the onset + maintenance insomnia subgroup and the non-insomnia group are reported in Di Marco et al. 2024 [23].

Spectral features

Spectral features were derived using the power in four frequency bands (delta: 0.5–4 Hz, theta: 4–8 Hz, alpha: 8–12 Hz, beta: 12–30 Hz). Band powers were estimated using a multi-taper spectrogram using 2-s windows (1-s overlap) within the recording [29–31]. Consecutive time slices overlapped by 1 s, i.e. there was a step of 1 s between the beginning of each time slice and the next. The frequencies resolved by the spectrogram falling into the same band were summed together, and across channel lateralization, resulting in powers for each band and each of the three longitudinal electrode positions (frontal, central, and occipital, except in analyses including SHHS data which only have the central position available). Different ratios were then computed from these powers for each 2-s time slice: relative powers (delta/*T*, theta/*T*, alpha/*T*, beta/*T*, where *T* = delta + theta + alpha + beta is the total power) and power ratios (delta/alpha, theta/alpha, alpha/beta, theta/beta, delta/beta, delta/theta). The final spectral features analyzed at the sleep stage and recording levels were the result of two further aggregations: (1) For each of the 4 relative band powers and 6 ratios in the 2-s windows, features were aggregated to 30-s epochs by computing their mean, standard deviation, minimum, and 95 per cent quantile values. The resulting 101 quantities over 30-s segments are similar to the features used for automated sleep staging in prior work [32]. These statistics over 30-s windows were then averaged over all the epochs of a recording in each sleep/wake stage (wake, N1, N2, N3, and REM) and over the entire lights-off period (which is equivalent to averaging over all five stages combined). Spectral features were finally log-transformed to reduce skew/outliers. In all, this resulted in 720 engineered spectral features (see Supplemental Materials for a detailed breakdown of these features).

Spindle features

Spindle features were calculated from sleep spindle events detected using the open-source Luna package (http://zzz.bwh.harvard.edu/luna/). In order to maintain consistency with the use of the Luna spindle analysis, filtering and artifact removal parameters were used as described by the Luna package. First, each recording was re-sampled to 128 Hz and a band-pass filter was applied restricting to 0.3 and 35 Hz. A Kaiser window was applied with a ripple of 0.01 and transition width of 0.5 Hz. Artifacts were removed by computing power within the delta band, rejecting epochs that had more than 2.5 the average delta power in a 15-epoch sliding window [34].

After artifact removal, spindles were detected within all N2 stages by first convolving a Morlet wavelet (13.5 Hz) over the signal and then smoothing the convolution's magnitude using a sliding window of 0.1 s. Spindles were detected from this convolution by thresholding: at least 0.3 s had to be over 4.5 the mean of all N2 epochs, and in a 0.5 s window around this region, power had to be at least twice this N2 epoch mean. These putative spindles were merged if they fell within 0.5 s of one another, and any that lasted longer than 3 s were rejected. This basic approach to spindle validation, via Morlet wavelets, has been validated against manual spindle annotation [35].

To derive features from the coupling between spindles and slow oscillations (SOs) during N2 sleep, SOs were detected by lowpass filtering the entire signal at 4.5 Hz and marking all consecutive positive-to-negative zero–crossings that fall between 0.8 and 2 s in length as a SO.

All spindle features were limited to stage N2 sleep. The following six features were derived from the detected spindles and their associated SO (when present): (1) amplitude: largest peak-to-peak amplitude in a band-pass-filtered signal (11–15 Hz); peak frequency (the modal frequency in the signal spectrum); oscillation count (number of oscillations per spindle); symmetry index (based on the relative location of the spindle's central "peak," which is the point of maximum peak-to-peak amplitude); SO phase peak: SO phase at spindle peak; duration (in s). Each spindle was categorized as being fast (peak frequencies ≥13 Hz) or slow (<13 Hz), and each of the six above features were then aggregated to the whole recording level using the following statistics: mean, standard deviation, kurtosis, skew, minimum, maximum, and 0.25, 0.5, and 0.75 quantile values. While computing these statistics, fast and slow spindles were kept separate for all six features except for peak frequency for which these same statistics were computed over fast and slow spindles combined. In addition, we computed and used the spindle density (number of spindles, without distinguishing between fast or slow spindles, per minute of N2 sleep) and dispersion (mean dispersion index of the spindle count over 30-s epochs, defined as the ratio of the variance to the mean of spindle counts, without distinguishing between fast and slow spindles). All spindle features were computed separately per channel. In all, this resulted in 606 engineered spindle features

(see Supplemental Materials for a detailed breakdown of these features).

Sleep macro-architecture features

All PSG recordings underwent conventional manual sleep stage scoring by qualified sleep technicians following American Academy of Sleep Medicine standards [33]. Each 30-s epoch is scored as 1 of the 5 stages: wake (W), rapid eye movement (REM), NREM stage 1 (N1), NREM stage 2 (N2), and NREM stage 3 (N3). Macro-architecture features were computed as the percentage of time spent in each non-awake sleep stage relative to the TST (time spent in N1, N2, N3, or REM but not in W) in each quarter of the lights-off period. This resulted in 16 engineered macro-architecture features.

Multivariate analyses: The IES

We created IES scores by training ML models to discriminate between brain activity from individuals without insomnia and individuals in any of the four insomnia subgroups. To compare the information contained in different aspects of sleep, we created separate IES scores based on our three types of engineered features: spectral features, spindle features, and macro-architecture features. In each case, the IES score is a weighted sum of feature values, where the weights are optimized using a sixfold cross-validation approach, with the following steps:

(1) Compute engineered features for each recording (see above)
(2) Normalize features to have zero mean and unit variance across all recordings
(3) Oversample recordings to create equal numbers of individuals with versus without insomnia
(4) Partition recordings into six random folds such that oversampled duplicates of recordings are placed in the same fold as the original recording
(5) Hold each fold out as an evaluation set, using the other five-fold as data to train L1-penalized logistic regression models for a range of different penalty strengths, λ [5, 10, 20, 50, 100, 200, 400, 1000]
(6) Select the penalty strength that produces the highest area under the curve (AUC) averaged over the folds
(7) Use the six resulting logistic regression models (one per fold) to compute the (held-out/cross-validated) predicted probability of being an individual without insomnia for each fold
(8) Interpret the output of these logistic regression models as an IES score

IES scores comprising features from all six PSG channels could only be trained on Beacon and Idorsia datasets, while those comprising features from central electrodes (C3, C4) could be trained on all three datasets (Beacon, Idorsia, and SHHS). This sixfold cross-validation methodology allows for the use of all data points for both training and validation and therefore avoids the general pitfall of the feature optimization methodology, which is that weights must be optimized on a held-out set of data (i.e. data not used to train the model). This methodology was applied for each of the three types of engineered features separately and for the three feature types combined. On 6-channel data, this methodology was also applied separately to the Beacon dataset only and on the Beacon and Idorsia datasets together. On 2-channel data, it was applied separately to all six combinations of the three datasets that have individuals with and without insomnia (this excludes using Idorsia on its own, since this dataset only has individuals with insomnia). All in all, this resulted in 32 IES scores.

Evaluation of feature discrimination

We use the area under the receiver operating characteristic curve (AUROC) to quantify how different the distributions of engineered and IES features are between individuals without insomnia and individuals with insomnia (in any of the four insomnia subgroups). AUROC measures the average true positive rate over the whole range of false positive rates: by varying a classification threshold applied to the feature, one obtains classifications with false positive rates ranging from 0.5 (no discrimination of groups) to 1 (maximum discrimination of groups), each with a corresponding true positive rate (sensitivity). The AUROC provides a more comprehensive account of model performance than can be conveyed by a simple accuracy measure, being the average sensitivity when weighing false positive rates equally as one varies the classification threshold. Note that AUROC values were expected to be elevated, given the underlying insomnia versus non-insomnia groups were defined in part by differences in objective sleep efficiency. To distinguish the significance of different AUROC scores, we performed DeLong's Test to determine the statistical significance (p-value) of the difference between curves [36]. DeLong's test uses the covariance between the AUC values of two ROC models to calculate a z-statistic, used to calculate a two-sided p-value for the null hypothesis that the models are equal. Bonferroni corrections were applied to each pairwise p-value to correct for multiple comparisons.

We also calculated precision recall (PR) curves, which emphasize the precision of the model, by deriving precision (true positives/(true positives + false positives)) and recall (true positives/(true positives + false negatives)) at varying model thresholds. The area under the PR curves (AUPR) was calculated and is reported with AUROC values.

Dependence on age and sex

In addition to the macroquantity, spindle and spectral features considered, all IES models include age and sex as inputs. Model AUROC statistics quoted here are therefore not adjusted for age or sex. To account for the impact of age and sex, we examined the AUROC of age or sex only models (0.62 AUC for age and 0.57 AUC for sex) and compared to test model AUROCs. Additionally, we selected the most accurate model and examined parameters included at a sweep of λ values to characterize which parameters would be included at varying penalties. This is characterized in the *Model specifications and coefficients* under Results section.

Results

Cohort characteristics

Characteristics of the cohort are shown in Figure 1. After applying inclusion and exclusion criteria, the numbers of individuals and recordings (n/N) were 1455/1575 (92 per cent) for the non-insomnia group, and 2248/2276 (99 per cent) for the insomnia group. In the insomnia group, the number of individuals and recordings available for analysis within each category was: subjective only: 19/19 (100 per cent); maintenance: 420/431 (97 per cent); onset: 108/108 (100 per cent); maintenance + onset: 1710/1718 (99.5 per cent). The median age of the non-insomnia group was 51 years, and 64 per cent of individuals were female. The median age of the insomnia group was 58 years and 50 per cent of the individuals were female ($p < .001$). Note that the populations were not intended to be matched between based on any variable for insomnia versus non-insomnia or between sites, with age and sex being controlled in the multivariate models.

It is therefore expected that age, sex, and TST were statistically different between sites (see Supplementary Table S4, all p-values <.001). This may also reflect fundamental "site" (the Idorsia clinical trial with patients rigorously selected as outlined, Beacon datastore with clinical PSGs from patients undergoing evaluation, and SHHS with subjects representing a cross section of the adult >40 population without sleep apnea) differences. Site level differences are shown in Supplementary Tables S4 and S5. All site level differences were statistically significant except that WASO was not different between the Beacon and SHHS insomnia groups, percent N2 was not statistically significantly different between the Beacon and SHHS non-insomnia groups, and age was not statistically significantly different between the Idorsia and Beacon with insomnia groups.

Univariate analyses. In Figure 2, we visually compared the ability of spectral (240 features), spindle (202 features), and macro-architecture (16 features) derived features and the IES scores (18 features) to correctly classify individuals using features derived from central channels only; Supplementary Figure S1 similarly shows all 720 spectral, 606 spindle, 16 macro-architecture, and 24 IES score features that can be derived from all six channels. Features are clustered along the X axis of the heat map by feature type, while insomnia and non-insomnia cohorts are clustered on the Y axis. Feature values are rescaled to a number between 0 and 1 so that all features could be visualized on the same plot despite differences in units and scales. Finally, because some features correlate positively and others correlate negatively with insomnia, features with negative correlation with insomnia are flipped ($1 - x$) so as to correlate positively with insomnia. An ideal feature set would therefore show mostly high values for insomnia individuals and low values for non-insomnia individuals.

Sleep macro-architecture features. The macroquantities considered for this study were sleep stage based features defined as the percentages of time spent asleep (spent in N1, N2, N3, or REM but not Wake) in each of the four non-Wake sleep stages in each quarter of the night (4 non-wake sleep stages over 4 quarters of the night, for a total of 16 features). These PSG-based macro-architecture features are the second column in Figure 2. Many of these features demonstrated statistically significant differences between insomnia and non-insomnia groups, as previously reported [23]. However, none of these features performed well in isolation as a classifier of insomnia. At best, the percentage of N1 time in the second quarter of the night demonstrated modest discrimination, with an area under the receiver operator curve (AUROC) of 0.72 (see Figure 3).

Spectral features. Results for the univariate analysis of spectral features are shown in the third column of the heat map in Figure 2. Consistent with prior reports, many of these features demonstrated statistically significant differences between insomnia and non-insomnia groups [23]. However, when considered individually, spectral features also poorly classified insomnia vs non-insomnia individuals, even when superior to macro and spindle features. The AUROC for the single best spectral feature (minimum relative delta over 30-s epochs averaged across the night) was 0.74. Precision recall AUC (AUPR) was poor at 0.51, suggesting that the model has some ability to differentiate between insomnia and non-insomnia, but its performance on classifying insomnia is weak (and this class is inherently less likely in the dataset).

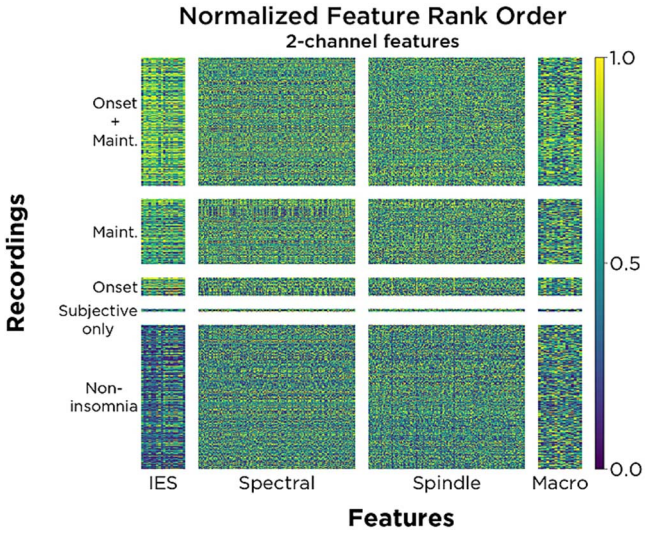


Figure 2. Heat map of normalized rank orders for all features derived from central channels, across all recordings, grouped by subgroup and feature type. To compute feature rank orders, feature values were mapped to their sorted rank among the set of non-missing values available for that feature, so that for example the smallest feature value turns into rank order 1, and the largest turns into the total number of recordings. These rank integers were then normalized to cover the range between 0 and 1. Features negatively correlated with insomnia are flipped ($1 - x$) so that for all features, larger values correspond to insomnia. Recordings missing values for any of these features are not shown. Recordings were randomly shuffled within each subgroup, and features shuffled randomly within each feature type. For macro-architecture (macro), spindle and spectral features, differences between insomnia and non-insomnia groups are subtle, whereas they are more apparent for IES features.

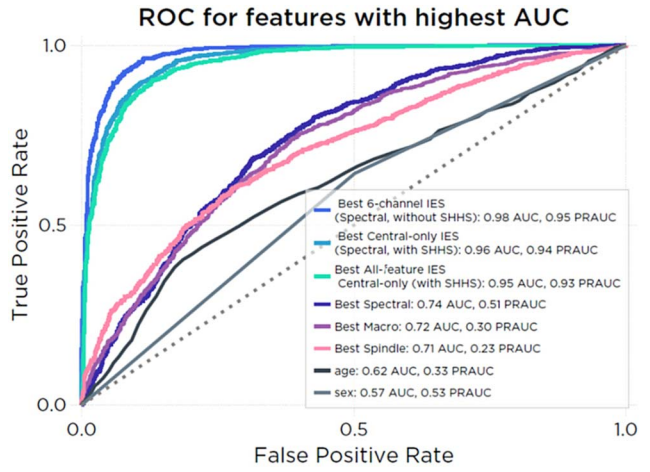


Figure 3. ROC curves and corresponding AUC scores for the macro, spindle, spectral, and IES features with the highest AUC scores, along with AUC scores for age and sex. Both age and sex were provided as inputs to compute IES, along with EEG features.

Spindle features. Results for the univariate analysis of spindle features are shown in the first column of the heat map in Figure 2. Many of these features demonstrated statistically significant differences between insomnia and non-insomnia groups [23]. Like sleep macro-architecture features, individual spindle features poorly distinguish individuals with insomnia from non-insomnia individuals, with an AUROC for the best spindle feature of 0.71 (the standard deviation of the peak spindle frequency on the O2 channel, see Figure 3). The AUPR was

0.23, again demonstrating that this univariate model had poor precision.

Multivariate analyses. As reported in prior literature, none of the individual EEG-based features examined reliably distinguished insomnia from non-insomnia individuals. The multivariate models (using combinations of macro, spindle, or spectral features) developed to distinguish insomnia from non-insomnia individuals provided output between 0 (definitely non-insomnia) and 1 (definitely insomnia) (IES). Two important points should be noted again: (1) insomnia and non-insomnia populations were defined such that they were inherently different based on sleep efficiency, which therefore results in high AUROC values, and (2) all IES models included age and sex in addition to any PSG-derived EEG features. For comparison, IES models using *only* age and sex in this population are shown in Figure 3 and Supplemental Figure S2.The IES model that used spectral features derived from all six PSG EEG channels from Beacon and Idorsia (excluding SHHS data since that dataset only has central channels), demonstrated the best ability to distinguish insomnia individuals from non-insomnia individuals (Figure 3, "Best 6-channel IES," AUROC: 0.98, AUPR, 0.95, Accuracy: 90 per cent). These AUROC and AUPR are not inherently informative given population differences discussed, but are useful when comparing to other models. Note that this model has the best discriminative value for insomnia and non-insomnia of the models we tested, and has excellent precision. The IES model incorporating spectral features from only central channels (therefore including SHHS data) performed nearly as well (Figure 3, "Best 2-channel IES," AUROC: 0.96, AUPR 0.94, Accuracy: 85 per cent). More complex models (incorporating a combination of spectral, macroquantity, and spindle features) performed slightly worse (AUROC: 0.95, AUPR 0.93, and Accuracy 84 per cent).

We then evaluated the best IES model performance when trained using only the first PSG for each participant (excluding repeat measures)—to determine the impact of repeat measures on model performance. There was only a minimal impact on performance, with a 0.01 reduction in AUROC and AUPR (reduced to 0.96 and 0.94, respectively—as shown in Supplementary Figure S3). Eliminating duplicates also did not change the list of features contributing most to model performance (Supplementary Table S7). Finally, we examined model performance when trained on limited subsets of data and applied to broader datasets, over all six PSG channels or using central channels only (in which case SHHS could be utilized). IES models trained only on Beacon and Idorsia datasets did not generalize well to SHHS data, and models trained on SHHS recordings did not generalize well to Beacon and Idorsia data, as evidenced by AUROC scores (see Supplementary Table S1). IES models using *only* age and sex performed most poorly. Using Delong's test, we evaluated statistical significance between the models shown in Figure 3. We found that all models were significantly different from each other (even with Bonferroni correction), except the spindle model versus age model (which had a corrected p-value of 1.0). This is shown in Supplementary Table S6.

Model specifications and coefficients. In order to characterize the importance of features for the best IES classifier model (6-channel spectral feature-based model), we retrained the model using λ (L1-penalties) ranging from 4000 (most restrictive, resulting in only one parameter being used) to 25 (least restrictive, resulting in 709 parameters being used). The AUROC for the resulting models was then characterized and is shown in Figure 4. Notably, spectral power related features were the six most important features (See Supplementary Table S7). These alone were able to generate a model with an AUROC of 0.78, which is already superior to the best univariate model. As λ was relaxed, additional features were accumulated in larger models with diminishingly improving AUROC (as expected). Age was the seventh most important feature, resulting in a model with an AUROC of 0.81. The best IES model reported here (again, the 6-channel spectral feature-based model) used a λ of 50, resulting in a model incorporating 301 model parameters. These results suggest that spectral features were more important than either age or sex (sex was not in the top 7).

IES scores across insomnia subtypes. The heat map in Figure 2 includes a column for every feature that can be computed with central channels. This includes the outputs of the 18 IES models that can be computed using central channels: IES-Spectral, IES-Macro, and IES-Spindle, trained on each of the six subsets of datasets other than the Idorsia dataset on its own, since the Idorsia dataset has no non-insomnia individuals. (Supplementary Figure S1 is a heat map showing all features computable using all six channels, but excluding SHHS since it is central channel only). This includes 6 IES model outputs in addition to the 18 IES model outputs shown in Figure 2: 6-channel IES-Spectral, IES-Macro, and IES-Spindle models trained on the Beacon dataset or on the Idorsia and Beacon datasets jointly. Note that all the IES models show improved ability to discriminate insomnia from non-insomnia individuals, with a gradation in scores depending on insomnia subtype. Mean IES scores were increased in individuals with insomnia and were (in increasing order of group mean): non-insomnia, 0.01; subjective-only insomnia, 0.21; sleep maintenance insomnia, 0.88; sleep onset insomnia, 0.95; sleep onset + maintenance insomnia, 0.996. This is shown in Figure 5 for the best performing 6-channel IES-Spectral model across Beacon and Idorsia datasets.

Finally, we examined the best performing IES model (6-channel spectral IES model) for patterns in misclassifications. We examined the five most misclassified insomnia subjects (IES score closest to 0) and compared their inputs to the five insomnia subjects with the highest IES scores. Similarly, we examined the five most misclassified non-insomnia subjects (IES score closest to 1) as compared to the five best classified non-insomnia subjects (IES score closest to 0). Notably, only two parameters showed a significant difference (uncorrected $p < .01$) in both misclassified non-insomnia subjects and misclassified insomnia subjects. Non-insomnia subjects who were misclassified as insomnia had 36 per cent lower log(mean relative alpha power) in occipital regions averaged across all states and 212 per cent more log(mean delta/theta power) in occipital regions in the wake state. In contrast, insomnia subjects who were misclassified as non-insomnia had 91 per cent lower log(mean relative alpha power) in occipital regions averaged across all states and 152 per cent less log(mean delta/theta power) in occipital regions in the wake state. Care must be taken in directly interpreting these parameters, which reflect only a small contribution to the overall model (and are difficult to interpret in isolation). Furthermore, Bonferroni correction for the 301 parameters used in the model eliminates the statistical significance of these differences. Nevertheless, it is notable that relative occipital alpha power in all states is the second most critical parameter for the 6-channel spectral model. Therefore, it is plausible that factors impacting relative occipital alpha without affecting sleep might increase the risk of model misclassification. Such risks might include certain medications

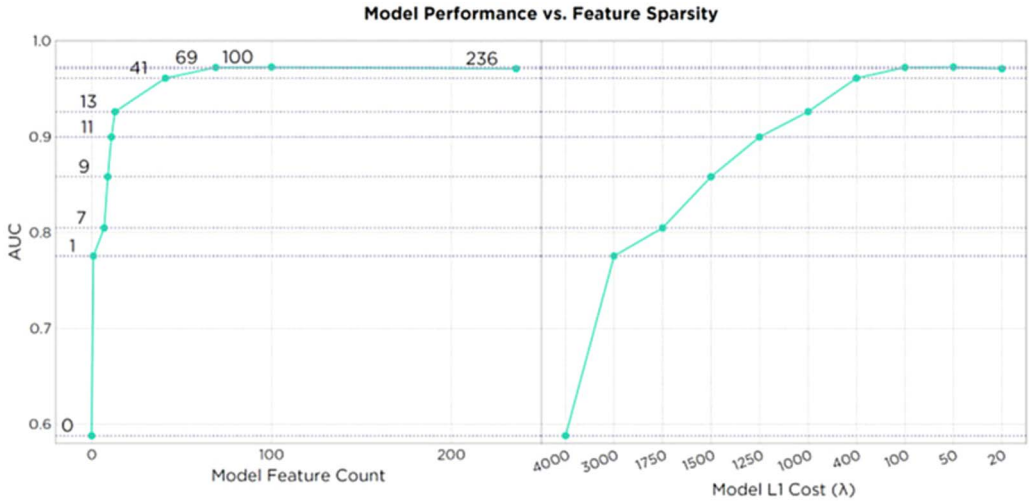


Figure 4. Performance of the 6-channel spectral feature IES model (the best performing IES model) was characterized at varying L1-penalty values, ranging from 4000 (which resulted in only 1 parameter being used) to 20 (resulting in 709 parameters). The first six parameters are related to spectral power content, and the seventh most important feature is age.

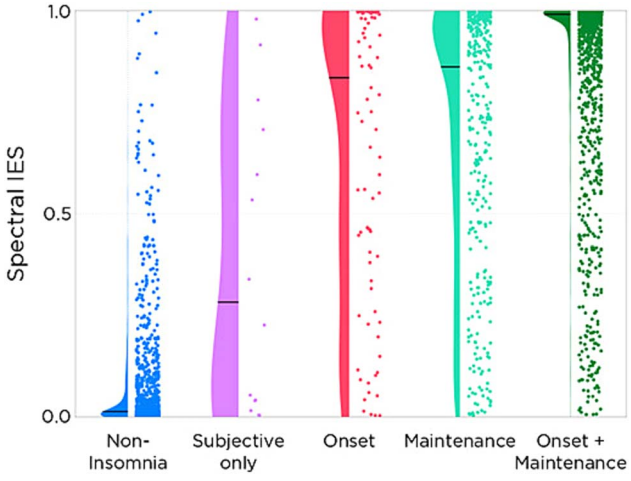


Figure 5. Distribution of spectral IES (trained over all six channels with beacon and Idorsia recordings) across all subgroups. Dots are all the individual recordings, offset randomly in the horizontal direction for visibility. Black horizontal tick marks are group medians.

(such as antiseizure medications, dopamine, and some antipsychotic medications), age extremes, encephalopathy of any cause, and underlying neurodegenerative disease.

Discussion

Classically, insomnia reflects subjective dissatisfaction with sleep initiation, maintenance, or inability to resume interrupted sleep and is associated with impairments in daytime function which leads to clinically significant distress [2]. Traditional interpretations of PSG-derived hypnograms from individuals with insomnia do not reveal major differences when compared to individuals without insomnia. Despite this, individual sleep characteristics reveal that the condition often presents itself with sleep features commonly linked to hyperarousal. This highlights the heterogeneity of insomnia and alludes to insomnia subtypes with differing EEG biomarkers [4, 14, 20]. However, it has previously been suggested that large datasets coupled to ML analytics have great potential in providing classifier models of insomnia capable of diagnosis [21].

Our findings show that PSG-derived EEG features, when examined at scale across large multi-site data, indeed identify reliable physiological differences between brain activity during sleep in individuals with insomnia relative to individuals without insomnia. It is not the presence of a simple, interpretable set of features that lead to good discrimination in our results. Rather it is the robust combination of many features that allows for good discrimination. For instance, a single feature combined with covariates such as age and sex are not sufficient, as demonstrated by our evaluation of model performance as a function of feature sparsity. With just a handful of features (e.g. 4 or 22), the model performs much worse than otherwise possible. Only when we allowed for less sparse models did we see the highest levels of insomnia versus non-insomnia discrimination.

Our prior research [23] supported the hypothesis that, at a population level, people with insomnia present with sleep spectral features (decreased relative delta and increased relative alpha spectral power in wake and N1) indicative of a hyperaroused brain state relative to people without insomnia. While these findings were apparent across a large population, they were not capable of distinguishing individuals with or without insomnia. Importantly, the current results demonstrated the ability of multivariate models (IES models) incorporating hundreds of quantitative PSG-based EEG-derived features to categorize insomnia at an individual level. These EEG spectral features were superior to age and sex in classifying insomnia versus non-insomnia, based on examination of model parameters included at different L1-penalties. Although no single spectral, spindle, or sleep macrofeature provided adequate classification performance, IES models that aggregated features within each category not only reliably classified individuals but were also able to distinguish the four insomnia subtypes examined here. The fact that the model with the most robust performance was trained on spectral features is notable given prior reports [22–24], suggesting that insomnia may be due to dysregulated wake-promoting neural circuitry, resulting in excess arousal and faster EEG patterns. As such, this may indicate that multivariate classification models may perform best when trained on features that may underlie some

of the key neurophysiological fingerprints present in people with insomnia.

While this study utilized the largest population of insomnia PSG data to date, several limitations remain. First, the population definitions were based on self-reported sleep problems consistent with an insomnia diagnosis and nocturnal PSG findings, but not necessarily meeting the full DSM-5 based insomnia diagnosis (which was not available for all data sources) and not including daytime sleep data. To account for the missing insomnia information, populations were in part defined by sleep efficiency. This results in two primary limitations. First, there is a risk that models using macrofeatures (linked to sleep efficiency) would outperform all others, given that sleep efficiency was used as part of the population definition. However, we note that microfeature models were not the best performing model (which were models based on spectral features). This result demonstrates potential value in spectral EEG content to assess insomnia, although this will need to be validated in additional clinically defined insomnia population. Reassuringly, the spectral models performed well in the Idorsia dataset, which did include pts with insomnia as defined by DSM-5 criteria. Second, the population differences resulted in models with high AUROC values (because the definitions result in populations that are more distinct than would be seen in a prospective data cohort). The strength of the models should therefore not be assessed based on the absolute AUROC or AUPR values, but rather on the values relative to the different models we present. This limitation cannot be overcome within the scope of the datasets available. Our results demonstrate substantially superior classification with PSG-derived EEG spectral features over sleep stage derived metrics (macroquantities) or sleep spindle-based features, even though the data are skewed to favor sleep stage derived metrics. We expect that additional training data will be needed for a clinically validated model, though these results highlight the potential use of models built on objective PSG-derived measures in sleep disorders.

The aggregation of diverse data sources also introduced additional study limitations. Inconsistent or lacking metadata regarding BMI, habitual bed time, shift work status, caffeine, alcohol, or tobacco use resulted in our models missing potentially informative features. This data would be valuable to include in future models and could account for significant performance variance. Furthermore, recording characteristics were different amongst the three core datasets, including minimum total recording. This might partially explain why model performance was not equivalent when tested on held-out data from each study site (Idorsia, Beacon, or SHHS). Performance between Beacon and Idorsia data were most comparable, with worse performance on SHHS data. However, missing metadata is likely less relevant to the worse performance for SHHS data than the fact that Beacon and Idorsia data were collected in sleep laboratory settings with at least six EEG electrodes during the first overnight period in the clinic, while SHHS data were collected in the individuals' homes with 2 electrodes. Therefore, in addition to the different electrode configurations between the datasets, recordings from the Beacon and Idorsia datasets may have been more susceptible to "first night effects" due to recording in an unfamiliar environment, leading to deviations from individuals' normal sleep patterns [37, 38]. A small subset of participants had multiple PSGs included in this study, which could alter first night effects for those participants. It must be noted that the majority of dependent, repeat PSGs fell into the control cohort, and thus could have artificially improved model performance. However, the percentage of repeat PSGs was small and retraining the best performant model without repeat data did not meaningfully alter the results. Future work should account for the impacts of recording natural sleep in the home environment and the impact of repeated recordings (including timing of repeat PSGs, cause of repetitions, and effects of differences in lifestyle or treatment parameters between repeat studies). Finally, individual insomnia cohorts—particularly subjective insomnia—were not well represented in our dataset. Although the IES-Spectral model appears to distinguish this cohort from non-insomnia individuals, the sample is too small to draw firm conclusions for this group.

Despite these potential limitations, IES models demonstrated similar patterns regardless of data source. These findings suggest that insomnia involves quantifiable differences in the EEG content during sleep, albeit manifesting across multiple EEG features and too subtle to be assessed by visual inspection. These multivariate changes appear to span multiple domains, as indicated by the relative success of three separate classification models. That the best performing model was based on spectral components (which we have previously shown as having more wake-like spectral patterns in individuals with insomnia [23]) are supportive of the hyperarousal model of insomnia. Individuals with only subjective insomnia (i.e. no issues with objective sleep onset or maintenance) scored closer to non-insomnia individuals, whereas individuals with combined sleep onset and maintenance insomnia segregated the most. The ability of these models to identify and classify individuals with insomnia demonstrates that IES may be a valuable quantitative neurophysiological assessment of insomnia. Future work will be required to validate EEG-based classifier models such as IES. Validation efforts will need to include more diverse data sets with additional clinical metadata, use of substances that impact sleep, mood, and lifestyle impacts on sleep, clinically defined insomnia diagnosis, and cohort selection to account for age and sex variability. Nevertheless, within the context of the IES models defined here (and within the limits of the insomnia definition used for this study), this work provides evidence of the feasibility of objective classification of insomnia. In theory, IES may offer an objective measure of insomnia presence that could be tracked longitudinally to characterize disease progression and aid in guiding the initiation of treatment strategies. In addition, alongside standard clinical measures of sleep, decreases in IES scores (i.e. becoming less insomnia-like) following initiation of treatment could be valuable in determining whether a treatment is efficacious in "normalizing" neurophysiological fingerprints of insomnia.

Supplementary material

Supplementary material is available at *SLEEP* online.

Disclosure statement

Financial disclosure: J.P., K.S., D.L., D.K., J.N., N.T., A.A., C.B.P., M.B.W., and J.D. are employees of Beacon BioSignals. A.O., T.D.M., and J.H. are employees of Idorsia Pharmaceuticals.

Non-financial disclosure: None declared.

Data availability

The Idorsia PSG database and Beacon Clinico-PSG database are proprietary and unable to be shared. The Sleep Heart Health Study PSG database [25] is publicly available.

References

1. Substance Abuse and Mental Health Services Administration. Impact of the DSM-IV to DSM-5 Changes on the National Survey on Drug Use and Health. CBHSQ Methodology Report. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, Rockville, MD. *Table 3.36, DSM-IV to DSM-5 Insomnia Disorder Comparison* 2016. https://www.ncbi.nlm.nih.gov/books/NBK519704/table/ch3.t36/
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013. https://doi.org/10.1176/appi.books.9780890425596.
3. Perlis ML, Giles DE, Mendelson WB, Bootzin RR, Wyatt JK. Psychophysiological insomnia: the behavioural model and a neurocognitive perspective. *J Sleep Res*. 1997;**6**(3):179–188. https://doi.org/10.1046/j.1365-2869.1997.00045.x
4. Trimmel K, Eder HG, Böck M, Stefanic-Kejik A, Klösch G, Seidel S. The (mis)perception of sleep: factors influencing the discrepancy between self-reported and objective sleep parameters. *J Clin Sleep Med*. 2021;**17**(5):917–924. https://doi.org/10.5664/jcsm.9086
5. Rezaie L, Fobian AD, McCall WV, Khazaie H. Paradoxical insomnia and subjective–objective sleep discrepancy: a review. *Sleep Med Rev*. 2018;**40**:196–202. https://doi.org/10.1016/j.smrv.2018.01.002
6. US Department of Health, Education, and Welfare. *Guidelines for the Evaluation of Hypnotic Drugs*. Food and Drug Administration (HEW Publication No. (FDA) 78-3051).
7. Riemann D, Espie CA, Altena E, *et al.* The European Insomnia Guideline: an update on the diagnosis and treatment of insomnia 2023. *J Sleep Res*. 2023;**32**(6):e14035. https://doi.org/10.1111/jsr.14035
8. Krystal AD, Edinger JD, Wohlgemuth WK, Marsh GR. NREM sleep EEG frequency spectral correlates of sleep complaints in primary insomnia subtypes. *Sleep*. 2002;**25**(6):630–640.
9. Normand M-P, St-Hilaire P, Bastien CH. Sleep spindles characteristics in insomnia sufferers and their relationship with sleep misperception. *Neural Plast*. 2016;**2016**:6413473. https://doi.org/10.1155/2016/6413473
10. Wei Y, Colombo MA, Ramautar JR, *et al.* Sleep stage transition dynamics reveal specific stage 2 vulnerability in insomnia. *Sleep*. 2017;**40**(9):zsaf155. https://doi.org/10.1093/sleep/zsx117
11. Edinger JD, Krystal AD. Subtyping primary insomnia: is sleep state misperception a distinct clinical entity? *Sleep Med Rev*. 2003;**7**(3):203–214. https://doi.org/10.1053/smrv.2002.0253
12. Baglioni C, Regen W, Teghen A, *et al.* Sleep changes in the disorder of insomnia: a meta-analysis of polysomnographic studies. *Sleep Med Rev*. 2014;**18**(3):195–213. https://doi.org/10.1016/j.smrv.2013.04.001
13. Buysse DJ, Germain A, Hall ML, *et al.* EEG spectral analysis in primary insomnia: NREM period effects and sex differences. *Sleep*. 2008;**31**(12):1673–1682. https://doi.org/10.1093/sleep/31.12.1673
14. Kang JM, Cho S-E, Moon JY, Kim SI, Kim JW, Kang S-G. Difference in spectral power density of sleep electroencephalography between individuals without insomnia and frequent hypnotic users with insomnia complaints. *Sci Rep*. 2022;**12**(1):2117. https://doi.org/10.1038/s41598-022-05378-6
15. Merica H, Blois R, Gaillard J-M. Spectral characteristics of sleep EEG in chronic insomnia: sleep EEG spectral characteristics in insomnia. *Eur J Neurosci*. 1998;**10**(5):1826–1834. https://doi.org/10.1046/j.1460-9568.1998.00189.x
16. Perlis ML, Merica H, Smith MT, Giles DE. Beta EEG activity and insomnia. *Sleep Med Rev*. 2001;**5**(5):365–376. https://doi.org/10.1053/smrv.2001.0151
17. Rezaei M, Mohammadi H, Khazaie H. Alpha-wave characteristics in psychophysiological insomnia. *J Med Signals Sens*. 2019;**9**(4):259. https://doi.org/10.4103/jmss.JMSS_51_18
18. Schwabedal JTC, Riedl M, Penzel T, Wessel N. Alpha-wave frequency characteristics in health and insomnia during sleep. *J Sleep Res*. 2016;**25**(3):278–286. https://doi.org/10.1111/jsr.12372
19. Edinger JD, Ulmer CS, Means MK. Sensitivity and specificity of polysomnographic criteria for defining insomnia. *J Clin Sleep Med*. 2013;**9**(5):481–491. https://doi.org/10.5664/jcsm.2672
20. Kao C-H, D'Rozario AL, Lovato N, *et al.* Insomnia subtypes characterised by objective sleep duration and NREM spectral power and the effect of acute sleep restriction: an exploratory analysis. *Sci Rep*. 2021;**11**(1):24331. https://doi.org/10.1038/s41598-021-03564-6
21. Andrillon T, Solelhac G, Bouchequet P, *et al.* Revisiting the value of polysomnographic data in insomnia: more than meets the eye. *Sleep Med*. 2020;**66**:184–200. https://doi.org/10.1016/j.sleep.2019.12.002
22. Zhao W, Van Someren EJW, Li C, *et al.* EEG spectral analysis in insomnia disorder: a systematic review and meta-analysis. *Sleep Med Rev*. 2021;**59**:101457. https://doi.org/10.1016/j.smrv.2021.101457
23. Di Marco T, Scammell TE, Sadeghi K, *et al.* Hyperarousal features in the sleep architecture of individuals with and without insomnia. *J Sleep Res*. 2025;**34**(1);e14256. https://doi.org/10.1111/jsr.14256
24. Riemann D, Spiegelhalder K, Feige B, *et al.* The hyperarousal model of insomnia: a review of the concept and its evidence. *Sleep Med Rev*. 2010;**14**(1):19–31. https://doi.org/10.1016/j.smrv.2009.04.002
25. Quan SF, Howard BV, Iber C, *et al.* The Sleep Heart Health Study: design, rationale, and methods. *Sleep*. 1997;**20**(12):1077–1085.
26. Zhang G-Q, Cui L, Mueller R, *et al.* The National Sleep Research Resource: towards a sleep data commons. *J Am Med Inform Assoc*. 2018;**25**(10):1351–1358. https://doi.org/10.1093/jamia/ocy064
27. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*. 2004;**27**(7):1255–1273. https://doi.org/10.1093/sleep/27.7.1255
28. Niedermeyer E, Lopes da Silva FH. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Philadelphia, PA: Lippincott Williams & Wilkins; 2005.
29. Babadi B, Brown EN. A review of multitaper spectral analysis. *IEEE Trans Biomed Eng*. 2014;**61**(5):1555–1564. https://doi.org/10.1109/TBME.2014.2311996
30. Thomson DJ. Spectrum estimation and harmonic analysis. *Proc IEEE*. 1982;**70**(9):1055–1096. https://doi.org/10.1109/PROC.1982.12433
31. Thomson D. Jackknifing multitaper spectrum estimates. *IEEE Signal Process Mag*. 2007;**24**(4):20–30. https://doi.org/10.1109/MSP.2007.4286561
32. Sun H, Jia J, Goparaju B, *et al.* Large-scale automated sleep staging. *Sleep*. 2017;**40**(10). https://doi.org/10.1093/sleep/zsx139
33. Berry R, Brooks R, Gamaldo CE, Harding SM, Marcus CL, Vaughn BV. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.2.2.2*. American Academy of Sleep Medicine: Darien, IL; 2015.

34. Buckelmüller J, Landolt H-P, Stassen HH, Achermann P. Trait-like individual differences in the human sleep electroencephalogram. *Neuroscience*. 2006;**138**(1):351–356. https://doi.org/10.1016/j.neuroscience.2005.11.005
35. Warby SC, Wendt SL, Welinder P, *et al*. Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nat Methods*. 2014;**11**(4):385–392. https://doi.org/10.1038/nmeth.2855
36. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;**44**(3):837–845. https://doi.org/10.2307/2531595
37. McCall C, McCall WV, Objective vs. subjective measurements of sleep in depressed insomniacs: first night effect or reverse first night effect? *J Clin Sleep Med*. 2012;**8**(1):59–65. https://doi.org/10.5664/jcsm.1664
38. Hu S, Shi L, Li Z, *et al*. First-night effect in insomnia disorder: a systematic review and meta-analysis of polysomnographic findings. *J Sleep Res*. 2024;**33**(1):e13942. https://doi.org/10.1111/jsr.13942

SLEEP, 2026, **49**, 1–11
https://doi.org/10.1093/sleep/zsaf353
Original Article