SLEEPJ, 2025, 00, 1–22
doi:10.1093/sleep/zsaf273
Advance Access publication 16 September 2025
Original Article

Original Article

The Boston Children's Hospital sleep corpus: a collection of 15695 annotated pediatric polysomnograms

Ayush Tripathi[1,2], Wolfgang Ganglberger[1,2], Haoqi Sun[1,2], Callison Alcott[1,2,3], Niels Turley[1,2], Rebecca Fitzgerald[3,4], Ayan Mitra[1,2], Samuel Waters[1,2], Arnav Gupta[1,5], Aditya Gupta[1,2,6], Manohar Ghanta[1,2,6], Valdery Moura Junior[1,2,6], Samaneh Nasiri[1,2,6,7], Bruce Nearing[1,2], Katie L. Stone[8,9], Emmanuel Mignot[10], Dennis Hwang[11], Matthew A. Reyna[7], Zuzana Koscova[7], Chad Robichaux[7], Zhiyong Zhang[10], Qiao Li[7], Gauri Ganjoo[10], Lynn Marie Trotti[7], Gari D. Clifford[7,12], Christine Tsien Silvers[13], Bharath Gunapati[13], Robert J. Thomas[2,14,†], M. Brandon Westover[1,2,*,†], Kiran Maski[2,3,4,†] and Umakanth Katwa[2,3,4,†]

[1]Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, United States, [2]Harvard Medical School, Boston, MA, United States, [3]Boston Children's Hospital, Boston, MA, United States, [4]Sleep Center, Boston Children's Hospital, Boston, MA, United States, [5]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, United States, [6]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States, [7]Emory University School of Medicine, Atlanta, GA, United States, [8]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, United States, [9]California Pacific Medical Center Research, CA, United States, [10]Stanford University, Palo Alto, CA, United States, [11]Kaiser Permanente, San Bernardino County Sleep Disorders Center, San Bernardino, CA, United States, [12]Georgia Institute of Technology, Atlanta, GA, United States, [13]Amazon Web Services and [14]Department of Medicine, Division of Pulmonary Critical Care & Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

*Corresponding author. M. Brandon Westover, 444 Kirstein, 330 Brookline Ave, Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA 02215, United States. Email: bwestove@bidmc.harvard.edu
†Co senior authors

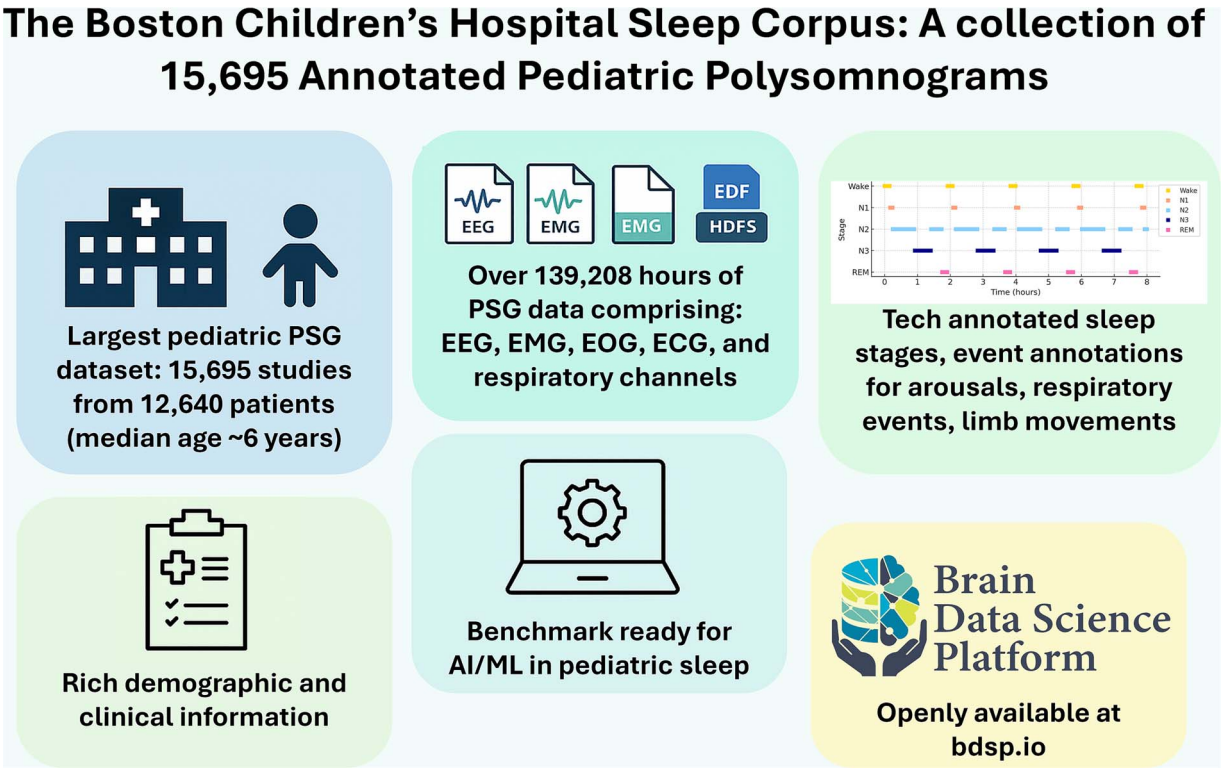
Abstract

Sleep is a fundamental biological process essential to health, particularly during early life when sleep patterns are developing and sleep disorders are common. Yet, pediatric sleep research is hindered by a lack of large-scale, high-quality polysomnography (PSG) datasets. To address this need, we introduce the Boston Children's Hospital Sleep Corpus—the largest pediatric PSG dataset available—comprising 15 695 overnight recordings from 12 640 unique patients (median age ~6 years). The dataset includes 16.7 million annotated sleep stages, 2.25 million respiratory, arousal, and limb movement events, and over 11 000 patient diagnoses linked through de-identified electronic health records. Each PSG has a median duration of 8.9 h, totaling 139 208 h of electroencephalography (EEG) data. Sleep staging follows American Academy of Sleep Medicine guidelines and reveals age-related trends: REM sleep decreases from 33.5% in neonates to 16.3% in teenagers, while N2 sleep increases from 21.7% to 35.4%. Central apneas decline with age, while obstructive hypopneas and respiratory effort-related arousals events rise. Limb movements are not scored in <1 year but remain at around 30 per PSG across older age groups. We also present age- and region-specific EEG spectral norms and respiratory event trends across the pediatric age range. The dataset is organized in Brain Imaging Data Structure format and publicly available via the Brain Data Science Platform. The dataset provides a valuable resource for improving our scientific understanding of pediatric sleep and developing automated PSG analysis with artificial intelligence tools.

Key words: pediatric sleep; polysomnography; sleep disorders; machine learning; EEG; respiratory events; sleep staging

Submitted: 17 June, 2025; **Revised:** 6 August, 2025; **Accepted:** 26 August, 2025

Graphical Abstract



Statement of Significance

Sleep plays a critical role in healthy brain and body development in children, yet pediatric sleep remains under-researched due to a lack of large-scale, high-quality data. This article introduces the Boston Children's Hospital Sleep Corpus, the largest pediatric polysomnography dataset ever released, comprising 15 695 annotated overnight studies from 12 640 patients. The dataset spans from neonates to adolescents, offering rich annotations of sleep stages, respiratory events, arousals, and limb movements, and is linked to de-identified clinical diagnoses. The dataset enables new opportunities for research in pediatric sleep physiology, disease prediction, and development of age-aware machine learning models. Its availability is poised to accelerate innovation in sleep science and improve diagnostic precision and outcomes in pediatric populations.

Introduction

Sleep, a fundamental biological necessity, underpins health and well-being by facilitating crucial restorative functions such as memory consolidation, immune regulation, hormonal balance, emotional stability, and cognitive performance [1–3]. Furthermore, sleep patterns serve as a key diagnostic indicator for various medical conditions, encompassing sleep disorders and more complex neurological and mental health issues [4, 5]. Untreated sleep disorders can lead to severe health complications, including neurocognitive, cardiovascular and metabolic dysfunctions, highlighting the importance of precise sleep analysis in healthcare. Poor sleep quality frequently coexists with chronic diseases, further diminishing patients' quality of life [6, 7].

In children, sleep disturbances present with diverse origins, including sleep-related breathing disorders, movement disorders, circadian rhythm disruptions, insomnia, hypersomnia, and parasomnias [8–10]. These disturbances impact daily functioning and can have long-lasting developmental consequences. For instance, insufficient sleep duration, frequent nighttime awakenings, or non-restorative sleep in children can impair learning, memory, emotional regulation, and physical growth [11–14].

Children's sleep patterns and circadian rhythms undergo substantial transformations throughout their development. Neonates exhibit polyphasic sleep patterns, sleeping for extended periods (14–16 h or more) without established day-night rhythms. As children progress from infancy into adolescence, shifts in sleep patterns correlate with neurodevelopmental changes and evolving cognitive and emotional regulation [15]. Monitoring sleep problems during these critical developmental stages allows for early detection and intervention, potentially preventing long-term negative impacts on a child's overall health and development.

The diagnosis of pediatric sleep disorders often involves polysomnography (PSG), an overnight sleep study that monitors a child's physiological signals [16]. During a PSG, trained sleep technologists attach sensors to record brain activity (EEG), heart activity (ECG), respiratory signals, gas exchange, and muscle activity (EMG). This data enables clinicians to evaluate sleep and identify abnormalities such as obstructive apnea (OA) events, limb movements, or irregular heart rhythms [17]. PSG data is essential for diagnosing sleep-related conditions and ensuring timely interventions for underlying disorders [18].

Despite considerable research on adult sleep, a significant gap exists in the pediatric domain. While numerous adult sleep

Table 1. Overview of existing pediatric sleep study datasets

Dataset	Age range (years)	Sample size	Data type	Sleep staging	Arousal	Respiratory event	Limb movement
Need for sleep [21]	14–19	110	PSG	✓			
Rise & SHINE [22]	0–2	433	Actigraphy				
CHAT [23]	5–9.9	464	PSG	✓	✓	✓	
CCSHS [24]	8–11	907	In-home sleep study			✓	
Chicago pediatric dataset [25]	5–7	1010	PSG	✓	✓	✓	✓
PATS [26]	3–12	555	PSG, Actigraphy	✓	✓	✓	✓
NCH Sleep DataBank [27]	0–58	3984	PSG	✓	✓	✓	✓
BCH Sleep Corpus	0–58	15 695	PSG	✓	✓	✓	✓

This table summarizes the datasets that are currently available for pediatric sleep studies, emphasizing the critical gap in data for understanding pediatric sleep patterns and disorders. The table highlights the limited availability of large-scale, high-quality datasets for children in comparison to adult sleep research. The need for specialized diagnostic tools like polysomnography (PSG) in pediatric populations contributes to this disparity, as children's sleep architecture varies significantly across developmental stages, often requiring more detailed analysis.

research datasets are available, large-scale, high-quality datasets for children are lacking. To address this deficiency, we introduce the Boston Children's Hospital (BCH) Sleep Study corpus, which includes 15 695 PSG recordings from 12 640 unique patients collected between 2010 and 2024. These PSGs, obtained in a clinical setting, are linked with de-identified electronic health records (EHR), providing clinical context. Each PSG recording includes annotations by trained sleep technologists for sleep stages, arousals, limb movements, and various respiratory events, such as central, mixed, and OA, respiratory effort-related arousals (RERAs), and hypopneas. This dataset offers a valuable resource for developing artificial intelligence algorithms for automated sleep analysis and, through the linked clinical data, enables the exploration of associations between sleep patterns and specific diagnoses, potentially facilitating early prediction and risk assessment of various health conditions [19, 20]. By providing a comprehensive collection of pediatric PSGs, the BCH Sleep Corpus has the potential to advance our understanding of sleep patterns and improve the diagnosis of sleep-associated disorders in the pediatric population. An overview of existing pediatric sleep datasets, including the BCH Sleep Corpus, is presented in Table 1 to highlight the unique scale and comprehensiveness of our dataset.

Methods

Data acquisition

The BCH Sleep Corpus is a comprehensive collection of overnight sleep studies conducted at Boston Children's Hospital between 2010 and 2024. The data were exported from the Natus SleepWorks [28, 29] system and include a diverse array of physiological signals collected during sleep. These signals include:

- Electroencephalography (EEG) data recorded from frontal, central, and occipital electrode locations, capturing brain activity.
- Electromyography (EMG) signals from the chin, and the left and right legs, which help to monitor muscle activity during sleep.
- Electrooculography (EOG) signals from both the left and right eyes, which track eye movements.
- Electrocardiography (ECG) data, acquired using a bipolar derivation between left and right chest electrodes (ECGL–ECGR).
- Respiratory monitoring using airflow, pressure transducer, end-tidal and transcutaneous capnography, and abdominal, and chest belt signals to track respiratory patterns and events.
- Oxygen saturation and photoplethysmography.
- Snoring sounds recorded throughout the night.

Each PSG was annotated by a trained and certified sleep technologist. Across the dataset, annotations were produced by a pool of approximately 10–12 registered sleep technologists at Boston Children's Hospital across the time span from 2010 to 2024. Each study was scored by a single rater, following the American Academy of Sleep Medicine (AASM) guidelines that were active at the time of acquisition. However, rater identifiers were not recorded during the de-identification process and are therefore not included in the released dataset. These annotations include sleep stages (WAKE, N1, N2, N3, REM) respiratory events, arousals, and limb movements. Annotations were made using both standard labels available within the Natus SleepWorks framework [28, 29], and free-text inputs provided by sleep technologists. The recordings adhere to guidelines set by the AASM that were active at the time of PSG collection. We did not retrospectively harmonize annotations to a single AASM version; instead, we retained the original annotations as scored by technologists following the standards in place during each recording period. After collection, data were converted into standard European Data Format (EDF) to facilitate broader accessibility and analysis.

In addition to the sleep recordings, the dataset also includes relevant EHR.

Data de-identification procedure

In compliance with the Health Insurance Portability and Accountability Act (HIPAA) guidelines [30], all PSG files and associated health information in the BCH Sleep Corpus were de-identified using the Safe Harbor method. The dataset is hosted on the Brain Data Science Platform (BDSP) for use by sleep researchers [31].

Each patient is assigned a unique identifier labeled "BDSP_ID" to maintain anonymity. To track multiple sleep studies from the same individual, a session-specific identifier "SESS_ID" is used to distinguish between different PSG recordings for the same patient. The original header information in the EDF files, which contained sensitive patient data, was removed. Additionally, annotations linked to the PSG recordings were de-identified. To achieve this, all unique free text entries from the annotations of all patients were extracted and compiled into a single text file. The file was then processed through PHILTER [32], a certified tool designed to detect and redact PHI from unstructured medical text. Once de-identified, the cleaned annotations were saved in CSV files accompanying the PSG data.

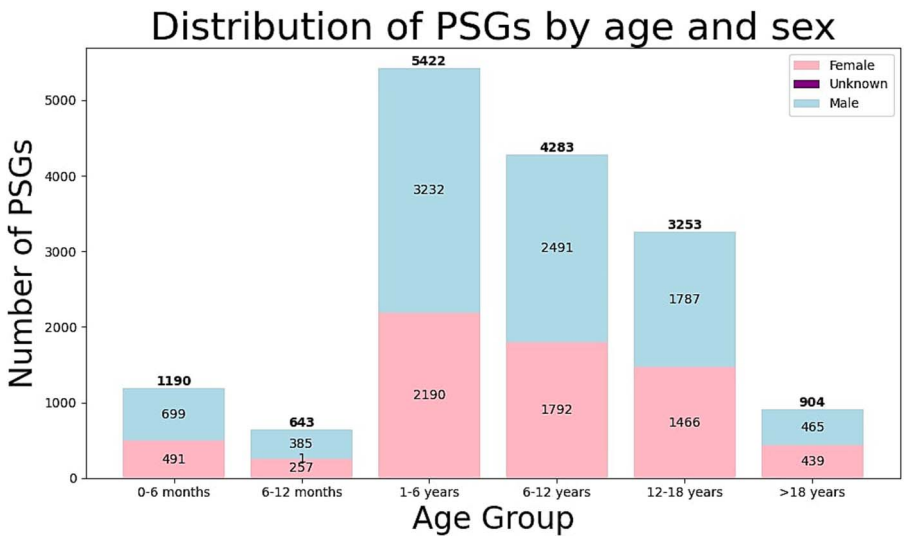


Figure 1. Distribution of the number of PSG recordings and sex distribution across different age groups in the BCH sleep study dataset. This figure presents a bar plot showing the distribution of the 15 695 PSG recordings across six distinct age groups, ranging from early infancy (0–6 months) to adulthood (18+ years). Each age group represents a critical stage in sleep development. The largest group is the toddler/preschool category (1–6 years), followed by primary school age (6–12 years) and adolescents (12–18 years). A small portion of the dataset consists of adult recordings from patients who continue to be followed at Boston Children's hospital. The distribution of recordings across various age groups ensures that the BCH sleep study dataset provides a comprehensive resource for studying pediatric sleep patterns and their evolution with age.

As part of data de-identification, all dates associated with the sleep studies and EHR data were randomly shifted. A random date shift was applied to all dates in each patient's data, ensuring that the temporal relationship between studies remained intact while removing the actual dates. This randomization was performed consistently for each patient so that all related data maintained accurate internal timing without revealing the original dates. These de-identification efforts ensure that the BCH Sleep Corpus complies with HIPAA regulations, safeguarding patient privacy while providing a rich dataset for research purposes. The study was conducted under an approved IRB protocol (#2022P000417) which granted a waiver of consent.

Data Records

The BCH Sleep Corpus includes comprehensive demographic data for all 15 695 recordings, which is stored in a CSV file (I0003_psg_metadata_2025_07_09.csv). This file contains essential information such as the unique patient identifier (BDSP_ID), session identifier (SESS _ID), age in years and days at the time of the recording, and the biological sex. These demographic data provide a foundation for analyzing sleep studies across various patient groups.

All data in the corpus are stored in the Brain Imaging Data Structure format. Each patient's PSG recordings and associated files are stored in individual folders, identified by their unique BDSP_ID. Within each patient folder, there is a .csv file named (BDSP_ID)_diseasediagnosis, which contains details about the ICD codes. Within each patient folder, subfolders labeled by the SESS_ID are included. These session-specific identifiers distinguish the different PSG recordings for patients who have undergone multiple sleep studies. Each SESS_ID folder contains several files:

- PSG recordings: The sleep recordings are stored as .edf files and are named in the format (BDSP_ID)_(SESS_ID)_taskpsg_eeg.edf. These files contain the raw EEG data collected during the sleep study.
- Sleep stage annotations: The sleep stages, such as REM, NREM, and wake, are annotated and saved in .csv format. These files are named (BDSP_ID)_(SESS_ID)_sleepannotations.csv.
- Event annotations: In addition to sleep stages, other events, such as apneas, hypopneas, and arousals, are also annotated and saved in a .csv file. These files are named (BDSP_ID)_(SESS_ID)_eventannotations.csv.
- Channel information: A Tab Separated Value (.tsv) file named (BDSP_ID)_(SESS_ID)_task-psg_channels.tsv provides details about the channels recorded during the PSG, along with their sampling frequencies.
- HDF5 files: A harmonized h5 file that contains both multichannel physiological signal data and synchronized per-sample annotations, uniformly resampled to 200 Hz. These files serve as an alternative to the raw EDF files, optimized for integration into machine learning pipelines and signal processing workflows.

Patient demographics

The BCH Sleep Corpus consists of a total of 15 695 PSG recordings from 12 640 patients. To illustrate how the dataset can be used to explore sleep pattern development throughout childhood and adolescence, we divided it into six distinct age groups based on the age of the patient at the time of the PSG. Figure 1 presents a bar plot depicting the number of PSGs in each group along with the sex distribution. The age groups are defined as follows:

- Early infancy (0–6 months): There are 1190 PSGs in this age range. Among these, 41.26% are female. The youngest patient was less than 24 h old at the time of the sleep study. This age group marks a critical stage of development where sleep pattern differs significantly from that of older age groups [33].
- Late infancy (6–12 months): This group contains 643 recordings, with 39.97% captured from female patients. Late infancy is characterized by further development of circadian rhythms and sleep cycles in children [33].

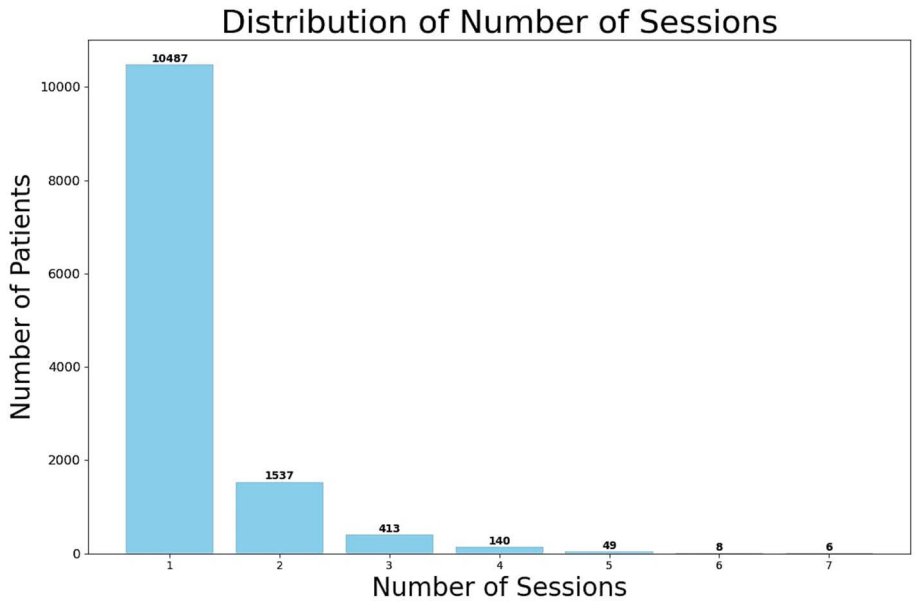


Figure 2. Distribution of the number of polysomnography (PSG) sessions per patient in the BCH sleep corpus. Out of 12 640 unique patients, 10 487 have a single PSG study, 1537 have two studies, and 616 have more than two studies. The maximum number of PSG studies for a single patient is seven. This distribution highlights the varying number of sleep studies conducted per patient, contributing to the comprehensive nature of the BCH sleep study dataset.

- Pre-school/toddlers (1–6 years): This group contains 5422 PSGs, 40.39% from female patients. This age range captures a significant portion of the early developmental changes in sleep, where sleep cycles and patterns start to stabilize [34].
- School-going (6–12 years): A total of 4283 PSGs were recorded in this age range, with 41.84% from females. The school-age period is a time when symptoms of sleep disorders may become more noticeable due to behavioral or learning issues and is often when children are first evaluated for such conditions [35, 36].
- Adolescents/teenagers (12–18 years): There are 3253 PSG recordings in this group, with 45.07% from females. Adolescence is a key developmental phase, where both sleep needs and patterns undergo substantial changes, influenced by hormonal shifts and lifestyle changes [37, 38].
- Adults (18+ years): Although the BCH Sleep Study dataset primarily focuses on pediatric sleep, it also includes 904 recordings from adult patients. Of these, 48.56% are from females. The oldest participant in this cohort is 58 years old.

The number of PSG studies per patient varies: 10487 patients have one study, 1537 patients have two studies, and 616 patients have more than two studies. The maximum number of PSGs for any one patient is seven. Figure 2 presents the distribution of the number of sessions with respect to the number of patients.

Clinical data

The BCH Sleep Corpus includes detailed information on diagnoses for 11 167 out of 12 640 patients (88.4% of patients in the dataset). This information is provided in de-identified .csv files stored in individual folders for each patient. These files contain both ICD codes and descriptions of the diagnoses. It is important to note that ICD codes were assigned at the time of the clinic visit, prior to the PSG, based on presenting symptoms. As such, a patient may carry an OSA ICD code despite not meeting diagnostic criteria during the subsequent sleep study.

To illustrate the potential utility of these data, we categorized patients into 22 clinically significant groups based on ICD codes. The categories include: Infectious diseases, Oncology, Hematology, Endocrinology, Psychiatry, Neurology, Ophthalmology, Otolaryngology (ENT), Cardiology, Pulmonology, Gastroenterology, Dermatology, Orthopedics, Urology, Obstetrics, Neonatology, Genetics, Internal Medicine, Emergency Medicine, External Morbidity Causes, Miscellaneous Diagnoses, and Special Diagnoses. In Figure 3, we present bar plots, sorting the categories in ascending order based on the number of patients in each to highlight the variation in prevalence among the different conditions. The top 20 most common ICD codes for each of the age groups are presented in Table 2. One notable finding is the high prevalence of dysphagia among infants. Specifically, the most common diagnosis for both early and late infants is "R13.12 Dysphagia, oropharyngeal phase," categorized under "R13 Aphagia and dysphagia." Dysphagia affects 62.18% of early infants and 57.08% of late infants in the dataset, who underwent sleep studies due to desaturation and apneas. This may reflect immature breathing which is commonly associated with oropharyngeal dysphagia as a comorbid condition as this function is controlled by the brainstem. In contrast, for all other age groups, the most common diagnosis is "G47.33 Obstructive sleep apnea (adult) (pediatric)," which falls under the broader category of "G47.3 Sleep apnea."

Polysomnography data

A diverse range of channels are included in the PSG recordings, ranging from 33 to 103 channels. The average number of channels across the dataset is 69, with a median of 50 channels and the 90th percentile being 91 channels. Table 3 provides a detailed summary of the channels of interest, along with the frequency of their use in the dataset.

Certain key channels are consistently present across all recordings. Specifically, central EEG channels (C3, C4, and Cz), occipital channels (O1 and O2), and frontal channels (F3 and F4) are found in every recording. In addition, channels for EOG, Chin EMG, respiratory effort monitoring of the abdomen and chest, and oxygen saturation (SpO_2/OSAT) are available in all the .edf files. The majority of recordings also include data from other

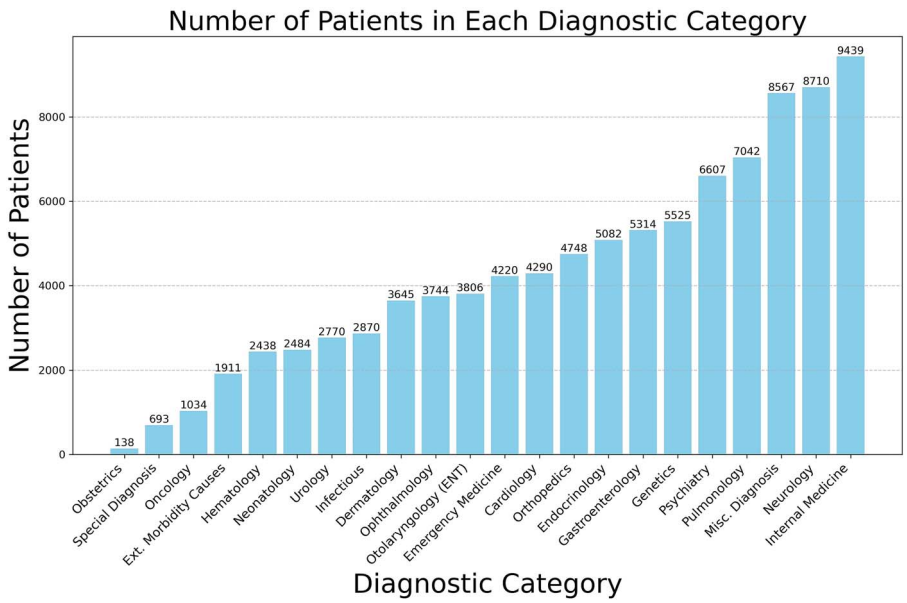


Figure 3. Distribution of patients across 22 diagnostic categories based on ICD-10-CM tabular list of diseases and injuries. Each bar represents the number of patients with at least one diagnosis in the corresponding category. Categories are arranged in ascending order by the number of patients. The height of each bar reflects the total number of patients, with exact counts displayed above the bars. Diagnoses were mapped to relevant categories using the ICD-10-CM tabular list of diseases and injuries. This figure highlights the prevalence of different disease categories in the patient cohort. The visualization provides an insightful overview of the diagnostic distribution, which is useful for clinical research and healthcare resource allocation.

EEG electrode placements, leg EMG, snore detection, and pressure and airflow measurements, offering a comprehensive view of the physiological state of the patient during sleep.

In total, the BCH Sleep Corpus provides about 139 208 h of PSG recording. The median PSG duration is 8.9 h. A boxplot of PSG duration, along with sleep duration, sleep fragmentation index (SFI), sleep efficiency, and wake after sleep onset (WASO) across different age groups, is shown in Figure 4. The plots highlight the distribution of sleep study duration and different standard sleep metrics across various demographics. A total of 11 497 recordings (representing 73.25% of the total) were sampled at 512 Hz. Another 3918 recordings (or 24.96%) were sampled at 256 Hz. The remaining 280 recordings (1.78%) were captured at a sampling rate of 200 Hz. During the de-identification process, all EDF files were loaded and re-saved using the MNE [39] library, which by default resamples all channels to the maximum sampling rate present in the recording. As a result, although original PSGs typically contain signals sampled at different native rates (e.g. SpO_2 at 1 Hz), the de-identified EDFs in our dataset represent all signals at a uniform rate equal to the highest channel's sampling rate.

Sleep stage annotations

The 15 695 PSG recordings in the BCH Sleep Corpus include associated sleep annotation files in CSV format. These files include a total of 16 704 946 annotations, with mean, median, and 90th percentile values of 1064.3, 1068, and 1172 annotations per PSG, respectively. The annotations classify sleep into various stages: WAKE, N1, N2, N3, N4, REM, and UNSCORED. The label "UNSCORED" typically refers to time periods that were not assigned a sleep stage by the technologist. This most commonly includes time before lights off and after lights on, but can also occur in the middle of the recording due to brief interruptions such as impedance checks, headbox disconnection, or when the patient temporarily leaves the bed. For trend analysis, stages N4 and N3 have been combined into a single N3 category. However, the original N4 annotations remain available in the CSV files for users who wish to use this information. In the dataset, the distribution of sleep stage annotations is as follows: WAKE: 2863046 (17.14%), N1: 817583 (4.89%), N2: 4912325 (29.41%), N3: 3817337 (22.85%), N4: 390118 (2.33%), REM: 2842367 (17.14%), and UNSCORED: 1062170 (6.36%). The breakdown of these annotations across different age groups is detailed in Table 4, while Figure 5 provides a visual representation of the variation in the percentage of different sleep stages across these age groups.

The analysis of sleep stage percentages across different age groups reveals significant changes in sleep architecture with development. A notable trend is the consistent decrease in the proportion of REM sleep as age progresses. This aligns with established literature [40]. This shift reflects developmental changes in the sleep cycle, with infants spending more time in REM sleep [41]. As REM sleep decreases with age, it is progressively replaced by an increase in NREM sleep, particularly stage N2 [40]. This increase in N2 sleep reflects the ongoing maturation of spindle activity (in thalamic region) and the stabilization of spindle frequency, which is crucial for consolidating sleep and cognitive processes [42]. The transition from predominantly REM sleep to a greater proportion of NREM sleep reflects the dynamic maturation of sleep architecture throughout development. The WAKE stage, on the other hand, remains relatively constant up until school age. However, a noticeable increase in wakefulness is observed in teenagers and adults. This increase may be attributed to a combination of shift in the circadian rhythms, hormonal changes, and lifestyle changes during adolescence. Although the sleep pattern in children may be affected by sleeping in the sleep lab environment, the transition from childhood to adulthood age group displays a distinct pattern of alterations in sleep patterns, including increased periods of wakefulness, increased N2 and N3 sleep and reduction in REM sleep which can be influenced by circadian, hormonal, developmental, psychological, and social factors.

Event annotations

In addition to the standard sleep stage annotations, the BCH Sleep Corpus includes an extensive collection of event annotations. In total, the dataset comprises 2 250 683 annotated events, with

Table 2. Most common ICD codes across age groups in the BCH Sleep Corpus

0–6 months			6–12 months		
ICD Code	**No. of PSGs**	**% PSGs**	**ICD Code**	**No. of PSGs**	**% PSGs**
R13.12: dysphagia, oropharyngeal phase	740	62.2	R13.12: dysphagia, oropharyngeal phase	367	57.1
R09.02: hypoxemia	660	55.5	R63.3: feeding difficulties	314	48.8
I36.1: nonrheumatic tricuspid (valve) insufficiency	610	51.3	G47.33: obstructive sleep apnea (adult) (pediatric)	311	48.4
I37.1: nonrheumatic pulmonary valve insufficiency	574	48.2	R09.02: hypoxemia	305	47.4
G47.33: obstructive sleep apnea (adult) (pediatric)	573	48.2	I36.1: nonrheumatic tricuspid (valve) insufficiency	299	46.5
Q21.1: atrial septal defect	522	43.9	K21.9: gastro-esophageal reflux disease without esophagitis	271	42.1
R63.3: feeding difficulties	519	43.6	Z71.3: dietary counseling and surveillance	271	42.1
Z20.822: contact with and (suspected) exposure to COVID-19	501	42.1	I37.1: nonrheumatic pulmonary valve insufficiency	252	39.2
P92.9: feeding problem of newborn, unspecified	484	40.7	R13.10: dysphagia, unspecified	246	38.3
K21.9: gastro-esophageal reflux disease without esophagitis	458	38.5	Z79.899: other long-term (current) drug therapy	243	37.8
Z71.3: dietary counseling and surveillance	454	38.2	Q21.1: atrial septal defect	235	36.5
R06.89: other abnormalities of breathing	410	34.5	R50.9: fever, unspecified	219	34.1
Z46.82: encounter for fitting and adjustment of non-vascular catheter	406	34.1	Z20.822: contact with and (suspected) exposure to COVID-19	219	34.1
R13.10: dysphagia, unspecified	398	33.4	R11.10: vomiting, unspecified	216	33.6
J98.11: atelectasis	374	31.4	J98.11: atelectasis	214	33.3
R91.8: other nonspecific abnormal finding of lung field	365	30.7	R91.8: other nonspecific abnormal finding of lung field	208	32.3
Z99.11: dependence on respirator (ventilator) status	365	30.7	K59.00: constipation, unspecified	207	32.2
R09.89: other specified symptoms and signs involving the circulatory and respiratory systems	355	29.8	R06.89: other abnormalities of breathing	205	31.9
R94.31: abnormal electrocardiogram (ECG) (EKG)	342	28.7	Z46.82: encounter for fitting and adjustment of non-vascular catheter	204	31.7
Z98.890: other specified postprocedural states	338	28.4	J98.4: other disorders of lung	197	30.6
1–6 years			**6–12 years**		
ICD code	**No. of PSGs**	**% PSGs**	**ICD code**	**No. of PSGs**	**% PSGs**
G47.33: obstructive sleep apnea (adult) (pediatric)	2664	49.1	G47.33: obstructive sleep apnea (adult) (pediatric)	1957	45.7
R06.83: snoring	2372	43.7	R06.83: snoring	1877	43.8
Z71.3: dietary counseling and surveillance	1636	30.2	Z71.3: dietary counseling and surveillance	1157	27.0
R63.3: feeding difficulties	1630	30.1	Z79.899: other long term (current) drug therapy	1039	24.3
R13.12: dysphagia, oropharyngeal phase	1551	28.6	K59.00: constipation, unspecified	990	23.1
Z01.10: encounter for examination of ears and hearing without abnormal findings	1490	27.5	G47.9: sleep disorder, unspecified	861	20.1
Z79.899: other long-term (current) drug therapy	1453	26.8	I36.1: nonrheumatic tricuspid (valve) insufficiency	854	19.9
R50.9: fever, unspecified	1438	26.5	Z23: encounter for immunization	847	19.8
K59.00: constipation, unspecified	1427	26.3	Z01.20: encounter for dental examination and cleaning without abnormal findings	843	19.7
Z20.822: contact with and (suspected) exposure to COVID-19	1416	26.1	Z20.822: contact with and (suspected) exposure to COVID-19	811	18.9
I36.1: nonrheumatic tricuspid (valve) insufficiency	1380	25.5	J45.909: unspecified asthma, uncomplicated	778	18.2
F80.9: developmental disorder of speech and language, unspecified	1377	25.4	Z68.54: body mass index (BMI) pediatric, greater than or equal to 95th percentile for age	744	17.4
H69.83: other specified disorders of eustachian tube, bilateral	1353	25.0	R05: cough	726	17.0
K21.9: gastro-esophageal reflux disease without esophagitis	1335	24.6	Z00.129: encounter for routine child health examination without abnormal findings	715	16.7
R05: cough	1333	24.6	Z20.828: contact with and (suspected) exposure to other viral communicable diseases	702	16.4
J35.3: hypertrophy of tonsils with hypertrophy of adenoids	1326	24.5	R63.3: feeding difficulties	679	15.9
R13.10: dysphagia, unspecified	1272	23.5	R62.50: unspecified lack of expected normal physiological development in childhood	678	15.8
Z98.890: other specified postprocedural states	1263	23.3	Z01.10: encounter for examination of ears and hearing without abnormal findings	678	15.8

(*Continued*)

Table 2. Continued

0–6 months			6–12 months		
ICD Code	**No. of PSGs**	**% PSGs**	**ICD Code**	**No. of PSGs**	**% PSGs**
R62.50: unspecified lack of expected normal physiological development in childhood	1233	22.7	J35.1: hypertrophy of tonsils	674	15.7
Z20.828: contact with and (suspected) exposure to other viral communicable diseases	1228	22.6	Z51.81: encounter for therapeutic drug level monitoring	674	15.7
12–18 years			**>18 years**		
ICD code	**No. of PSGs**	**% PSGs**	**ICD code**	**No. of PSGs**	**% PSGs**
G47.33: obstructive sleep apnea (adult) (pediatric)	1596	49.1	G47.33: obstructive sleep apnea (adult) (pediatric)	449	49.7
R06.83: snoring	1102	33.9	Z79.899: other long term (current) drug therapy	408	45.1
Z71.3: dietary counseling and surveillance	984	30.2	Z51.81: encounter for therapeutic drug level monitoring	299	33.1
Z79.899: other long term (current) drug therapy	942	29.0	Z71.3: dietary counseling and surveillance	290	32.1
F41.9: anxiety disorder, unspecified	820	25.2	I36.1: nonrheumatic tricuspid (valve) insufficiency	279	30.9
Z23: encounter for immunization	794	24.4	Z23: encounter for immunization	270	29.9
I36.1: nonrheumatic tricuspid (valve) insufficiency	724	22.3	R94.31: abnormal electrocardiogram (ECG) (EKG)	261	28.9
K59.00: constipation, unspecified	722	22.2	R06.83: snoring	256	28.3
Z68.54: body mass index (BMI) pediatric, greater than or equal to 95th percentile for age	703	21.6	K59.00: constipation, unspecified	245	27.1
E66.9: obesity, unspecified	678	20.8	F41.9: anxiety disorder, unspecified	239	26.4
Z51.81: encounter for therapeutic drug level monitoring	651	20.0	E55.9: vitamin D deficiency, unspecified	236	26.1
E55.9: vitamin D deficiency, unspecified	633	19.5	J98.4: other disorders of lung	233	25.8
Z20.822: contact with and (suspected) exposure to COVID-19	627	19.3	F79: unspecified intellectual disabilities	231	25.6
I49.9: cardiac arrhythmia, unspecified	608	18.7	K21.9: gastro-esophageal reflux disease without esophagitis	231	25.6
J45.909: unspecified asthma, uncomplicated	604	18.6	I49.9: cardiac arrhythmia, unspecified	224	24.8
Z00.129: encounter for routine child health examination without abnormal findings	577	17.7	I37.1: nonrheumatic pulmonary valve insufficiency	214	23.7
I37.1: nonrheumatic pulmonary valve insufficiency	563	17.3	Z20.822: contact with and (suspected) exposure to COVID-19	204	22.6
K21.9: gastro-esophageal reflux disease without esophagitis	557	17.1	R50.9: fever, unspecified	200	22.1
Z01.20: encounter for dental examination and cleaning without abnormal findings	555	17.1	Z20.828: contact with and (suspected) exposure to other viral communicable diseases	200	22.1
R10.9: unspecified abdominal pain	540	16.6	F88: other disorders of psychological development	197	21.8

This table displays the top 20 ICD codes for 11 167 patients (88.4% of the dataset). Linking the PSG data to medical diagnoses allows for investigating how various diseases are associated with disordered sleep. Oropharyngeal dysphagia is the most common diagnosis for infants undergoing polysomnography, affecting over half of these patients. For older age groups, obstructive sleep apnea is the leading diagnosis. Diagnoses without corresponding ICD codes have been omitted from this table but remain in the dataset preserving the full range of diagnostic data for further analysis.

mean, median, and 90th percentile values of 143.4, 113, and 275 events per PSG, respectively. These events are categorized into arousal, limb movement, and respiratory events. The respiratory event annotations covering central apnea (CA), mixed apnea (MA), OA, central hypopnea (CH), mixed hypopnea (MH), obstructive hypopnea (OH), and RERA. Additionally, some annotations for uncategorized apnea and hypopnea events are included. Apnea–hypopnea index values have been computed for each patient and are included in the Demographics metadata file. It is to be noted that at Boston Children's Hospital, all hypopneas were scored using the criterion of ≥30% reduction in airflow accompanied by either ≥3% oxygen desaturation and/or an arousal, in accordance with AASM guidelines. This standard was applied consistently across both pediatric and chronically ill adult patients included in the dataset. Users who wish to apply alternative definitions—such as the ≥4% desaturation-only threshold—may do so by reprocessing the raw PSG signals using the accompanying SpO_2 data available in both the EDF and HDF5 files. A breakdown of the total counts and per-PSG distributions of these event annotations across different age groups is provided in Table 5. To better understand the variation in event annotations with age, the trends are visualized in Figure 6. Key observations include:

- The frequency of CA events per PSG declines progressively with age, reflecting developmental changes in respiratory control mechanisms. This reduction is evident from the mean number of CA events per PSG, which decreases from 37.4 in early infants and 20 in late infants to just 4.5 in adults. CA occurs when there is a cessation or significant reduction in airflow due to dysregulation of the central respiratory drive. In infants, this phenomenon is linked to the immaturity of the brainstem's control over breathing, resulting in a higher occurrence of CA events. The elevated number of CA events in infants reflects their immature and developing respiratory control systems, which are prone to respiratory instability. As children grow, the maturation of the brainstem and the respiratory regulatory mechanisms lead to a decrease in CA frequency and periodic breathing, contributing to fewer

Table 3. Frequency and availability of channels across the BCH Sleep Study dataset

Channel	Percentage	Channel	Percentage	Channel	Percentage
EEG channels		**EEG channels**		**EMG channels**	
C3	100.00	T4	53.56	CHINz	46.44
C4	100.00	T5	53.36	CHIN1	100.00
Cz	99.96	T6	53.36	CHIN2	100.00
O1	100.00	T7	46.44	RLEG+	46.44
O2	100.00	T8	46.44	RLEG-	8.26
Oz	46.16	A1/M1	100.00	LLEG+	46.44
F3	100.00	A2/M2	100.00	LLEG-	8.26
F4	100.00	**Respiratory channels**		LAT1	53.55
F7	99.48	FLOW/DIF4+	100.00	LAT2	53.55
F8	99.48	CHEST/Chest/DIF1+	100.00	RAT1	53.55
Fz	99.52	ABD/Abdomen/DIF2+	100.00	RAT2	53.55
Fp1	99.48	SNORE/DIF5+	99.49	**EOG channels**	
Fp2	99.48	SpO2 (named DC7)	100.00	LOC/E1	100.00
Fpz	99.68	PWAVE (named DC8)	100.00	ROC/E2	100.00
P3	99.52	EtCO2 (named DC5)	100.00	**ECG channels**	
P4	99.52	TcCO2 (named DC3)	99.49	ECG-LA/ECGL	100.00
Pz	99.52	Pressure/DIF2+/DIF6	100.00	ECG-RA/ECGR	100.00
P7	46.16	CAPNO (named DC9)	100.00		
P8	46.16	CFLOW (named DC4)	99.49		
T3	53.56	CPRES (named DC10)	99.32		

This table provides a summary of the channels available in the BCH Sleep Study dataset, detailing the frequency with which each channel appears across the 15 695 polysomnography (PSG) recordings. It highlights that central EEG channels (C3, C4, Cz), occipital channels (O1, O2), and frontal channels (F3, F4) are consistently present in all recordings, along with respiratory effort, SpO_2, EOG, and chin EMG channels. Other channels, such as leg EMG and airflow measurements, are included in the majority of recordings, offering a comprehensive view of sleep physiology for each patient. This detailed distribution of channels enhances the dataset's versatility for diverse sleep-related analyses and research.

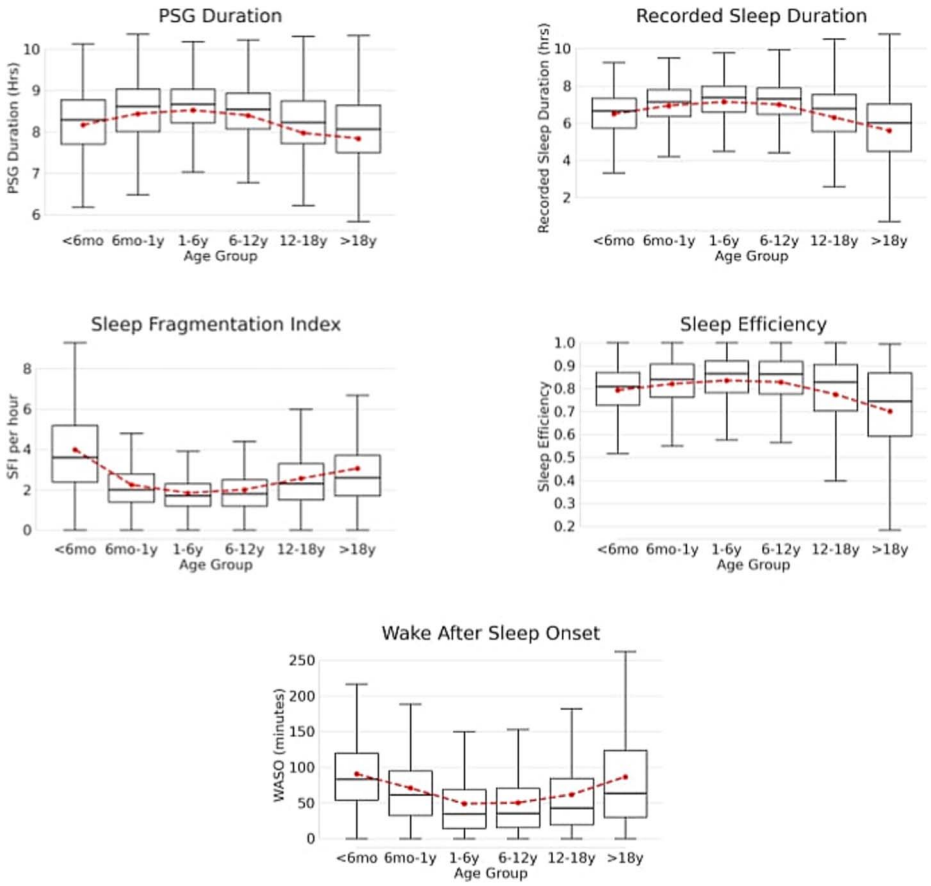


Figure 4. Boxplot of sleep study duration and sleep metrics across age groups. This figure presents a comparative boxplot showing the distribution of key sleep metrics from the BCH sleep study dataset, including PSG duration, sleep duration, SFI, sleep efficiency, and WASO across different age groups. The plots highlight age-related trends in sleep quality and fragmentation, with notable variations in sleep duration and efficiency between age groups. These visualizations provide insights into developmental differences in sleep patterns, offering a valuable resource for understanding pediatric sleep health.

Table 4. Distribution of sleep stage annotations across different age groups

	0–6 months		6–12 months		1–6 years	
Event	**Total**	**Per PSG**	**Total**	**Per PSG**	**Total**	**Per PSG**
N1	59 149	49.7 ± 62.7	41 186	64.1 ± 64.2	294 742	54.4 ± 56.8
N2	194 326	163.3 ± 198.2	153 396	238.6 ± 128.0	1 529 005	282.0 ± 124.0
N3	46 738	39.3 ± 97.7	172 079	267.6 ± 132.5	1 715 821	316.5 ± 133.0
N4	251 034	211.0 ± 209.7	10 251	15.9 ± 77.5	60 234	11.1 ± 54.9
REM	375 569	315.6 ± 99.9	158 250	246.1 ± 88.5	1 050 048	193.7 ± 74.5
WAKE	239 517	201.3 ± 106.1	115 989	180.4 ± 115.9	897 118	165.5 ± 114.5
UNSCORED	55 488	46.6 ± 75.4	28 406	44.2 ± 72.4	270 139	49.8 ± 77.5
	6–12 years		**12–18 years**		**>18 years**	
Event	**Total**	**Per PSG**	**Total**	**Per PSG**	**Total**	**Per PSG**
N1	190 884	44.6 ± 43.2	173 286	53.3 ± 47.6	58 336	64.5 ± 70.3
N2	1 433 426	334.7 ± 127.8	1 263 078	388.3 ± 158.4	339 094	375.1 ± 175.2
N3	1 230 264	287.2 ± 125.8	551 560	169.6 ± 102.0	100 875	111.6 ± 89.1
N4	45 876	10.7 ± 51.7	19 560	6.0 ± 31.7	3163	3.5 ± 22.4
REM	697 797	162.9 ± 70.4	454 828	139.8 ± 83.9	105 875	117.1 ± 82.9
WAKE	717 770	167.6 ± 120.0	649 202	199.6 ± 152.0	243 450	269.3 ± 193.9
UNSCORED	294 305	68.7 ± 266.5	327 611	100.7 ± 195.8	86 221	95.4 ± 164.1

This table provides a detailed breakdown of sleep stage annotations from the BCH Sleep Study dataset, showing the frequency of each sleep stage (WAKE, N1, N2, N3, N4, REM, and UNSCORED) across various age groups. The data highlights significant developmental trends in sleep architecture, with a gradual decrease in REM sleep and an increase in N2 and N3 sleep stages as age progresses. This comprehensive summary of sleep stages across age demographics enables a deeper understanding of age-related changes in sleep patterns.

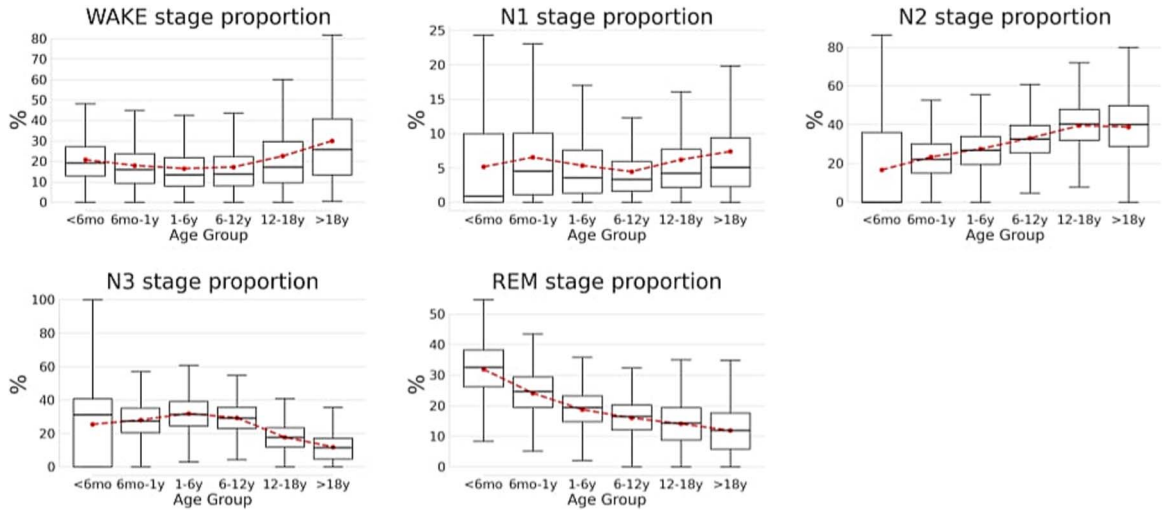


Figure 5. Percentage distribution of sleep stages across age groups. This figure illustrates the variation in the proportion of different sleep stages (WAKE, N1, N2, N3, REM) as a percentage of total sleep time across different age groups. Key trends include a marked decline in the percentage of REM sleep with age, offset by an increase in N2 and N3 stages, reflecting the maturation of sleep cycles. The figure also reveals increased wakefulness in teenagers and adults, providing insight into how sleep patterns evolve from infancy to adulthood.

events in older children and adults [43, 44]. CH events follow a similar age-related decline, mirroring the trend observed with CA events. However, a small increase in the number of CH events is noted in adults, which may be attributed to the presence of underlying medical or genetic/neuromuscular conditions in our adult population who are followed since childhood at BCH and are a part of this dataset.

- The analysis of OA events in the pediatric population presents noteworthy patterns in breathing behavior as children develop. Infants, with their immature respiratory systems, tend to experience a higher frequency of OA events per PSG study. This is primarily due to their underdeveloped airway control, upper airway muscle tone, and laryngomalacia which often results in breathing difficulties during sleep. As children grow older, the number of OA events decreases, reflecting improvement in upper airway muscle tone, laryngomalacia, and maturation in breathing pattern resulting in stabilization of their breathing patterns [45–47]. However, an increase in OA events is observed post-adolescence, commonly linked to obesity. The increased body mass can lead to airway obstructions during sleep, causing a resurgence in the frequency of OA events. The occurrence of OH events, however, follows a different trend. Younger children exhibit fewer OH events, as apnea is more common in infancy due to immature breathing mechanisms. Furthermore, hypopnea is more difficult to score in infants because of their rapid breathing rate. It becomes more noticeable as the child matures and their breathing slows down and stabilizes. This

Table 5. Distribution of event annotations in the BCH Sleep Corpus

	0–6 months		6–12 months		1–6 years	
Event	**Total count**	**Per PSG**	**Total count**	**Per PSG**	**Total count**	**Per PSG**
Central apnea	44 491	37.4 ± 64.2	12 861	20.0 ± 29.7	51 508	9.5 ± 28.3
Mixed apnea	1384	1.2 ± 5.1	293	0.5 ± 1.5	2980	0.5 ± 3.8
Obstructive apnea	21 735	18.3 ± 42.4	7139	11.1 ± 31.6	49 604	9.1 ± 29.9
Uncategorized apnea	13	0.0 ± 0.1	6	0.0 ± 0.1	89	0.0 ± 0.2
Central hypopnea	5984	5.0 ± 16.1	1220	1.9 ± 7.8	2248	0.4 ± 3.7
Mixed hypopnea	28	0.0 ± 0.2	11	0.0 ± 0.2	212	0.0 ± 0.6
Obstructive hypopnea	12 224	10.3 ± 26.5	7332	11.4 ± 34.7	69 069	12.7 ± 35.4
Uncategorized hypopnea	77	0.1 ± 0.6	116	0.2 ± 1.4	665	0.1 ± 1.0
RERA	555	0.5 ± 2.6	646	1.0 ± 3.7	16 742	3.1 ± 8.2
Arousal	121 901	102.4 ± 40.6	50 324	78.3 ± 44.5	412 829	76.1 ± 47.9
Limb movement	3368	2.8 ± 22.5	2029	3.2 ± 17.0	194 907	35.9 ± 60.1
	6–12 years		**12–18 years**		**>18 years**	
Event	**Total count**	**Per PSG**	**Total count**	**Per PSG**	**Total count**	**Per PSG**
Central apnea	28 590	6.7 ± 18.1	15 258	4.7 ± 13.7	4128	4.6 ± 17.2
Mixed apnea	2698	0.6 ± 8.1	1776	0.5 ± 4.5	868	1.0 ± 9.4
Obstructive apnea	21 841	5.1 ± 22.9	27 091	8.3 ± 34.3	9325	10.3 ± 38.6
Uncategorized apnea	36	0.0 ± 0.1	57	0.0 ± 0.3	7	0.0 ± 0.1
Central hypopnea	1766	0.4 ± 4.9	1576	0.5 ± 5.4	1448	1.6 ± 14.8
Mixed hypopnea	206	0.0 ± 0.8	116	0.0 ± 0.5	50	0.1 ± 0.6
Obstructive hypopnea	48 860	11.4 ± 32.7	56 657	17.4 ± 41.6	18 889	20.9 ± 40.3
Uncategorized hypopnea	630	0.1 ± 1.7	608	0.2 ± 1.5	178	0.2 ± 1.2
RERA	16 096	3.8 ± 7.8	18 151	5.6 ± 11.2	4565	5.0 ± 11.3
Arousal	294 113	68.7 ± 43.0	231 191	71.1 ± 54.9	62 912	69.6 ± 64.2
Limb movement	145 021	33.9 ± 53.3	112 042	34.4 ± 67.4	29 343	32.5 ± 73.6

This table provides a detailed breakdown of the total counts and per-PSG distributions of event annotations, including arousal, limb movement, and respiratory events, across different age groups. Key patterns include a higher prevalence of CA events in infants, a marked increase in obstructive apnea and respiratory effort related arousal (RERA) events with age, and a significant number of arousals in younger children. The table highlights the variation in sleep-related events across the developmental spectrum.

shift allows for more accurate identification of hypopnea events as the child progresses into later childhood and adolescence.

- The BCH Sleep Study dataset includes event annotations for RERA events, with a total of 56 755 such annotations. Analysis of the occurrence of RERA per PSG shows that these events increase in frequency with age. Scoring RERA events in infants under one year of age is challenging due to the immature respiratory mechanics and frequent variability in breathing patterns. In early infancy, flow limitation is often less clearly defined and arousals may be subtler or more difficult to associate with increased respiratory effort. As children grow older and their upper airway physiology and respiratory patterns mature, the characteristics required for identifying RERAs—such as discernible flow limitation and effort-related arousals—become more apparent, leading to an increase in scored RERA events with age [48].
- The dataset contains 1 173 270 annotated arousal events. The data show that younger patients tend to experience more arousals. As the brain matures with age, sleep becomes more consolidated, leading to a reduction in the frequency of arousals [49].
- The dataset contains a total of 486 710 annotated limb movements. This is particularly important for identifying disorders such as periodic limb movement disorder (PLMD) [50, 51]. It is to be noted, however, that the number of limb movements recorded per PSG is relatively low for patients under one year of age. This is because, as per guidelines at Boston Children's Hospital, limb movements are not typically scored for infants below one year. Beyond this age group, the average number of limb movements per PSG increases to about 30 per recording across all other age ranges. This makes the dataset particularly useful for studying limb movements in older children and adolescents, and can further the development of tools for early detection and better understanding of movement-related sleep disorders in the pediatric population.

HDF5 derivative files

To support efficient computational access and large-scale analysis, we provide a parallel collection of derivative files in the Hierarchical Data Format version 5 (HDF5). Each of these files corresponds to a single PSG session and contains both multichannel physiological signal data and synchronized per-sample annotations, uniformly resampled to 200 Hz. These files serve as a standardized, compressed alternative to the raw EDF recordings, optimized for integration into machine learning pipelines and signal processing workflows.

The structure of each HDF5 file comprises two main components: a signals group containing the physiological time series, and an annotations group containing aligned categorical and binary labels. Metadata attributes stored at the file level include the sampling rate (fixed at 200 Hz), unit of measurement for voltage-based channels (microvolts), and the age of the patient in days at the time of the recording. The signals group includes a curated set of 24 channels spanning EEG, EOG, EMG, ECG, respiratory, oxygenation, and auxiliary modalities. EEG is represented using six bipolar montages—C3-M2, C4-M1, O1-M2, O2-M1, F3-M2, and F4-M1—chosen to reflect common clinical derivations. Ocular activity is recorded using E1-M2 and E2-M1, while

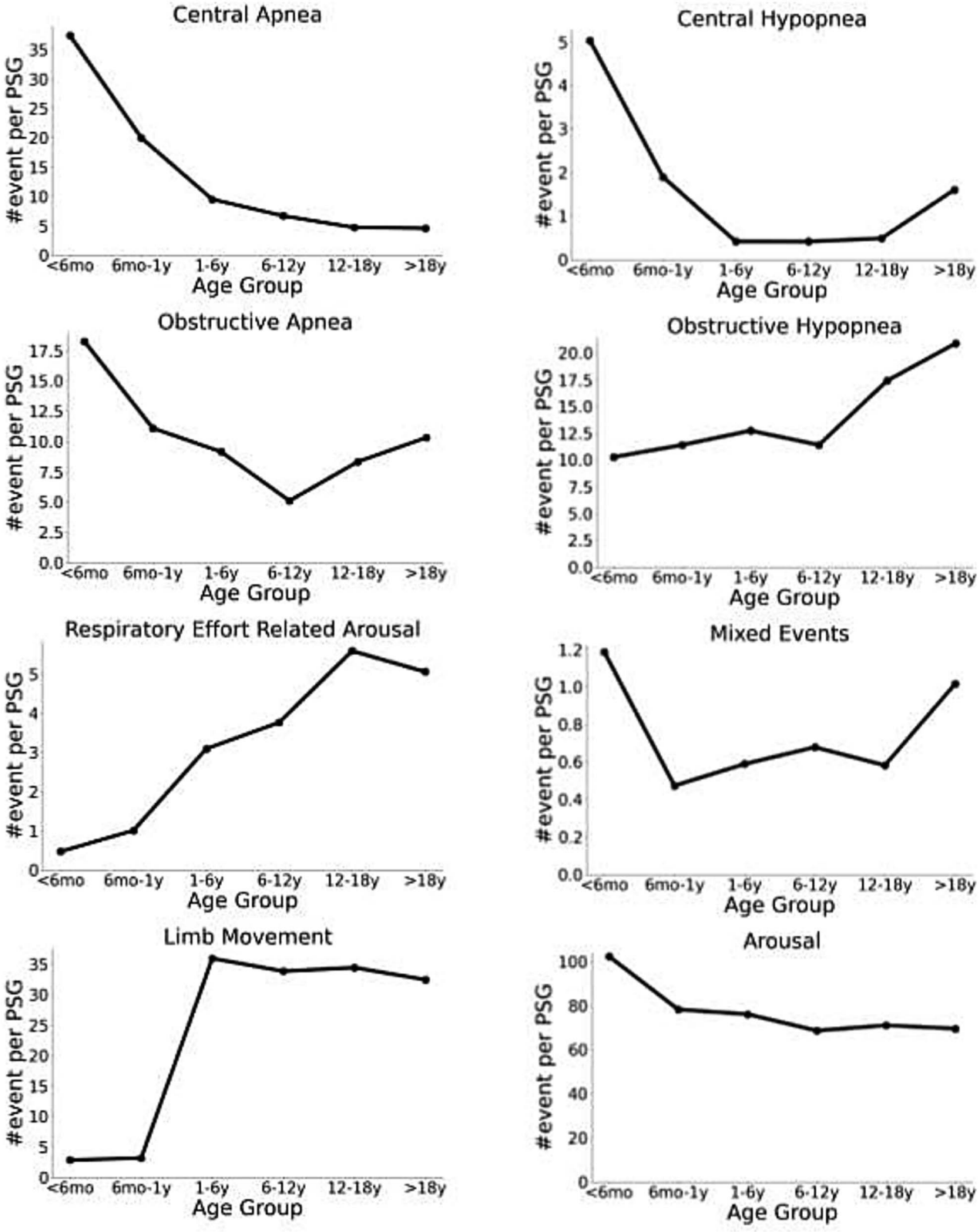


Figure 6. Event annotation distribution per PSG across age groups. This figure visualizes the distribution of various event annotations per PSG (including arousal, limb movement, and respiratory events) across different age groups in the BCH sleep study dataset. Notable trends include a decrease in central apnea (CA) and central hypopnea (CH) events with age, reflecting the maturation of respiratory control mechanisms, while obstructive apnea (OA) and obstructive hypopnea (OH) events show distinct age-dependent patterns. Respiratory effort-related arousal (RERA) events increase with age, aligning with changes in breathing patterns during sleep. The distribution of arousal events and limb movements also reflects developmental changes, offering valuable insights into sleep maturation and related disorders.

chin muscle activity is captured via the CHIN1–CHIN2 derivation. ECG is provided in a bipolar format as ECGL–ECGR. Thoracic and abdominal respiratory effort are represented by CHEST and ABD channels, respectively, while airflow and pressure-related channels include AIRFLOW, Pressure, PLETH, SNORE, CFLOW, and CPRES. Oxygenation and capnography are captured through the SpO_2 (oxygen saturation), $EtCO_2$ (end-tidal carbon dioxide), and $TcCO_2$ (transcutaneous carbon dioxide) channels. In addition, leg electromyography is represented using bipolar derivations for each leg: LAT1–LAT2 (left leg, stored as lat) and RAT1–RAT2 (right leg, stored as rat). All channels are rescaled and stored as single-channel arrays using 32-bit floating point precision with gzip compression.

Aligned with these signals, the annotations group stores four sample-level time series. The first is stage, which encodes the sleep technologist-scored sleep stage at each time point using integer labels: 0 for N1, 1 for N2, 2 for N3, 3 for REM, 4 for wake, and 9 for unknown or unscored segments. These labels are derived by expanding the 30-s epoch annotations to match the 200 Hz temporal resolution. The second key, arousal, is a binary mask that denotes the presence (1) or absence (0) of arousals at each sample, based on sleep technologist annotations. The third, resp, contains integer-coded respiratory event types, including CA (1), unspecified apnea (2), CH (3), unspecified hypopnea (4), MH (5), OH (6), MA (7), OA (8), and RERA (9), with 0 indicating no respiratory event. Lastly, limb provides a binary mask indicating the presence or absence of limb movement events.

All annotations are aligned in time with the physiological signals, enabling direct indexing for supervised learning or temporal analysis. This derivative format offers several advantages over the raw EDF files, including reduced storage footprint, consistent signal sampling and naming, and sample-level access to multimodal annotations. The .h5 files are fully compatible with popular scientific computing libraries, making them particularly well-suited for training neural networks and building automated analysis tools. By providing this harmonized representation of the data, we aim to lower technical barriers for research and enable scalable, reproducible analyses of pediatric sleep physiology.

Technical Validation

Validation of EDF files and annotations

All EDF files included in the BCH Sleep Corpus underwent a thorough error-checking process before inclusion in the dataset. To ensure the data integrity, we read all the files into Python using the MNE library [39]. Files that encountered any errors during this process were excluded. In addition to this automated check, a visual inspection of several physiological signals and their corresponding sleep stage or event annotations was conducted to verify correctness.

Figure 7 presents a representative plot illustrating how sleep stage annotations were visually inspected for quality verification. This figure shows a hypnogram plotted for a full night of sleep, accompanied by spectrograms for frontal, central, and occipital EEG channels. The spectrograms were computed using the multitaper method with a 2-s window, 1-s step size (50% overlap), a time-half bandwidth product of 2, and a frequency range of 0–20 Hz. Spectral power was converted to decibels (dB), and adaptive weighting was enabled to improve spectral resolution. The process for creating these plots involved first computing individual spectrograms from pairs of channels (F3, F4 for frontal, C3, C4 for central, and O1, O2 for occipital, each referenced to contralateral mastoids), which were then averaged and displayed alongside the hypnogram. The plot also includes markers for arousal events, allowing for a detailed examination of sleep progression throughout the night. One notable observation from this visual inspection is the high prevalence of alpha rhythm during the wake stage, which is particularly pronounced in the occipital region—a finding consistent with established literature [52]. Similarly, the presence of slow delta activity during REM sleep and spindles in deep non-REM sleep, especially visible in the central electrodes, further validates the accuracy of the EEG data and sleep staging annotations [53–55].

To complement the sleep data validation, we created additional visualizations to examine physiological changes during breathing events. Several 40-min-long segments were plotted, featuring signals from the abdomen and chest belts, SpO_2, and airflow and pressure transducer data. Figure 8 presents a sample plot, where different respiratory events are highlighted using distinct colors for easier visualization. Arousals are also overlaid on the plot, providing a view of the physiological responses to various breathing abnormalities, such as CA and hypopnea. The characteristic patterns associated with these events can be clearly inferred from the plot.

Lastly, another key validation involved examining limb movement annotations alongside EMG signals. Figure 9 shows the left and right EMG signals plotted with limb movement annotations from the event annotation file. The plot reveals a clear variation in the EMG signals during periods of limb movement, illustrating the accuracy of the limb movement annotations.

Age-associated spectral norms

To illustrate the utility of the dataset for studying EEG-age relationships during sleep, we examined the spectral characteristics of EEG signals across various sleep stages across the pediatric age range. We employed a parametric modeling approach that leverages a neural network to fit a model of the mean spectral characteristics as a function of age [56]. The first step involved calculating the power spectral density (PSD) for each PSG recording, focusing on 30-s epochs corresponding to the five standard sleep stages: Wake, N1, N2, N3, and REM. We used EEG data from multiple regions of the brain, specifically the frontal (F3, F4), central (C3, C4), and occipital (O1, O2) channels. We applied the Welch periodogram method with a 2-s window to compute the power spectrum. All EEG recordings were downsampled to 200 Hz to standardize the frequency resolution before power spectrum computation. To focus on the most relevant spectral information, we restricted the analysis to the frequency range of 0–20 Hz, which encompasses key frequencies of interest. This step resulted in a 40-dimensional spectral representation for each 30-s window of EEG data. For each PSG recording, we computed the mean of this 40-dimensional vector separately for each electrode and sleep stage. As a final step, we averaged the power spectrum computed from electrodes of the same brain region, providing a robust summary of the spectral characteristics across different brain regions and sleep stages. This process yielded a unique spectral profile for each combination of sleep stage (Wake, N1, N2, N3, REM) and brain region (frontal, central, occipital).

To model the age-related trends in spectral characteristics, we developed a simple neural network regression model that takes as input the terms (age,age^2) to account for potential nonlinear effects of age on the power spectrum. We designed the network with two hidden layers, each comprising 50 neurons, and an output layer with 40 neurons corresponding to the 40-dimensional power spectrum representation. All layers employ the Exponential Linear Unit activation function. The model is individually trained for each of the five sleep stages (WAKE, N1, N2, N3, REM) and for the three brain regions (frontal, central, occipital). The

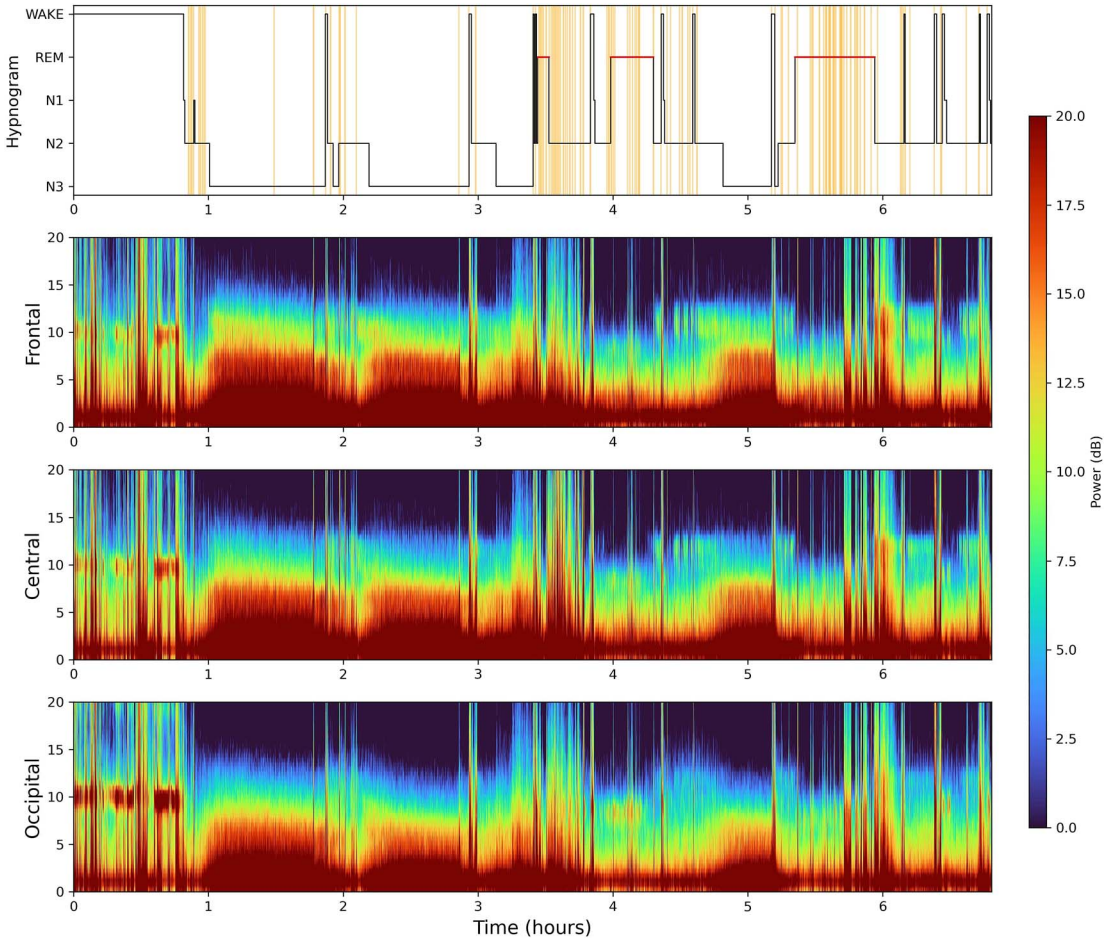


Figure 7. Hypnogram and EEG spectrograms for a full night of sleep. This figure presents a comprehensive validation of sleep staging using a hypnogram overlaid with EEG spectrograms from average frontal (F3-M2, F4-M1), central (C3-M2, C4-M1), and occipital (O1-M2, O2-M1) channels. The spectrograms capture distinct neural activity across sleep stages, showing pronounced alpha activity during WAKE, delta power during deep sleep (N3), and spindle activity in N2 sleep. Arousal events are marked, providing further context for sleep dynamics throughout the night. This plot serves as visual validation, confirming the accuracy of both the EEG data and sleep stage annotations in the BCH sleep corpus.

training/fitting process optimizes the mean squared error loss function using the Adam optimizer. The model is trained for 10 000 epochs to ensure convergence, and we observed that the loss function stabilized as the training progressed, indicating that the network effectively learned the underlying patterns in the dataset. After training, we fed continuous age values ranging from 0 to 20 years into the model to capture fine-grained variations in the spectral patterns across the pediatric age spectrum. This approach enabled us to generate plots illustrating how the spectral characteristics evolve with age for different sleep stages and brain regions, as shown in Figure 10. To facilitate comparison, we also computed and plotted the ground truth averaged power spectrum by averaging across all age groups, as presented in Figure 11. This comparison provides a clear visual representation of the relationship between age and spectral dynamics during sleep across different brain regions.

Age-associated sleep-stage norms

To analyze variation in sleep stages across different age groups, particularly during the developmental phase, we modeled the percentage distribution of sleep stages using a parametric approach. First, for each PSG recording, the corresponding sleep annotations were processed to calculate the percentage of time spent in each of the five sleep stages: WAKE, N1, N2, N3, and REM. Note that for this analysis, annotations for stage N4 were combined with N3, resulting in a five-dimensional vector for each recording. These vectors, with elements summing to 1, quantify the proportions of time spent in each of the five sleep stages during a given recording. Next, we modeled this five-dimensional sleep stage vector as a function of age, using (age,age^2) as input. The network consisted of a single layer with an input dimension of 2 (corresponding to the linear and quadratic age terms) and an output dimension of 5 (corresponding to the sleep stage probabilities). To ensure that the outputs formed valid proportions, the softmax activation function was applied to the output layer, constraining the output values to sum to 1. We employed the categorical cross-entropy loss function and optimized this loss using the Adam optimizer. We trained the model for 10 000 epochs, and visual inspection of the loss curve indicated that the loss converged to a minimum, suggesting that the model had successfully learned the relationship between age and sleep stage distributions.

Once the model was trained, we fed it continuous age values within the range of 0–20 years, covering the developmental phase of interest. Using a continuous age range allows us to capture fine variations in sleep stages that might be missed in analyses that group children into broad age categories. The model output for these age values was used to plot the variation in sleep stage proportions across this age range. Figure 12 presents this sleep stage variation, with each sleep stage color-coded for better visualization: WAKE is represented in Gold, N1 in Light Blue, N2 in Blue, N3 in Dark Blue, and REM in Purple. This visualization

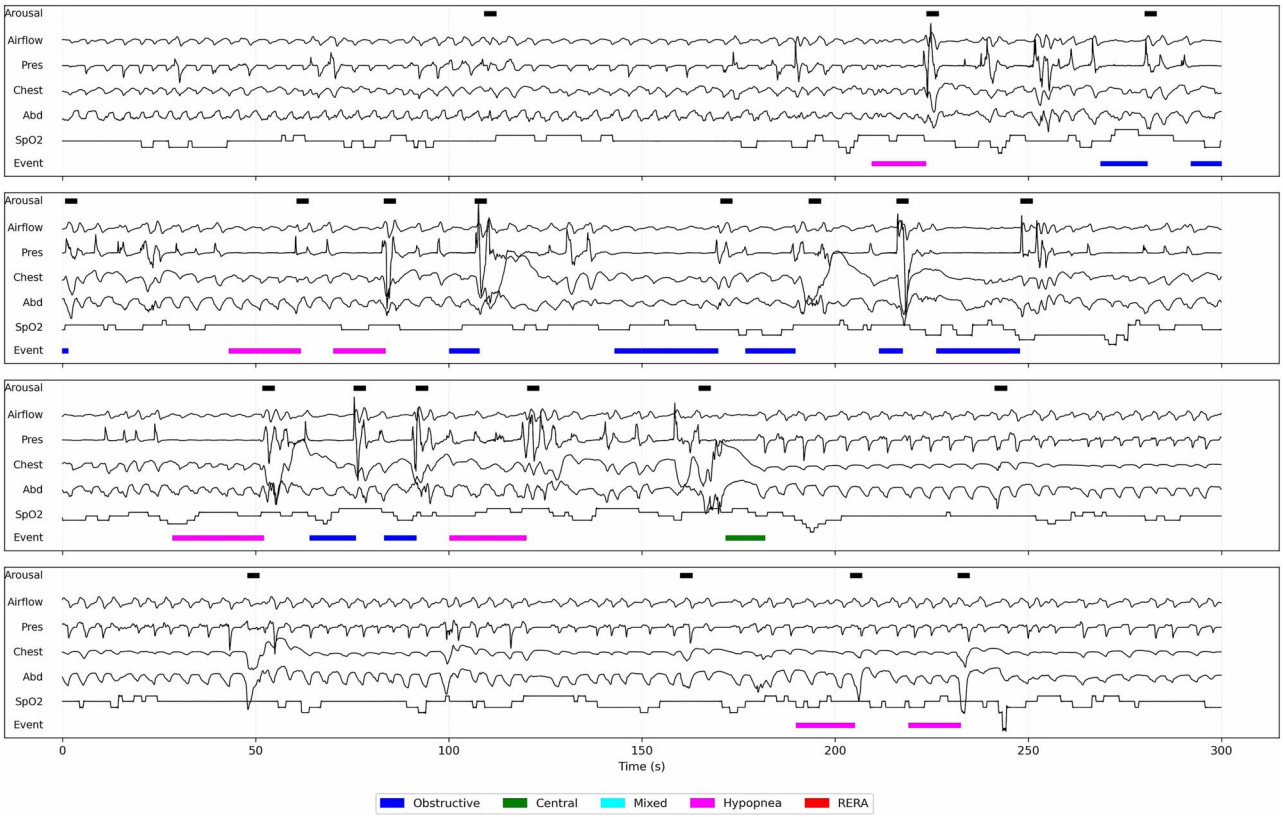


Figure 8. Physiological signals during respiratory events. This figure illustrates a representative 20-minute segment of physiological data, showing signals from the abdomen and chest belts, SpO_2, and airflow/pressure sensors. Each row in the plot corresponds to 5-min of data. Various respiratory events, such as central apnea and hypopnea, are highlighted in different colors, with arousal events overlaid. Characteristic physiological patterns, such as changes in airflow, are clearly visible during these events, verifying the accuracy of the respiratory event annotations. This plot provides visual validation of the respiratory event data in the dataset.

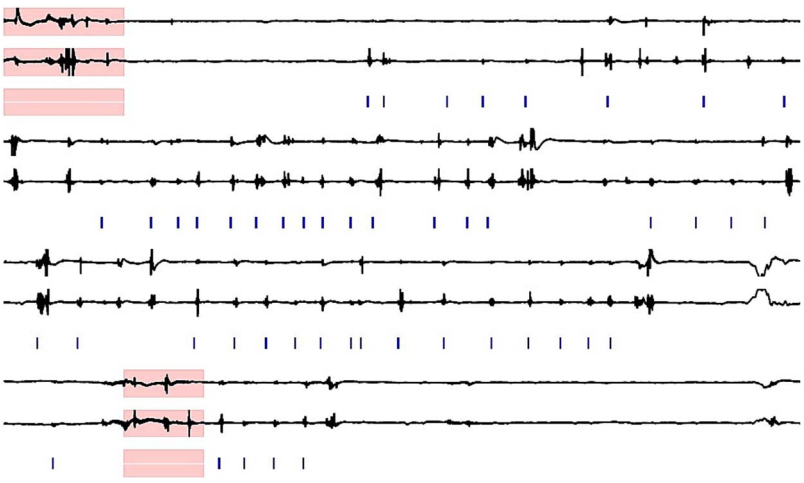


Figure 9. EMG signals and limb movement annotations. This figure shows right and left EMG signals alongside limb movement annotations from the event file. Each row corresponds to 10 minutes and the regions during which patient was in sleep stage WAKE are highlighted. The variation in EMG activity during annotated limb movements is evident, highlighting the reliability of the limb movement annotations. This visualization plays a key role in verifying that limb movement events are correctly captured and aligned with the corresponding physiological signals, further demonstrating the integrity of the BCH sleep corpus.

provides insights into how sleep evolves during the critical years of pediatric development.

Age-associated respiratory-event norms

We next explore the utility of the BCH Sleep Corpus for studying age trends in respiratory events, including CA, CH, MA and hypopnea, OA, OH, and RERA. Since MA and MHs events are less frequent compared to other events, they were grouped together into a single mixed events category, resulting in a six-dimensional target vector for the analysis. To model the trends, we applied a parametric approach similar to what was used in the sleep stage analysis. The six-dimensional vector representing the percentage of time spent in each of the six respiratory events was modeled as

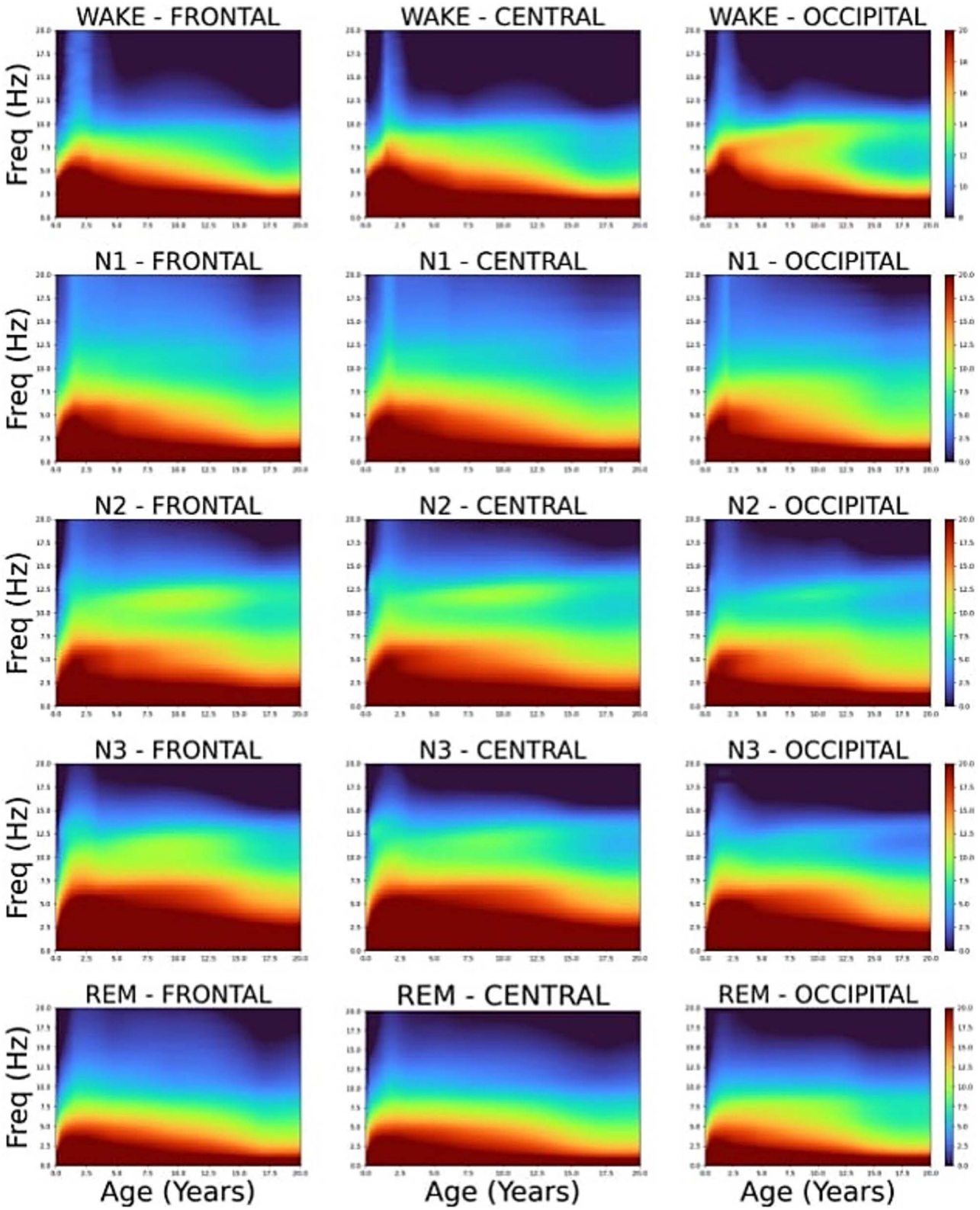


Figure 10. Age-related trends in EEG spectral power across sleep stages and brain regions. This figure presents the modeled spectral power distribution across various sleep stages—WAKE, N1, N2, N3, and REM—over the pediatric age range, focusing on three brain regions: frontal, central, and occipital. Each subplot shows how power in the 0–20 Hz frequency range evolves with age. Notable trends include an increase in low-frequency delta power, particularly during N3 sleep, as children grow older. Alpha activity in the occipital region is observed during the WAKE stage, and spindle activity becomes more prominent with age during N2 sleep. These trends reflect the maturation of sleep-related neural processes during childhood and adolescence.

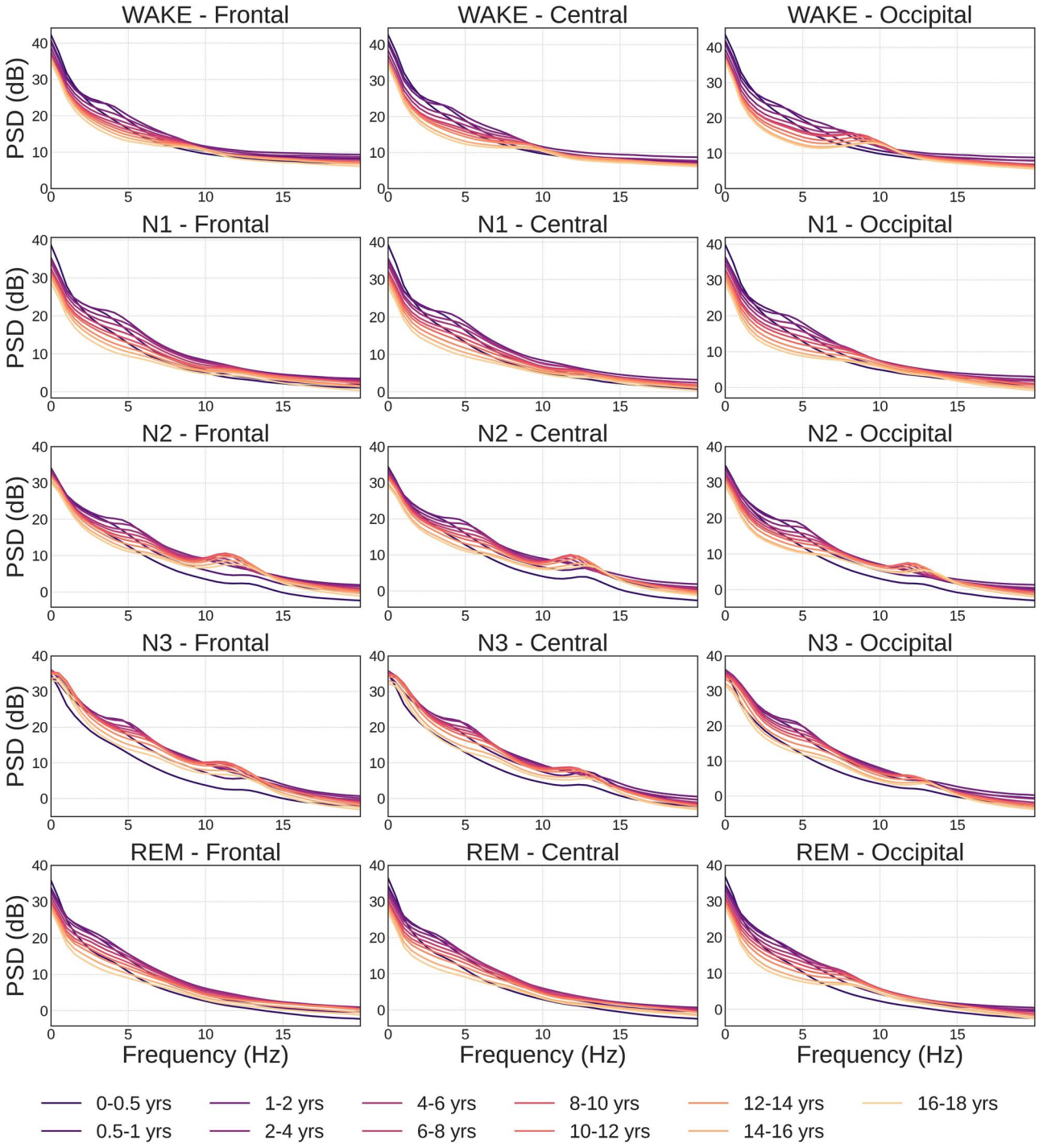


Figure 11. Averaged EEG power spectrum across sleep stages and brain regions. This figure shows the ground truth averaged PSD for the five sleep stages—WAKE, N1, N2, N3, and REM—across three brain regions: frontal, central, and occipital. The PSD is averaged across all age groups to provide a baseline comparison for the age-related trends. Each sleep stage exhibits characteristic spectral features, such as alpha activity during WAKE, spindle activity during N2, and dominant delta power during N3. This figure serves as a reference for understanding how spectral power varies across different brain regions and sleep stages in pediatric populations.

a function of age using (age,age^2) as the input variables. The neural network consisted of a single input layer with two nodes and a single output layer with six nodes. The model parameters were learned by minimizing the categorical cross-entropy loss function and optimized using the Adam optimizer. The model was trained on the dataset, and after training, continuous age values ranging from 0 to 20 years were fed into the model to capture the fine-grained variations in respiratory event patterns across this critical developmental period. This allowed us to generate a detailed plot illustrating the changing trends of respiratory events over time. The resulting respiratory event pattern, plotted against age, is shown in Figure 13. Each respiratory event is color-coded: light green for CA, dark green for CH, purple for mixed events, light blue for OA, dark blue for OH, and red for RERA. The plot offers valuable insights into the occurrence of different respiratory events across the pediatric population.

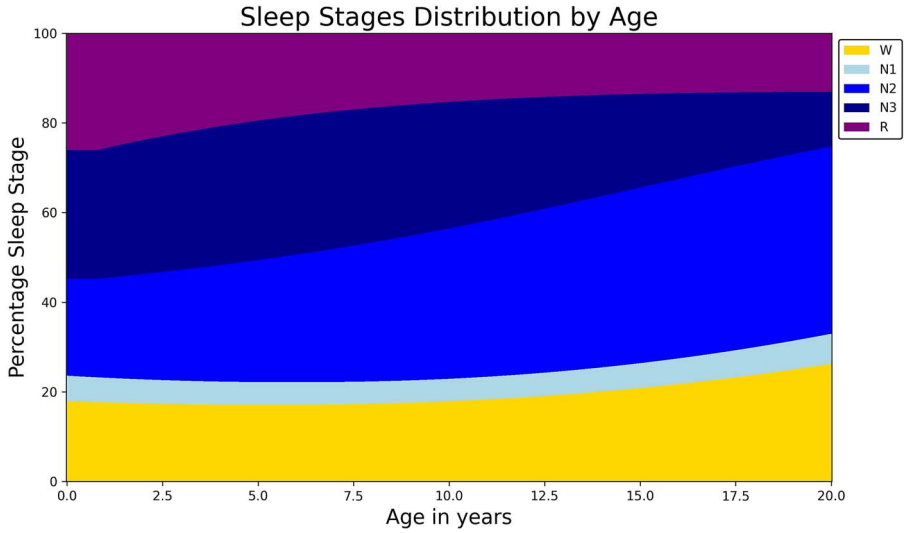


Figure 12. Age-dependent trends in sleep stage distribution. This figure depicts the modeled percentage distribution of sleep stages—WAKE (gold), N1 (light blue), N2 (blue), N3 (dark blue), and REM (purple)—as a function of age from infancy to 20 years. The plot highlights key developmental trends in sleep patterns, including a decrease in REM sleep with age, an increase in N2 sleep, and a gradual decline in N3 sleep. WAKE percentage is highest in infants, decreases in early childhood, but rises again during adolescence. This visualization provides insight into how sleep architecture evolves during childhood and adolescence.

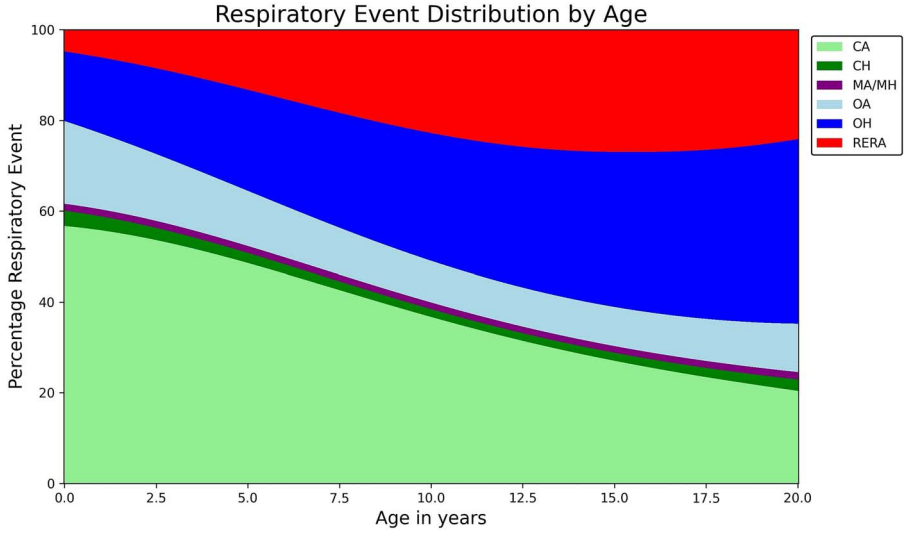


Figure 13. Age-related trends in respiratory event occurrence. This figure shows the modeled age-dependent changes in the percentage of time spent in each of six respiratory events relative to the other respiratory events: central apnea (CA), central hypopnea (CH), mixed apnea and hypopnea (MA/MH), obstructive apnea (OA), obstructive hypopnea (OH), and respiratory effort-related arousal (RERA). The plot highlights a decrease in CA with age, higher OA in infancy, a rise in OH during adolescence, and increasing RERA as respiratory patterns stabilize with age. These trends provide insight into changing vulnerabilities in respiratory physiology during childhood and adolescence.

Discussion

The BCH Sleep Corpus is a comprehensive publicly available pediatric PSG dataset that offers an expansive resource for both clinical and computational sleep research. Its scale, richness of annotations, and demographic diversity position it as a foundational dataset for advancing our understanding of pediatric sleep physiology and for enabling the development of age-aware, machine learning-based diagnostic tools. One of the most immediate applications of this dataset is the automation of sleep staging. Manual scoring of sleep stages remains the clinical standard but is time-consuming, subjective, and resource-intensive. While adult-focused sleep staging algorithms have seen success [57, 58], pediatric sleep staging poses unique challenges due to the dynamic nature of sleep architecture across developmental stages. The BCH Sleep Corpus spans from neonates to young adults, enabling the creation and validation of algorithms that generalize across age groups and support transfer learning techniques where annotated data are sparse—particularly in early infancy.

A notable observation in the dataset is the high percentage of REM sleep in infancy, which progressively declines with age. This pattern is well-established in the literature and reflects the unique role of REM sleep in early neurodevelopment, including synaptic plasticity and brain maturation during the first years of life. It has been shown in literature [59, 60] that early childhood is marked by high cortical activity during REM, which diminishes as sleep–wake rhythms stabilize and cortical structures mature. As REM sleep decreases, there is a corresponding increase in

N2 sleep, which becomes the dominant stage in later childhood and adolescence. This trend is likely related to the development and consolidation of sleep spindles—transient bursts of sigma activity that emerge with age and are prominent markers of thalamocortical maturation and cognitive development. In parallel, N3 (slow-wave) sleep also shows a gradual decline across age, consistent with prior EEG studies indicating a reduction in delta power and changes in cortical synchronization as adolescence progresses [61, 62]. These developmental changes in NREM sleep reflect the broader trajectory of cortical pruning and reorganization occurring throughout early life and adolescence. Sleep stage N1 remains the least prevalent across all age groups, as expected. Infants in the dataset also spend a relatively larger proportion of time in the WAKE stage during PSG recordings, likely due to developmental sleep fragmentation, technical interventions, and the unfamiliar sleep environment. Interestingly, a modest increase in wakefulness re-emerges during adolescence. This reversion may reflect well-documented shifts in circadian timing, hormonal changes, and increased psychosocial demands during the teenage years—all of which have been shown to influence sleep timing and continuity [62].

Furthermore, the sleep spectrograms analysis for all five sleep stages reveal distinct developmental trends. A rapid increase in total power is observed up until around 2 years of age, followed by stabilization with decreased power across subsequent ages. Notably, there is an initial increase in power within the low theta frequency range until approximately 2 years, after which delta activity becomes the dominant feature across all sleep stages [56]. During the WAKE stage, alpha activity is evident, particularly in the occipital region, and significant changes in the posterior dominant rhythm (PDR) are observed. The PDR, which begins to emerge in late infancy at a frequency between 5 and 7 Hz, gradually stabilizes and shifts into the alpha range of approximately 8 Hz with increasing age. This maturation of the PDR reflects the neurodevelopmental phase of sleep in children [63]. In the N1 sleep stage, the spectral patterns are generally slower compared to WAKE, with faint spindle activity detected around the 10–12 Hz frequency range. N2 sleep is characterized by pronounced spindles in the sigma frequency band (approximately 12–15 Hz), with spindle activity being most prominent in the central brain region. In infants, where the spindle-generating mechanisms are still developing, the spindle frequency is observed to be higher than in older children [64]. Stage N3 sleep consistently exhibits higher delta power compared to N2 across all age groups, underscoring the importance of slow-wave activity during this deeper sleep stage [65]. In the REM sleep stage, slow-wave delta power remains prevalent, contributing to the unique spectral characteristics of REM sleep in pediatric populations. These findings highlight the dynamic evolution of sleep-related brain activity and developmental changes in the EEG spectrum across childhood and early adolescence [42].

In parallel, the corpus provides a foundation for developing models to improve detection of respiratory events—such as central, obstructive, and MAs. Current machine learning models for apnea detection often perform suboptimally in pediatric populations due to physiological variability with age. Moreover, most focus on binary classification and ignore clinically meaningful distinctions between event subtypes. The breadth of annotated respiratory events in this corpus allows for development of multiclass, age-adaptive models that could enhance diagnostic precision in pediatric sleep apnea and related conditions. The occurrence of CA events is highest in infants and steadily declines with age—a pattern reflecting the maturation of the brainstem and the gradual stabilization of respiratory control mechanisms during early childhood. This trend has been well-documented in normative pediatric PSG studies, which consistently report higher rates of CAs in the first year of life, followed by a marked reduction during toddlerhood and beyond [66–68]. CHs and mixed events are relatively rare across all age groups in both the BCH dataset and prior literature. OA is also more frequent in infancy, a pattern associated with anatomical and developmental factors such as adeno-tonsillar hypertrophy and laryngomalacia—common causes of airway obstruction in young children. The BCH Sleep Corpus captures this trend clearly, with OA events declining as children age, aligning with reported normative values for airway maturation [67]. In contrast, OHs increase with age, particularly during adolescence, likely reflecting the rising prevalence of obesity and other structural risk factors that emerge during this developmental stage—findings that are similarly reported in large-scale pediatric PSG cohorts [68]. RERA events are infrequently detected in infants and toddlers, due to both physiological characteristics—such as high respiratory rates and variable breathing patterns—and technical limitations in scoring subtle flow-limited events at younger ages [48]. As children mature and their respiratory rhythms stabilize, RERAs become easier to identify, resulting in a gradual increase in frequency with age. This age-dependent detectability is well aligned with established clinical observations and further supports the validity of the respiratory event patterns observed in the BCH Sleep Corpus.

Another area of clinical relevance is the detection of limb movements during sleep. Limb movements are under-reported in clinical scoring—particularly when they occur adjacent to respiratory events—despite their potential diagnostic value in conditions such as PLMD [50, 51]. The inclusion of detailed EMG signals and expert annotations in this dataset offers an opportunity to revisit current scoring guidelines and explore novel detection frameworks that incorporate under-recognized movement patterns. The dataset also supports the development of automated arousal detection systems. Arousals fragment sleep and are frequently implicated in disorders such as sleep apnea, PLMS, and insomnia. However, their subtle and variable nature makes them difficult to detect consistently. With over one million annotated arousal events, the BCH Sleep Corpus enables the creation of robust, generalizable arousal detection algorithms [69, 70], which could improve diagnostic accuracy and help quantify sleep fragmentation in pediatric populations.

Importantly, this dataset holds significant promise for exploring sleep as a biomarker for systemic health and disease. The inclusion of de-identified clinical data linked to each PSG enables researchers to model associations between sleep metrics and a wide range of medical diagnoses. This facilitates investigations into early predictors of disease and the use of sleep patterns as potential phenotypic markers for neurological, developmental, and cardiopulmonary conditions. However, the following limitation should be noted when interpreting findings. The dataset originates from a single, high-resource academic medical center, and the cohort consists of pediatric patients referred for sleep evaluation—many with suspected or diagnosed sleep disorders or underlying health conditions. As such, the population may not be representative of the general pediatric population, and findings may not generalize to healthy children or those from more diverse community-based samples. Additionally, all recordings were obtained in a sleep laboratory setting, which may not fully capture natural sleep behavior. Despite these constraints, the BCH Sleep Corpus represents a critical advance in pediatric sleep research infrastructure. It offers unmatched breadth, annotation

quality, and clinical linkage, providing a foundational resource for advancing clinical and machine learning approaches in pediatric sleep medicine. The corpus supports a wide range of investigations—from the refinement of scoring algorithms to the discovery of sleep-derived biomarkers—and is poised to catalyze innovation across both research and clinical domains.

Conclusion

The BCH Sleep Corpus represents the most comprehensive annotated collection of pediatric PSG data to date. By spanning a wide range of ages, diagnoses, and physiological recordings, this dataset enables unprecedented opportunities to study the developmental trajectory of sleep and to develop robust machine learning tools for pediatric sleep analysis. It is a foundational resource for advancing sleep science, particularly in the context of age-specific physiology and clinical diagnosis. Future research leveraging this corpus may transform pediatric sleep diagnostics and deepen our understanding of sleep-related neurodevelopmental processes.

Disclosure statement

Financial disclosures: This work was funded by grants from the NIH (R01HL161253).

Non-financial disclosures: Dr. Westover is a co-founder, scientific advisor, and consultant to, and has a personal equity interest in Beacon Biosignals. Dr. Clifford has received research funding from the NSF, NIH, and LifeBell AI, and unrestricted donations from AliveCor Inc., Amazon Research, the Center for Discovery, the Gates Foundation, Google, the Gordon and Betty Moore Foundation, MathWorks, Microsoft Research, NextSense Inc., One Mind Foundation, and the Rett Research Foundation. Dr. Trotti is a member of the Board of Directors of the American Academy of Sleep Medicine; any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the American Academy of Sleep Medicine. Dr. Clifford has advisory roles and financial interests in AliveCor Inc. and NextSense Inc. He is also the CTO of MindChild Medical with significant stock. These relationships are unconnected to the current work. Dr. Thomas is co-inventor of: (1) Cardiopulmonary sleep spectrogram to assess sleep stability/quality and sleep apnea, licensed by the Beth Israel Deaconess Medical Center to MyCardio, LLC; (2) Patent for Enhanced Expiratory Rebreathing Space to treat high loop gain sleep apnea; (3) Patent for estimating respiratory self-similarity for detection of high loop gain sleep apnea. (4) General sleep medicine consulting: GLG Councils, Guidepoint, Beacon Biosignals, Jazz Pharmaceuticals. Dr. Stone reports grant funding from Eli Lilly and is consultant for Axsome Therapeutics. Dr. Maski (1) is consultant for Alkermes, Avadel, Harmony Biosciences, Jazz Pharmaceuticals, Takeda Pharmaceuticals, (2) has grant funding from Harmony Biosciences and Jazz Pharmaceuticals, (3) is DSMB chair for Idorsia, (4) collaborator on clinical trials sponsored by Alkermes and Takeda. These relationships are unconnected to the current work.

Data availability

The BCH Sleep Study Corpus is hosted on the BDSP, a resource supported by multiple NIH grants (RF1AG064312, RF1NS120947, R01AG073410, R01HL161253, R01NS126282, R01AG073598, R01NS131347, R01NS130119) to facilitate the sharing of de-identified, brain-related clinical data. BDSP houses several major repositories, including the Human Sleep Project (HSP), the International Cardiac Arrest Research Consortium (I-CARE), and the Harvard Electroencephalography Database (HEEDB). On November 21, 2024, the NIH formally recognized BDSP as an approved data-sharing repository, aligning with its Data Management and Sharing (DMS) policy and reinforcing its commitment to open science and collaborative neuroscience research. BDSP's infrastructure is supported by the AWS Open Data Sponsorship Program, providing secure, scalable, and cost-free access to datasets. The platform also promotes community-driven contributions through initiatives like BDSPC3, encouraging the sharing of de-identified data to advance research in neurology, critical care, and cognitive neuroscience. To learn more, visit bdsp.io or explore BDSP resources at bdsp.io/content. In particular, the Boston Children's Hospital Sleep Study Corpus can be accessed at https://bdsp.io/content/l8c86mgywuneuy74ae71/1.0.1/. Researchers can request access and download the dataset through AWS using the instructions provided on the website. The dataset includes full polysomnography recordings in EDF format, covering EEG, EOG, EMG, ECG, respiratory effort, airflow, oxygen saturation, and additional channels. A harmonized and downsampled HDF5 version is also provided for efficient data processing. To support usability, we have released a GitHub repository containing example scripts to read and load the files, as well as to generate plots and perform basic analyses. The code is available at https://github.com/bdsp-core/BCH-PSG-dataset.

References

1. Worley SL. The extraordinary importance of sleep: the detrimental effects of inadequate sleep on health and public safety drive an explosion of sleep research. *Pharm Ther*. 2018;**43**:758.
2. Vyazovskiy VV. Sleep, recovery, and metaregulation: explaining the benefits of sleep. *Nat Sci Sleep*. 2015;**7**:171–184. https://doi.org/10.2147/NSS.S54036
3. Scott AJ, Webb TL, Martyn-St James M, Rowse G, Weich S. Improving sleep quality leads to better mental health: a meta-analysis of randomised controlled trials. *Sleep Med Rev*. 2021;**60**:101556. https://doi.org/10.1016/j.smrv.2021.101556
4. Lumeng JC, Chervin RD. Epidemiology of pediatric obstructive sleep apnea. *Proc Am Thorac Soc*. 2008;**5**(2):242–252. https://doi.org/10.1513/pats.200708-135MG
5. Beebe DW, Wells CT, Jeffries J, Chini B, Kalra M, Amin R. Neuropsychological effects of pediatric obstructive sleep apnea. *J Int Neuropsychol Soc*. 2004;**10**(7):962–975. https://doi.org/10.1017/S135561770410708X
6. Malow BA, McGrew SG. Sleep disturbances and autism. *Sleep Med Clin*. 2008;**3**(3):479–488. https://doi.org/10.1016/j.jsmc.2008.04.004
7. Bruni O, Novelli L. Sleep disorders in children. *BMJ Clin Evid*. 2010;**2010**:2304.
8. Lam P, Hiscock H, Wake M. Outcomes of infant sleep problems: a longitudinal study of sleep, behavior, and maternal well-being. *Pediatrics*. 2003;**111**(3):e203–e207. https://doi.org/10.1542/peds.111.3.e203
9. Scher MS, Steppe DA, Banks DL. Prediction of lower developmental performances of healthy neonates by neonatal EEG-sleep measures. *Pediatr Neurol*. 1996;**14**(2):137–144. https://doi.org/10.1016/0887-8994(96)00013-6
10. Touchette É, Petit D, Séguin JR, Boivin M, Tremblay RE, Montplaisir JY. Associations between sleep duration patterns

and behavioral/cognitive functioning at school entry. *Sleep*. 2007;**30**(9):1213–1219. https://doi.org/10.1093/sleep/30.9.1213
11. Grigg-Damberger M. Neurologic disorders masquerading as pediatric sleep problems. *Pediatr Clin North Am*. 2004;**51**(1):89–115. https://doi.org/10.1016/S0031-3955(03)00180-9
12. Nunes ML, Bruni O. Insomnia in childhood and adolescence: clinical aspects, diagnosis, and therapeutic approach. *J Pediatr (Rio J)*. 2015;**91**(6):S26–S35. https://doi.org/10.1016/j.jped.2015.08.006
13. Liu J, Ji X, Pitt S, *et al*. Childhood sleep: physical, cognitive, and behavioral consequences and implications. *World J Pediatr*. 2024;**20**(2):122–132. https://doi.org/10.1007/s12519-022-00647-w
14. Mercante A, Owens J, Bruni O, *et al*. International consensus on sleep problems in pediatric palliative care: paving the way. *Sleep Med*. 2024;**119**:574–583. https://doi.org/10.1016/j.sleep.2024.05.042
15. Dahl RE, Lewin DS. Pathways to adolescent health: sleep regulation and behavior. *J Adolesc Health*. 2002;**31**(6):175–184. https://doi.org/10.1016/S1054-139X(02)00506-2
16. Beck SE, Marcus CL. Pediatric polysomnography. *Sleep Med Clin*. 2009;**4**(3):393–406. https://doi.org/10.1016/j.jsmc.2009.04.007
17. Church GD. The role of polysomnography in diagnosing and treating obstructive sleep apnea in pediatric patients. *Curr Probl Pediatr Adolesc Health Care*. 2012;**42**(1):2–25. https://doi.org/10.1016/j.cppeds.2011.10.001
18. Stowe RC, Afolabi-Brown O. Pediatric polysomnography—a review of indications, technical aspects, and interpretation. *Paediatr Respir Rev*. 2020;**34**:9–17. https://doi.org/10.1016/j.prrv.2019.09.009
19. Biswal S, Sun H, Goparaju B, Westover MB, Sun J, Bianchi MT. Expert-level sleep scoring with deep neural networks. *J Am Med Inform Assoc*. 2018;**25**(12):1643–1650. https://doi.org/10.1093/jamia/ocy131
20. Fiorillo L, Puiatti A, Papandrea M, *et al*. Automated sleep scoring: a review of the latest approaches. *Sleep Med Rev*. 2019;**48**:101204. https://doi.org/10.1016/j.smrv.2019.07.007
21. Lo JC, Ong JL, Leong RL, Gooley JJ, Chee MW. Cognitive performance, sleepiness, and mood in partially sleep-deprived adolescents: the need for sleep study. *Sleep*. 2016;**39**(3):687–698. https://doi.org/10.5665/sleep.5552
22. Yu X, Quante M, Rueschman M, *et al*. Emergence of racial/ethnic and socioeconomic differences in objectively measured sleep–wake patterns in early infancy: results of the rise & SHINE study. *Sleep*. 2021;**44**(3). https://doi.org/10.1093/sleep/zsaa193
23. Marcus CL, Moore RH, Rosen CL, *et al*. A randomized trial of adenotonsillectomy for childhood sleep apnea. *N Engl J Med*. 2013;**368**(25):2366–2376. https://doi.org/10.1056/NEJMoa1215881
24. Rosen CL, Larkin EK, Kirchner HL, *et al*. Prevalence and risk factors for sleep-disordered breathing in 8- to 11-year-old children: association with race and prematurity. *J Pediatr*. 2003;**142**(4):383–389. https://doi.org/10.1067/mpd.2003.28
25. Hunter SJ, Gozal D, Smith DL, Philby MF, Kaylegian J, Kheirandish-Gozal L. Effect of sleep-disordered breathing severity on cognitive performance measures in a large community cohort of young school-aged children. *Am J Respir Crit Care Med*. 2016;**194**(6):739–747. https://doi.org/10.1164/rccm.201510-2099OC
26. Wang R, Bakker JP, Chervin RD, *et al*. Pediatric adenotonsillectomy trial for snoring (PATS): protocol for a randomised controlled trial to evaluate the effect of adenotonsillectomy in treating mild obstructive sleep-disordered breathing. *BMJ Open*. 2020;**10**(3):e033889. https://doi.org/10.1136/bmjopen-2019-033889
27. Lee H, Li B, DeForte S, *et al*. A large collection of real-world pediatric sleep studies. *Sci Data*. 2022;**9**(1):421. https://doi.org/10.1038/s41597-022-01545-6
28. Natus Medical Incorporated. *SleepWorks 8 Reference Manual*. 2017.
29. Natus Medical Incorporated. *SleepWorks 9 Reference Manual*. 2017.
30. U.S. Department of Health and Human Services. *Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the HIPAA Privacy Rule*. 2012. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html. Accessed June 16, 2025.
31. Sun H, Ganglberger W, Nasiri S, *et al*. *The Human Sleep Project (version 2.0)*. Brain Data Science Platform. 2023. https://doi.org/10.60508/qjbv-hg78
32. Radhakrishnan L, Schenk G, Muenzen K, *et al*. A certified de-identification system for all clinical text documents for information extraction at scale. *JAMIA Open*. 2023;**6**(3):ooad045. https://doi.org/10.1093/jamiaopen/ooad045
33. Weinraub M, Bender RH, Friedman SL, *et al*. Patterns of developmental change in infants' nighttime sleep awakenings from 6 through 36 months of age. *Dev Psychol*. 2012;**48**(6):1511–1528. https://doi.org/10.1037/a0027680
34. Magnusdottir S, Hill EA. Prevalence of obstructive sleep apnea (OSA) among preschool aged children in the general population: a systematic review. *Sleep Med Rev*. 2024;**73**:101871. https://doi.org/10.1016/j.smrv.2023.101871
35. Blader JC, Koplewicz HS, Abikoff H, Foley C. Sleep problems of elementary school children: a community survey. *Arch Pediatr Adolesc Med*. 1997;**151**(5):473–480. https://doi.org/10.1001/archpedi.1997.02170420043007
36. Nixon GM, Thompson JMD, Han DY, *et al*. Short sleep duration in middle childhood: risk factors and consequences. *Sleep*. 2008;**31**(1):71–78. https://doi.org/10.1093/sleep/31.1.71
37. McKnight-Eily LR, Eaton DK, Lowry R, Croft JB, Presley-Cantrell L, Perry GS. Relationships between hours of sleep and health-risk behaviors in US adolescent students. *Prev Med*. 2011;**53**(4-5):271–273. https://doi.org/10.1016/j.ypmed.2011.06.020
38. Keyes KM, Maslowsky J, Hamilton A, Schulenberg J. The great sleep recession: changes in sleep duration among US adolescents, 1991–2012. *Pediatrics*. 2015;**135**(3):460–468. https://doi.org/10.1542/peds.2014-2707
39. Gramfort A, Luessi M, Larson E, *et al*. MEG and EEG data analysis with MNE-python. *Front Neurosci*. 2013;**7**:267. https://doi.org/10.3389/fnins.2013.00267
40. 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
41. Dereymaeker A, Pillay K, Vervisch J, *et al*. Review of sleep-EEG in preterm and term neonates. *Early Hum Dev*. 2017;**113**:87–103. https://doi.org/10.1016/j.earlhumdev.2017.07.003
42. Campbell IG, Feinberg I. Longitudinal trajectories of non-rapid eye movement delta and theta EEG as indicators of adolescent brain maturation. *Proc Natl Acad Sci U S A*. 2009;**106**(13):5177–5180. https://doi.org/10.1073/pnas.0812947106
43. Ghirardo S, Amaddeo A, Griffon L, Khirani S, Fauroux B. Central apnea and periodic breathing in children with underlying conditions. *J Sleep Res*. 2021;**30**(6):e13388. https://doi.org/10.1111/jsr.13388

44. Trachsel D, Erb TO, Hammer J, von Ungern-Sternberg BS. Developmental respiratory physiology. *Paediatr Anaesth*. 2022;**32**(2):108–117. https://doi.org/10.1111/pan.14362
45. Ratanakorn W, Brockbank J, Ishman S, Tadesse DG, Hossain MM, Simakajornboon N. The maturation changes of sleep-related respiratory abnormalities in infants with laryngomalacia. *J Clin Sleep Med*. 2021;**17**(4):767–777. https://doi.org/10.5664/jcsm.9046
46. Tan HL, Gozal D, Kheirandish-Gozal L. Obstructive sleep apnea in children: a critical update. *Nat Sci Sleep*. 2013;**5**:109–123. https://doi.org/10.2147/NSS.S51907
47. Schwengel DA, Dalesio NM, Stierer TL. Pediatric obstructive sleep apnea. *Anesthesiol Clin*. 2014;**32**(1):237–261. https://doi.org/10.1016/j.anclin.2013.10.012
48. Marcus CL, Brooks LJ, Draper KA, *et al*. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics*. 2012;**130**(3):576–584. https://doi.org/10.1542/peds.2012-1671
49. Grigg-Damberger M, Gozal D, Marcus CL, *et al*. The visual scoring of sleep and arousal in infants and children. *J Clin Sleep Med*. 2007;**3**(2):201–240. https://doi.org/10.5664/jcsm.26819
50. Hornyak M, Feige B, Riemann D, Voderholzer U. Periodic leg movements in sleep and periodic limb movement disorder: prevalence, clinical significance and treatment. *Sleep Med Rev*. 2006;**10**(3):169–177. https://doi.org/10.1016/j.smrv.2005.12.003
51. DelRosso LM, Picchietti DL, Sharon D, *et al*. Periodic limb movement disorder in children: a systematic review. *Sleep Med Rev*. 2024;**76**:101935. https://doi.org/10.1016/j.smrv.2024.101935
52. Halgren M, Ulbert I, Bastuji H, *et al*. The generation and propagation of the human alpha rhythm. *Proc Natl Acad Sci U S A*. 2019;**116**(47):23772–23782. https://doi.org/10.1073/pnas.1913092116
53. Malerba P, Whitehurst L, Mednick SC. The space-time profiles of sleep spindles and their coordination with slow oscillations on the electrode manifold. *Sleep*. 2022;**45**(8). https://doi.org/10.1093/sleep/zsac132
54. Armitage R, Roffwarg HP. Distribution of period-analyzed delta activity during sleep. *Sleep*. 1992;**15**(6):556–561.
55. Bernardi G, Betta M, Ricciardi E, Pietrini P, Tononi G, Siclari F. Regional delta waves in human rapid eye movement sleep. *J Neurosci*. 2019;**39**(14):2686–2697. https://doi.org/10.1523/JNEUROSCI.2298-18.2019
56. Sun H, Ye E, Paixao L, *et al*. The sleep and wake electroencephalogram over the lifespan. *Neurobiol Aging*. 2023;**124**:60–70. https://doi.org/10.1016/j.neurobiolaging.2023.01.006
57. Einizade A, Nasiri S, Sardouie SH, Clifford GD. ProductGraph-SleepNet: sleep staging using product spatio-temporal graph learning with attentive temporal aggregation. *Neural Netw*. 2023;**164**:667–680. https://doi.org/10.1016/j.neunet.2023.05.016
58. Perslev M, Darkner S, Kempfner L, Nikolic M, Jennum PJ, Igel C. U-sleep: resilient high-frequency sleep staging. *NPJ Digit Med*. 2021;**4**(1):72. https://doi.org/10.1038/s41746-021-00440-5
59. Kurth S, Ringli M, Geiger A, LeBourgeois M, Jenni OG, Huber R. Mapping of cortical activity in the first two decades of life: a high-density sleep electroencephalogram study. *J Neurosci*. 2010;**30**(40):13211–13219. https://doi.org/10.1523/JNEUROSCI.2532-10.2010
60. Blumberg MS, Gall AJ, Todd WD. The development of sleep-wake rhythms and the search for elemental circuits in the infant brain. *Behav Neurosci*. 2014;**128**(3):250–263. https://doi.org/10.1037/a0035891
61. Tarokh L, Carskadon MA. Developmental changes in the human sleep EEG during early adolescence. *Sleep*. 2010;**33**(6):801–809. https://doi.org/10.1093/sleep/33.6.801
62. Colrain IM, Baker FC. Changes in sleep as a function of adolescent development. *Neuropsychol Rev*. 2011;**21**(1):5–21. https://doi.org/10.1007/s11065-010-9155-5
63. Freschl J, Azizi LA, Balboa L, Kaldy Z, Blaser E. The development of peak alpha frequency from infancy to adolescence and its role in visual temporal processing: a meta-analysis. *Dev Cogn Neurosci*. 2022;**57**:101146. https://doi.org/10.1016/j.dcn.2022.101146
64. Scholle S, Zwacka G, Scholle HC. Sleep spindle evolution from infancy to adolescence. *Clin Neurophysiol*. 2007;**118**(7):1525–1531. https://doi.org/10.1016/j.clinph.2007.03.007
65. Kurth S, Jenni OG, Riedner BA, Tononi G, Carskadon MA, Huber R. Characteristics of sleep slow waves in children and adolescents. *Sleep*. 2010;**33**(4):475–480. https://doi.org/10.1093/sleep/33.4.475
66. Schlüter B, Buschatz D, Trowitzsch E. Polysomnographic reference curves for the first and second year of life. *Somnologie*. 2001;**5**(1):3–16. https://doi.org/10.1046/j.1439-054x.2001.01148.x
67. Marcus CL, Omlin KJ, Basinki DJ, *et al*. Normal polysomnographic values for children and adolescents. *Am Rev Respir Dis*. 1992;**146**(5_pt_1):1235–1239. https://doi.org/10.1164/ajrccm/146.5_Pt_1.1235
68. Scholle S, Wiater A, Scholle HC. Normative values of polysomnographic parameters in childhood and adolescence: cardiorespiratory parameters. *Sleep Med*. 2011;**12**(10):988–996. https://doi.org/10.1016/j.sleep.2011.05.006
69. Sforza E, Chapotot F, Lavoie S, Roche F, Pigeau R, Buguet A. Heart rate activation during spontaneous arousals from sleep: effect of sleep deprivation. *Clin Neurophysiol*. 2004;**115**(11):2442–2451. https://doi.org/10.1016/j.clinph.2004.06.002
70. Taylor KS, Murai H, Millar PJ, *et al*. Arousal from sleep and sympathetic excitation during wakefulness. *Hypertension*. 2016;**68**(6):1467–1474. https://doi.org/10.1161/HYPERTENSIONAHA.116.08212

SLEEPJ, 2025, **00**, 1–22
https://doi.org/10.1093/sleep/zsaf273
Original Article