Reviewer Information 12/5/20

Number of tables : 4

Number of figures : 2

Abstract word count : 234

Statement of Significance word count : 90

Word count: 6311

Title: Stability of Nocturnal Wake and Sleep Stages Defines CNS Disorders of Hypersomnolence

Authors: Kiran P. Maski, MD MPH[1*]; Alicia Colclasure, MS[2*]; Elaina Little, BA[1]; Erin Steinhart, MA[1]; Thomas E. Scammell, MD [1,4]; William Navidi, PhD[2]; Cecilia Diniz Behn, PhD[2,3]

[1]Department of Neurology, Boston Children's Hospital, Boston, MA

[2]Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO

[3]Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO

[4]Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA.

Corresponding Author:

Kiran Maski, MD MPH
300 Longwood Ave
Boston, MA 02115
Kiran.maski@childrens.harvard.edu

[*]These authors contributed equally to the manuscript

Abstract:

Objective: We determine if young people with narcolepsy type 1 (NT1), narcolepsy type 2 (NT2) and idiopathic hypersomnia (IH) have distinct nocturnal sleep stability phenotypes compared to subjectively sleepy controls.

Methods: Participants were 5-21 years old and drug-naïve or drug free: NT1 (n=46), NT2 (n=12), IH (n=18) and subjectively sleepy controls (n=48). We compared the following sleep stability measures from polysomnogram (PSG) recording between each hypersomnolence disorder to subjectively sleepy controls: number of wake and sleep stage bouts, Kaplan Meier survival curves for wake and sleep stages, and median bout durations.

Results: Comparisons are made to subjectively sleepy controls group. NT1 participants had more bouts of wake and all sleep stages ($p \leq 0.005$) except stage N3. NT1 participants had worse survival of nocturnal wake, stage N2 , and REM bouts ($p<0.005$). In the first 8 hours of sleep, NT1 participants had longer stage N1 bouts but shorter REM and stage N3 bouts (all p's <0.013). IH participants had a similar number of bouts but better survival of stage N2 bouts ($p=0.001$) and, in the first 8 hours of sleep, shorter duration of stage N3 bouts ($p=0.003$). In contrast, NT2 participants showed better stage N1 bout survival ($p=0.006$) and longer stage N1 bouts ($p=0.02$).

Conclusions: NT1, NT2 and IH have unique sleep pathology compared to subjectively sleepy controls, with only NT1 demonstrating clear nocturnal wake and sleep instability. Overall, sleep stability measures may aid in diagnoses and management of these CNS disorders of hypersomnolence.

Keywords: narcolepsy, idiopathic hypersomnia, sleep, sleep stability, disrupted nighttime sleep

Statement of Significance: As standard PSG endpoints do not easily distinguish between CNS disorders of hypersomnolence, we used PSG measures of sleep stability and found that NT1 has nocturnal sleep state instability, NT2 has more stable stage N1 (lighter, unrefreshing sleep), and IH has evidence of overly stable stage N2 sleep and less stable stage N3 sleep. Sleep stability measures hold great promise to demarcate NT1, NT2, and IH phenotypes, could be incorporated into future hypersomnia research to help define more homogeneous study populations, and provide objective endpoints for future clinical research.

Introduction:

Central nervous system (CNS) disorders of hypersomnolence are characterized by excessive daytime sleepiness despite sufficient nocturnal sleep. The most common of these disorders typically begin in the first two decades of life[1, 2]: narcolepsy type 1 (NT1, narcolepsy with cataplexy), narcolepsy type 2 (NT2, narcolepsy without cataplexy) and idiopathic hypersomnia (IH).

The daytime symptoms of NT1, NT2, and IH overlap and include daytime sleepiness, fatigue, and related cognitive difficulties. However, these disorders differ by subjective and objective aspects of sleep disturbance and sleep quality. [3-5]. For example, disrupted nighttime sleep is common in people with narcolepsy[6] and refers to the presence of frequent brief awakenings from sleep [7]. NT1 is caused by near-complete loss of the orexin-producing neurons, and orexins are essential for stabilizing wake and sleep. [8] Specifically, people with NT1 have more wakings on nocturnal polysomnography (PSG) than controls [9-12], and fragmented sleep also occurs in mice lacking the orexin neurons[13]. People with NT2 also report disrupted sleep[14], but they seem to have the same number of wakings on PSG as controls[9], perhaps because they do not lack orexins[15]. Last, people with IH generally report long sleep times, few wakings, profound sleep inertia, and non-restorative sleep. [6, 16] Objectively, people with IH have fewer arousals[17] and higher sleep efficiency[18] than controls on PSG, but the number of nocturnal wakings does not differ[9]. The pathophysiology of IH is unknown but researchers hypothesize it could be due to potentiation of GABA receptor signaling[19-21] or a prolonged circadian period length[22]. In either case, sleep would not be expected to

be disrupted by excessive arousals or wakings. Thus, it would seem that the underlying disease mechanisms for NT1, NT2, and IH may produce distinct nocturnal sleep phenotypes.

Knowledge of the nocturnal sleep phenotypes of these CNS disorders of hypersomnolence could be clinically useful. It is often challenging to determine if a patient has NT2 or IH as the MSLT is not reliable [23, 24]. We propose that differences in nocturnal sleep stability may provide objective, distinguishing characteristics to distinguish between NT1, NT2 and IH, aiding the clinical diagnosis. A better understanding of which features of sleep stability differ from subjectively sleepy controls and between hypersomnia conditions may aid in the evolving nosology and clinical diagnosis of NT1, NT2 and IH. For instance, there is disagreement in the field regarding the diagnoses of NT2 and IH whether these conditions should be collapsed together or if their phenotypes are distinct enough to warrant separate names[25, 26]. Last, measures of nocturnal sleep stability may enable identification of distinct subtypes of disorders that may benefit from sleep-based therapeutics in addition to wake promoting treatments.

However, it is unclear which measures are best to clinically evaluate sleep stability. Clinical measures such as arousal index, wake after sleep onset, number of wakings, and sleep efficiency can reflect sleep quality and sleep fragmentation, but alone, these do not detail the dynamics of wake and sleep bout stability. Thus, it is currently unclear which sleep stages are unstable, how long wakings typically last, and how sleep changes across the night. Basic scientists have employed measures including sleep/wake bout number, median duration and survival curves to define sleep and wake

stability in animal models of narcolepsy in light and dark periods[13,27], and these approaches may offer sensitive metrics to assess nocturnal wake and sleep stability in clinical populations.

In this study, we compared nocturnal PSG sleep stability measures in young participants with NT1, NT2 and IH to subjectively sleepy controls ages 5-21 years old. We hypothesized that each condition would be characterized by distinct changes in sleep stability compared to subjectively sleepy controls as measured by number, median duration, and survival curves of wake and sleep stage bouts across the night. We predicted that compared to subjectively sleepy controls, 1) NT1 participants would have unstable NREM and REM sleep; 2) NT2 participants would have normal stability of NREM and REM sleep; 3) IH participants would have more stable stage N2 and stage N3 sleep as they generally have high sleep efficiency and sleep inertia. We also examined sleep stability between NT1, NT2 and IH to identify findings unique to each hypersomnia. As some hypersomnia patients sleep for very long periods, we analyzed data reflecting both the full sleep time recordings and the first 8 hours of sleep to appropriately characterize the full sleep episode within each group as well as normalizing sleep time to compare across groups. In addition, to account for changes in sleep with age, developmental stage, and sex, we also examined the effects of age and sex on wake and sleep bout survival curve measures.

Methods

Study design

We conducted a retrospective, cross-sectional study of consecutive PSGs and multiple sleep latency tests (MSLTs) ordered for evaluation of suspected hypersomnia conditions at Boston Children's Hospital Pediatric Sleep Laboratory from 2015-2018. Nocturnal PSGs were re-scored blinded to diagnoses in 30 second epochs as wake and sleep stages (N1, N2 , N3, or REM) according to AASM guidelines[28]. For each stage, bouts were defined as successive epochs of the same stage. We analyzed features of these bouts to characterize nocturnal sleep behavior for each diagnosis group. We focused the sleep stability analysis (number, survival curves, and median duration of wake and sleep stage bouts across the night) on the first 8 hours of PSG recordings to compare groups using a standard reference period. However, we also analyzed data using full sleep time recording as sleep stability may change across the sleep period, especially among those with longer sleep durations.

Patients

Patients were either drug naïve or weaned from potential sedating, alerting, and/or REM sleep-suppressing medications 2 weeks prior to the sleep study date, slept a minimum of 6 hours during the PSG, and had no findings of obstructive sleep apnea (AHI ≥ 1/hour). [6] Of the 125 patients screened, 107 also had routine urine drug screening in accordance with AASM guidelines [29], testing for common substances of abuse (marijuana, caffeine, nicotine, stimulants, PCP, cocaine, benzodiazepines). One control patient with positive drug screen results was excluded from analysis, but patients with missing or negative results were included. Patient age, sex, and Epworth sleepiness score were obtained on the sleep study night, and body mass index (BMI) and symptom

duration were extracted from the clinical visit records of the ordering sleep study provider.

The current diagnostic criteria for NT1 includes the presence of excessive daytime sleepiness >3 months, the presence of cataplexy and sleep study testing showing the MSLT mean sleep latency of ≤8 minutes and ≥2 sleep onset REM periods (REM latency within 15 minutes from sleep onset). Alternatively, NT1 can be diagnosed by CSF orexin <110 pg/ml. In this study, 2 patients were diagnosed by CSF orexin testing as their sleep studies did not meet diagnostic cut off values. NT2 diagnosis requires the presence of excessive daytime sleepiness >3 months and sleep study testing showing a MSLT mean sleep latency ≤8 minutes and ≥2 sleep onset REM sleep periods (including REM sleep within 15 minutes of sleep onset on PSG). IH has broader diagnostic criteria but requires the presence of subjective excessive daytime sleepiness. IH diagnosis can be supported by objective evidence of daytime sleepiness (MSLT mean sleep latency ≤8 minutes) or increased sleep need as shown by total 24 hour sleep duration ≥ 660 minutes on a 24 hour PSG or actigraphy data averaged across at least 7 days. In our study, all IH patients were diagnosed by MSLT criteria except two patients were diagnosed by long sleep times (but had normal MSLT values). One of these patients slept 19.6 hours on an extended PSG. Outlier effects of this participant were assessed in our analysis. Subjectively sleepy controls in this study were patients who reported problematic daytime sleepiness but they did not meet PSG and MSLT criteria for NT1, NT2, or IH. Participants with chronic medical conditions (including neurologic, respiratory, genetic, or immunological) that could explain hypersomnolence were not selected for the study.

We analyzed nocturnal PSG data for 124 pediatric participants evaluated for hypersomnia including 48 subjectively sleepy controls, 46 with NT1, 12 with NT2, and 18 with IH (**Table 1**). The patients' mean age was 13.7 (3.6) years. Age and sex demographics were similar across groups with the exception of the IH group which was older and predominantly female and the NT1 group which had a higher BMI. Several different races/ethnicities were represented. However, the majority of the participants were Caucasian, and small sample sizes for other races/ethnicities limited our power to determine race/ethnicity effects. Mean sleep duration on nocturnal PSG was 9.1 (1.6) hours and ranged from 5.6 to 19.6 hours (the latter being an extended PSG for the IH patient noted above).

Clinical and Neurophysiological Evaluation

We collected patients' age, sex, race, body mass index (BMI), self-reported daytime sleepiness by adapted Epworth Sleepiness Scale (ESS)[30, 31], HLA DQB1*06:02 status, symptom duration, and medication naïve or medication withdrawal status from the sleep clinic visit proceeding the PSG/MSLT night. Data collected from MSLT testing included mean sleep latency (MSL) and total number of sleep onset REM sleep periods (SOREMPs). Data collected from the PSG testing included total nocturnal sleep time (TST), sleep onset latency (SOL), REM sleep onset latency, nocturnal arousal index, nocturnal sleep efficiency (amount of sleep relative to time in bed), wake after sleep onset (WASO), and nocturnal sleep stage percentages of the sleep period.

Analysis of PSG measures

Demographic and standard sleep measure analysis was conducted using SPSS for Windows (version 19; IBM Corp, Armonk, NY, USA). All other data analysis and modeling was performed with MATLAB (Mathworks, Natick, MA).

We compared standard PSG sleep quality measures across the full night of sleep (TST, SOL, arousal index, sleep efficiency, WASO, nocturnal wake and sleep stage percentages, and REM sleep onset latency) using analysis of variance (ANOVA) adjusted for sex (male, female), race/ethnicity (Caucasian, African American, other, unknown), and age at time of sleep study testing.

We define wake stability as maintenance of a continuous series of wake bouts before transitioning to sleep and sleep stability as maintenance of a particular sleep stage before waking or transitioning to another sleep stage. To assess stability of sleep/wake states, we compared the number of bouts and median bout duration across groups. We compared the number of bouts in the first 8 hours of PSG recordings for wake and each sleep stage using ANOVA adjusted for demographic variables as described above. Bout number was analyzed using 8h truncated PSG recordings because patients with longer sleep times would be expected to have more bouts. We conducted pairwise tests to directly compare groups.

We further analyzed median bout duration for each wake and sleep stage (N1, N2, N3, and REM) in the full and 8h truncated PSG recordings. Median bout durations for wake and sleep stages (N1, N2, N3 and REM) for each group were calculated as the median of the individual participants' medians. Medians for each group were compared using Mood's median test[32]. As we cannot adjust for demographic factors in Mood's median

test, we tested if results differed in subgroup analysis of sex (male, female) and age (≤14 years; >14 years).

Survival analysis of bouts

To quantify differences in the distribution of the wake and sleep bouts between groups, we used Kaplan-Meier survival curves to describe the percentage of bouts that "survive" through a given duration. We analyzed data from both full and 8h truncated PSG recordings, but we focus the presentation of results on analysis of the full sleep time recordings because they represent sleep/wake bout durations across each participant's total sleep time. To compare the curves while adjusting for potential demographic effects, we used the Cox proportional hazards model (MATLAB's built-in *coxphfit* function). The hazard function h(t) describes the rate at which bouts end as a function of bout duration and is given by

$$h(t) = -\frac{d \ln S(t)}{dt}$$

where S(t) is the survival function.

The Cox proportional hazards model assumes a hazard function of the following form

$$h(t) = h_0(t)\exp\left[\sum_{j=1}^{p} x_j b_j\right]$$

where $h_0(t)$ is the baseline hazard function, p is the number of variables, x_j is the value of the jth variable, and b_j is the jth coefficient. Since the hazard function describes the rate at which bouts end, greater values of the hazard imply decreased survival of bouts. Thus, a negative coefficient is associated with an increase in survival, and a positive coefficient is associated with a decrease in survival.

Model variables included hypersomnia condition (subjectively sleepy controls, NT1, NT2, IH) and demographic variables (age, sex, race/ethnicity). Each variable in the model was compared to a baseline status with demographic baselines taken to be age 14, male, and Caucasian. These baselines were selected because 14 years was the median age of all participants, and the male and Caucasian subgroups were the largest subgroups by sex and race/ethnicity, respectively. We separately considered each CNS disorder of hypersomnolence as a baseline to assess group comparisons. Older and younger subgroups were defined to include participants >14 and ≤14 years of age, respectively, and additional analyses used the upper (age 16) and lower (age 11) quartiles as the baselines for older and younger subgroups, respectively. For analyses of IH participants, we did not separately consider the age 11 baseline because all participants were older than age 11.

To investigate potential effects of demographics on the manifestation of diagnosed CNS hypersomnia condition (subjectively sleepy controls, NT1, NT2, IH), we included interaction terms in the Cox proportional hazards models. Models were fit to survival data using an iterative process to determine which interaction terms should be retained in the final model. First, we included terms in the model to represent the interaction of each demographic variable (age, sex, race/ethnicity) with CNS hypersomnia condition for each sleep/wake stage. Next, we further investigated interactions that were considered potentially significant ($p<0.2$) by fitting models that included all possible pairs of these interactions. Terms that remained potentially significant ($p<0.2$) when considered pairwise were retained in the model. Given the low number of potentially significant interaction terms identified in our data, this process reduced the number of

potentially significant interaction terms to at most two for each model. Therefore, it was not necessary to consider possible combinations of three or more interactions. Finally, the model was run with all remaining potentially significant interaction terms, and the significance of the interaction of each demographic effect with CNS hypersomnia condition was assessed based on the p-values associated with the appropriate interaction term.

For each Cox proportional hazards model, we report the hazard ratio and 95% confidence interval for the significant main effects of CNS hypersomnia condition controlling for age, sex, and race as well as the age interaction.

Statistical significance was taken at P = 0.05 level for all output. P-values <0.0005 were reported as ~0.

Standard protocol approvals, registrations, and patient consents

This study was approved by the Institutional Review Board (IRB) at Boston Children's Hospital and the need for patient consent/assent was waived by the IRB for this retrospective analysis of de-identified data.

Results

Characterization of standard sleep measures

We present sleep quality measures of the full sleep time recording and ANOVA analysis with all group comparisons in Table 2. Compared to subjectively sleepy controls, NT1 participants had shorter sleep onset latency (B=-22.3, p=0.006), shorter REM sleep onset latency (B=-62.38, p~0), higher arousal index (B=4.1, p~0), higher stage N1 percentage (B=3.7, p=0.01), and lower stage N2 percentage (B=-3.0, 0.026). NT1 participants also had more WASO (B=16.13, p=0.021) but better sleep efficiency

(B=3.14, p=0.045) than the subjectively sleepy controls group. NT2 participants showed increased sleep efficiency (B=7.1, p=0.019), shorter REM sleep onset latency (B-39.57, p=0.033), and higher REM sleep percentage (B=6, p=0.014) compared to subjectively sleepy controls. Group differences in total sleep time (F=4.9, p=0.003) reflected longer sleep time in IH participants compared to subjectively sleepy controls (B=1.6, p~0). IH participants also had higher sleep efficiency (B=4.41, p=0.045), lower amount of stage N3 sleep (B-5.73, p=0.009), and there was trend for IH participants to have higher amounts of stage N2 sleep (B=4.7, p=0.05) than subjectively sleepy controls. After excluding the IH participant with >19 hours of total sleep time, only the reduction in stage N3 retained significance.

PSG wake and sleep median bout duration and bout number

We performed group comparisons of median bout durations using both full sleep time and 8h truncated recordings obtained from PSG. Full results of demographic subgroup analysis are presented in Supplementary Materials.

In the analysis of the full sleep time recordings, subjectively sleepy controls and NT1 participants had similar median bout durations of wake and sleep stages. NT2 participants had longer median stage N1 bouts than subjectively sleepy controls (B=0.13, p=0.02) and IH participants (B=0.13, p=0.02). The IH group showed no differences in median wake or sleep stage bouts compared to any group, but subgroup analysis showed that older IH participants (>14 years) had longer median stage N2 bouts than older subjectively sleepy controls (B=2.88, p=0.04) and older NT1 participants (B=4.75, p=0.01).

In the analysis of the first 8 hours of PSG recordings (**Table 3**), we found that NT1 participants had longer stage N1 bouts (B=0.25, p=0.004), shorter REM bouts (B=-3.87, p~0), and shorter stage N3 bouts (B=-5.5, p=0.013) compared to subjectively sleepy controls. Consistent with analysis of the full PSG recording data, NT2 participants showed longer stage N1 bouts than subjectively sleepy controls (B=0.38, p=0.006). IH participants had shorter stage N3 bouts compared to subjectively sleepy controls (B=-6.88, p=0.003).

Results for analyses of the number of bouts in the first 8 hours of PSG recordings are summarized in **Figure 1**. We found group differences in the number of wake (F=8.4, p~0), stage N1 (F=6.9, p~0), stage N2 (F=4.6, p=0.005), and REM (F=9.9, p~0) bouts driven by an increase in number of bouts in NT1 participants. Notably, the number of stage N3 bouts (F=1.25, p=0.297) did not differ across groups. On subgroup analysis, NT1 participants had more bouts of wake (B=10.5, p=0.001), stage N1 sleep (B=13.3, p=0.001), stage N2 sleep (B=6.6, p=0.006), and REM sleep (B=5.6, p~0) than subjectively sleepy controls. The NT1 group had more wake bouts (B=15.09, p=0.004), stage N1 sleep bouts (B=14.14, p=0.043) and stage N2 bouts (B=8.6 p=0.032) than NT2, but NT1 and NT2 groups had similar numbers of REM bouts. The number of bouts did not differ between subjectively sleepy controls vs. NT2, subjectively sleepy controls vs. IH or NT2 vs. IH for any stage.

PSG Survival analysis of bouts

We further analyzed the stability of wake and sleep stages across CNS hypersomnia conditions using Kaplan-Meier survival curves for bout durations (**Figure 2)** and Cox proportional hazards regression analysis. Based on full PSG sleep time recordings and

compared to subjectively sleepy controls, NT1 participants showed worse survival of wake bouts (hazard ratio 1.15, 95% CI [1.04, 1.27], p=0.002), stage N2 bouts (hazard ratio 1.31, 95% CI [1.21, 1.42], p~0) and REM bouts (hazard ratio 1.60, 95% CI [1.39, 1.83], p~0). REM sleep was less stable in the NT1 group based on pairwise comparisons to NT2 (hazard ratio 1.65, 95% CI [1.33, 2.05], p~0) and IH (hazard ratio 1.54, 95% CI [1.83, 1.29], p~0) as well. Stage N1 and stage N3 survival curves did not differentiate the NT1 group from subjectively sleepy controls (p's >0.18).

Duration of PSG analysis did not affect sleep survival curve findings, but the finding that survival of wake bouts was worse in NT1 was not observed in the analysis of the 8h truncated PSG recordings, suggesting that wake bout fragmentation is greatest at the end of the full sleep period.

For NT2 participants, the survival of stage N1 bouts was increased compared to subjectively sleepy controls (hazard ratio 0.81, 95% CI [0.69, 0.95], p=0.006) as well as compared to NT1 participants (hazard ratio 0.86, 95% CI [0.75, 0.99], p=0.036) and IH participants (hazard ratio 1.63, 95% CI [1.32, 2.03], p~0). We did not find that stage N3 or REM sleep differed between the NT2 group and subjectively sleepy controls and observed differences in wake and stage N2 were mediated by interactions between NT2 status and sex (described in Supplementary Material). Group comparisons with NT2 did not differ with 8 hour PSG analyses.

Lastly, the IH group showed increased stage N2 sleep stability compared to subjectively sleepy controls (hazard ratio 0.83, 95% CI [0.74, 0.93], p=0.001) and the NT1 group (hazard ratio 0.63, 95% CI [0.56, 0.71], p~0) in full PSG recordings. The additional features of survival of sleep/wake stages in IH participants were largely

mediated by interactions with age as described below. Results retained significance in assessing only the first 8 hours of PSG recordings and excluding the outlier with total sleep time >19 hours.

Group differences compared to subjectively sleepy controls are summarized in **Table 4**.

Demographic Effects on Sleep/Wake Survival Curves Across and Between Groups

Survival curve analyses also identified several features of sleep/wake stage stability associated with demographic variables (age, sex) across groups, with age yielding the strongest effects. The influence of sex across and within groups is reported in Supplemental Material. Stability of stage N3 sleep worsens with age for all groups (hazard ratio 1.04, 95% CI [1.02, 1.06, p=0.001) with an even greater decrease in stability associated with IH due to an interaction between IH status and age. All groups, except for IH participants, had worse wake stability with age (hazard ratio 1.02, 95% CI [1, 1.04], p=0.003), and all groups, except for NT1 participants, had improved REM stability with age (hazard ratio 0.97, 95% CI [0.88, 1.07], p=0.016). Interactions between group status and age offset the main effects of age on the stability of wake and REM for IH and NT1 participants, respectively.

Within the IH group, features of survival of wake, stage N1 and stage N3 were largely mediated by interactions with age (significance of interaction terms: p=0.004 for wake; p~0 for stage N1; and p=0.078 for stage N3). IH participants <16 years of age had less stable wake (hazard ratio 1.28, 95% CI [1.06, 1.56], p=0.012) and stage N1 sleep compared to subjectively sleepy controls <16 years of age (hazard ratio 1.32, 95% CI

[1.11, 1.58], p=0.001), but these differences were not detected in older IH patients. In contrast, older IH participants (≥16 years) had less stable stage N3 sleep compared to older subjectively sleepy controls (hazard ratio 1.25, 95% CI [1.00, 1.55], p=0.049) and older NT2 participants (hazard ratio 1.40, 95% CI [1.07, 1.85], p=0.02). These differences were not significant at median age 14. We did not see differences in stage N3 survival curves between older IH and NT1 participants.

Discussion

In this retrospective, cross-sectional study, we compared sleep/wake state stability measures in young participants with NT1, NT2 and IH to subjectively sleepy controls (results summarized in **Table 4**). Among other methods, we used survival analysis to compare distributions of bout lengths for each sleep/wake stage to distinguish groups' abilities to sustain sleep/wake states. Consistent with our hypothesis, we found that each hypersomnia disorder has differences in sleep/wake bout analysis, suggesting that sleep quality and sleep stability differ in each group. For example, across the full sleep time recordings, NT1 participants had clear sleep fragmentation, with increased arousal index and WASO compared to subjectively sleepy controls. However, NT1 participants had better sleep efficiency than controls, probably because of their rapid sleep onset latency and poorer survival of wake bouts emerging from the full sleep period. In comparing the sleep stability measures in the first 8 hours of PSG recordings to subjectively sleepy controls, NT1 patients had 1) longer stage N1 bouts and shorter REM and stage N3 bouts of sleep, 2) increased bout number of wake and all sleep stages except stage N3, and 3) worse survival of stage N2 and REM bouts and

additionally worse survival of wake bouts in the full sleep time recording. In contrast, the NT2 group seemed to have better sleep efficiency than subjectively sleepy controls, though sleep onset latency, WASO, and number of wake bouts were comparable. On sleep stability analysis, NT2 participants showed better survival and longer median duration of stage N1 bouts compared to subjectively sleepy controls (and all other groups). This suggests that while this group has better sleep continuity, it is because they tend to linger in lighter non-refreshing sleep. IH participants showed longer sleep time with high sleep efficiency compared to subjectively sleepy controls, but they had shorter stage N3 bouts in the first 8 hours of PSG recordings and better survival of stage N2 bouts compared to subjectively sleepy controls. With older age (≥ 16 years), the IH participants showed decreased survival of stage N3 bouts compared to older subjectively sleepy controls. Though speculative, this finding may suggest that IH participants have a stage N3 deficiency and compensate with overly stable stage N2 sleep. Counter to our predictions, we also found that: 1) stage N3 sleep stability seems relatively preserved in NT1 in contrast to stage N2 and REM sleep, 2) stage N1 stability defines NT2 sleep, and 3) stage N2 but not stage N3 sleep is more stable in IH.

In exploring differences unique to each condition, we found that NT1 shows worse REM sleep fragmentation and NT2 shows more stable stage N1 sleep compared to all other CNS disorders of hypersomnolence and subjectively sleepy controls.

Intrinsic Sleep Instability in NT1

Consistent with previous reports, our analyses show that nocturnal sleep in NT1 is disrupted by frequent wake bouts and arousals[9, 10, 33], and here, we further detail

nocturnal wake and sleep stability among young NT1 participants. Barateau et al. recently examined sleep stability in adults with NT1 in relation to CSF orexin levels[34]. Similar to their findings, our NT1 participants (who are presumably orexin deficient) had a higher number of wake and sleep bouts across the nocturnal period compared to subjectively sleepy controls. Thus, bout number seems like a reliable measure of sleep stability and reflects increased sleep-wake transitions as described previously[10, 11 9]. Building on the Barateau study, we find that NT1 mainly increases the number of stage N1 and to a lesser extent the number of stage N2 and REM bouts . This increase in stage N1 bout number and median duration (in the first 8 hours of sleep) likely accounts for the increase in stage N1 percentage of total sleep time in the NT1 group described here and by others[9, 35, 36]. This increase in the amount of stage N1 sleep seems to come at the expense of decreased stage N2 sleep but not stage N3 sleep relative to total sleep time.

In NT1, stage N3 sleep was relatively stable. We found that stage N3 bouts are shorter in the first 8 hours of sleep but there were no other signs of stage N3 instability such as increased stage N3 bout number or worse survival of stage N3 bouts. There are few data regarding stage N3 stability in narcolepsy. Data from electroencephalogram spectral analysis showed declining slow wave intensity across successive NREM-REM sleep in NT1 [36] which may manifest as less durable stage N3 sleep as the night progresses. It is possible that NT1 effects on stage N3 sleep were not apparent in our study because 1) we studied stage N3 data collated across the first 8 hours of sleep/ full recordings of sleep rather than comparing successive stage N3 bouts and 2) our subjects were young (mean age 12.9 years). Younger participants in our study showed

more stable stage N3 sleep overall, and, in other work, young people have more slow wave activity and more time in slow wave sleep than adults[37]. Future work could investigate if a temporal dissipation of slow wave power is associated with greater stage N3 sleep stage instability later in the sleep period in pediatric and adult NT1 cohorts.

When participants with NT1 transitioned to wake during sleep, these wake periods often proved to be unstable resulting in rapid return to sleep. Worse survival of wake bouts was only apparent in our full PSG recording analysis so it is likely that individuals with NT1 return to sleep more rapidly from wakings at the end of the night than non-hypersomnia controls. Surprisingly, the median wake bout duration in NT1 did not differ from the subjectively sleepy controls group, possibly because wake bouts were short in both groups (median 1 minute for NT1 and subjectively sleepy controls, Table 3). The finding that survival of wake bouts was worse in NT1 participants is consistent with mouse models of NT1 [13] in which poorer survival of wake bouts was associated with orexin neuronal degeneration compared to wild-type mice. Worse survival of wake bouts in NT1 is also consistent with the clinical effect of orexin antagonist drugs[38], sleep aids used to reduce nocturnal wakefulness to treat insomnia. Clinically, the increase in the number of wake bout that occurs concurrently with the decreased survival of wake bouts likely accounts for the increase in WASO in NT1.

Of all the sleep stages, REM sleep shows the greatest instability in NT1 with shorter but more numerous REM sleep bouts and decreased survival of REM sleep bouts in NT1 participants compared to subjectively sleepy controls. These findings highlight the prominent REM sleep dysregulation that emerges in the setting of orexin neuron loss. [9,

[39] Conversely, Branch et al. showed that mice with induced orexin neuronal degeneration had increased survival of REM bouts compared to baseline in the light period. The differences in REM stability seen in the two studies are likely due to differences in acute vs. chronic orexin loss but may be caused by differences in scoring or species physiology. REM sleep dysfunction also occurs in NT2[14], and we found that REM bout duration and REM bout survival curves were similar in NT1 and NT2 participants, suggesting some overlapping physiology.

Overall, these new analyses detail the problem of intrinsic sleep instability in young people with NT1. Though we did not measure CSF orexin levels in our subjects, prior work has shown that unstable sleep is associated with orexin deficiency in NT1[10, 33, 34]. Just how orexin deficiency fragments sleep remains unknown. Mathematical modeling of mouse sleep/wake neuronal networks reveal that in the absence of orexins, there is reduced inhibition of the sleep-promoting neural population consequently destabilizing the model's sleep/wake flip-flop mechanism.[40, 41] This results in reduced consolidation of sleep with sleep bouts interrupted by brief bouts of wakefulness.

Narcolepsy type 2 shows more stable stage N1 sleep

Our results show that stage N1, a light stage of sleep, is more stable for NT2 participants and may account for the differences in sleep quality between NT2 and NT1 participants. In one survey of patients with self-reported narcolepsy, respondents with NT2 reported less disrupted nighttime sleep but they felt less rested in the morning compared to NT1 respondents.[3] Clinically, it will be important to determine if such

feelings of unrefreshing sleep relate to overly stable stage N1 sleep in NT2 as this might be improved with sedating medications.

The pathophysiological basis of our NT2 findings are unclear. Experiments on rats indicate that a loss of approximately 75% of the orexin neurons decreases CSF orexin levels about 50%[42]. This would not meet the clinical criteria for diagnosing NT1 [6] but this partial loss of orexin neurons may underlie at least some of the cases of NT2. Plausibly, our findings that NT2 participants have more stable stage N1 sleep but do not transition to a full wake state or show sleep stage instability as seen in NT1 participants support differential sleep effects of orexin insufficiency vs. orexin deficiency. It is difficult to create accurate mouse models of NT2 and accurately capture stage N1 sleep to test this theory. Supporting differential sleep phenotypes between NT1 and NT2, Black et al. reported that the sleep/wake of mice with 71% orexin-A positive cell loss resembles that of orexin-intact controls more than mice with near complete neurodegeneration.[43] Conversely, mice with 56.6% orexin-A positive cell loss had sleep/wake phenotypes that were similar to those of mice with near complete orexin-A positive cell loss.

<u>More stable stage N2 sleep in IH</u>

Participants with IH showed better survival of stage N2 sleep bouts but worse survival of stage N3 bouts (in participants ≥ 16 years of age). Coupled with the decrease in stage N3 sleep, this raises the possibility that in IH, the reductions in stage N3 are offset by more stable stage N2 sleep. As stage N3 declines through adulthood[44], this unstable stage N3 sleep may be more pronounced in adults with IH. Short and unstable stage N3 bouts may contribute to the reduced amount of stage N3 percentage of total sleep time

we see in our study and reported in a recent meta-analysis of 9 studies with adult IH patients [18]. In IH, findings of more stable stage N2 may be the consequence of pathologically high sleep pressure, lack of adequate arousal mechanisms or malfunctioning circadian processes. In any case, sleep stability measures could be used as objective endpoints in future studies with larger and older cohorts than ours to study underlying mechanisms and whether an apparent shift from stage N3 to stage N2 sleep contributes to the common symptoms of less restorative sleep, sleep inertia, fatigue and sleepiness in people with IH[6, 45].

CNS hypersomnia conditions interact with developmentally-mediated changes in stability of sleep/wake behavior

In youth with CNS hypersomnias, we found better survival of stage N2 bouts and fewer stage N3 bouts, which is consistent with developmental sleep literature [37, 46]. Importantly, these analyses showed relatively few age interactions, suggesting that most aspects of sleep dysregulation in these hypersomnias are present at all ages. Two exceptions were notable: 1) the stability of REM sleep bouts improved with age, except in NT1 and 2) the stability of stage N3 bouts decreases with age for all groups with even greater age-related decreases in stability for IH participants. In both IH and NT1, longitudinal studies are needed to understand the evolution of sleep stability in the progression to adulthood.

Limitations

There are several limitations to this study that may affect the interpretation of some of our results. The main limitation is reflected by the small sample sizes of the NT2 and IH

groups as well as some demographic groups. It is possible that we are underpowered to detect differences between all groups (and especially NT2 vs. IH) and so we focused our analyses on contrasting CNS disorders of hypersomnolence vs. subjectively sleepy controls. Also, our subjectively sleepy controls are young people with self-reported sleepiness, but they are not healthy controls recruited from the community, possibly limiting our ability to detect group differences. The subjectively sleepy participants were not systematically followed in our clinics to reliably state the cause of their perceived excessive daytime sleepiness. This subjectively sleepy population is a cohort commonly encountered in clinical sleep medicine practice and studied as a reference group in other hypersomnia clinical research studies[9, 47-49]. Thus, our comparisons with subjectively sleepy controls offers broader generalizability and ability to compare findings across studies. Age, sex, and demographic composition varied among the groups as well. In particular, most of the IH participants were female, and older than the overall mean age but this largely reflects the demographics of IH patients which typically begins at a latter age than NT1 or NT2[1, 5]. Nonetheless, our findings in IH may be less applicable to males and younger IH patients. Our study was also limited by confounding due to developmentally-mediated changes and/or time since disease onset. For example, although we considered age effects in the linear and Cox proportional hazards models, we were unable to assess potential effects of pubertal stage due to the absence of Tanner stage data in our retrospective study. Pubertally-mediated changes in sleep/wake behavior are well documented, therefore, studies of CNS hypersomnia conditions that control for pubertal stage are needed. Similarly, we could did not assess the effect of time since disease-onset on sleep/wake behavior because it was collinear

with age. Last, our analyses did not correct for multiple comparisons. For this reason, we report corresponding 95% confidence limits when reporting hazard ratios for sleep stability analysis in addition to the p-values. Our regression analyses included covariate adjustments (age, sex, race) to reduce confounding bias when comparing groups.

Significance and implications

Our work demonstrates that the nocturnal sleep of young patients with NT1, NT2 and IH has many unique features that may aid in diagnosis and might provide insights into the underlying pathophysiology of these disorders. If unique nocturnal sleep profiles are confirmed in larger studies and in adults, diagnostic testing and clinical trials could utilize these objective sleep stability measures. Currently, it is challenging to distinguish between NT2 and IH[23, 24] and these findings may help to establish distinct disease phenotypes or perhaps provide a rationale to collapse them into a single disease if differences do not exist[26]. Machine-learning based approaches using polysomnography data have been leveraged for NT1 diagnosis[50] and our results suggest that further classification schemes could be elicited from other CNS disorders of hypersomnolence using sleep stability analytics especially among the young and potentially closer to disease onset. The differences we find between NT1 and NT2 when compared to subjectively sleepy controls highlight that wake/sleep instability may be more unique to NT1 and possibly a biomarker of orexin deficiency. It would be of great interest to see if sleep stability measures improve with orexin agonists in the NT1 population. The lingering of NT2 in lighter phases of sleep requires further investigation into associations with CSF orexin and would likely need to be studied in human subjects as stage N1

sleep is difficult to assess in mouse models. As a subset of NT2 patients convert to NT1 over time[51], it would be important to assess NT2 sleep longitudinally to determine if a more unstable NT1-like sleep pattern emerges suggesting underlying progressive orexin neuronal loss. Last, future clinical trials could include treatments for sleep instability in NT1 and assess if increasing stage N3 sleep in IH produces more restorative sleep[3, 5]. Overall, we suggest sleep stability measures be incorporated into future research on CNS hypersomnias to help define more homogeneous study populations and provide objective endpoints for future clinical research.

Disclosure Statement: This study was funded by NINDS K23 NS104267-01A1 and Jazz Pharmaceuticals Investigator-initiated grant to Dr. Maski and NSF DMS 1853511 to Dr. Diniz Behn. Authors have the following financial disclosures: Dr. Maski has received prior consulting fees from Harmony Biosciences, Alkermes, and Jazz Pharmaceuticals. She received an investigator initiated grant from Jazz Pharmaceuticals. Dr. Scammell has consulted for Avadel, Consynance, Harmony Biosciences, Idorsia, Jazz Pharmaceuticals, Eisai, and Takeda and has received research grants from Merck, Jazz, and Takeda. All other authors have no financial disclosures.

REFERENCES

1. Ali M, Auger RR, Slocumb NL, Morgenthaler TI. Idiopathic hypersomnia: clinical features and response to treatment. Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine 2009;5:562-568.
2. Thorpy MJ, Krieger AC. Delayed diagnosis of narcolepsy: characterization and impact. Sleep medicine 2014;15:502-507.
3. Maski K, Steinhart E, Williams D, et al. Listening to the Patient Voice in Narcolepsy: Diagnostic Delay, Disease Burden, and Treatment Efficacy. J Clin Sleep Med 2017;13:419-425.

4. Khan Z, Trotti LM. Central Disorders of Hypersomnolence: Focus on the Narcolepsies and Idiopathic Hypersomnia. Chest 2015;148:262-273.
5. Arnulf I, Leu-Semenescu S, Dodet P. Precision Medicine for Idiopathic Hypersomnia. Sleep Med Clin 2019;14:333-350.
6. American Academy of Sleep Medicine. International Classification of Sleep Disorders - Third Edition (ICSD-3). Darien, IL: American Academy of Sleep Medicine, 2014.
7. Roth T, Dauvilliers Y, Mignot E, et al. Disrupted nighttime sleep in narcolepsy. Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine 2013;9:955-965.
8. Sakurai T. The neural circuit of orexin (hypocretin): maintaining sleep and wakefulness. Nat Rev Neurosci 2007;8:171-181.
9. Pizza F, Vandi S, Iloti M, et al. Nocturnal Sleep Dynamics Identify Narcolepsy Type 1. Sleep 2015;38:1277-1284.
10. Sorensen GL, Knudsen S, Jennum P. Sleep transitions in hypocretin-deficient narcolepsy. Sleep 2013;36:1173-1177.
11. Hansen MH, Kornum BR, Jennum P. Sleep-wake stability in narcolepsy patients with normal, low and unmeasurable hypocretin levels. Sleep medicine 2017;34:1-6.
12. Barateau L, Lopez R, Chenini S, et al. Association of CSF Orexin-A levels and nocturnal sleep stability in patients with hypersomnolence. Neurology 2020.
13. Branch AF, Navidi W, Tabuchi S, et al. Progressive Loss of the Orexin Neurons Reveals Dual Effects on Wakefulness. Sleep 2016;39:369-377.
14. Baumann CR, Mignot E, Lammers GJ, et al. Challenges in diagnosing narcolepsy without cataplexy: a consensus statement. Sleep 2014;37:1035-1042.
15. Mignot E, Lammers GJ, Ripley B, et al. The role of cerebrospinal fluid hypocretin measurement in the diagnosis of narcolepsy and other hypersomnias. Arch Neurol 2002;59:1553-1562.
16. Trotti LM. Waking up is the hardest thing I do all day: Sleep inertia and sleep drunkenness. Sleep medicine reviews 2017;35:76-84.
17. Vernet C, Arnulf I. Idiopathic hypersomnia with and without long sleep time: a controlled series of 75 patients. Sleep 2009;32:753-759.
18. Plante DT. Nocturnal sleep architecture in idiopathic hypersomnia: a systematic review and meta-analysis. Sleep medicine 2018;45:17-24.
19. Rye DB, Bliwise DL, Parker K, et al. Modulation of vigilance in the primary hypersomnias by endogenous enhancement of GABAA receptors. Sci Transl Med 2012;4:161ra151.
20. Dauvilliers Y, Evangelista E, Lopez R, et al. Absence of gamma-aminobutyric acid-a receptor potentiation in central hypersomnolence disorders. Annals of neurology 2016;80:259-268.
21. Trotti LM, Saini P, Koola C, LaBarbera V, Bliwise DL, Rye DB. Flumazenil for the Treatment of Refractory Hypersomnolence: Clinical Experience with 153 Patients. Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine 2016;12:1389-1394.
22. Materna L, Halfter H, Heidbreder A, et al. Idiopathic Hypersomnia Patients Revealed Longer Circadian Period Length in Peripheral Skin Fibroblasts. Front Neurol 2018;9:424.
23. Lopez R, Doukkali A, Barateau L, et al. Test-Retest Reliability of the Multiple Sleep Latency Test in Central Disorders of Hypersomnolence. Sleep 2017;40.
24. Ruoff C, Pizza F, Trotti LM, et al. The MSLT is Repeatable in Narcolepsy Type 1 But Not Narcolepsy Type 2: A Retrospective Patient Study. Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine 2018;14:65-74.
25. Fronczek R, Arnulf I, Baumann CR, Maski K, Pizza F, Trotti LM. To Split or to Lump? Classifying the Central Disorders of Hypersomnolence. Sleep 2020.
26. Lammers GJ, Bassetti CLA, Dolenc-Groselj L, et al. Diagnosis of central disorders of hypersomnolence: A reappraisal by European experts. Sleep medicine reviews 2020;52:101306.

27. Perez-Atencio L, Garcia-Aracil N, Fernandez E, Barrio LC, Barios JA. A four-state Markov model of sleep-wakefulness dynamics along light/dark cycle in mice. PLoS One 2018;13:e0189931.
28. Iber C A-IS, Chesson A, Quan S. The AASM Manual for the Scoring of Sleep and Associated Events. Westchester, IL: American Academy of Sleep Medicine, 2007.
29. Littner MR, Kushida C, Wise M, et al. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. Sleep 2005;28:113-121.
30. Murali H, Kotagal S. Off-label treatment of severe childhood narcolepsy-cataplexy with sodium oxybate. Sleep 2006;29:1025-1029.
31. Janssen KC, Phillipson S, O'Connor J, Johns MW. Validation of the Epworth Sleepiness Scale for Children and Adolescents using Rasch analysis. Sleep medicine 2017;33:30-35.
32. Tang S, Jeong JH. Median tests for censored survival data; a contingency table approach. Biometrics 2012;68:983-989.
33. Maski K, Pizza F, Liu S, et al. Defining Disrupted Nighttime Sleep and Assessing its Diagnostic Utility for Pediatric Narcolepsy Type 1. Sleep 2020.
34. Barateau L RL, S Chenini, A Rassu, S Scholz, M Lotierzo, J Cristol, I Jaussent, Y Dauvilliers. Nocturnal Sleep Stability and Cerebrospinal Fluid Orexin-A Levels: Sleep and Wake Bouts. Sleep 2020;43:A285.
35. Antelmi E, Pizza F, Vandi S, et al. The spectrum of REM sleep-related episodes in children with type 1 narcolepsy. Brain : a journal of neurology 2017;140:1669-1679.
36. Khatami R, Landolt HP, Achermann P, et al. Insufficient non-REM sleep intensity in narcolepsy-cataplexy. Sleep 2007;30:980-989.
37. Gaudreau H, Carrier J, Montplaisir J. Age-related modifications of NREM sleep EEG: from childhood to middle age. J Sleep Res 2001;10:165-172.
38. Svetnik V, Snyder ES, Tao P, et al. Insight Into Reduction of Wakefulness by Suvorexant in Patients With Insomnia: Analysis of Wake Bouts. Sleep 2018;41.
39. Christensen JA, Carrillo O, Leary EB, et al. Sleep-stage transitions during polysomnographic recordings as diagnostic features of type 1 narcolepsy. Sleep medicine 2015;16:1558-1566.
40. Diniz Behn CG, Kopell N, Brown EN, Mochizuki T, Scammell TE. Delayed orexin signaling consolidates wakefulness and sleep: physiology and modeling. J Neurophysiol 2008;99:3090-3103.
41. Fulcher BD, Phillips AJ, Postnova S, Robinson PA. A physiologically based model of orexinergic stabilization of sleep and wake. PLoS One 2014;9:e91982.
42. Gerashchenko D, Murillo-Rodriguez E, Lin L, et al. Relationship between CSF hypocretin levels and hypocretin neuronal loss. Exp Neurol 2003;184:1010-1016.
43. Black SW, Sun JD, Santiago P, et al. Partial ablation of the orexin field induces a sub-narcoleptic phenotype in a conditional mouse model of orexin neurodegeneration. Sleep 2018;41.
44. Carskadon MAaDW. Monitoring and staging human sleep, 6th ed. St. Louis: Elseviers Saunders, 2011.
45. L M Trotti JO, D Plante, D Powell, D L Bliwise. Self-Reported Sleep Inertia in the Hypersomnia Foundation Registry. Sleep 2018;41:A228-229.
46. 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;27:1255-1273.
47. Barateau L, Lopez R, Chenini S, et al. Association of CSF orexin-A levels and nocturnal sleep stability in patients with hypersomnolence. Neurology 2020;95:e2900-e2911.
48. Andlauer O, Moore H, Jouhier L, et al. Nocturnal rapid eye movement sleep latency for identifying patients with narcolepsy/hypocretin deficiency. JAMA Neurol 2013;70:891-902.
49. Ferri R, Pizza F, Vandi S, Iloti M, Plazzi G. Decreased sleep stage transition pattern complexity in narcolepsy type 1. Clin Neurophysiol 2016;127:2812-2819.

50. Stephansen JB, Olesen AN, Olsen M, et al. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. Nat Commun 2018;9:5229.
51. Andlauer O, Moore Ht, Hong SC, et al. Predictors of hypocretin (orexin) deficiency in narcolepsy without cataplexy. Sleep 2012;35:1247-1255F.