Digital phenotyping of sleep patterns among heterogenous samples of Latinx adults using unsupervised learning

Center for Health Analytics
Data Science Institute

Data Science Institute
Columbia University

Ipek Ensari,[1] Billy A. Caceres,[2] Kasey B. Jackman,[2] Niurka Suero-Tejeda,[2] Ari Shechter,[3] Michelle L. Odlum,[2] Suzanne Bakken[2]

[1]Columbia University Data Science Institute, [2]Columbia University School of Nursing, [3]Columbia University Irving Medical Center

Background

Sleep health is recognized as a public health concern in the United States. Risk of poor sleep health is higher among certain populations, such as individuals with chronic conditions associated with circadian rhythm disturbance (e.g., HIV). In addition, stigmatized populations, such as racial, ethnic, or sexual and gender minority (SGM) adults, may experience poor sleep health due to chronic exposure to social stressors. There is scarce research using objectively-estimated (e.g., accelerometer-based) data to investigate sleep health patterns and between-individual heterogeneity that could be indicative of sleep disturbance subtypes (i.e.,"phenotypes"). Identification of possible sleep phenotypes constitutes a starting point for developing individualized prevention and treatment strategies, providing precedence for undertaking this investigation.

Aim

This study investigates the research question "Are there homogeneous sub-groups of individuals with distinct sleep patterns who have a risk of worsened sleep outcomes but are otherwise heterogeneous with respect to age, co-morbidities, lifestyle, and self-reported psychological health?". To identify possible phenotypes, we leverage a flexible unsupervised learning technique that relies on the mixture of multivariate generalized linear mixed models (MMGLMMs) that can accommodate inherent data complexities.

Sample

Sample included 494 person-level days from 118 Latinx adults (80% female sex) in Precision in Symptom Self-Management (PriSSM) Center studies at Columbia University investigating sleep health and disturbance among adults across the life span at increased risk for sleep disturbance, including midlife women (N=53), SGM adults (N=39), and adults living with HIV (N=26). Median age was 55.5 years (MAD=15.57, 19-77), median weight was 160.5 lbs (MAD=31.13, 102-335). Participants on average had 4.2 days of data (SD=1.47)

Study Data

- Objective sleep: using Actilife Actigraph accelerometers, self-reported sleep disturbance (PROMIS T-score)
- Raw sleep data collected in 60-sec epochs, validated/processed in Actilife software with Cole-Kripke algorithm
- Potential cluster-relevant sleep variables: total sleep time (TST), time spent in bed (TIB), bed time, awakening time, wake after sleep onset (WASO), number of awakenings after sleep onset (NOA), sleep efficiency (SE; 100×sleep duration/time between sleep onset and awakening time), fragmentation index (FI: 100 × number of groups of consecutive mobile 30-s epochs/total number of immobile epochs)

Data Analysis

Principal Component Analysis (PCA)
- dimensions of variability (i.e., >80%),
- cluster-specific sleep variable relevance

→

Cluster Analysis
- Fit MMGLMMs using PCA-identified variables
- estimate individual component probability (ICP) Markov Chain Monte Carlo-based Bayesian inference

→

Model Fit Assessment
- Penalized Expected Deviance values (PEDΔ)
- Likelihood ratios of posterior deviances (Pdiff)

Results

A 3-cluster resolution provided the best fit for the data (i.e.,PEDΔ~ -21 from 2 to 3-cluster models, P_{diff} ~0.92 model likelihood ratio). Overall cluster membership probability was high (Median individual probability component =0.98). TIB, SE, WASO and FI provided the greatest discrimination without penalty on model complexity and were retained in the final model. There were further statistically significant differences in the other sleep parameters between phenotypes (Table 1).

Table 1. Descriptives on sleep parameters for each phenotype. *p<0.05.

Sleep Parameter	Median (MAD)	Range
TST		
Phenotype 1	5.17 (1.63)	3.14-8.8
Phenotype 2	6.61 (0.53)	4.36-7.38
Phenotype 3*	7.03 (1.04)	2.91-9.78
NOA*		
Phenotype 1	12.0 (5.56)	3-32
Phenotype 2	21.0 (4.82)	14-27
Phenotype 3	16.0 (7.04)	3-30
FI*		
Phenotype 1	9.31 (4.73)	0-20.0
Phenotype 2	15.29 (3.82)	10.00-25.00
Phenotype 3	26.75(11.4)	7.06-62.26
Awakening time		
Phenotype 1*	8:00 (2.22)	04:00-13:00
Phenotype 2	7:00 (0.00)	02:00-09:00
Phenotype 3	7:00 (1.48)	02:30-19:30
PROMIS T-score		
Phenotype 1	45.80 (8.45)	26.90-71.30
Phenotype 2	44.0 (8.23)	35.70-65.60
Phenotype 3	45.20 (10.45)	26.90-71.30

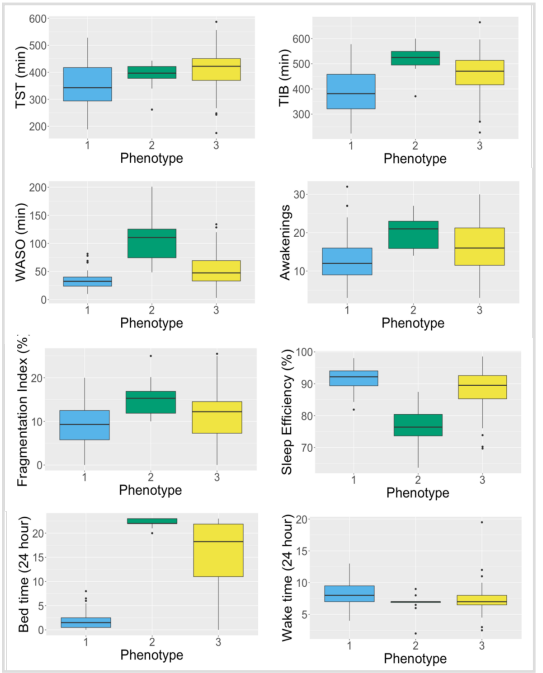

Figure 1. Boxplots showing sleep parameter central tendency measures for each phenotype (N1=64, N2=12, N3=42).

Table 2. Adjusted t-tests comparing sleep parameter variability among phenotypes.*p<0.05,*p <0.0001.**

Comparisons	TIB SD *t* (df)	WASO SD *t* (df)	SE SD *t* (df)
Phenotypes 1-2	6.58 (33.3)***	-2.92 (14.6)*	-2.62 (12.1)*
Phenotypes 1-3	0.49 (93.4)	-2.49 (61.0)*	-2.71 (61.8)*
Phenotypes 2-3	-5.74 (36.1)***	0.71 (26.7)	1.26 (15.4)

Discussion

Digital data-driven approaches with appropriate analytic techniques can aid in early detection and management of sleep problems. Our limited understanding of how sleep patterns unfold among different patient populations hinders the development of tailored interventions. These analyses accordingly constitute a first step to informing the design of culturally-tailored and individualized sleep health interventions.

Acknowledgments: This work is supported by a postdoctoral fellowship from the Data Science Institute at Columbia University (Ensari) and an award from the National Institutes of Health (P30NR016587, MPI: Bakken).