Data-Driven Identification of Comorbidities and Pharmacological Patterns in Patients with Sleep Disorders

Praveen Kumar[1], Kristan A Schneider[1], Fariha Moomtaheen[1], Rajesh Upadhayaya[1], Scott A. Malec[1], Jeremy J. Yang[1], Cristian G. Bologa[1], Yiliang Zhu[1], Mauricio Tohen[2], Gerardo Villarreal[2,3], Douglas J. Perkins[1], Elliot M. Fielstein[4], Sharon E. Davis[4], Michael E. Matheny[4,5], Christophe G. Lambert[1]

[1]University of New Mexico, Department of Internal Medicine, Albuquerque, NM, USA
[2]University of New Mexico, Department of Psychiatry & Behavioral Sciences, Albuquerque, NM, USA
[3]VA New Mexico Healthcare System, Albuquerque, NM, USA
[4]Vanderbilt University Medical Center, Department of Biomedical Informatics, Nashville, TN, USA
[5]Tennessee Valley Healthcare System VA, Nashville, TN, USA

Background

Sleep is a fundamental biological activity essential for overall health and well-being. Sufficient, uninterrupted sleep is critical for maintaining both physical and mental health. Insufficient or disrupted sleep can have detrimental effects, leading to a range of physical and mental health conditions[1,2]. Sleep disorders, i.e., conditions that impact the timing, quantity, or quality of sleep, are associated with the development of numerous health problems, including depression, obesity, type 2 diabetes, heart disease, dementia, and certain types of cancer[1,3,4]. Sleep loss and disorders are among the most prevalent yet often overlooked health concerns. It is estimated that 50 to 70 million Americans suffer from chronic sleep or wakefulness disorders[1,4], and over 100 million Americans of all ages report not obtaining sufficient sleep[5]. Each year, these sleep-related conditions contribute significantly to national healthcare costs. For instance, in the U.S., the annual cost related to undiagnosed sleep apnea alone is estimated to amount to $150 billion; this figure does not include additional costs related to associated health problems, loss in productivity, and accidents, highlighting sleep disorders as a significant public health concern[1].

Given the substantial health and economic burden of sleep disorders, identifying at-risk individuals is essential to mitigate serious health consequences. Existing research has relied primarily on supervised classification methods to predict sleep disorders[2,6,7]. These methods typically treat uncoded patients as negative examples in machine learning (ML) models. However, the absence of a diagnosis code in a health record does not necessarily imply that the person is free of the condition. In fact, underdiagnosis and undercoding is common in healthcare records.

In this study, our objective is to characterize the clinical patterns of patients diagnosed with sleep disorders, allowing us to identify undiagnosed individuals at risk who share similar profiles. Specifically, we aim to identify common comorbidities and medications among individuals diagnosed with any type of sleep disorder by employing the non-negative matrix factorization (NMF) algorithm[8]. This approach could ultimately facilitate earlier intervention and help prevent the severe outcomes associated with sleep disorders.

Materials and Methods

In this study, we utilized the *All of Us* electronic health record (EHR) data mapped to the OMOP Common Data Model[9] for cohort identification and feature selection. To define the study cohort, we selected individuals with at least 730 days of continuous observation beginning on or after January 1, 2017. The sleep disorder phenotype was defined using International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes "F51.*" and "G47.*". In clinical practice, patients with sleep disturbances may receive either of these codes depending on whether their symptoms are considered to have a medical or psychological basis. Consequently, we included both G47 (physiological or medical origin) and F51 (psychological or behavioral origin) codes to define the sleep disorder phenotype.

We identified a total of N=83,610 unique patients with any recorded sleep disorder code. For these individuals, we collected all associated conditions (ICD-10-CM codes only) and drug exposures (RxNorm codes) recorded prior to their first sleep disorder diagnosis, resulting in 41,068 unique covariates. Using these patients and covariates, we constructed a compressed sparse row (CSR) matrix X of size 83,610 × 41,068. In this matrix, each cell was set to 1 if the patient had the corresponding condition or medication, and 0 otherwise.

In general, the NMF algorithm factorizes a non-negative n×m-matrix X into two lower-rank non-negative matrices: a basis matrix W (n×k) and a coefficient matrix H (k×m), such that $X \approx WH$, where $k \ll \min(n,m)$ denotes the number of latent components. We applied NMF to the CSR matrix X with k=4, selected based on minimizing the reconstruction error measured by the Frobenius norm. The coefficient matrix H captures the contribution of each covariate to each latent component (with rows of H representing components and columns representing covariates). For each component, we selected the top 15 covariates to determine the most common conditions and medications associated with different clusters of patients coded for sleep disorders.

Results

Table 1 presents the top 15 covariates with the highest coefficients for each component of the NMF coefficient matrix (H). These components reflect distinct patterns of co-occurring conditions and medication exposures that were prevalent among patients prior to their sleep disorder diagnosis.

NMF Component	Top 15 covariates
Component 1	1) Other Chronic Pain; 2) Anxiety Disorder, Unspecified; 3) Gastro-Esophageal Reflux Disease Without Esophagitis; 4)Low Back Pain; 5) Major Depressive Disorder, Single Episode, Unspecified; 6) 2 Ml Ondansetron 2 Mg; 7) Unspecified Abdominal Pain; 8) Cervicalgia; 9) Acetaminophen 325 Mg; 10) Cough; 11) Fluticasone Propionate 0.05 Mg; 12) 2 Ml Fentanyl 0.05 Mg; 13) Other Fatigue; 14) Other Specified Postprocedural States; 15) Acute Upper Respiratory Infection, Unspecified
Component 2	1) Sodium chloride 9 MG; 2) Fentanyl 0.05 MG; 3) Midazolam 1 MG; 4) Propofol 10 MG; 5) Naloxone hydrochloride 0.4 MG; 6) Lidocaine hydrochloride 20 MG; 7) Potassium chloride 20 MEQ Extended Release Oral Tablet; 8) Diphenhydramine

	hydrochloride 50 MG; 9) Calcium chloride 0.0014 MEQ; 10)1 ML promethazine hydrochloride 25 MG; 11) Dexamethasone phosphate 4 MG; 12) Ketorolac tromethamine 30 MG; 13) Morphine sulfate 4 MG; 14) Rocuronium bromide 10 MG; 15) 50 ML magnesium sulfate 40 MG
Component 3	1) Sodium chloride 9 MG; 2) Fentanyl; 3) Ondansetron 2 MG; 4) Midazolam; 5) Lidocaine; 6) Calcium chloride 0.0014 MEQ; 7) Ondansetron; 8) Propofol; 9) 10 ML sodium chloride 9 MG; 10) Midazolam 1 MG; 11) Acetaminophen 325 MG Oral Tablet; 12) Dexamethasone; 13) Oxycodone hydrochloride 5 MG Oral Tablet; 14) Oxycodone; 15) Hydromorphone
Component 4	1) Essential (primary) hypertension; 2) Hyperlipidemia, unspecified; 3) Type 2 diabetes mellitus without complications; 4) Atherosclerotic heart disease of native coronary artery without angina pectoris; 5) Obesity, unspecified; 6) Mixed hyperlipidemia; 7) Gastro-esophageal reflux disease without esophagitis; 8) Shortness of breath; 9) Chest pain, unspecified; 10) Type 2 diabetes mellitus with hyperglycemia; 11) Atorvastatin 40 MG Oral Tablet; 12) Pure hypercholesterolemia, unspecified; 13) Aspirin 81 MG Delayed Release Oral Tablet; 14) Vitamin D deficiency, unspecified; 15) Hypothyroidism, unspecified

Table 1. Top 15 covariates with the highest coefficients in each latent component of the NMF coefficient matrix (*H*), representing conditions and medications prior to the first sleep disorder diagnosis.

Because NMF assigns patients different membership weights across multiple components, each patient can be partially associated with all components. The dominant component of a patient is the component with the highest weight. Figure 1 shows the absolute frequency of dominant components, i.e., number of patients assigned primary to each component.

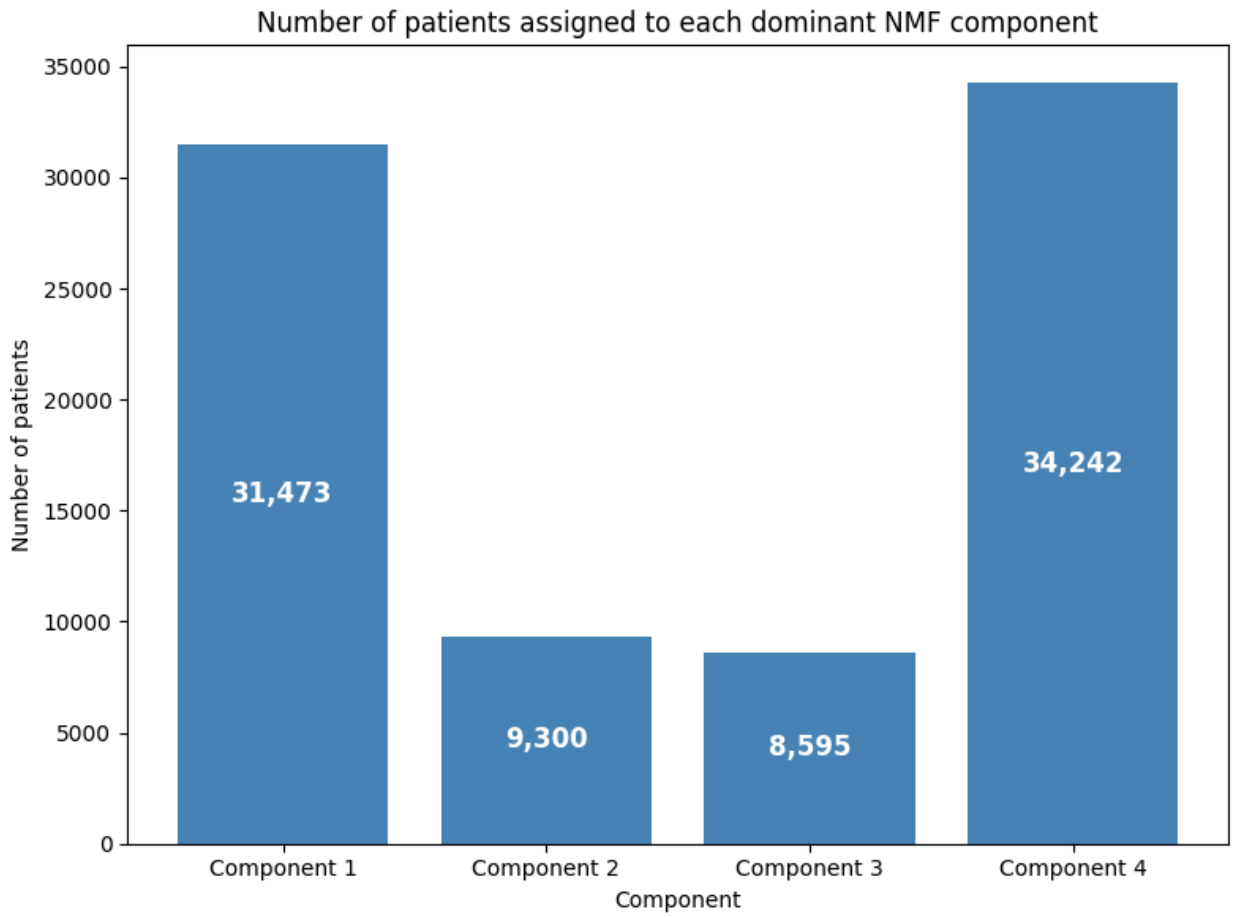

Figure 1: Number of patients assigned to each component based on their largest membership weight.

Discussion and Conclusion

The four identified components revealed meaningful and clinically interpretable patterns of pre-diagnosis conditions and medication exposures among patients who were subsequently diagnosed with a sleep disorder. Chronic pain and mood disorders (such as anxiety and depression)[10] , along with cardiometabolic conditions (including hypertension, hyperlipidemia, diabetes, obesity, and cardiovascular disease)[11], are well-known contributors to poor sleep quality and an increased risk of sleep disorders. Furthermore, exposure to opioids (e.g., fentanyl, morphine), sedatives (e.g., midazolam, propofol), opioid antagonists (e.g., naloxone), and medications commonly used for pain and nausea (e.g., ondansetron, acetaminophen, oxycodone) has been linked to disrupted sleep patterns[12]. These findings indicate that unsupervised methods such as NMF can effectively identify patterns of conditions and medication exposures preceding a sleep disorder diagnosis. While this study applied NMF to conditions and medications, the approach is extendable to any class of covariates to discover important features for a given phenotype. As such, these models can support the early identification of undiagnosed patients with similar clinical profiles who are at elevated risk of developing sleep disorders, thereby enabling earlier interventions to help prevent the progression of severe sleep-related problems.

Acknowledgment
This research was supported by funding from the US National Institutes of Health, specifically, the National Institute of Mental Health grant R01MH129764, the National Library of Medicine grant R00LM013367, and infrastructure support from the National Center for Advancing Translational Sciences grant UL1TR001449. We gratefully acknowledge *All of Us* participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's *All of Us* Research Program for making available the participant data examined in this study.

References

1. Sleep Science and Sleep Disorders: https://www.nhlbi.nih.gov/science/sleep-science-and-sleep-disorders
2. Phan, D. V., Yang, N. P., Kuo, C. Y., & Chan, C. L. (2020). Deep learning approaches for sleep disorder prediction in an asthma cohort. Journal of Asthma, 58(7), 903–911. https://doi.org/10.1080/02770903.2020.1742352
3. Xu S, Faust O, Seoni S, Chakraborty S, Barua PD, Loh HW, Elphick H, Molinari F, Acharya UR. A review of automated sleep disorder detection. Computers in biology and medicine. 2022 Nov 1;150:106100.
4. Institute of Medicine (US) Committee on Sleep Medicine and Research; Colten HR, Altevogt BM, editors. Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem. Washington (DC): National Academies Press (US); 2006. 3, Extent and Health Consequences of Chronic Sleep Loss and Sleep Disorders. Available from: https://www.ncbi.nlm.nih.gov/books/NBK19961/
5. Sleep Disorders: https://my.clevelandclinic.org/health/diseases/11429-sleep-disorders
6. Kim WP, Kim HJ, Pack SP, Lim JH, Cho CH, Lee HJ. Machine learning–based prediction of attention-deficit/hyperactivity disorder and sleep problems with wearable data in children. JAMA network open. 2023 Mar 1;6(3):e233502-.
7. Ma EY, Kim JW, Lee Y, Cho SW, Kim H, Kim JK. Combined unsupervised-supervised machine learning for phenotyping complex diseases with its application to obstructive sleep apnea. Scientific Reports. 2021 Feb 24;11(1):4457.
8. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. nature. 1999 Oct 21;401(6755):788-91.
9. Voss EA , Makadia R, Matcho A, et al. . Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc 2015; 22 (3): 553–64.
10. Morin, Charles M. Ph.D.; Gibson, Douglas Psy.D.; Wade, James Ph.D.. Self-Reported Sleep and Mood Disturbance in Chronic Pain Patients. The Clinical Journal of Pain 14(4):p 311-314, December 1998.
11. Xu, Zhisen BMa; Qian, Xuanfang MDb,c; Xu, Junyao PhDb,*. The cardiometabolic index as a predictor of sleep disorders and mortality: A cross-sectional study. Medicine 104(14):p e42029, April 04, 2025. | DOI: 10.1097/MD.0000000000042029

12. Angarita, G.A., Emadi, N., Hodges, S. et al. Sleep abnormalities associated with alcohol, cannabis, cocaine, and opiate use: a comprehensive review. Addict Sci Clin Pract 11, 9 (2016). https://doi.org/10.1186/s13722-016-0056-7