

Analysis of Sleep and Depression as Risk Factors for Diabetes

Abstract

The purpose of our analysis is to investigate the interaction between sleep and depression frequency by sex as risk factors for diabetes, contributing to the existing research that has established each of these risk factors independently. Using 2018 National Health Interview Survey data from the IPUMS database, we fit four logistic regression models with diabetes status as the response. Although our results in relation to sleep were inconclusive, daily or weekly depression was found to be a significant predictor of diabetes risk, particularly for women. This research adds to the growing evidence of a link between mental and physical health.

Background and Significance

With over 34 million Americans, or more than 1 in 10, diagnosed with the condition, diabetes is one of the most prevalent chronic diseases in the US ¹. Diagnosed diabetics face an overall 60% greater risk of early death than non-diabetics, underlining the importance of identifying the most potent risk factors ². In addition to other modifiable factors such as diet and physical activity, sleep and mental health have been explored as possible mediators of diabetes risk.

Several studies examining sleep and diabetes determined both short (<6 hrs) and long (>8 hrs) sleep duration to be as strongly associated with diabetes risk as traditional risk factors, such as family history of diabetes and overweightness ^{3,4}. Furthermore, a meta-analysis reported that individuals with depression are 37% more likely to develop type 2 diabetes than their depression-free counterparts ⁵. Another review found that this increased risk varied between sexes ⁶.

Despite having well-established relationships with diabetes independently, sleep and depression have not been thoroughly examined simultaneously as predictors of the condition. Because lack of sleep could contribute to depression, and depression could also affect sleep patterns, a potential interaction effect between sleep and depression may be of interest when analyzing the relationship between sleep and diabetes risk. In this report we examine sleep and depression in tandem to diabetes risk.

Methods

Data

This study uses data from IPUMS-NHIS, the National Health Interview Survey from 2018. The data was collected by obtaining a random sample of 35,000 U.S. households and randomly selecting one adult and one child (if any were present) from each. Data analysis was restricted to those 18 years and older. An observational unit comprised of an individual adult in the U.S. surveyed by the CDC’s National Center for Health Statistics. The population to which we intend to generalize our analysis is the non institutionalized adult population of the United States (roughly 300 million individuals total).

Variables

The variables in our analysis include a binary response variable indicating whether or not an individual has diabetes, with average number of hours of sleep per night as the primary explanatory variable, a three-level categorical variable for the frequency an individual reports experiencing depression as the secondary explanatory variable, and sex as the tertiary explanatory variable. The three levels for depression frequency are “often” (defined as daily or weekly), “monthly”, and “rarely” (defined as a few times a year or never). In order to include this categorical variable in our model, we separated it into two binary variables for being depressed monthly or not and being depressed often or not, with being depressed rarely as the reference group.

Our covariates include age, BMI, activity level, income, and race. We created a numerical variable for activity score using the Oncology Nursing Society formula $\text{Weekly leisure activity score} = (9 \times \text{Strenuous}) + (5 \times \text{Moderate}) + (3 \times \text{Light})$ from the Godin Leisure-Time Exercise Questionnaire. We substituted strength exercise for light exercise due to the availability of the IPUMS data, and we think strength training is analogous to light exercise for the purposes of our analysis. For simplicity and based on an analysis of faceted scatterplots, we collapsed the categorical race variable into 3 binary variables indicating black/African American or not, Asian or not, and Native American/Alaska Native or not, with white race as the reference group. For income level, we created two binary variables indicating high income (annual family income over \$75,000 or not) and middle income (annual family income between \$35,000 and \$75,000 or not), with low income (annual family income below \$35,000) as the reference group. Since our data is based on survey responses, all values are self-reported. Additionally, we did not include in our analysis any values that corresponded to non-response, inavailability, or extreme values that had been top or bottom coded.

Analysis

In order to analyze our primary, secondary, and tertiary hypothesis, we fit 4 different logistic regression models with the binary variable for diabetes or not as the response. Our first model includes only our primary and secondary explanatory variables in order to observe these relationships without accounting for the covariates. Our second model includes all of our explanatory variables and covariates, including the interaction terms between hours of sleep and depression frequency. Our third and fourth models represent males and females separately in order to observe differences by sex. Both include all covariates.

In order to test the significance of each coefficient, we performed individual z-tests. In order to evaluate and compare the overall models, we performed likelihood ratio tests and calculated the percentage of concordant pairs for each model.

Results

After filtering out non-applicable observations, we performed our analysis with a sample size of 21,229, with 46.4% males and a median age of 51 years. 1.9% ($n = 407$) of the individuals in the sample are classified as having chronic diabetes. The median number of hours of sleep per night is 7. The majority of individuals (83.78%) reported experiencing depression rarely, with 6.48% experiencing depression monthly, and 10.21% daily or weekly. The median BMI is 27.11, and the median activity score is 896. The income distribution is fairly even, with 38.54% classified as high income, 28.61% classified as middle income, and 32.87% classified as low income. The majority of individuals (81.09%) identified as white, with 5.51% identifying as Asian, 11.97% identifying as black, and 1.43% identifying as Native American or Alaska Native.

Based on individual z-tests, the coefficients for all 3 explanatory variables were significant for the simple model (model 1) ($p < 0.05$), and yet the coefficients for sleep and monthly depression frequency became insignificant once the covariates were added (model 2). For models 3 and 4 (separated by sex), the coefficient for often depression frequency was found to be significant only for females ($p < 0.05$). The likelihood ratio tests for all models yielded significant p-values ($p < 0.05$) indicating that all four models were significant overall. However, it is clear that the simple model (model 1) is much less accurate than the models including the covariates (models 2-4) as shown by the percentages of concordant pairs: 53.6%, 85.6%, 85.7%, and 86.0% respectively.

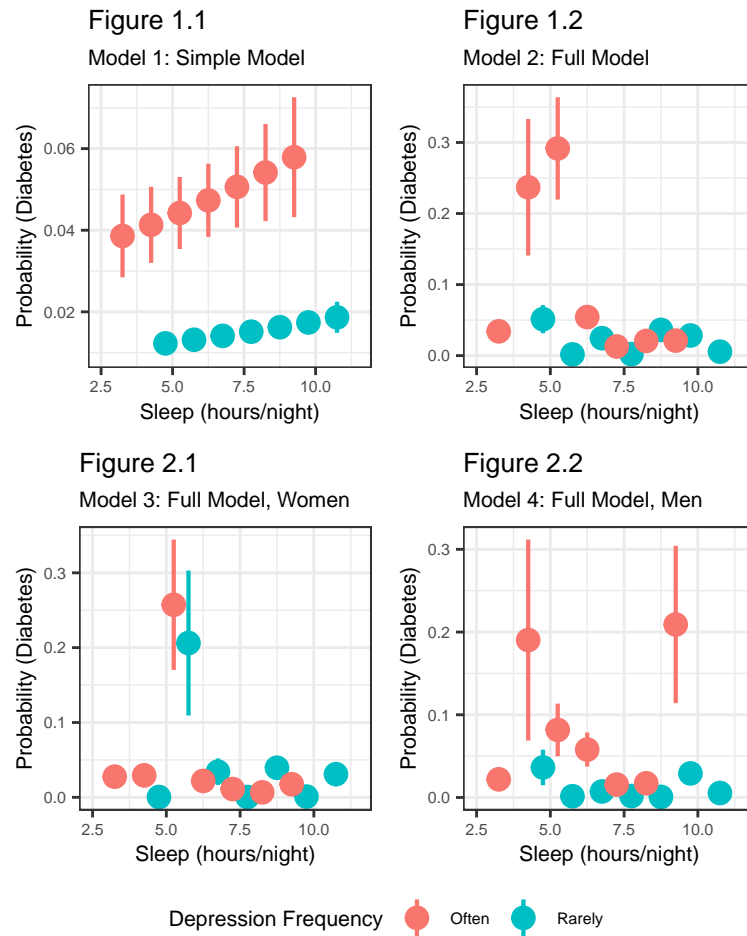
Discussion

Our analysis aimed to assess the relationship between sleep duration and depression in relation to diabetes risk. As shown by the figures below, if we do not account for the covariates, we see a clear positive relationship between hours of sleep and the probability of having chronic diabetes, which is surprising because the majority of the literature has shown a negative relationship between these two variables (figure 1.1). However, once we account for the covariates, there is no clear relationship between sleep and diabetes for either sex (figures 1.2, 2.1, and 2.2). When considering our secondary explanatory variable, depression frequency, without the covariates it is clear that experiencing depression often is associated with a much higher probability of having diabetes compared to experiencing depression monthly or rarely (figure 1.1). This observation supports our hypothesis as well as the pre-existing literature. However, the patterns appear random when accounting for the covariates.

The lack of significance for sleep duration was surprising and initially appeared to not match previous findings on sleep and diabetic risk. However, several studies have reported that while sleep duration is a significant predictor of diabetes, short and long sleeping times are both associated with higher risk, suggesting a U-shaped relationship between sleep and diabetes. This non-linear relationship was not taken into account during our analysis and is likely the reason for sleep duration's insignificance.

Daily/weekly depression was a significant predictor of depression in the overall population; when analyzed separately by sex, this variable was found to be significant only for women. This finding corroborates previous literature documenting sex differences in rates of depression among diabetics.

Probability of Having Diabetes by Sleep and Depression Frequency



Limitations

The limitations of our analysis include the inability to account for additional covariates including diet, genetics, sleep quality and other health conditions, as well as the fact that all values are self-reported, which could result in less accurate data. Future analyses should take into account the fact that abundant literature has identified both short and long sleep as significant risk factors for diabetes. Our analysis was focused on short sleep, but it could be beneficial to divide the numerical sleep variable into categories in order to compare the individual relationships of short, average, and long sleep with diabetes risk. Furthermore, we cannot draw any causal inferences due to the nature of the data. Future studies in a controlled setting ought to examine the direction of the relationship between diabetes, sleep, and depression.

Conclusion

Although our findings in relation to sleep were somewhat inconclusive, it is clear from our analysis that experiencing depression daily or weekly is significantly associated with developing chronic diabetes, particularly for women. This research adds to the growing evidence of a link between mental and physical health.

References

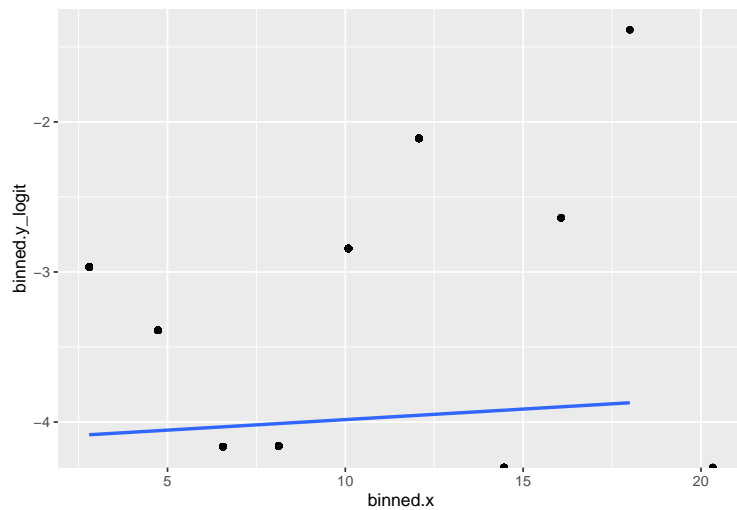
1. “Diabetes | Social Media | Resource Center | Diabetes | CDC,” February 13, 2020. <https://www.cdc.gov/diabetes/library/socialmedia/infographics/diabetes.html>.
2. Anothaisintawee, Thunyarat, Sirimon Reutrakul, Eve Van Cauter, and Ammarin Thakkinstian. “Sleep Disturbances Compared to Traditional Risk Factors for Diabetes Development: Systematic Review and Meta-Analysis.” *Sleep Medicine Reviews* 30 (2016): 11–24. <https://doi.org/10.1016/j.smrv.2015.10.002>.
3. Cappuccio, Francesco P., Lanfranco D’Elia, Pasquale Strazzullo, and Michelle A. Miller. “Quantity and Quality of Sleep and Incidence of Type 2 Diabetes: A Systematic Review and Meta-Analysis.” *Diabetes Care* 33, no. 2 (February 2010): 414–20. <https://doi.org/10.2337/dc09-1124>.
4. Cappuccio, Francesco P., and Michelle A. Miller. “Sleep and Cardio-Metabolic Disease.” *Current Cardiology Reports* 19, no. 11 (2017). <https://doi.org/10.1007/s11886-017-0916-0>.
5. Holt, Richard I. G., Mary de Groot, and Sherita Hill Golden. “Diabetes and Depression.” *Current Diabetes Reports* 14, no. 6 (June 2014): 491. <https://doi.org/10.1007/s11892-014-0491-3>.
6. Ogilvie, Rachel P., and Sanjay R. Patel. “The Epidemiology of Sleep and Diabetes.” *Current Diabetes Reports* 18, no. 10 (August 17, 2018): 82. <https://doi.org/10.1007/s11892-018-1055-8>.

Data Appendix

Checking Regression Assumptions

Linearity

The empirical logit plot seems approximately linear.



Randomness

Obviously, individuals are not assigned BMIs by a spinner model, but since we have a random sample, the randomness assumption is met.

Independence

As established above for the linear regression model, the independence assumption is reasonable for this data set.

Checking Collinearity between Explanatory Variables

	chronic	depfreq_often	depfreq_monthly	HRSLEEP	AGE	BMI	activity_score
chronic	1.00	0.07	0.01	0.01	0.11	0.10	0.10
depfreq_often	0.07	1.00	-0.09	-0.06	-0.00	0.06	0.11
depfreq_monthly	0.01	-0.09	1.00	-0.02	-0.06	0.01	-0.01
HRSLEEP	0.01	-0.06	-0.02	1.00	0.10	-0.05	0.03
AGE	0.11	-0.00	-0.06	0.10	1.00	0.02	0.26
BMI	0.10	0.06	0.01	-0.05	0.02	1.00	0.13
activity_score	0.10	0.11	-0.01	0.03	0.26	0.13	1.00
SEX	-0.01	0.04	0.05	0.00	0.04	-0.02	0.08
race_asian	0.00	-0.04	-0.02	-0.02	-0.06	-0.12	-0.02
race_black	0.04	-0.01	-0.02	-0.02	-0.04	0.09	0.06
race_na	0.03	0.01	-0.02	-0.01	-0.02	0.03	0.00
income_high	-0.07	-0.13	-0.03	-0.01	-0.08	-0.08	-0.23
income_middle	-0.01	-0.01	0.01	-0.00	-0.00	0.05	0.02

	SEX	race_asian	race_black	race_na	income_high	income_middle
chronic	-0.01	0.00	0.04	0.03	-0.07	-0.01
depfreq_often	0.04	-0.04	-0.01	0.01	-0.13	-0.01
depfreq_monthly	0.05	-0.02	-0.02	-0.02	-0.03	0.01
HRSLEEP	0.00	-0.02	-0.02	-0.01	-0.01	-0.00
AGE	0.04	-0.06	-0.04	-0.02	-0.08	-0.00
BMI	-0.02	-0.12	0.09	0.03	-0.08	0.05
activity_score	0.08	-0.02	0.06	0.00	-0.23	0.02
SEX	1.00	-0.01	0.04	0.00	-0.06	-0.02
race_asian	-0.01	1.00	-0.09	-0.03	0.04	-0.02
race_black	0.04	-0.09	1.00	-0.04	-0.13	-0.00
race_na	0.00	-0.03	-0.04	1.00	-0.05	-0.01
income_high	-0.06	0.04	-0.13	-0.05	1.00	-0.50
income_middle	-0.02	-0.02	-0.00	-0.01	-0.50	1.00

The largest magnitude correlation in the correlation matrix above is 0.50. Collinearity therefore appears to not be a concern.

Summary Statistics (Post-data wrangling)

	variable	min	max	median	mean	std_dev
1	HRSLEEP	1	22	7.00	7.07	1.42
2	BMI	14.98	85.78	27.11	28.17	6.38
3	AGE	18	85 (top coded)	51.00	50.86	18.18
4	activity_score	17	1615	896.00	862.11	593.64

	variable	yes	no
1	chronic	407 (1.9%)	20822 (98.1%)
2	depfreq_often	2167 (10.21%)	19062 (89.79%)
3	depfreq_monthly	1376 (6.48%)	19853 (93.52%)
4	SEX	(female) 11369 (53.6%)	(male) 9860 (46.4%)
5	income_high	8181 (38.54%)	13048 (61.46%)
6	income_middle	6073 (28.61%)	15156 (71.39%)
7	race_black	2542 (11.97%)	18687 (88.03%)
8	race_asian	1169 (5.51%)	20060 (94.49%)
9	race_na	303 (1.43%)	20926 (98.57%)

Full Logistic Model Coefficient Summary

	coeff	value	LCI	UCI	p_val
1	HRSLEEP	0.01	0.94	1.09	0.72
2	depfreq_often**	1.31	1.54	8.94	0.00
3	depfreq_monthly	0.33	0.34	5.82	0.65
4	I(HRSLEEP * depfreq_often)	-0.06	0.84	1.07	0.36
5	I(HRSLEEP * depfreq_monthly)	0.04	0.86	1.26	0.64

** = significant at the 0.01 level

Likelihood Ratio Test

	model	degf	log_likelihood	chisq	p_val
1	Full	15	-1635.1	—	—
2	Control	1	-2012.5	754.74	<2.2e-16