Does Sleep Deprivation Put Your Health At Risk?

Authors: Jason Chen, Shahreer Al Hossain

Professor Abdelwahed

SS-340-1 Cause and Effect

Introduction

A: About Our Question

We wanted to investigate the relationship between sleep and health. So we asked the question: does sleeping less cause you to become less healthy? This particular question interested us because we are always told about how important sleep is for you. Additionally, from our own experiences, we often have to give up on sleeping well in order to finish an important assignment and meet a deadline. There has been a lot of literature discussing the impacts of not sleeping enough on people's health too. For example, the CDC website lists all the negative results of not sleeping enough. It claims that sleep deprivation can cause poor health and even early death"[1] The CDC also recommends that adults sleep at least 7 hours and teenagers should sleep at least 8 hours. Often Americans (including us) are unable to meet this minimum. It's estimated that sleep deprivation costs over $300 billion dollars in losses in the United States each year.[2] This tells us that sleep is a major issue in the United States because it is both dangerous for health and as a consequence it results in economic loss.

We wanted to find out for ourselves if we could also observe the same claims that medical literature says about sleep deprivation. Since running a study or experiment is extremely costly and time consuming, we are going to analyze our problem using data from the CDC BRFSS system. We will process the data to find the causal relationship between sleeping with mental and physical wellbeing.

---

[1] CDC. "Sleep for Good Health." *Centers for Disease Control and Prevention*, 19 Sep. 2022.
[2] Hafner M, Stepanek M, Taylor J, Troxel WM, van Stolk C. Why Sleep Matters-The Economic Costs of Insufficient Sleep: A Cross-Country Comparative Analysis. Rand Health Q. 2017 Jan 1;6(4):11. PMID: 28983434; PMCID: PMC5627640.

B: Information About the Data Set Used

As previously mentioned, we are going to use the CDC Behavioral Risk Factor Surveillance System (BRFSS) data of 2018. Notably, we made sure to avoid 2021 because it is missing the sleep variable and also avoided 2020 because of the pandemic. We chose 2018 because it is the most recent data without being during COVID so it will not have biases or different results due to COVID. This data is collected through a telephone survey. A government employee calls phone numbers randomly and then asks questions to the caller which they reply with. The dataset comprises over 400,000 responses. Importantly, this data excludes several categories of people. For example, looking in the survey questions provided, we see that the survey rejects all people under the age of 18. This means we will only be analyzing adults in our dataset. Additionally, it excludes homeless people because it requires the person to answer yes when asked if they live in a private home or in college housing. As such when we interpret our results, we make sure to understand the limitations of our dataset when making claims and also we will address any issues with our data not being a completely random sample of the population.

C: Data Processing

Importantly, the data we get from the CDC BRFSS is not cleaned for us. They provide us with a code book of all the variables and an ASCII or SAS file of the dataset. We first used R to import the SAS file and export a CSV file. We then did all of our analysis using such CSV files in Python 3. We used the codebook to find what the values of each of the variables meant. When we found invalid data such as the question not being answered or the question being answered with unrealistic values we made sure to drop the values. We chose to simply drop the invalid

data points as that would be the simplest method and still yields relatively good accuracy. This is because even after dropping the values we still have a large sample size of over several hundred thousand and we want to weight all data points equally for simplicity. Additionally, some variables could be answered in two different ways which meant we would have to remap the values so it would make sense. Many of the categorical variables were remapped into a dummy variable of 0 and 1 to simplify the interpretation of the regression results. The Table in Appendix C shows how we processed each variable.

Then we ran a simple OLS regression because our dependent variables (bad mental and physical health days in the past 30 days) and our independent variables (sleep hours) were all discrete variables. We made sure to select as many good controls as possible and a few useless controls to improve our results and prove that there is a causal relationship between sleep and health.

Summary of Data

A: Summary Table

Summary Statistics Table					
Count = 314694					
Variable	**Mean**	**STDev**	**Min**	**Median**	**Max**
Number_of_days_drinks	5.405	8.586	0	1	30
Income	0.518	0.5	0	1	1
Weight	182.342	45.716	50	177	663
Age	0.329	0.47	0	0	1
Age Categorical	7.49	3.443	1	8	13
Gender	0.484	0.56	0	0	9
Race	0.77	0.421	0	1	1
Asthma	0.861	0.346	0	1	1
CHD or MI	0.089	0.285	0	0	1
Exercise	0.767	0.423	0	1	1
Mental Health Bad	**3.623**	**7.827**	**0**	**0**	**30**
Physical Health Bad	**4.192**	**8.649**	**0**	**0**	**30**
Smoker	0.433	0.495	0	0	1
Cancer	0.095	0.294	0	0	1
Diabetes	0.142	0.349	0	0	1
Kidney	0.036	0.187	0	0	1
Depressive	0.19	0.392	0	0	1

Stroke	0.04	0.196	0	0	1
Sleep	**7.022**	**1.458**	**1**	**7**	**24**
Bronchitis	0.081	0.272	0	0	1
Married	0.533	0.499	0	1	1

The table tells us about the average person surveyed after dropping invalid data points. On average, people in the sample slept 7.022 hours. The sample had an average of 3.623 bad mental health days in the last 30 days. Additionally, the sample had an average of 4.192 bad physical health days in the last 30 days. The average person has drunk 5.405 days in the last 30 days and the minimum drinking days in the past 30 days is 0 and the maximum is 30. The income variable is a dummy and the average is 0.518 which means that income below and above the median is about the same since that is how we re-coded the variable. The average weight was 183.24 pounds with a minimum weight of 50 lb and a maximum weight of 758 lb. Age is a dummy variable so the average of 0.329 does not tell us what is the average age of the sample. The gender variable is a dummy and the average is 0.484 which means there are more women than men. Race is another dummy with an average of 0.77 which tells us most of the sample is white. Asthma is a dummy variable and since the average is 0.861 it means that 86% do not have asthma. Heart Disease/Heart Attack is a dummy variable and the average of 0.089 means that only 8.9% of the sample has heart disease or suffered from a heart attack. Exercise is a dummy and the average of 0.767 means that 77% of people have exercised in the last 30 days outside of their job. Smoking is a dummy variable which we divided such that the average of 0.433 means that 43% of the sample never smoked before. Cancer is a dummy which shows that

9.5% of the sample has had some kind of cancer before. Diabetes is a dummy which shows that 14.2% of the sample has had diabetes. Kidney disease is another dummy and it shows that 3.6% of the sample has kidney disease. Depression is another dummy and 19% of our sample has it. 4% of our sample has suffered a stroke before. 8.1% of the sample has bronchitis which is a dummy variable. Marriage is another dummy variable and 53.3% of the sample is married.

B: Discussion About Data

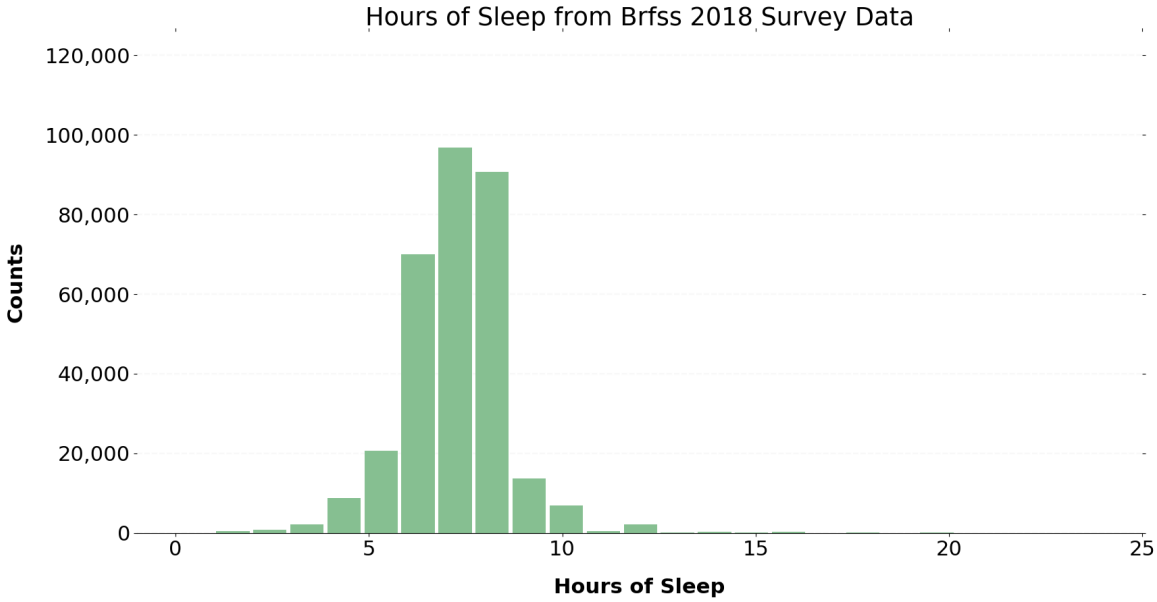

Figure 1: Distribution of Sleep Hours in Data Set

The distribution of the sleep hours in figure 1 appears similar to the normal distribution while it is centered at around 7 hours. The mean is 7.033 and the median is 7 which shows that the average is not affected by outliers. All the sleep values fall between 24 and 1 hours as nobody can survive without any sleep. There are some values very far from the median but there are so few. The standard deviation of about 1.5 hours shows that there is actually a large variance in how much people sleep and there is a substantial amount of people sleeping less than the recommended. Having this spread is good as it means we will have more data points of differing hours of sleep and we can observe their health as a result of their different hours of sleep.

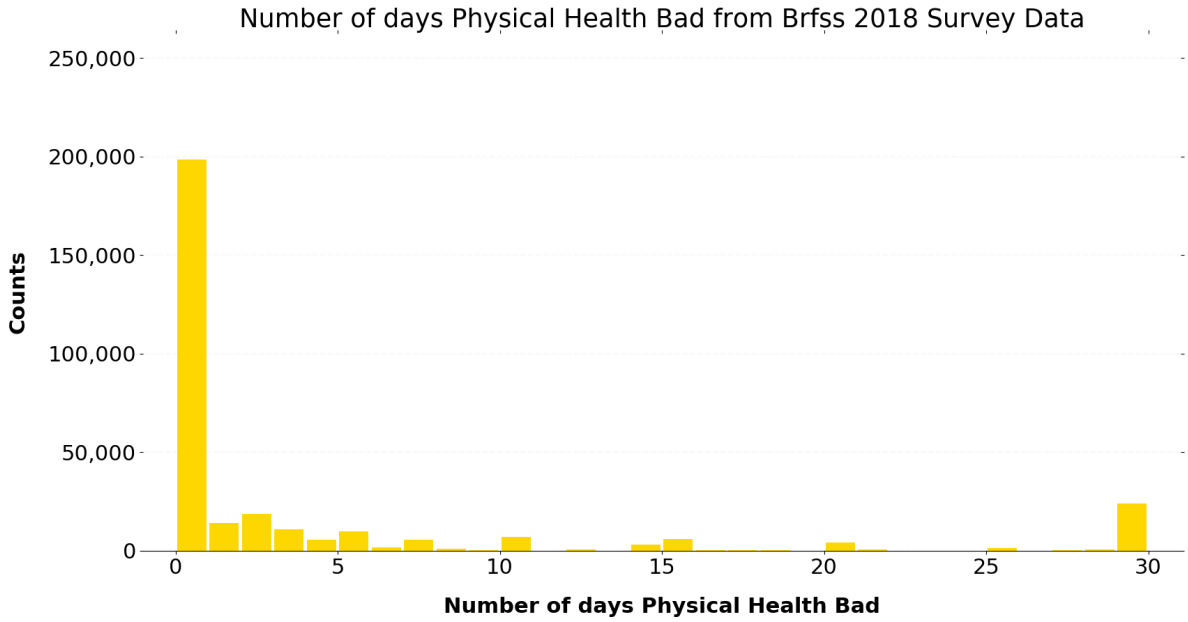

Figure 2: Distribution of Bad Physical Days in the Last 30 days

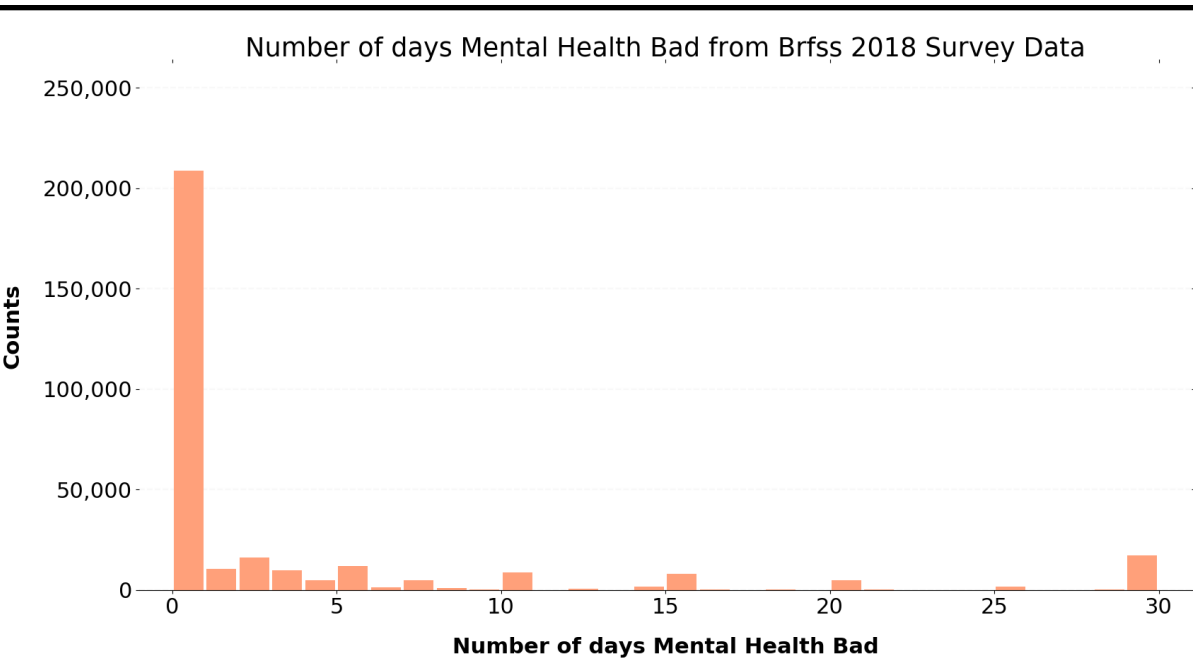

Figure 3: Distribution of Bad Mental Health Days in the Last 30 Days

We can see that there is substantial difference in the distribution for the bad health days as compared to the sleep distribution in figure 1. This is reflected in the summary statistics

table. The median days mental health bad is 0 and the median days bad physical health is also 0. However, the mean is different as it is 3.6 and 4.1 bad health days for mental health and physical health respectively. This means that the data is not normally distributed as it is skewed left towards 0 bad health days. Additionally, their standard deviations are 7.8 and 8.6 days respectively. This means there is a broad spread of how many bad health days there are. This isn't an issue for us since it is expected that some people will have worse or better health and we will control for the covariates. However, we did not drop the values. This is because we are interested in the effect of sleep deprivation on the average person. It doesn't make sense to drop a lot of the values just because they are skewed to the left since they do represent the average person.

In figures 2 and 3, we can see the distribution of all the bad mental and physical health days in the past 30 days. It shows why there is the difference between the mean and median. Also it is notable to see that there clusters at multiples of 5 because it is not expected for people to exactly know the precise number of days they don't feel well in the past 30 days. This is not an issue because a person's estimate will not be that wrong by an order of magnitude. And since they are estimating it's expected that they will underestimate and overestimate about the same amount so it would not introduce much bias. We will still be able to prove a causal relationship between sleep deprivation and bad health days because their estimations are random rather than a biased estimation.

Regression Strategy

At first, the Brfss dataset was split into two datasets based on how long a person slept. According to the summary statistics, on average, a person in the US sleeps for 7.022 hours. Thus, if a person sleeps less than 7.022 hours, the person is designated to be part of the group that does not get enough sleep. On the other hand, if a person sleeps more than 7.022 hours, the person is designated to be part of the group that gets enough sleep.

Table of Covariates for 2-Sample Hypothesis Test

Covariate	Mean (Full Data Set)	Mean (Not Enough Sleep)	Mean (Enough Sleep)	P Value (T-test)
Income (binary)	0.518	0.539	0.483	0.00
Gender (binary)	0.484	0.493	0.468	0.00
Exercise (binary)	0.767	0.775	0.752	0.00
Smoker (binary)	0.433	0.431	0.436	0.006
Weight (pounds)	182.342	184.008	179.442	0.00
Number of days a person drinks in past 30 days (days)	5.405	5.362	5.479	0.0002

In theory, one should be able to identify the effect of sleep by comparing the unadjusted difference in mean number of days mental health bad and mean number of days physical health bad between the two groups (Not Enough Sleep and Enough Sleep), when there is no selection bias between the sample of smoking mothers and non-smoking mothers. In other words, to be able to make a conclusion about the causal effect of sleep on mental health and physical health,

one must control for everything that is correlated with sleep and mental health and physical health.

For all the covariates included in the table above, the p-values obtained from the 2-sample hypothesis test are less than 0.05. Since a p-value of less than 0.05 is usually considered to be statistically significant, we can say with 95% confidence that the means for people who get enough sleep and people who do not get enough sleep are different. This proves that we can not assume there is no selection bias between the sample of Not Enough Sleep group and Enough Sleep group.

Therefore, using a simple linear regression, we first try to estimate the impact of the amount of sleep a person gets on their mental health and physical health. Then, we try to control for all the good controls provided in the data set to diminish bias due to omitted factors that vary from person to person. Additiontionally, we control for several useless controls to reduce the standard error of the OLS regression analysis and reduce the effect of heteroskedasticity along with using robust standard errors.

The following two regression equations are being estimated:

$$Number\ of\ days\ Mental\ Health\ Bad_i = \beta_0 + \beta_1 Sleep_i + \beta_2 Number\ of\ days\ drinks_i + \beta_3 Weight_i + \beta_4 Age_i + \beta_5 Exercise_i + \beta_6 Smoking_i + \beta_7 Income_i + \beta_8 Race_i + \beta_9 Gender_i + \delta X'_i + \epsilon_i$$

$$Number\ of\ days\ Physical\ Health\ Bad_i = \beta_0 + \beta_1 Sleep_i + \beta_2 Number\ of\ days\ drinks_i + \beta_3 Weight_i + \beta_4 Age_i + \beta_5 Exercise_i + \beta_6 Smoking_i + \beta_7 Income_i + \beta_8 Race_i + \beta_9 Gender_i + \delta X'_i + \epsilon_i$$

In this regression equation, the treatment variable, $Sleep_i$ is the number of hours of sleep a person gets in the US . $Number\ of\ days\ Mental\ Health\ Bad_i$ and

$Number\ of\ days\ Physical\ Health\ Bad_i$ is how many days in the past 30 days the individual's mental health was not good and the individual's physical health was not good respectively. The regression also controls for many covariates that are most likely responsible for omitted variable bias such as the number of days in the past 30 days a person drinks, if the person is a smoker or not, if the person is high income or not, if the person is participates in any physical exercise or not, the weight of the person, etc. The regression equations also controls for various demographic variables such as race, gender, and age, which would be useless controls in our regression analysis. The regression also controls for many physical health issues, mental health issues, and diseases, which are good controls (X'_i) in this regression analysis. This is because physical health issues, mental health issues, and diseases are not only correlated with a person's physical health, but they are also correlated with the amount of sleep a person gets.

$X'_i = [Asthma_i,\ CHD\ or\ MI_i,\ Cancer_i,\ Diabetes_i,\ Kidney\ disease_i,\ Depressive\ disorder_i,$

$Stroke_i,\ Bronchitis_i]$

When the effect of one independent variable depends on the level of another independent variable, we have an interaction, and an interaction term should be included in the regression equation. However, in our regression analysis, we do not include interaction terms between two independent variables because we assume that no two independent variables depend on each other. Looking at the list of variables in Appendix C, it is a reasonable and justifiable assumption.

Results

A: Physical Health and Mental Health OLS Regression Results

OLS Regression Coefficients Before and After Controls				
Standard Errors in Parenthesis, Complete Results in Appendix				
	Physical Health		Mental Health	
	Before	After	Before	After
Intercept	6.5022 (0.110)	7.0215 (0.13)	8.0433 (0.106)	7.5656 (0.119)
Sleep	-0.3289 (0.015)	-0.2356 (0.013)	-0.6295 (0.015)	-0.4252 (0.012)
Income		-1.5365 (0.03)		-1.3935 (0.027)
Gender		-0.1958 (0.029)		-0.3021 (0.027)
Exercise		-3.099 (0.043)		-1.0865 (0.035)
Smoker		0.6467 (0.03)		0.6384 (0.026)
Weight		0.0065 (0.00037)		0.0005 (0.00033)
Number of days drinks		-0.0267 (0.002)		0.0147 (0.002)
Heart Attack or Heart Disease		2.345 (0.069)		0.443 (0.0552)
Diabetes		1.6906 (0.052)		0.2012 (0.041)

From the regression results, it can be observed that if the average number of hours of sleep a person gets in a 24 hour period increases by 1 hour, the number of bad physical health days during the past 30 days decreases by 0.3289 days on average. Furthermore, when the average number of hours of sleep a person gets in a 24 hour period is 0 hours, the number of bad physical health days during the past 30 days is on average 6.5022 days. These values can be visualized very easily using the plot below:

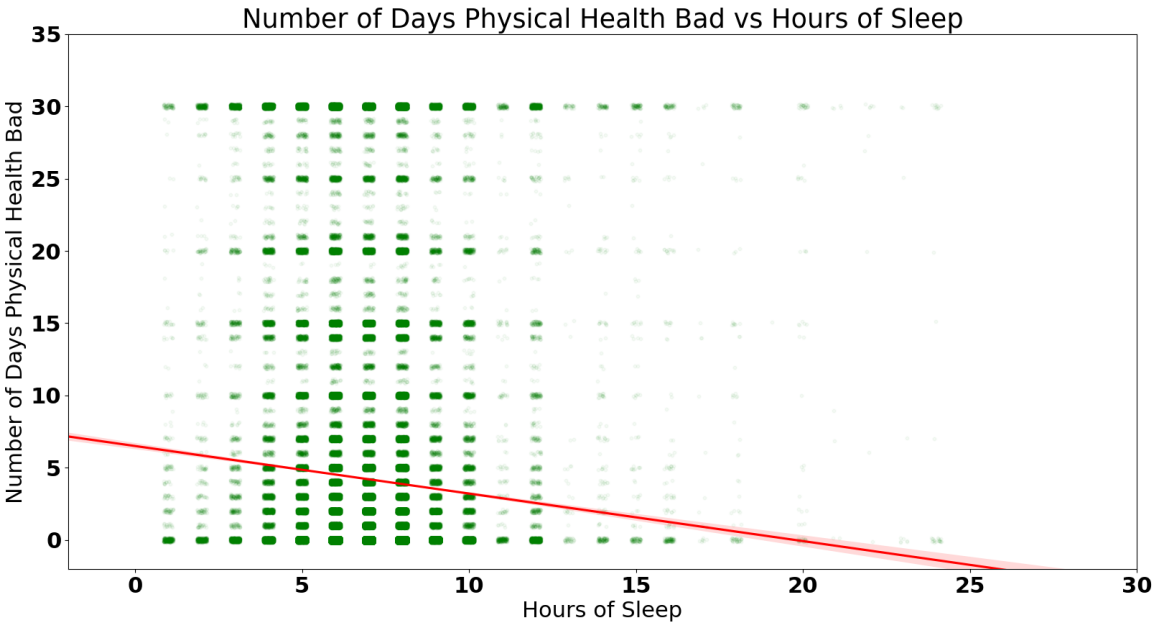

Figure 4: OLS Regression of Sleep On Bad Physical Health Days

It can also be observed that if the average number of hours of sleep a person gets in a 24 hour period increases by 1 hour, the number of bad mental health days during the past 30 days decreases by 0.6295 days on average. Furthermore, when the average number of hours of sleep a person gets in a 24 hour period is 0 hours, the number of bad mental health days during the past 30 days is on average 8.0433 days. Similarly, these values can be visualized very easily using the plot below:

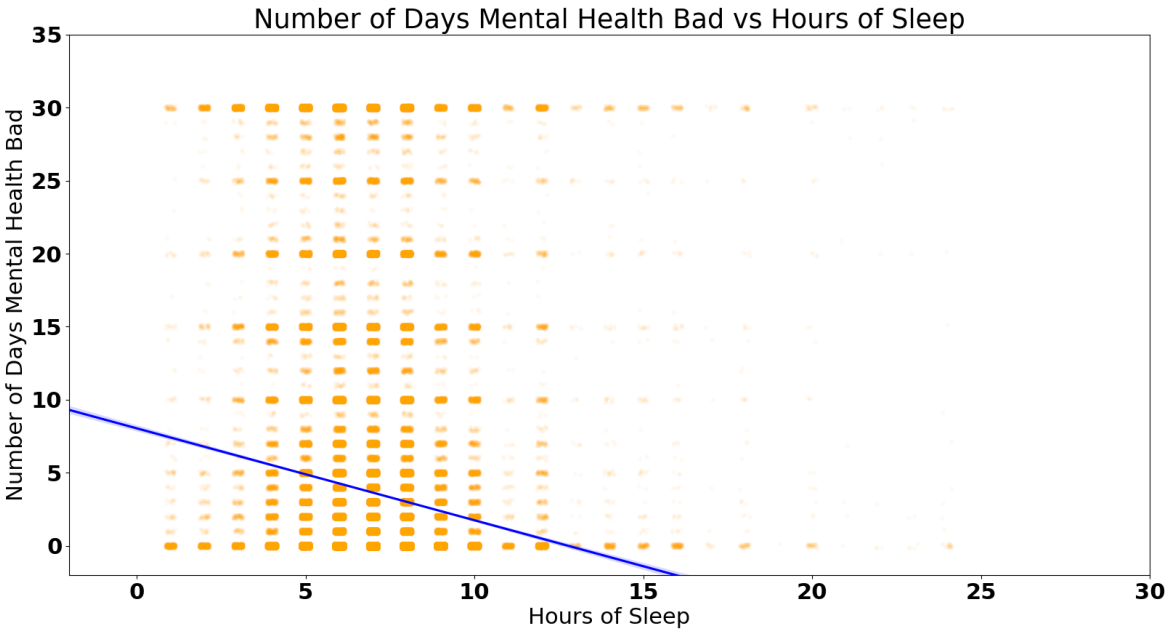

Figure 5: OLS Regression of Sleep On Bad Mental Health Days

If the average number of hours of sleep a person gets in a 24 hour period increases by 1 hour, the number of bad physical health days during the past 30 days decreases by 0.2356 days on average, holding fixed all the control variables mentioned above including age, race, income, etc. If the average number of hours of sleep a person gets in a 24 hour period increases by 1 hour, the number of bad mental health days during the past 30 days decreases by 0.4252 days on average, holding fixed all the control variables mentioned above and including age, race, and income, etc. Almost all of the beta coefficients of the control variables are significantly different from 0 because the P-values are extremely small (less than 0.05).

Discussion

We know that regression analysis can be used to describe correlations between two variables in the dataset. However, we also care about the causal effect between the amount of sleep a person gets and the number of days in the past 30 days a person's mental and physical health was bad, not necessarily just the correlations between them. We know correlation does not necessarily imply causality. If we can take all variables that could possibly affect our outcome variable (number of day mental health and physical health bad) and hold everything constant except for the amount of sleep a person gets and vary the amount of sleep and if our outcome variable changes, then there is a causal effect. But, there are always some variables we're not observing. To solve this issue, in social science, we can make the assumption that if we observe and control for enough variables and then the treatment (in our case not getting enough sleep and getting enough sleep) is as good as randomly assigned. This is known as the selection on observables assumption[3]. For this project, we have measured almost all possible variables that could be confounding factors. ***Assuming we have observed enough variables, we can say the correlation between the amount of sleep a person gets and the number of days in the past 30 days a person's mental and physical health was bad, obtained from OLS regression analysis, is actually the causal effect.***

[3] https://scholar.princeton.edu/sites/default/files/jmummolo/files/selection_on_observables_jm.pdf

Choice of Controls

We chose a variety of both good and useless controls while avoiding bad controls. We wanted to make sure that our regression was analyzing the effect of sleeping rather than any other omitted variable bias which caused the change in health. This meant we needed to find good controls which were correlated with sleeping but not caused by it. We thought income would be a good control because busy and stressed low income people would sleep less. That means increasing sleep would mean the person is more likely to have higher income and have fewer bad health days. This led us to predict a negative omitted variable bias which is also reflected in the results.

We used a similar train of thought for the other variables we selected for good controls. A lot of diseases such as asthma, cancer, diabetes, kidney disease, depression, stroke, and bronchitis. These diseases would be correlated with sleeping but also obviously being sick correlates with health. Diseases aren't also caused by sleeping so they aren't bad controls. For example if someone has cancer, there is a higher chance they will sleep more and we will want to get rid of hits omitted variable bias. They make for good controls as they are correlated with the independent and the dependent variable.

We also selected several useless controls such as gender and race. These will be correlated with the dependent but not with the independent. We include these because they decrease the standard errors which is good and improves the significance of our results.

Confounding Factors and Biases

We made sure to avoid the 2020 data year because that was when COVID started which would skew our results. People suffering from COVID and the economic impact of COVID may show a different rate of bad health days for the amount of sleep. We are interested in the normal impact of sleep rather than during an unprecedented pandemic so we avoided 2020 data and used 2018 data instead. This avoids anything about COVID. It is also still fairly recent so its results would not be out of touch with our current time. We initially considered using 2021 data but that did not have the data for sleep so it was useless to us.

Additionally to avoid selection bias, we make sure that the data collection is random. Since the data is collected by calling phone numbers in the US randomly, we can expect that there isn't a bias in choosing who to survey. There may be some bias in who responds to the survey though. For instance, the survey only selects people who live in a private home or in a college dorm. This isn't a concern for us because we are interested in what the average person's health impacts from sleeping less. Only a tiny 0.2% of people in the US are homeless which means we can safely assume that excluding them is not changing our results.[4] Additionally, there is no bias from having a phone or not having a phone because almost everyone in the US is able to own a phone. Notably, an issue is that the survey requires that the surveyee responds from a home landline phone. It is becominging increasingly rare for people to own a home landline phone, especially younger people. This is reflected in figure 6 which shows the age distribution. We address the issue by controlling for age so we look at the specifically the result

[4] *Homelessness statistics in the US for 2021 | Policy Advice*. Accessed 17 Dec. 2022.

of sleeping rather than age. Additionally, the CDC recommends the same amount of sleep (7 hours)[5] for all adults so the age bias will not matter nor impact our findings about sleep.

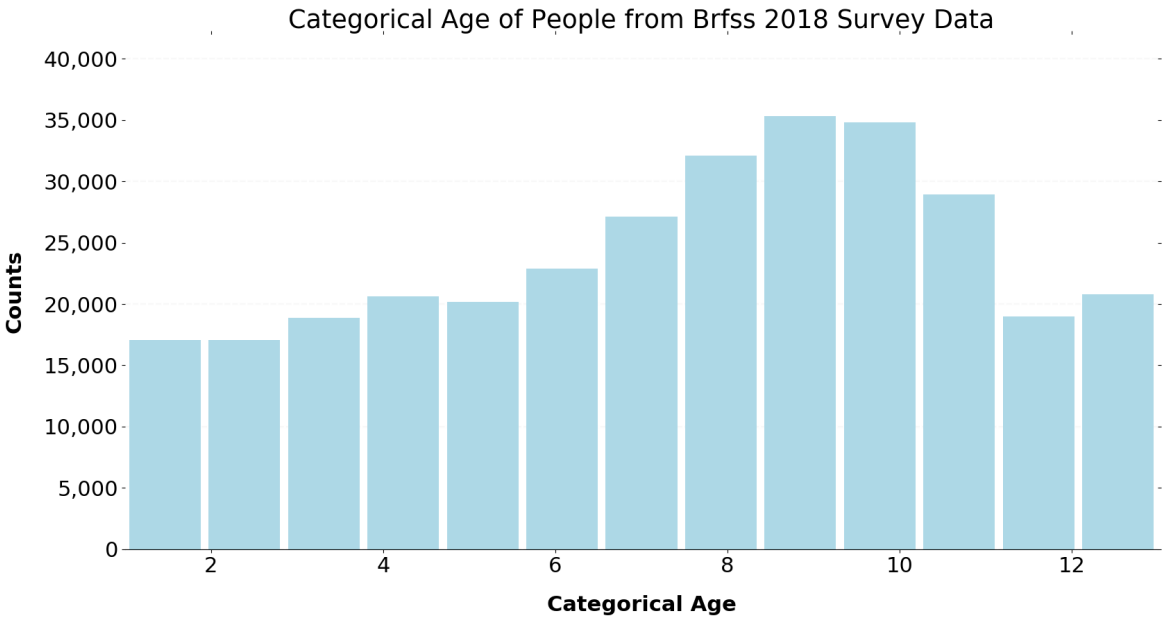

Figure 6: Distribution of Age by Category

[5] CDC. "Sleep for Good Health." *Centers for Disease Control and Prevention*, 19 Sep. 2022.

Conclusion

Our results lead us to claim that sleeping less negatively impacts your health. This is because we were able to get rid of our omitted variable bias and confounding factors by using all the major good controls provided in the Brfss 2018 Dataset. Additionally, our data is adequately randomly selected for there to be an insignificant amount of bias for our analysis. We were able to control for socioeconomic status, chronic illnesses, and age which are all related to sleep and as such we are only looking at what is the impact of sleeping on one's mental and physical health. We saw that sleeping more by 1 hour will result, on average, in a 0.23 reduction in bad physical health days and a 0.42 reduction in days of bad mental health. These results are significant on the 95% confidence level, as both have a P-value of less than 0.05. We were able to answer our original research question. Conclusively, we found that sleep causes your health to improve, and sleeping less causes your health to worsen.

Appendix

A: Regression Table for Physical Health

Regression Table for Physical Health				
OLS Robust Standard Errors HC3				
	coefficient	Standard Error	z	P>|z|
Intercept	7.0215	0.13	54.174	0
Sleep	-0.2356	0.013	-18.102	0
Number_of_days_drinks	-0.0267	0.002	-16.277	0
Weight	0.0065	0	17.292	0
Gender	-0.1958	0.029	-6.737	0
Asthma	-1.1733	0.047	-24.872	0
CHD_or_MI	2.345	0.069	33.844	0
Exercise	-3.099	0.043	-72.611	0
Smoker	0.6467	0.03	21.639	0
Cancer	1.9646	0.061	32.103	0
Diabetes	1.6906	0.052	32.278	0
Kidney	3.16	0.107	29.545	0
Depressive	3.1135	0.045	69.206	0
Stroke	2.6976	0.103	26.195	0
Bronchitis	4.0799	0.078	52.474	0
Income	-1.5365	0.03	-50.868	0
Race	0.1213	0.036	3.411	0.001
Age	-0.3047	0.034	-8.833	0

B: Regression Table for Mental Health

Regression Table for Mental Health				
OLS Robust Standard Errors HC3				
Variable	Coefficient	Standard Error	z	P>|z|
Intercept	7.5656	0.119	63.636	0
Sleep	-0.4252	0.012	-35.054	0
Number_of_days_drinks	0.0147	0.002	9.737	0
Weight	0.0005	0.0003	1.511	0.131
Gender	-0.3021	0.027	-11.204	0
Asthma	-0.8324	0.043	-19.412	0
CHD_or_MI	0.443	0.052	8.504	0
Exercise	-1.0865	0.035	-31.356	0
Smoker	0.6384	0.026	24.139	0
Cancer	0.2612	0.045	5.751	0
Diabetes	0.2012	0.041	4.898	0
Kidney	0.6293	0.082	7.693	0
Depressive	7.616	0.047	162.319	0
Stroke	1.112	0.082	13.575	0
Bronchitis	1.261	0.063	20.021	0
Income	-1.3935	0.027	-52.279	0
Race	-0.106	0.032	-3.276	0.001
Age	-1.8614	0.028	-66.758	0

C: Variable Processing

SAS Variable Name	Retitled Variable	Variable Description	Recoded Variable
alcday5	Number of Days Drinks	**Question:** During the past 30 days, how many days per week or per month did you have at least one drink of any alcoholic beverage such as beer, wine, a malt beverage or liquor?	⇨ 0 → 0 days in past 30 had alcoholic beverage ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
_incomg	Income	**Question:** Income categories	⇨ 0 → less than $50,000 ⇨ 1 → $50,000 or more ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
weight2	Weight	**Question:** About how much do you weigh without shoes?	⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
_ageg5yr	Age categorical	**Question:** Fourteen-level age category	⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
_age65yr	Age	**Question:** Two-level age category	⇨ 0 → Subject is less than 65 years old ⇨ 1 → Subject is older than 65 years old ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
chcocncr	Cancer	**Question:** (Ever told) you had any other types of cancer?	⇨ 0 → Subject did not have cancer ⇨ 1 → Subject had cancer ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
sex1	Gender	**Question:** What is your sex? or What was your sex at birth?	⇨ 0 → Subject is female ⇨ 1 → Subject is male

			⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
diabete3	Diabetes	**Question:** Have you had a test for high blood sugar or diabetes within the past three years?	⇨ 0 → Subject did not have diabetes ⇨ 1 → Subject had diabetes ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
chckdny1	Kidney	**Question:** (Ever told) you have kidney disease?	⇨ 0 → Subject did not have kidney disease ⇨ 1 → Subject had kidney disease ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
addepev2	Depressive	**Question:** (Ever told) you have a depressive disorder (including depression, major depression, dysthymia, or minor depression)?	⇨ 0 → Subject did not have depressive disorder ⇨ 1 → Subject had depressive disorder ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
cvdstrk3	Stroke	**Question:** (Ever told) you had a stroke?	⇨ 0 → Subject did not have a stroke ⇨ 1 → Subject had a stroke ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
sleptim1	Sleep	**Question:** On average, how many hours of sleep do you get in a 24-hour period?	⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
chccopd1	Bronchitis	**Question:** (Ever told) you have chronic obstructive pulmonary disease, C.O.P.D., emphysema or chronic bronchitis?	⇨ 0 → Subject did not have bronchitis ⇨ 1 → Subject had bronchitis ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are

			removed
_raceg21	Race	**Question:** White non-Hispanic race group and Non-white or Hispanic race group?	⇨ 0 → Subject is Non-White or Hispanic ⇨ 1 → Subject is White and Non-Hispanic ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
_ltasth1	Asthma	**Question:** Adults who have ever been told they have asthma	⇨ 0 → Subject did not have asthma ⇨ 1 → Subject had asthma ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
_michd	CHD or MI	**Question:** Respondents that have ever reported having coronary heart disease (CHD) or myocardial infarction (MI)	⇨ 0 → Subject did not have heart attack or chronic heart disease ⇨ 1 → Subject had heart attack or chronic heart disease ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
_totinda	Exercise	**Question:** Adults who reported doing physical activity or exercise during the past 30 days other than their regular job	⇨ 0 → Subject did not do any exersise in the past 30 days ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
menthlth	Mental Health Bad	**Question:** Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?	⇨ 0 → Subject did not have any bad mental health day in the past 30 ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
physhlth	Physical Health Bad	**Question:** Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?	⇨ 0 → Subject did not have any bad physical health day in the past 30 ⇨ Don't know /Not Sure observations are removed

			⇨ Refused observations are removed
smoke100	Smoker	**Question:** Four-level smoker status: Everyday smoker, Someday smoker, Former smoker, Non-smoker	⇨ 0 → Subject is not a smoker ⇨ 1 → Subject is a smoker ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed
marital	Married	**Question:** Are you: (marital status)	⇨ 0 → Subject is not married ⇨ 1 → Subject is married ⇨ Don't know /Not Sure observations are removed ⇨ Refused observations are removed

D: Regressions Without Controls

Regression Table for Physical Health				
OLS Robust Standard Errors HC3				
	Coefficient	**Standard Error**	**z**	**P>|z|**
const	6.5022	0.11	58.989	0
Sleep	-0.3289	0.015	-21.269	0

Regression Table for Mental Health				
OLS Robust Standard Errors HC3				
	Coefficient	**Standard Error**	**z**	**P>|z|**
const	8.0433	0.106	75.989	0
Sleep	-0.6295	0.015	-42.69	0

Works Cited

CDC. “Sleep for Good Health.” *Centers for Disease Control and Prevention*, 19 Sep. 2022.

Hafner M, Stepanek M, Taylor J, Troxel WM, van Stolk C. Why Sleep Matters-The Economic Costs of Insufficient Sleep: A Cross-Country Comparative Analysis. Rand Health Q. 2017 Jan 1;6(4):11. PMID: 28983434; PMCID: PMC5627640.

Homelessness statistics in the US for 2021 | Policy Advice. Accessed 17 Dec. 2022.