The Relationships among Sleep duration, Mental Health, and BMI by Basic Demographic Factors

Group 7

Liancheng Lu, Zibo Yu, He Zhang, Zhengrui Huang

December 8, 2023

Abstract

Background

The study focuses on the intricate interplay between physical health (represented by BMI), sleep duration, and mental health. Previous research has explored these relationships individually, but a comprehensive understanding of their combined effects remains elusive. Utilizing data from the National Health and Nutrition Examination Survey (NHANES), we employ multiple linear regression models to analyze these associations in adults aged 18-60 in the United States.

Methods

Multiple linear regression and variable transformation were conducted to explore relationships among sleep duration, mental health, and Body Mass Index (BMI) across basic demographic factors. The first model was fitted based on primary interest of variables and confounding variables, and interaction terms were considered in another model.

Results

For individuals aged 18-60, an increase of one hour in sleep duration is associated, on average, with a 0.03 unit decrease in logBMI, a finding that remains significant even after adjusting for other factors ($p = 0.003$). An additional day of poor mental health is significantly associated with a 0.0012 unit decrease in logBMI on average among individuals aged 18-60, controlling for other factors (95% CI: -0.0023, -0.00017, $p = 0.023$). Females have a significantly higher BMI than males, with a difference of 0.115 units ($p = 0.009$). The impact of sleep on BMI diminishes with age, decreasing by 0.0005 units for each additional year ($p = 0.04$).

Conclusions

Our findings reveal a significant inverse relationship between sleep duration and log-transformed BMI, highlighting that an increase in sleep duration correlates with a decrease in BMI. Additionally, we uncover a noteworthy association between poor mental health days and reduced BMI, with this relationship varying across different demographic groups.

1. Introduction

The complex relationship among physical well-being, indicated by Body Mass Index (BMI), sleeping time amount, and mental health, is being actively researched and shows complex, often contradictory results. While previous studies have looked at these factors independently, the relationship between them remains unclear.

Research indicates that there may be a link between lack of sleep and elevated BMI, which could result in a serious public health issue, obesity. It is thought that this correlation is due to changes in metabolism, altered appetite regulation and disruption of energy balance caused by lack of sleep. However, not every study supports this idea, and some have not found a significant relationship between sleep duration and changes in body mass index. (1). Mental health's role is also very important. Bad mental health can badly affect both sleeping patterns and choices in diet, possibly leading to weight increase and BMI changes. (2) On the other hand, the psychological stress associated with obesity can negatively impact mental health, exacerbating the cycle of poor health. (3).

This complexity increases when demographic factors such as age and gender are included. Age-related physiological changes will affect sleep patterns, metabolism, and mental health in different ways at different stages of life. Also, gender differences will have significant effects on hormones, stress responses, and social roles.

Our study, recognizing these gaps in current research, intends to examine again at the relationship between BMI, sleep duration, and mental health, paying close attention to age and gender differences. Our goal is to get a deeper understanding of these complex relationships in adults between the ages of 18 and 60, so we can inform future public health problems.

2. Method

2.1 Study population

The study sample was comes from the National Health and Nutrition Examination Survey (NHANES) database, managed by the Centers for Disease Control and Prevention (CDC) (4). This sample includes people aged 18-60 living in the United States. Beginning in the 1960s, the NHANES program switched to a continuous data collection model every two years in 1999. Each year, approximately 5000 individuals of various ages are selected for in-home interviews and comprehensive health examinations.

2.2 Variables

This section details the variables extracted from the NHANES dataset.

2.2.1 Primary Interest

The variables of primary interest included hours study participant usually gets at night on weekdays or workdays (SleepHrsNight), number of days participant's mental health was not good out of the past 30 days (DaysMentHlthBad), age and gender.

2.2.2 Outcome

The primary outcome variable for our analysis is Body Mass Index (BMI), which ranges from 15 to 69. High BMI values are indicative of overweight or obesity.

2.2.3 Covariates

Additional covariates, such as demographic information, total cholesterol levels, combined diastolic blood pressure readings, and physical health status, were included in the final analysis. These were selected based on bivariate analyses and a review of relevant literature.

2.3 Statistical Analysis

Descriptive statistics were performed for the dataset, the basic characteristics of primary interests and confounding variables were concluded in Table 1, and complete and incomplete cases were compared with graphs and t/χ^2 tests. Multiple linear regression was fitted to our primary interests of variables and some confounding variables (1). Diagnosis of the adjusted model before transformation suggested that some model assumptions were violated in this model. Thus, data transformation was performed based on model diagnosis (2).

$$\begin{aligned}\text{BMI} = {} & \beta_0 + \beta_1 \times \text{SleepHrsNight} + \beta_2 \times \text{Age} + \beta_3 \times \text{Gender} + \beta_4 \times \text{DaysMentHlthBad} \\ & + \beta_5 \times \text{Poverty} + \beta_6 \times \text{Race1} + \beta_7 \times \text{BPDiaAve} + \beta_8 \times \text{BPSysAve} \\ & + \beta_9 \times \text{AlcoholYear} + \beta_{10} \times \text{Smoke100} + \beta_{11} \times \text{UrineFlow1} \\ & + \beta_{12} \times \text{TotChol} + \beta_{13} \times \text{DaysPhysHlthBad} + \beta_{14} \times \text{HealthGen} \\ & + \beta_{15} \times \text{PhysActive} + \epsilon \quad (1)\end{aligned}$$

$$\begin{aligned}\log(\text{BMI}) = {} & \beta_0 + \beta_1 \times \text{SleepHrsNight} + \beta_2 \times \text{Age}\beta_3 \times \text{Gender} + \beta_4 \times \text{DaysMentHlthBad} \\ & + \beta_5 \times \text{Poverty} + \beta_6 \times \text{Race1}\beta_7 \times \text{BPDiaAve} + \beta_8 \times \text{BPSysAve} \\ & + \beta_9 \times \text{AlcoholYear} + \beta_{10} \times \text{Smoke100} + \beta_{11} \times \text{UrineFlow1} + \beta_{12} \times \text{TotChol} \\ & + \beta_{13} \times \text{DaysPhysHlthBad} + \beta_{14} \times \text{HealthGen} + \beta_{15} \times \text{PhysActive} + \epsilon \quad (2)\end{aligned}$$

Then, to examine the interaction terms in this model, SleepHrsNight with age and SleepHrsNight with gender were considered as the final model to investigate the association between BMI and the two primary interests, SleepHrsNight and DaysMentHlthBad (3). After three diagnoses, including model assumption diagnosis, influence diagnosis and multicollinearity diagnosis, this model served as the final model to investigate the association between the primary interests. Finally, to investigate the interaction terms, the adjusted model and the final model were compared with each other.

$$
\begin{aligned}
\log(\text{BMI}) = \beta_0 &+ \beta_1 \times \text{SleepHrsNight} + \beta_2 \times \text{Age} + \beta_3 \times \text{Gender} \\
&+ \beta_4 \times \text{DaysMentHlthBad} + \beta_5 \times \text{Poverty} + \beta_6 \times \text{Race1} + \beta_7 \times \text{BPDiaAve} \\
&+ \beta_8 \times \text{BPSysAve} + \beta_9 \times \text{AlcoholYear} + \beta_{10} \times \text{Smoke100} + \beta_{11} \times \text{UrineFlow1} \\
&+ \beta_{12} \times \text{TotChol} + \beta_{13} \times \text{DaysPhysHlthBad} + \beta_{14} \times \text{HealthGen} + \beta_{15} \times \text{PhysActive} \\
&+ \beta_{16} \times \text{SleepHrsNight} \times \text{Age} + \beta_{17} \times \text{SleepHrsNight} \times \text{Gender} + \epsilon \quad (3)
\end{aligned}
$$

2.4 Hypothesis Testing for Multiple Regression

Within the multiple linear regression framework, our objective is to predict the log-transformed Body Mass Index (logBMI). We present the hypotheses concerning the effects of various predictors on logBMI as follows:

Hypothesis 1: Effect of Sleep Duration on logBMI

- **Null Hypothesis (H_{01})**: Sleep duration, measured as the number of hours slept per night (SleepHrsNight), does not predict logBMI ($\beta_1 = 0$).
- **Alternative Hypothesis (H_{A1})**: Sleep duration predicts logBMI, such that there is a non-zero effect ($\beta_1 \neq 0$).

Hypothesis 2: Effect of Mental Health on logBMI

- **Null Hypothesis (H_{02})**: The number of days a participant experiences poor mental health (DaysMentlHlthBad) does not predict logBMI ($\beta_4 = 0$).
- **Alternative Hypothesis (H_{A2})**: The number of days with poor mental health predicts logBMI, indicative of a non-zero effect ($\beta_4 \neq 0$).

Hypothesis 3: Interaction Effect of Sleep Duration and Age on logBMI

- **Null Hypothesis (H_{03})**: There is no interaction effect between sleep duration and age on logBMI ($\beta_{16} = 0$).
- **Alternative Hypothesis (H_{A3})**: There exists an interaction effect between sleep duration and age on logBMI ($\beta_{16} \neq 0$).

Hypothesis 4: Interaction Effect of Sleep Duration and Gender on logBMI

- **Null Hypothesis (H_{04})**: There is no interaction effect between sleep duration and gender on logBMI ($\beta_{17} = 0$).
- **Alternative Hypothesis (H_{A4})**: There is an interaction effect between sleep duration and gender on logBMI ($\beta_{17} \neq 0$).

These hypotheses are formulated under the assumption that all model assumptions are met and will be tested using an appropriate level of significance. The results will determine the predictive capacity of sleep patterns and mental health on the variation in log-transformed BMI, as well as the potential moderating roles of age and gender.

In the context of regression analysis, these hypotheses would be tested using t-tests for each coefficient. The resulting p-values would determine whether there is statistically significant evidence to reject the null hypotheses in favor of the alternative hypotheses.

3. Result

3.1 Descriptive Statistics

Table 1 and table 4 presented the descriptive results of the data. A total of 4211 participants were included in the analysis, with 2710 having complete data and 1501 having some variables with missing data. Table 4 also demonstrated a significant difference between complete and incomplete cases across several covariates, including BMI. We employed t-tests (χ^2 tests) to compare the differences in means (or proportions) of variables between complete and incomplete cases; p-values from the t or χ^2 tests are reported, with significant differences ($\alpha = 0.05$). Moreover, the P-values for most characteristics are below the 0.05 threshold, indicating statistical significance, with the exception of BMI, whose P-value is marginally above the significance level. Thus, there is a significant difference between complete cases and incomplete cases in BMI.Table 1 presents participant characteristics stratified by BMI category (<18, 18-30, >30), Comparing participants across BMI categories, sleep hours per night appear to be slightly lower on average in the highest BMI category, although the difference is statistically significant across categories, indicating a possible relationship between sleep duration and BMI. Days with poor mental health do not significantly differ across BMI categories.

These findings illustrate that individuals with a higher BMI (>30) tend to have markers indicative of potential health challenges, including higher blood pressure, cholesterol levels, and an increased incidence of diabetes, consistent with the broader literature on the health impacts of obesity. Additionally, socioeconomic factors such as education level and wealth appear to be inversely related to BMI.This analysis substantiates that the missing data would be excluded in the following analysis.

Table 5 provided the descriptive results of the outcome, primary interests, and demographic variables without missing data. BMI had a mean value (SD) of 28.75 (6.72), ranging from 15 to 69. SleepHrsNight had a mean of 6.80 (1.31), with a range of 2 to 12 hours. DaysMentHlthBad showed a mean of 4.45 (8.00), ranging from 0 to 30 days. The demographic variable Age had a mean of 39.12 (11.40), ranging from 20 to 59 years. Gender distribution was 1199 (53.4%) male and 1048 (46.6%) female.

3.2 Multiple Linear Regression Analysis

In multiple linear regression analysis, adjusted models without interaction terms and full model with interaction terms were mentioned above, and the results were shown in Table 2.

In the final regression model, the coefficient of determination (R^2) is 15.87%, indicating that approximately 15.87% of the variance in log-transformed BMI (logBMI) is explained by the model.

Variable	All participants N = 2247[1]	<18 N = 16 (0.71%)[1]	18-30 N = 1449 (64.5%)[1]	>30 N = 782 (34.8%)[1]	p-value[2]
SleepHrsNight	6.80 (1.31)	6.69 (1.54)	6.85 (1.27)	6.69 (1.37)	0.019
BMI	28.75 (6.72)	17.04 (0.86)	25.01 (3.01)	35.91 (5.73)	<0.001
Age	39.12 (11.40)	30.00 (13.29)	38.79 (11.51)	39.94 (11.05)	<0.001
Gender					0.058
Female	1,048 (46.6%)	12 (75.0%)	665 (45.9%)	371 (47.4%)	
Male	1,199 (53.4%)	4 (25.0%)	784 (54.1%)	411 (52.6%)	
Race1					<0.001
Black	301 (13.4%)	4 (25.0%)	155 (10.7%)	142 (18.2%)	
Hispanic	155 (6.9%)	0 (0.0%)	105 (7.2%)	50 (6.4%)	
Mexican	250 (11.1%)	2 (12.5%)	144 (9.9%)	104 (13.3%)	
White	1,370 (61.0%)	10 (62.5%)	916 (63.2%)	444 (56.8%)	
Other	171 (7.6%)	0 (0.0%)	129 (8.9%)	42 (5.4%)	
TotChol	5.08 (1.06)	4.39 (0.85)	5.07 (1.08)	5.10 (1.01)	0.008
BPDiaAve	71.24 (11.75)	64.31 (11.63)	70.01 (11.29)	73.67 (12.19)	<0.001
BPSysAve	117.60 (14.57)	109.63 (12.79)	116.05 (13.80)	120.65 (15.45)	<0.001
AlcoholYear	70.43 (94.41)	45.94 (54.34)	79.84 (99.82)	53.50 (81.49)	<0.001
Poverty	2.81 (1.69)	2.05 (1.66)	2.89 (1.70)	2.70 (1.67)	0.008
DaysMentHlthBad	4.45 (8.00)	9.31 (12.71)	4.26 (7.76)	4.69 (8.29)	0.229
UrineFlow1	1.08 (0.97)	0.79 (0.77)	1.11 (1.08)	1.01 (0.81)	0.070
PhysActive	1,304 (58.0%)	3 (18.8%)	906 (62.5%)	395 (50.5%)	<0.001
DaysPhysHlthBad	3.17 (7.20)	2.63 (5.18)	2.77 (6.61)	3.93 (8.16)	0.020
Smoke100	1,032 (45.9%)	10 (62.5%)	675 (46.6%)	347 (44.4%)	0.249
HealthGen					<0.001
Excellent	252 (11.2%)	1 (6.3%)	220 (15.2%)	31 (4.0%)	
Vgood	725 (32.3%)	2 (12.5%)	539 (37.2%)	184 (23.5%)	
Good	885 (39.4%)	11 (68.8%)	508 (35.1%)	366 (46.8%)	
Fair	335 (14.9%)	2 (12.5%)	167 (11.5%)	166 (21.2%)	
Poor	50 (2.2%)	0 (0.0%)	15 (1.0%)	35 (4.5%)	

[1]p-values from Kruskal-Wallis rank sum test

[2]Kruskal-Wallis rank sum test; Pearson’s Chi-squared test

Table 1: Participant Characteristics, by BMI Category

For Sleep Duration, in analysis, sleep duration was significantly associated with logBMI after adjusting for other covariates in the final model ($p = 0.003$). Specifically, for individuals aged 18-60, an increase of one hour in sleep duration is associated, on average, with a 0.03 unit decrease in logBMI, a finding that remains significant even after adjusting for other factors ($p = 0.003$). The adjusted estimates for the effect of sleep duration are attenuated compared to the crude, unadjusted estimates.

For days of poor mental health, an additional day of poor mental health is significantly associated with a 0.0012 unit decrease in logBMI on average among individuals aged 18-60, controlling for other factors (95% CI: -0.0023, -0.00017, (p = 0.023)). This provides sufficient evidence of an association between the number of days with poor mental health and logBMI.

For the interaction terms, the interaction terms in our model uncovers significant heterogeneity in the relationship between sleep duration and logBMI across gender and age cohorts. There is a statistically significant gender modification effect on the sleep-BMI relationship, evidenced by an adjusted mean difference in logBMI of 0.115 units between females and males (p = 0.009). This indicates that the regression coefficient for sleep duration is significantly higher for females than for males when predicting logBMI, holding other variables constant.

Additionally, the interaction term between sleep duration and age is significant (p = 0.04), demonstrating a differential age gradient in the sleep-BMI association across genders. The negative coefficient for the sleep-duration-age interaction term suggests a negative trend, with the effect of sleep duration on logBMI diminishing by 0.0005 units per year. This negative effect is more evident in males and indicates a significant interaction between sleep duration and age in predicting logBMI. Thus, the data provide evidence of both gender and age as effect modifiers in the relationship between sleep duration and BMI.

Moreover, the adjusted main effect of other covariates was similar in magnitude and signs regardless of which interaction terms were included in the model. It is important to note that the interaction term is statistically significant while the main effect is not significant. The hierarchical principle of model building recommends including the main effects in the model regardless of their individual statistical significance.

In conclusion, the findings indicate a significant association between sleep duration, mental health and logarithmic BMI (logBMI) among individuals aged 18-60, aligning with previous research.

	Adjusted Model			Full Model		
Variable	β	SE	P	β	SE	P
SleepHrsNight	-0.005	0.003	0.148	-0.034	0.012	< 0.05
DaysMentHlthBad	-0.001	0.001	< 0.05	-0.001	0.001	< 0.05
Age	0.000	0.000	0.337	-0.003	0.002	0.082
Gender (male)	-0.002	0.009	0.838	-0.115	0.044	< 0.05
SleepHrsNight:Age				0.001	0.000	< 0.05
SleepHrsNight:Gender (male)				0.017	0.006	< 0.05

Table 2: Multiple Linear regression Analysis

3.3 Model Diagnostics

Diagnostic plots for LINE assumption before variable transformation were shown in Fig. 1(a). The scatter in the Q-Q plot didn' t follow the 45° line, and the plots showed a cone-shaped scatter. These suggested that the normality assumption and constant variance assumption were violated. Hence, log transformation was conducted on BMI to remedy the violation. Diagnostic plots fitted well for model assumption after transformation in Fig. 1(b).

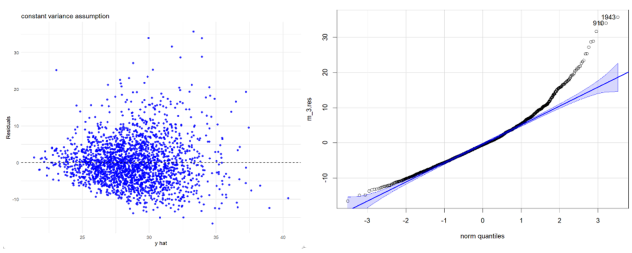

Fig. 1(a)

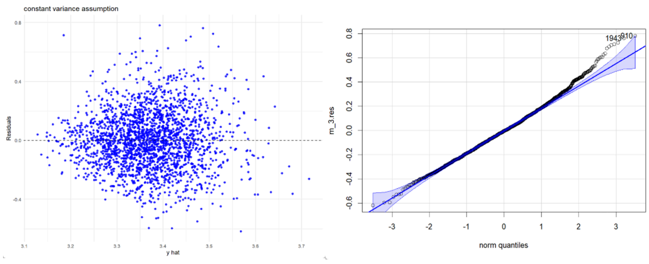

Fig. 1(b)

Diagnostic plots of the final model were shown in Fig. 2(a). In terms of linearity assumption, the regression linear plot shows that the model works well. In terms of normality assumption, all data points follow a straight line in the Q-Q plot. In terms of constant variance assumption, both Fig. 2(a) and Fig. 2(b) suggested a random scatter of residuals. In terms of Independence assumption, the D-W test yielded a p-value of 0.225, which did not give sufficient evidence to reject the constant variance assumption.

To examine high leverage and influential points, Fig. 2(c) suggested that some points fall outside Cook's distance bound. There are 7 observations (1048, 1769, 1684, 74, 72, 1689, 1311) located in the intersection areas of both outlier and leverage, which is to say, those observations have both the leverage and the externally studentized residual exceeding their respective thresholds. Due to its large DIFFITS and Cook's D, they are potentially influential observations. We choose to remove these 7 observations in the re-fitted model. The estimated regression coefficients don't change slightly after removing these observations, with 5 of the estimates having changed by more than 10% after calculation. Thus all outliers would be retained in the dataset.

To detect multicollinearity, the variance inflation factor (VIF) was presented in Table 3. For the final model, no large effects of multicollinearity were found except for the interaction terms, as well as the corresponding main effect terms. Compared to the VIF results from model A, we suggested that the large effects of multicollinearity in the final model could be due to adding the interaction terms.

Upon thorough examination of residuals from models using log-transformed BMI adjusted for other predictors, I concur with the decision to use log-transformed BMI and this model will be considered as the final model for further conclusion and discussion.

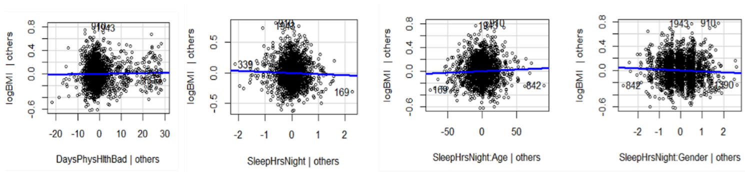

Fig. 2(a)

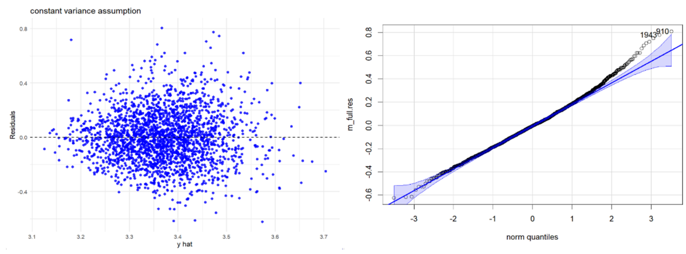

Fig. 2(b)

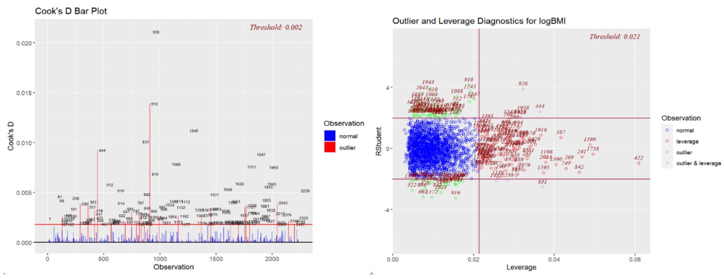

Fig. 2(c)

Group	Df	Adjusted Model	Full Model
SleepHrsNight	1	1.03	3.68
Age	1	1.16	5.30
Gender	1	1.07	5.34
DaysMentHlthBad	1	1.07	1.07
SleepHrsNight:Age	1		6.11
SleepHrsNight:Gender	1		5.48

Table 3: VIF of Each Covariate in 2 Models

4. Conclusions and Discussion

4.1 Conclusion

Our analysis focuses on four key hypotheses examining the association between sleep duration (*SleepHrsNight*), mental health (*DaysMentHlthBad*), and Body Mass Index (*BMI*), further adjusted by demographic variables such as age and gender. We applied a logarithmic transformation to BMI data due to the results of model diagnostics and then investigated its relationship with other variables.

In conclusion, the findings indicate a significant association between sleep duration, mental health and logarithmic BMI (*logBMI*) among individuals aged 18-60, aligning with previous research. In summary, our analysis confirms the significant roles of sleep duration and mental health in BMI variation, while emphasizing the intricate association with demographic factors.

4.2 Discussion

In addition, there are also some limitations of our study including the following. First, self-reported data is likely to cause bias in the result. For example, participants may not be able to accurately recall past behaviors or situations and may overstate or understate their behavior, such as food intake, amount of physical activity, or length of sleep (5). Further, different participants may have different understandings of the questions, resulting in their responses not being comparable (6). Second, the demographic variables that may not be included, such as education level and marital status are not included. Therefore, the study may have omitted some demographic variables and missed important associations and patterns. Third, in cross-sectional survey studies, since each data collection is independent and reflects the situation at that particular time, it can only provide information about a specific time point, ignoring dynamics over time (7). Fourth, after diagnosing the model, we have kept the high leverage points and outliers, since they are in an appropriate range. However, it is necessary to apply some approaches to handle these outliers better (8).

For future work of this project, there are some suggestions of improvement measures. To address the problem of self-reported data bias, the questionnaire design could be improved. Memory bias over-or under-reporting and differences in understanding can be reduced through the use of more precise question formulations and avoidance of leading questions. Furthermore, by adding additional demographic variables, more in-depth multidimensional analyses can be conducted to identify specific patterns in particular groups. Consequently, the study could provide a better understanding of the association between sleep, mental health, and BMI. Then, if data from multiple points in time are available, hierarchical models combined with multiple study designs, such as a before-and-after test design can help explore changes over time in the association between variables. Eventually, to handle the outliers in a more practicable way, the robust regression methods could reduce the weight of outliers by weighing them less (9). Additionally, sensitivity analysis can help researchers identify whether outliers in the dataset have a significant impact on the results of the study. By comparing the results of the analysis when outliers or leverage points are included and excluded, it is possible to assess the extent to which these points influence the final conclusion (10).

Tables

Table 4: Participant Characteristics aged 18-60 in this study

Characteristic	Complete cases N=2710		Non-complete Cases	P-value
	Value	N	Value (SD)	
BMI	28.66	1473	28.53 (7.45)	0.58
SleepHrsNight	6.77	1490	6.90 (1.39)	< 0.05
DaysMentHlthBad	4.43	1006	4.72 (8.54)	0.35
HealthGen				
Poor	1.99%	32	3.2%	–
Fair	14.2%	154	15.4%	–
Good	39.6%	418	41.5%	–
Vgood	33.0%	282	28.2%	–
Excellent	11.2%	118	11.7%	–
Age	39.42	1501	35.77 (12.51)	< 0.05
TotChol	5.06	1276	4.85 (1.01)	< 0.05
BPDiaAve	71.45	1323	69.78 (12.00)	< 0.05
BPSysAve	117.70	1323	116.26 (15.09)	< 0.05
AlcoholYear	71.77	581	64.91 (88.51)	0.09
Race (Race_4Cat)				
Black	11.8%	265	17.2%	–
Hispanic	6.3%	199	17.2%	–
Mexican	10.2%	228	24.3%	–
White	64.6%	398	41.3%	–
Other	7.1%	–	–	–
DaysPhysHlthBad	3.09	1007	3.28 (7.17)	0.46
Poverty	2.86	1181	2.38 (1.69)	< 0.05
UrineFlow1	1.08	1232	0.98 (1.08)	< 0.05
Gender	1257.98	840.0	0.55 (–)	< 0.05
Smoke100	1241.99	452.0	0.34 (–)	< 0.05
PhysActive	1594.02	779.0	0.5189873 (–)	< 0.05

*SD = Standard Deviation

– Data not available

Table 5: Descriptive results of interest variables

Variable	Mean (SD)	Range	N (%)
BMI	28.75 (6.72)	15-69	
SleepHrsNight	6.80 (1.31)	2-12	
DaysMentHlthBad	4.45 (8.00)	0-30	
Age	39.12 (11.40)	20-59	
Gender			
Male			1199 (53.4%)
Female			1048 (46.6%)

Bibliography

[1] V. Garfield, "The association between body mass index (bmi) and sleep duration: Where are we after nearly two decades of epidemiological research?" *International Journal of Environmental Research and Public Health*, vol. 16, no. 22, 2019. [Online]. Available: https://www.mdpi.com/1660-4601/16/22/4327

[2] G. Ambrósio, F. N. Kaufmann, L. Manosso, N. Platt, G. Ghisleni, A. L. S. Rodrigues, D. K. Rieger, and M. P. Kaster, "Depression and peripheral inflammatory profile of patients with obesity," *Psychoneuroendocrinology*, vol. 91, pp. 132–141, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306453017312799

[3] R. He, R. Zheng, J. Zheng, M. Li, T. Wang, Z. Zhao, S. Wang, H. Lin, J. Lu, Y. Chen, Y. Xu, W. Wang, M. Xu, Y. Bi, and G. Ning, "Causal Association Between Obesity, Circulating Glutamine Levels, and Depression: A Mendelian Randomization Study," *The Journal of Clinical Endocrinology Metabolism*, vol. 108, no. 6, pp. 1432–1441, 12 2022. [Online]. Available: https://doi.org/10.1210/clinem/dgac707

[4] R. Pruim, *NHANES: Data from the US National Health and Nutrition Examination Study*, 2015, r package version 2.1.0. [Online]. Available: https://CRAN.R-project.org/package=NHANES

[5] A. S. Nicole Black, David W Johnston, "Justification bias in self-reported disability: New evidence from panel data," *Journal of health economics*, vol. 54, p. 124–134, 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28558294/

[6] R. S. J. P. A. P. J. A. T. Maria Del Mar Bibiloni, Josep Ll Coll, "Ten-year trends (2000-2010) in bias of self-reported weight, height and body mass index in a mediterranean adult population," *Nutricion hospitalaria*, vol. 33, no. 6, p. 1367–1371, 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28000467/

[7] M. S. Setia, "Methodology series module 3: Cross-sectional studies," *Indian journal of dermatology*, vol. 61, no. 3, pp. 261–264, 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4885177/

[8] H. X. M. N. H. M. Md Alamin, Most Humaira Sultana, "Robustification of linear regression and its application in genome-wide association studies," *Frontiers in genetics*, vol. 11, p. 549, 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32582288/

[9] J. Z. B. X. J. L. W. L. Yunlu Jiang, Yan Wang, "Outlier detection and robust variable selection via the penalized weighted lad-lasso method," *Journal of applied statistics*, vol. 48, no. 2, p. 234–246, 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/35707691/

[10] P. N. P. K. Manabesh Nath, Shubham Misra, "Carotid intima-media thickness and risk of mild cognitive impairment: A systematic review and meta-analysis," *Acta neurologica Scandinavica*, vol. 145, no. 2, p. 139–150, 2022. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/34672362/

Contributions

Liancheng Lu:

- Conducted in-depth research on the project's background.
- Cleaned and processed data using R programming.
- Drafted the Introduction and Methods sections of the report.

He Zhang:

- Utilized R to diagnose models and refine data analysis.
- Responsible for drafting the Methods and Results sections of the report.
- Created and prepared data visualization charts and graphs for the report.

Zhengrui Huang:

- Extensively researched and integrated relevant literature and references.
- Applied diagnostic techniques to models using R.
- Drafted the Conclusion and Discussion sections of the report.

Zibo Yu:

- Prepared data visualization charts and graphs for the report.
- Contributed to drafting the Methods and Results sections.
- Managed the editing and formatting of the final report, ensuring consistency and clarity.

Each group member played a pivotal role in the successful completion of the final project. The diverse skill sets and dedicated efforts brought together by the team were instrumental in achieving the project objectives.