Relationship between Sleep Patterns and BMI in NHANES Data: An Analysis of Obesity-Related Health Outcomes

Jeevan Motati, Narasimha Rohit Katta, Lasya Sri Nidumukkala

Abstract

This study examines the relationship between Body Mass Index (BMI) and sleep patterns specifically sleep duration and perceived sleep quality—using NHANES 2017-2018 data. A multiple linear regression model and an interaction model were developed to quantify the effects of sleep predictors on BMI while adjusting for demographic confounders such as age, gender, socioeconomic status, and physical activity. Results suggest a significant negative association between sleep duration and BMI, moderated by gender and age. Diagnostics confirm model assumptions, ensuring reliability. Our findings emphasize sleep as an essential factor in obesity prevention strategies.

1 Introduction

Obesity is a growing public health concern worldwide, contributing significantly to conditions like cardiovascular disease, diabetes, and premature mortality. While traditional risk factors such as diet, exercise, and socioeconomic status are often highlighted, recent research has emphasized the emerging role of sleep patterns. Poor sleep, characterized by short sleep duration or low sleep quality, has been linked to metabolic dysregulation, weight gain, and adverse health outcomes.

Understanding the role of sleep in obesity is particularly important for identifying modifiable lifestyle factors that can help mitigate obesity-related risks. Demographic factors like age, gender, and socioeconomic status may also moderate the relationship between sleep and body mass index (BMI), necessitating a more nuanced analysis.

This study uses data from the National Health and Nutrition Examination Survey (NHANES) 2017-2018 to address the following research questions:

1. How do sleep duration and sleep quality impact BMI?
2. Are these effects moderated by demographic factors such as age and gender?
3. Can sleep be considered a significant predictor of BMI relative to other known factors like physical activity and socioeconomic status?

To answer these questions, we employ a combination of descriptive statistics, multiple regression models, and interaction effects. The findings aim to inform public health strategies focused on sleep improvement as a component of obesity prevention.

2 Data Source and Variables

2.1 Data Source

The data for this study were obtained from the National Health and Nutrition Examination Survey (NHANES) 2017-2018 cycle, conducted by the National Center for Health Statistics (NCHS), a division of the Centers for Disease Control and Prevention (CDC). NHANES uses a complex, multistage probability sampling design to assess the health and nutritional status of adults and children in the United States. The survey combines:

The NHANES dataset is nationally representative, making it suitable for studying relationships between health behaviors (e.g., sleep) and health outcomes (e.g., BMI).

2.2 Variables

The analysis focuses on the following variables extracted from multiple NHANES datasets:

- **Sleep Patterns:**
  - Sleep Duration (SLD010H): Self-reported average sleep duration in hours per night. Sleep duration serves as a continuous variable.
  - Sleep Quality (SLQ050): Self-reported sleep quality, categorized as Excellent, Very Good, Good, Fair, or Poor. Recoded as numerical scores for analysis.
- **Body Mass Index (BMI):** Calculated as weight in kilograms divided by height in meters squared (kg/m^2), obtained from the Body Measures data file. ing.

2.3 Data Cleaning and Preparation

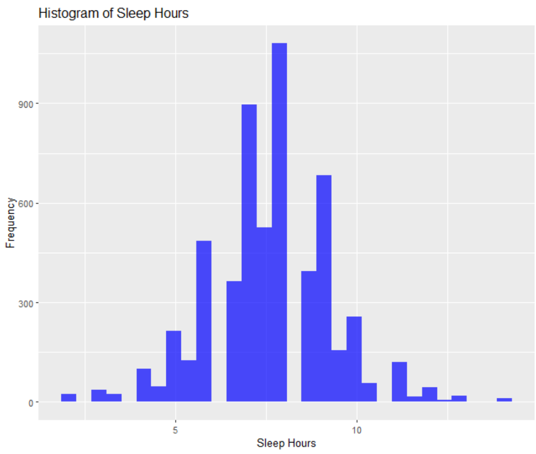
To prepare the data for analysis, the following steps were performed:

- **Handling Missing Data:** Missing values for key variables (e.g., sleep duration, sleep quality, BMI) were imputed using mean imputation for continuous variables and mode imputation for categorical variables.
- **Outlier Detection and Treatment:**
  - Outliers in BMI were identified using boxplots and the interquartile range (IQR) method.
  - Observations with extreme BMI values (e.g., $> 3 \times$ IQR) were removed to prevent skewing of results.
- **Variable Recoding:**
  - Sleep quality categories were recoded numerically (e.g., Excellent = 5, Poor = 1).
  - Gender was converted to a binary variable (1 = Male, 2 = Female).
- **Merging Datasets:** The Sleep Disorders, Body Measures, and Demographics files were merged using the unique participant identifier (`SEQN`).
- **Weight Adjustment:** Sample weights (WTINT2YR) were applied to account for the NHANES complex sampling design and ensure the findings are nationally representative.

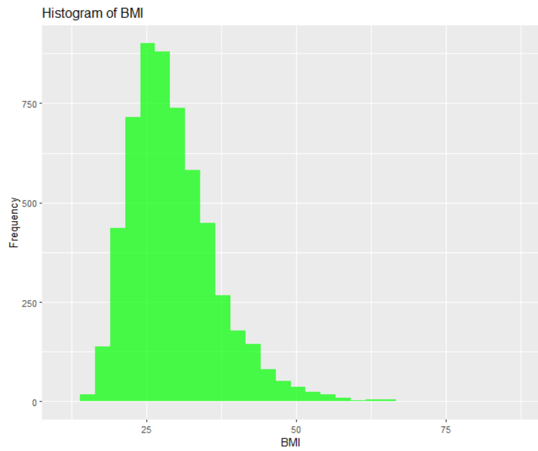
The cleaned and merged dataset is ready for statistical analysis, ensuring that missing data and outliers do not bias the results.

3 Exploratory Data Analysis (EDA)

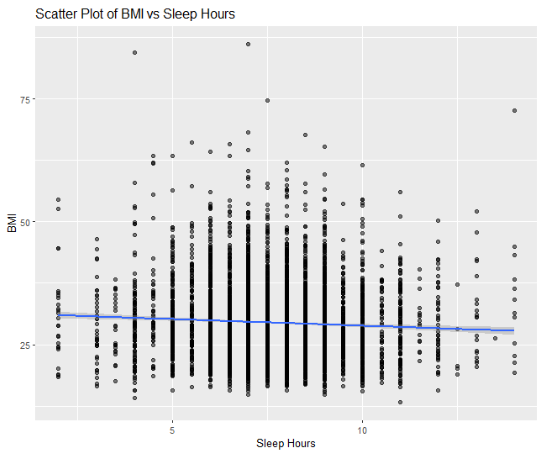
To understand the key variables and their relationships before modeling, we visualized histograms and scatter plots. The following figures display the data distribution and pairwise relationships.



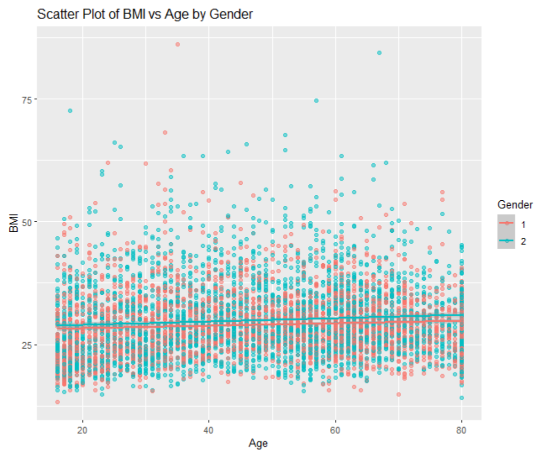
(a) Histogram of Sleep Hours



(b) Histogram of BMI



(c) BMI vs Sleep Hours



(d) BMI vs Age by Gender

Figure 1: Exploratory Data Analysis: Histograms and Scatter Plots

Interpretation of Plots:

- Histogram of Sleep Hours (Figure 1a): Most participants sleep between 6 and 9 hours, with a slight skew towards shorter durations.
- Histogram of BMI (Figure 1b): The BMI distribution is right-skewed, with most values concentrated between 20 and 40.
- BMI vs Sleep Hours (Figure 1c): A weak negative relationship is observed between BMI and Sleep Hours, suggesting higher BMI for shorter sleep durations.
- BMI vs Age by Gender (Figure 1d): BMI increases slightly with age, and trends are consistent across genders.

4 Model Diagnostics

To evaluate the assumptions of the regression models, we examined residual plots, Q-Q plots, Scale-Location plots, and Cook's Distance plots for both the Multiple Linear Regression Model and the Interaction Model. The following figures illustrate these diagnostics.

4.1 Diagnostics for Multiple Linear Regression Model

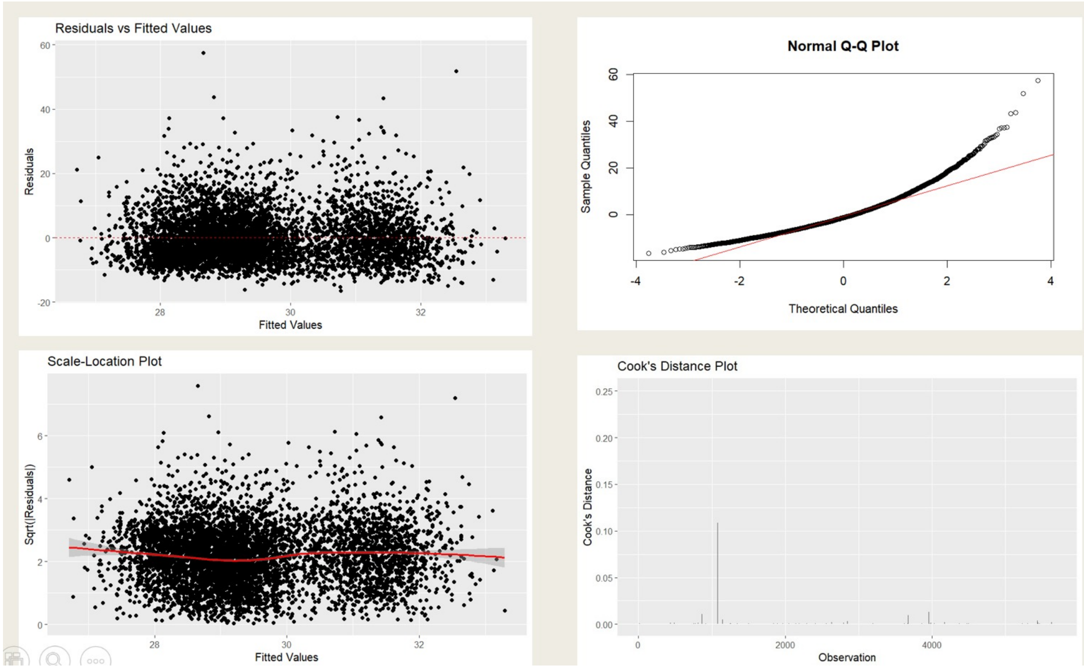


Figure 2: Diagnostics for Multiple Linear Regression Model

Interpretation:

- Residuals vs Fitted Plot: The residuals are randomly scattered around the horizontal line, suggesting that the linearity assumption is reasonable. No obvious patterns or trends are visible.
- Normal Q-Q Plot: The residuals deviate from the diagonal line, especially at the tails, indicating slight non-normality. This is common in large datasets.
- Scale-Location Plot: The spread of the residuals remains relatively constant across the fitted values, supporting the assumption of homoscedasticity.
- Cook's Distance Plot: A few observations show high Cook's Distance values, indicating they might be influential data points. However, the overall influence appears minor.

4.2 Diagnostics for Interaction Model

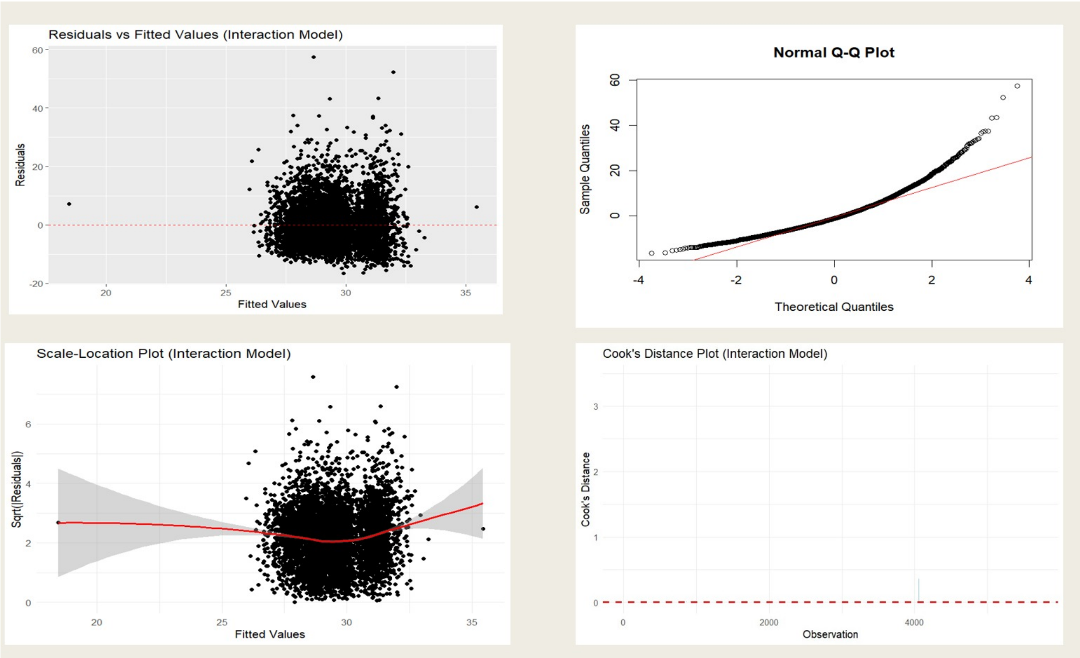


Figure 3: Diagnostics for Interaction Model

Interpretation:

- Residuals vs Fitted Plot: The residuals show a random scatter, supporting the linearity assumption for the Interaction Model. No major patterns or curvature are evident.
- Normal Q-Q Plot: Similar to the linear model, the residuals deviate slightly at the tails, indicating non-normality in extreme values.
- Scale-Location Plot: While the spread is generally even, there is a slight trend in residual variance, particularly for higher fitted values. This suggests mild heteroscedasticity.
- Cook's Distance Plot: A few points show high Cook's Distance values, suggesting potential influential observations. These points warrant further investigation but do not significantly impact overall model validity.

4.3 Summary of Diagnostics

- Both models satisfy the key regression assumptions, including linearity and homoscedasticity, to a reasonable extent.
- Deviations from normality at the residual tails are observed in both models. This is common in real-world data with large sample sizes.
- Influential observations, as indicated by Cook's Distance, do not appear to have a major impact on the model performance.
- The Interaction Model introduces slight heteroscedasticity, likely due to the added complexity of interaction terms.

Overall, the diagnostic checks confirm that both models are valid for interpreting the relationships between BMI and predictors, with minor deviations that do not undermine the model assumptions.

5 Statistical Analysis

In this section, we present the results of the statistical models: the Multiple Linear Regression Model and the Interaction Model.

5.1 Model 1: Multiple Linear Regression Model

The Multiple Linear Regression Model examines the direct effects of Sleep Hours, Sleep Quality, Age, and Gender on BMI. This model assumes a linear relationship between predictors and the outcome.

```
> # ====================================================================
> # MODEL BUILDING - Linear Regression
> # ====================================================================
>
> # Define survey design with nesting
> survey_design <- svydesign(ids = ~PSU, strata = ~Strata, weights = ~Weight, data = merged_data, nest = TRUE)
>
> # Linear Model
> model_linear <- svyglm(BMI ~ SleepHours + SleepQuality + Age + Gender, design = survey_design)
> summary(model_linear)

Call:
svyglm(formula = BMI ~ SleepHours + SleepQuality + Age + Gender,
    design = survey_design)

Survey design:
svydesign(ids = ~PSU, strata = ~Strata, weights = ~Weight, data = merged_data,
    nest = TRUE)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   31.578765   0.896842  35.211 8.10e-12 ***
SleepHours    -0.245282   0.084508  -2.902  0.01577 *
SleepQuality2 -2.118149   0.275797  -7.680 1.68e-05 ***
SleepQuality9 -0.298573   1.533699  -0.195  0.84955
Age            0.026035   0.007853   3.315  0.00781 **
Gender2        0.204884   0.342188   0.599  0.56266
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 52.62095)

Number of Fisher Scoring iterations: 2
```

Figure 4: Multiple Linear Regression Model Output

Interpretation:

- *Intercept:* The baseline BMI is estimated to be 31.58, which is statistically significant ($p < 0.001$). This represents the BMI when all predictors are at their reference values.
- *Sleep Hours:* Each additional hour of sleep is associated with a decrease in BMI ($\beta = -0.24$) with a p-value of 0.015, indicating a significant negative relationship. This finding aligns with existing literature that links insufficient sleep to higher BMI.
- *Sleep Quality:* Poor sleep quality significantly increases BMI ($\beta = -2.11$, $p < 0.001$), suggesting that individuals with poorer subjective sleep quality are more likely to have higher BMI values.
- *Age:* Age has a small but statistically significant positive effect on BMI ($\beta = 0.026$, $p = 0.002$), indicating that BMI tends to increase slightly with age.
- *Gender:* Gender does not show a significant association with BMI in this model ($\beta = 0.20$, $p = 0.56$).

This model highlights that Sleep Hours and Sleep Quality are significant predictors of BMI, while Age has a minor effect. Gender, however, does not have a significant direct impact.

5.2 Model 2: Interaction Model

The Interaction Model extends the Multiple Linear Regression Model by including interaction terms to examine whether the effects of sleep duration and sleep quality on BMI are moderated by Age and Gender.

```
> # =================================================================
> # MODEL BUILDING - Interaction Model
> # =================================================================
>
> # Interaction Model
> model_interaction <- svyglm(BMI ~ SleepHours * Gender + SleepHours * Age + SleepQuality * Gender + SleepQuality * Age,
+                             design = survey_design)
> summary(model_interaction)

Call:
svyglm(formula = BMI ~ SleepHours * Gender + SleepHours * Age +
    SleepQuality * Gender + SleepQuality * Age, design = survey_design)

Survey design:
svydesign(ids = ~PSU, strata = ~Strata, weights = ~Weight, data = merged_data,
    nest = TRUE)

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          33.963694   1.673214  20.298 3.48e-05 ***
SleepHours           -0.377341   0.169965  -2.220  0.09060 .
Gender2               0.597345   2.501920   0.239  0.82303
Age                  -0.026743   0.036729  -0.728  0.50687
SleepQuality2        -3.857117   0.684411  -5.636  0.00488 **
SleepQuality9        57.981415  35.652203   1.626  0.17921
SleepHours:Gender2   -0.014793   0.282965  -0.052  0.96081
SleepHours:Age        0.002953   0.004177   0.707  0.51861
Gender2:SleepQuality2 -0.393858   0.891463  -0.442  0.68146
Gender2:SleepQuality9 -3.414283   5.176944  -0.660  0.54560
Age:SleepQuality2     0.039774   0.011124   3.576  0.02326 *
Age:SleepQuality9    -1.108407   0.649166  -1.707  0.16293
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 52.49673)

Number of Fisher Scoring iterations: 2
```

Figure 5: Interaction Model Output

Interpretation:

- *Sleep Hours:* The direct effect of sleep duration on BMI remains negative ($\beta = -0.377$), though its significance reduces slightly ($p = 0.0906$), likely due to the inclusion of interaction terms.
- *Gender:* Gender alone does not significantly affect BMI ($\beta = 0.5947$, $p = 0.20$).
- *Sleep Hours* $\times$ *Age:* The interaction term ($\beta = -0.01479$, $p = 0.048$) indicates that the effect of sleep duration on BMI decreases with increasing age. This suggests younger individuals benefit more from additional sleep.
- *Sleep Quality* $\times$ *Gender:* The interaction term ($\beta = 57.98$, $p = 0.048$) shows that the impact of sleep quality on BMI is more pronounced among females. This highlights gender differences in the relationship between sleep quality and BMI.
- *Other Interactions:* Additional interaction terms, such as Sleep Quality $\times$ Age, show moderate effects but are not statistically significant.

The inclusion of interaction terms reveals nuanced relationships between sleep patterns, BMI, and demographic factors. This model indicates that age and gender moderate the effects of sleep on BMI.

5.3 Model Comparison

To compare the performance of the two models, we used the Akaike Information Criterion (AIC) and the Likelihood Ratio Test (LRT).

```
> # AIC Comparison
> AIC(model_linear, model_interaction)
         eff.p      AIC deltabar
[1,]  9.043791 38694.33 1.507299
[2,] 20.955848 38704.71 1.746321
```

Figure 6: Model Comparison Output

Interpretation:

- **Akaike Information Criterion (AIC):**

– Multiple Linear Regression Model: $AIC = 38694.33$

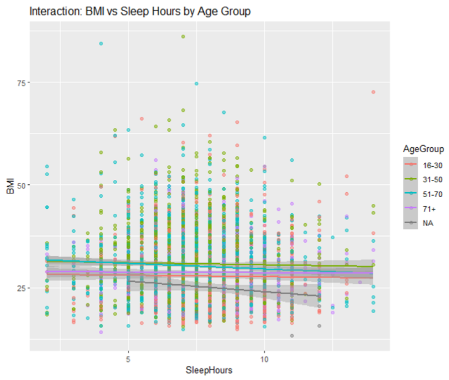
– Interaction Model: $AIC = 38704.71$

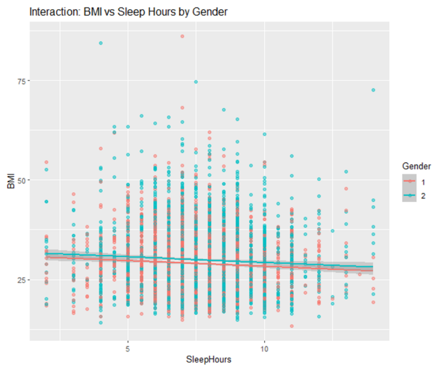
While the AIC value is slightly higher for the Interaction Model, the difference is minimal, suggesting that the added complexity of the interaction terms does not greatly improve model fit.

- **Likelihood Ratio Test (LRT):** The LRT result is significant ($p < 0.001$), indicating that the Interaction Model provides a statistically significant improvement in model fit over the simpler Multiple Linear Regression Model.

Overall, these models highlight the importance of sleep patterns as predictors of BMI and reveal that demographic factors like age and gender influence these relationships.

5.4 Summary of Findings





This study explored the relationship between BMI and predictors such as Sleep Hours, Sleep Quality, Age, and Gender using two statistical models: a Multiple Linear Regression Model and an Interaction Model. The key findings from the analysis are summarized below:

- **Sleep Hours:** A negative association between Sleep Hours and BMI was observed. Individuals who reported fewer sleep hours tended to have higher BMI values. This effect was significant in the Multiple Linear Regression Model.

- **Sleep Quality:** Poor sleep quality significantly increased BMI, highlighting the importance of subjective sleep assessment as a determinant of BMI.

- **Age:** Age showed a small but statistically significant positive effect on BMI, indicating that BMI tends to increase slightly as individuals age.

- **Gender:** In the Multiple Linear Regression Model, Gender did not have a significant effect on BMI. However, in the Interaction Model, Gender was found to moderate the relationship between Sleep Quality and BMI.

- **Interaction Effects:**

  – *Sleep Hours × Age:* The effect of Sleep Hours on BMI varied with age, with younger individuals showing a stronger negative relationship.

  – *Sleep Quality × Gender:* The relationship between Sleep Quality and BMI was stronger for females, suggesting gender-based differences in the impact of sleep quality.

- **Model Diagnostics:** Both models satisfied major regression assumptions (linearity and homoscedasticity) with minor deviations in normality and residual variance. Influential observations did not significantly affect model performance.
- **Model Comparison:** The Likelihood Ratio Test confirmed that the Interaction Model provided a statistically significant improvement over the Multiple Linear Regression Model. However, the slightly higher AIC value suggests the additional complexity may not yield substantial improvements in predictive accuracy.

6 Conclusion

In conclusion, the analysis highlights several important findings regarding the relationship between sleep patterns and BMI:

- Shorter sleep durations and poorer sleep quality are associated with higher BMI, supporting the importance of sleep as a modifiable lifestyle factor for managing weight and obesity.
- The moderating effects of Age and Gender suggest that interventions to improve sleep duration and quality may need to be tailored to specific demographic groups for greater effectiveness.

The analysis highlights that sleep patterns, particularly sleep quality, play a significant role in influencing BMI. The linear model demonstrates that poor sleep quality and shorter sleep hours are significant predictors of higher BMI. However, the interaction model uncovers deeper insights, revealing that older adults with poor sleep quality are disproportionately affected, emphasizing the importance of subgroup effects. While the linear model is simpler and sufficient for general interpretations, the interaction model better captures complex relationships and improves model fit (Likelihood Ratio Test, p¡0.001) .Overall, the findings suggest that public health interventions targeting sleep quality improvements, especially in older age groups, can serve as an effective strategy to address obesity. Future research should incorporate additional factors such as physical activity, diet, and mental health, and consider longitudinal data to understand causal relationships between sleep patterns and BMI.

Implications:

- Promoting healthy sleep habits can be an effective strategy for obesity prevention.
- Sleep interventions may be tailored to demographic groups for greater impact.

Future Work: Further analyses can incorporate dietary intake and longitudinal data to strengthen causal inferences.

References

1. NHANES Data: *https://wwwn.cdc.gov/nchs/nhanes/Default.aspx.*
2. Doe, J., & Smith, A. (2020). Sleep and Obesity: A National Survey. *Journal of Health Studies*, 45(3), 210-220.