

Overview

There has been extensive research on the relationship between physical activity and sleep. Prior research on physical activity and sleep have relied heavily on self-reported measures so it lacked the precision needed to establish the association between physical activity and sleep quality in clinical trials. In this research, we solve the precision issue by using objective measures of physical activity, such as accelerometry data collected from wearable devices. We attempt to build the linkage between physical activity levels and sleep quality. There are two specific aims in this study. Specific aim 1 is to determine if participants with different sleep conditions have different activity levels and scores. Specific aim 2 is to evaluate whether the inclusion of accelerometry data can enhance the precision of predicting sleeping conditions.

Specific Aim 1

The objective of this aim is to investigate the relationship between sleep conditions and physical activity levels/scores in participants. The hypothesis is that participants with different sleep conditions will have different activity levels/scores.

Currently, we have completed our first aim. Initially, we conducted some data-preprocessing and exploratory data analysis for the full dataset. On our web page for data visualization, a table including all the variables and patients is shown. A data dictionary is also created including paths, categories, brief description, and value types. We also checked the missing values for all the whole dataset and summarized the results under different levels and stages. For example, here we chose to filter on the field variable, which has 470 different values. In the histogram, the x axis represents the stage for the study, and the y axis shows the percentage of missing values at each stage. Data labels were also presented. Since we only considered the baseline data, we also summarized the baseline's counts and percentages of missing values by category. Next, we generated box plots to show the distributions of activity scores among different sleep conditions for each categorical variable, and each subplot represents each intensity level. The results from the box plots showed that there was a significant difference in physical activity scores across the different sleep condition groups. We also visualized accelerometry data to transform numerous accelerometry data into 3 activity values. The ranges for the activity scores are: 1 - 19 milli-gravities for Low-intensity Activity, 20 - 95 milli-gravities for Mid-intensity Activity, 100 - 1900 milli-gravities for High-intensity Activity.

As for other data-preprocessing steps, we transformed the original accelerometry data into probability mass on each acceleration value, and visualized the results for each participant. Based on the acceleration cutoffs, the total probability within each intensity level can measure the amount of activities. Then, we chose the categorical variables related to sleep conditions and conducted several tests to evaluate our hypothesis.

Based on our goal and purpose, we have selected one-way ANOVA to compare differences in physical activity performance between different sleep condition groups. Our null hypothesis is: there is no difference in means of sleep quality between the 3 activity groups. Our alternative

hypothesis is: there is at least one difference in means of sleep quality score between 3 activity levels.

Before running the ANOVA tests, we have checked the three assumptions of ANOVA tests: normality, homogeneity, and independence. To validate the assumptions of anova tests, we also visualized the distribution of mean of activity amount for participants with multiple sleep conditions, where we were able to validate whether activity amount can distinguish sleep condition effectively. After checking the assumptions, we proceeded to build the ANOVA models. We will use the p-value and R-square score to generate the final result and to evaluate the model. The results will be explained in detail in our final report.

Specific Aim 2

The objective of this aim is to identify statistically significant variables based on the full model. We hypothesize that activity level scores measured by accelerometers will be significant predictors of sleep quality-related variables, including snoring status, sleep efficiency, and wake time. We chose specific aim 2 because this is a key step to understanding the relationship between physical activity and sleep quality. Furthermore, this aim has the potential to inform interventions and strategies aimed at improving sleep health.

To study the relationship between activity level and sleep quality related, we plan to use a random forest model to construct a prediction model of sleep profile variables. Before training the model, we will conduct necessary exploratory analysis and visualization. We will check the missing data, the correlation between all included outcome variables, as well as distributions of all outcome variables to see if the data have enough observations. The random forest model will include all features from the study as independent variables (e.g., activity level scores using accelerometers, age, gender, health status), and the sleep profile variable created in Aim 1 as the dependent variable. The model will be used to identify which features are most strongly associated with sleep quality and further, develop a predictive model that can predict sleep quality based on variables such as age, gender, and health status. The decision tree used can also be visualized. Further validation of the model can be done using test data.

One potential limitation of the study is that the sample may be biased. The recruited participants are people who have accelerometers so these people may be individuals who have a greater awareness of their health status and therefore, more likely to monitor their sleep quality. Some alternative approaches for this study could be to study the impact of different types of physical activity (e.g., aerobic vs. strength training) on sleep quality, or investigate the mechanisms underlying the relationship between physical activity and sleep health.

We expect to complete Aim 2 by April 21st.