- Specific Aims: 1 page

Background

Research question

The relationship between physical activity and human sleep has been extensively studied from various perspectives. Prior research that relied on self-reported measures of activity frequency and duration lacked the precision needed to accurately establish the association between physical activity and sleep quality in clinical trials. The difficulty in quantifying total physical activity across different levels of intensity further complicates the issue. To overcome these limitations, objective measures of physical activity have been developed and complement self-reported assessments. One such method is the use of accelerometry with wearable devices, which has gained widespread acceptance in recent years. Accelerometers are sensors that detect the acceleration of objects in motion along reference axes, and they provide a reliable and objective means of measuring physical activity.

Our hypothesis is that there is a substantial variation in the frequency and intensity of physical activity among individuals with varying sleep conditions. A further hypothesis is that using accelerometry data can be a supplement to predict human sleep conditions in conjunction with other self-reported biomedical variables. These hypotheses will be further investigated and analyzed in the following Specific Aims.

> Specific Aim 1. To determine if participants with different sleep conditions have different activity levels/scores?
>
> Specific Aim 2. To evaluate whether the inclusion of accelerometry data can enhance the precision of predicting sleeping conditions.

- Research Strategy ~3 pages

Significance: explain background/research gap

The ability to predict human sleeping conditions is of significant importance. First, sleep is a critical component of overall health and well-being, and disruptions in sleep patterns have been linked to a range of negative health outcomes, including cardiovascular disease, obesity, and mental health issues. Accurately predicting sleep patterns could therefore help identify individuals who are at increased risk of these health problems and allow for early intervention and prevention strategies.Second, sleep disorders such as sleep apnea and insomnia are common and can have a significant impact on quality of life. By predicting sleep conditions, healthcare providers can more accurately diagnose and treat these disorders, improving patients' quality of life and reducing the risk of negative health outcomes.

Innovation: literature review

Accelerometry can be employed with specific signal processing techniques to extract physical activity information, which can then be utilized to characterize activity levels (Willetts et al., 2018; Ravi et al., 2005; Yang & Hsu, 2010). This activity information has the potential to be associated with various health outcomes. Several studies have confirmed the associations between accelerometry-derived measures of physical activity or sleep and certain diseases

such as obesity, cardiovascular disease, cardio-metabolic disease, breast cancer, psychiatric disorder, and Parkinson's disease (Guo et al., 2019; Guo et al., 2020; Ramakrishnan et al., 2021; Barker et al., 2019; Cassidy et al., 2018; Dennison et al., 2021; Nikbakhtian et al., 2021). The proposed research will combine statistical analysis and machine learning models to relate accelerometry-derived activity with sleeping conditions.

Research plan

UK Biobank data provides a bunch of self-reported biomedical variables with summary acceleration data, facilitating the identification of statistical significance and the prediction of sleeping conditions. In this study, we employ a hybrid approach that combines statistical analysis and machine learning techniques, which will be applied to the available UK Biobank dataset.

Specific aims:

Aim 1: does participants with different sleep conditions have different activity levels/scores

Hypothesis

We hypothesize that people with different levels of sleep conditions will have significantly different activity scores in each intensity of accelerometry-derived physical activity.

Rationale

As described in the Research Strategy section, numerous studies have explored the link between physical activity and sleep, and most conclude that certain types of physical activity improve sleep quality and duration. However, different exercise intensities may have opposite effects on sleep quality. For instance, some studies have found that moderate exercise training over the course of several weeks can improve sleep quality and duration for adolescents, whereas vigorous exercise during the same timespan has been shown to decrease sleep duration for some teens. In this Aim, we propose to determine whether different sleep qualities will have significantly different performance in each intensity of accelerometry-derived physical activity.

Experimental Approach

In order to quantify the intensity of physical activity, only patients with accelerometers will be considered in our experiments. Accelerometers will record the values for each patient during activities, which can be used to classify activity intensities. Multiple two-sample t tests will be conducted to compare differences of physical activity performance between different sleep condition groups. In each test, a categorical variable which measures sleep quality will be used as the outcome variable, and the independent variable will be the sum of values at each intensity for each patient.

Correlation between Accelerometry and Activity: As the intensity of activity is positively correlated to acceleration, acceleration can be used to classify low, moderate and high intensity activities. By summing up all the fractions in each intensity, the total fraction depicted the

percentage of acceleration triggered by the different kinds of activities. For example, the total fraction of acceleration from 1 to 19 milli-gravities is proportional to the amount of low-intensity.

Acceleration	Activity
1 - 19 milli-gravities	Low-intensity Activity
20 - 95 milli-gravities	Mid-intensity Activity
100 - 1900 milli-gravities	High-intensity Activity
1 - 1900 milli-gravities	Total Activity

Independent variables (one at a time): high activity scores, low activity scores, medium activity scores
Dependent variables (one at a time): "getting_up_in_morning", "morningevening_person_chronotype", "nap_during_day", "sleeplessness_insomnia", "snoring", "daytime_dozing_sleeping_narcolepsy", "trouble_falling_or_staying_asleep_or_sleeping_too_much" , "did_your_sleep_change", "trouble_falling_asleep", "sleeping_too_much" , "waking_too_early", "degree_bothered_by_trouble_sleeping_in_the_last_3_months"

Interpretation of Results
As explained in the Experimental approach, we will be utilizing several two-sample t-tests to evaluate whether there are differences in physical activity levels among distinct sleep condition groups. For each t-test, our hypothesis assumes that there is no significant difference in the average physical activity value between the two groups being compared. To illustrate, we will be conducting three t-tests concerning "snoring", which will involve contrasting individuals who snore in the high-intensity activity group with those who don't snore in the high-intensity activity group, as well as comparisons in the low and medium intensity activity groups. Once a two-sample t-test is performed, we will make a conclusion using the t-statistics and p-value. For example, if the p-value for comparing individuals who snore and those who don't snore in the high-intensity activity group is less than the standard alpha level of 0.05, we will reject the null hypothesis of equal means and conclude that the difference in high-intensity activity between individuals who snore and those who don't snore is statistically significant at the 0.05 level. By analyzing the outcomes of these t-tests, we can draw inferences about the specific range of activity intensity that may have an impact on particular sleep activities.

Potential Problems and Alternative Approaches
The t-test assumes that the data is normally distributed and that the variances of the two groups being compared are equal. However, inour case, these assumptions may not hold true, which can result in potential problems. If the data is not normally distributed, the t-test may not provide reliable results. For example, if the data is skewed or has outliers, the t-test may be biased towards the group with the larger sample size. If the variances of the two groups are significantly different, the t-test may not provide accurate results. When the variances are

unequal, the t-test can become less powerful and biased towards the group with the larger variance.

To address these issues, there are alternative tests that can be used. For example, if the data is not normally distributed, non-parametric tests such as the Wilcoxon rank-sum test or the Kruskal-Wallis test can be used. If the variances are unequal, Welch's t-test or the Mann-Whitney U test can be used instead of the traditional t-test. It is important to assess the assumptions of the t-test before conducting the analysis and to choose the appropriate test accordingly to ensure the validity and reliability of the results.

Aim 2: additional statistical significant variables based on the new model

- Hypothesis

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 too .

- Rationale

Overall, the proposed hypothesis represents an important step towards understanding the complex interplay between physical activity and sleep quality, and has the potential to inform interventions and strategies aimed at improving sleep health. This relationship may be explained by previous research indicating that different physical activity is related with sleep quality, or by potential confounding factors such as physical discomfort or difficulty relaxing that may lead to both higher activity levels and poorer sleep quality. By analyzing data collected from a sample of individuals and using statistical methods, we will test whether there is a statistically significant relationship between activity level scores and sleep quality-related variables, while incorporating all of the other factors that could contribute to sleep qualities. In this aim, we would thus propose to determine the statistical significance of the activity level scores under the presence of all the other features included inside the study, which is about… features.

- Experimental Approach

To investigate the relationship between activity level scores and sleep quality-related variables, we will use a random forest model to construct a prediction model of sleep profile variables. This model will include all of the features included in the study as independent variables, with the sleep profile variable created by Aim 1 serving as the dependent variable. This approach will allow us to identify which independent variables are most strongly associated with sleep quality, and to develop a predictive model that can accurately estimate sleep quality based on these variables.

Specifically, we will collect data on activity level scores using accelerometers, as well as other features such as age, gender, and health status. We will then use these features as inputs to a random forest model, which is a machine learning algorithm that can effectively model complex relationships between variables. By analyzing the output of this model, we will be able to identify

which features are most strongly associated with sleep quality, and to develop a predictive model that can accurately estimate sleep quality based on these variables.

- Interpretation of Results

The aim of this study is to investigate the importance of activity level scores as a predictor of sleep quality-related variables, compared to other features included in the sleep profile prediction model. Specifically, we will use a random forest model to identify the feature importance associated with each independent variable, including activity level scores, age, gender, and health status. By analyzing the feature importance scores, we will be able to determine the extent to which activity level scores contribute more to the prediction of sleep quality-related variables, compared to other variables.

This analysis will allow us to evaluate the relative importance of activity level scores in predicting sleep quality, and to determine whether this variable is a key driver of the relationship between physical activity and sleep health. By understanding the specific contributions of activity level scores and other variables to the prediction of sleep quality-related variables, we can identify potential targets for interventions aimed at improving sleep health.

- Potential Problems and Alternative Approaches

One potential limitation of our study is that the population sample may be biased. Because we recruited participants who had accelerometers, it is possible that these individuals have a greater awareness of their health status and are more likely to monitor their sleep quality, potentially resulting in a population bias.

Provided with these limitations, some alternative approaches that could be used to further explore the relationship between physical activity and sleep quality. For example, future studies could examine 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. Such studies could provide valuable insights into the specific interventions and strategies that can be used to improve sleep health through physical activity.