1. Group Membership: Vishwajith Reddy Anagandula, Scott Callahan, Alvaro J. Castro Rivadeneira, Jusny Louis-Jean, Lindsey Russo, Faith English
2. Target population and study inclusion criteria: The target population is a representative sample of the U.S. population of all ages from the National Health and Nutrition Examination Survey (NHANES) biannual surveys between 2007-2018. We will include subpopulation analyses by age. Inclusion criteria are all respondents who provided information for survey questions related to our exposure and outcome of interest (exercise and short sleep duration).
3. Research question: We will examine the association between meeting 2008 U.S. Physical Activity guidelines and short sleep duration (<=6 hours).
4. Target Causal parameter: We are interested in the causal risk difference between exercising >=150 minutes per week (a=1) and exercising <150 minutes per week (a=0) on short sleep duration:



In words, the target causal parameter is the difference in the counterfactual risk of short sleep duration if all participants exercised >=150 minutes per week vs if all participants did not meet the >=150 minutes of exercise per week criteria.

1. Brief description: We will be using data from the National Health and Nutrition Examination Survey (NHANES), which is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The NHANES program began in the early 1960s but became a continuous one in 1999. Both exposure and outcome are measured for the NHANES through an interview, an examination component with lab tests. Hence NHANES being a cross-sectional study used to determine the prevalence of major diseases and risk factors for diseases. For our analysis, we will compile data from 2007 to 2018, and will study the effect of PAD660 - Minutes vigorous recreational activities on the number of hours of sleep.

The sample for the survey is selected to represent the U.S. population of all ages. To produce reliable statistics, NHANES over-samples persons 60 and older, African Americans, and Hispanics. The sample size is 38,120 participants from NHANES biannual surveys between 2007-2018.

There was extensive data cleaning, and the details can be found in the R script, but broadly, we compiled NHANES data from 2007-2018, for participants for whom data was available. We computed the total number of minutes of moderate and vigorous recreational activities done weekly by participants. From this we created a binary variable, where zero means less than 150 minutes a week, and 1 means 150 or more minutes. Similarly, we computed the average number of hours of sleep per weeknight and created a binary variable where six or less is zero, and more than six is one. In between there was a lot of data cleaning and table joining, and even simple imputations for some of the missing data points. Finally, here is a simple contingency table with the marginal distributions:

|  |  |  |
| --- | --- | --- |
|  | Six hours or less of sleep | More than six hours sleep |
| >150 min exercise/wk | 8,431 | 17,609 |
| 150 or more min exercise | 3,480 | 8,600 |

1. Anticipated challenges and how we will address them: Given the cross-sectional design of NHANES, we will not be able to assess causal questions; but rather, descriptive associations only. This presents a particular challenge in that we will have to think carefully about the timing of occurrence for confounders (which include alcohol use, smoking status, SES, BMI, chronic stress, and the use of sleep medication) that we include in our multivariable models. Further, we will be assessing waves of study participants rather than individuals from the same cohort throughout.