Project 2
Grant Development: Sample Size Calculations
Interim Data Analysis Plan Due 10/16
Final Report Due 10/28

Description

The purpose of this project is to develop the analysis and sample size sections for a hypothetical grant application. Sample size calculations must be done using the simulation based framework discussed in class.

The goal of this research is to understand the effect of physical activity on 5-year all-cause mortality. The grant application proposes a randomized trial to investigate a novel intervention named "ACTUP" which increases physical activity by a fixed amount of 30% (person-specific) among sedentary older adults (aged 60-75). Physical activity is measured by wrist-worn accelerometers, with participants instructed to wear the devices continuously for 7 days. There are initial estimates suggesting that the intervention reduces the probability of 7-year all-cause mortality risk by 7%. However, due to a data handling error, preliminary data are no longer directly available. Fortunately, a publicly available dataset containing a nationally representative sample with both physical activity and mortality data are available in the National Health and Nutrition Survey (NHANES) 2011-2014 waves. In the NHANES 2011-2014 study, physical activity data are publicly available for participants aged 5 and over. Analyses of the preliminary data (before it was lost) suggested the possibility that the intervention was more efficacious in men than in women. The aims of the proposed grant are presented below.

Aim 1. Determine the efficacy of ACTUP in sedentary adults aged 60-75. <u>Hypothesis 1</u>: ACTUP will lead to a reduction in risk of 5-year all-cause mortality (primary endpoint) at the group-average level.

Aim 2. Investigate whether gender affects ACTUP treatment efficacy. <u>H2:</u> Male gender is associated with a greater reduction in risk of all-cause mortality as compared to female gender.

You are responsible for drafting the statistical design and consideration components for this grant submission (sample size/power calculations and statistical analysis). A description of the data (available on Canvas) is provided below.

Data description: NHANES 2011-2014

- SEQN: unique participant identifier
- SDDSRVYR: indicator for NHANES wave (7 = 2011-2012, 8 = 2013-2014)
- SDMVSTRA: survey strata (stratified sampling design component)
- SDMVPSU: primary sampling unit (stratified sampling design component)
- WTMEC2YR: 2-year survey weight associated with the participant. This number corresponds to how many people this individual "represents" nationally
- permth_exm: follow-up time in months for mortality data
- mortstat: indicator for whether the person experienced a mortality event (0=alive, 1=deceased at follow up time, permth_exm)
- gender: binary indicator of gender (male/female)
- age_years_interview: age of the participant in years
- education_adult: education variable (only available in adults)
- race: variable indicating race of the participant
- BMI_cat: categorical BMI assigned using CDC cut-points (underweight, normal, overweight, obese)
- overall_health_combined: self reported overall health
- diabetes: self-reported history of diabetes diagnosis
- heart_attack: self-reported history of heart attack
- CHF: self-reported history of congestive heart failure
- CHD: self-reported history of coronary heart disease
- stroke: self-reported history of stroke diagnosis
- cancer: self-reported history of cancer diagnosis
- mobility_problem: self-reported mobility difficulty
- alcohol_consumption_fac: self-reported alcohol consumption
- cigarette_smoking: self-reported cigarette smoking
- **TMIMS_mean**: total MIMS, averaged across days
- **TLMIMS_mean**: total log(1+MIMS), averaged across days
- **M10_MIMS_mean**: average MIMS during the most 10 active hours of the day, averaged across days
- **M10_IMIMS_mean**: average log(1+MIMS) during the most 10 active hours of the day, averaged across days
- **L5_MIMS_mean**: average MIMS during the least 5 active hours of the day, averaged across days
- **L5_IMIMS_mean**: average log(1+MIMS) during the least 5 active hours of the day, averaged across days
- **RA_MIMS_mean**: Relative amplitude (M10 – L5)/(M10 + L5), averaged across days
- **RA_IMIMS_mean**: Relative amplitude (M10 – L5)/(M10 + L5), averaged across days calculated using log(1+MIMS)
- **n_good_days**: number of valid days of accelerometry data used for deriving wearable accelerometer related variables