Project 2 Final report

Shuai Zhu

2024-10-28

1 INTRODUCTION

To understand the effect of physical activity on 7-year all-cause mortality, the proposed grant application outlines a randomized trial to investigate a novel intervention called "ACTUP." This intervention is designed to increase physical activity by a fixed 30% (individual-specific) among sedentary older adults. Participants will be asked to wear wrist-worn accelerometers for 7 days to objectively measure their physical activity levels. Activity will be evaluated using the Total Monitor Independent Movement Summary (TMIMS), calculated as the mean value over the 7-day period.

The study has two primary aims. First Aim is to determine whether the ACTUP intervention leads to a reduction in the risk of 7-year all-cause mortality (the primary endpoint) in sedentary adults aged 60-75 at the group-average level. The second Aim is to explore whether the efficacy of the ACTUP intervention is moderated by gender, assessing if there are gender-specific differences in the treatment effect.

2 METHOD

2.1 Data cleaning

Since preliminary data from ACTUP are no longer available, we will use data from the National Health and Nutrition Examination Survey (NHANES) for the sample size and power calculations. The NHANES dataset was restricted to participants aged 60–75 and further subset to include only those with average MIMS values below the 25th percentile. Additionally, follow-up data censored before seven years were excluded,resulting in a final sample size of 2,068 participants. Deaths occurring after the seven-year follow-up period were recoded as alive within the seven-year period.

2.2 Simulation

For Aim 1, we will assess whether the ACTUP intervention reduces the risk of 7-year all-cause mortality in sedentary adults aged 60–75 using a logistic regression model. The effect of gender, with females as the reference group, will be estimated through a logistic regression model with mortality as the outcome and gender as the predictor. A grid search will be conducted to identify the treatment coefficient corresponding to a 7% reduction in mortality.

Each simulated sample will maintain a 1:1 ratio of men to women and a 1:1 ratio of participants in the control and treatment groups. Mortality in the simulated data will be modeled as a binary variable, with the probability estimated using a logit link function that incorporates coefficients for both gender and treatment group. Sample sizes will range from 100 to 2000, increasing in increments of 50. For more precise results, the sample size range will then narrow to 1450–1550, increasing in increments of 10. The simulated data will be analyzed using a multiple logistic regression model, with mortality as the outcome and treatment and gender as predictors. This process will be repeated 10,000 times, and the power will be calculated as the proportion of iterations yielding statistically significant results out of 10,000.

The minimal sample size required to achieve 80% power at a significance level of 0.05 will be used for the simulations in Aim 2.

To determine whether the efficacy of the ACTUP intervention is moderated by gender, we will fit a second logistic regression model with mortality as the outcome and treatment, gender, and their interaction term as predictors. This model will evaluate if there are significant gender-specific differences in the treatment effect. Each simulated sample will maintain a 1:1 ratio of men to women and a 1:1 ratio of participants in the control and treatment groups. Mortality in the simulated data will be modeled as a binary variable, with the probability estimated using a logit link function that incorporates the coefficients for gender, treatment, and their interaction. The sample sizes used will align with the optimal results identified from Aim 1. The interaction coefficient will range from -1 to 0, increasing in increments of 0.01. The simulated data will be analyzed using a multiple logistic regression model, with mortality as the outcome and treatment, gender, and their interaction as predictors. This process will be repeated 10,000 times, and power will be calculated as the proportion of iterations yielding statistically significant interaction effects across the 10,000 simulations.

Parallel computing was implemented to accelerate the simulations, with all analyses conducted in R version 4.3.3.

3   RESULTS

3.1   Aim 1

Mortality in the simulated data will be modeled as a binary variable, with the probability estimated using a logit link function that incorporates coefficients for the intercept, gender, and treatment group, which are -0.48, 0.227, and -0.3004, respectively. The left side of Figure 1 shows that a sample size of 1,510 is the minimum required to achieve 80% power at a significance level of 0.05, assuming the intervention reduces the probability of 7-year

3

all-cause mortality risk by 7%.

3.2 Aim 2

The right side of Figure 1 shows that an interaction coefficient of -0.63 is the maximum required to achieve 80% power at a significance level of 0.05, assuming the intervention reduces the probability of 7-year all-cause mortality risk by 7%.

4 CONCLUSION

The power analyses indicate that a sample size of 1,510 is sufficient for Aim 1 and interaction coefficient of -0.63 is sufficient for Aim2 to achieve 80% power at a significance level of 0.05. By integrating parallel computing into this project, we can achieve faster, more efficient, and more comprehensive findings. One issue I encountered while running the code is that I couldn't implement a progress bar for parallel computing.
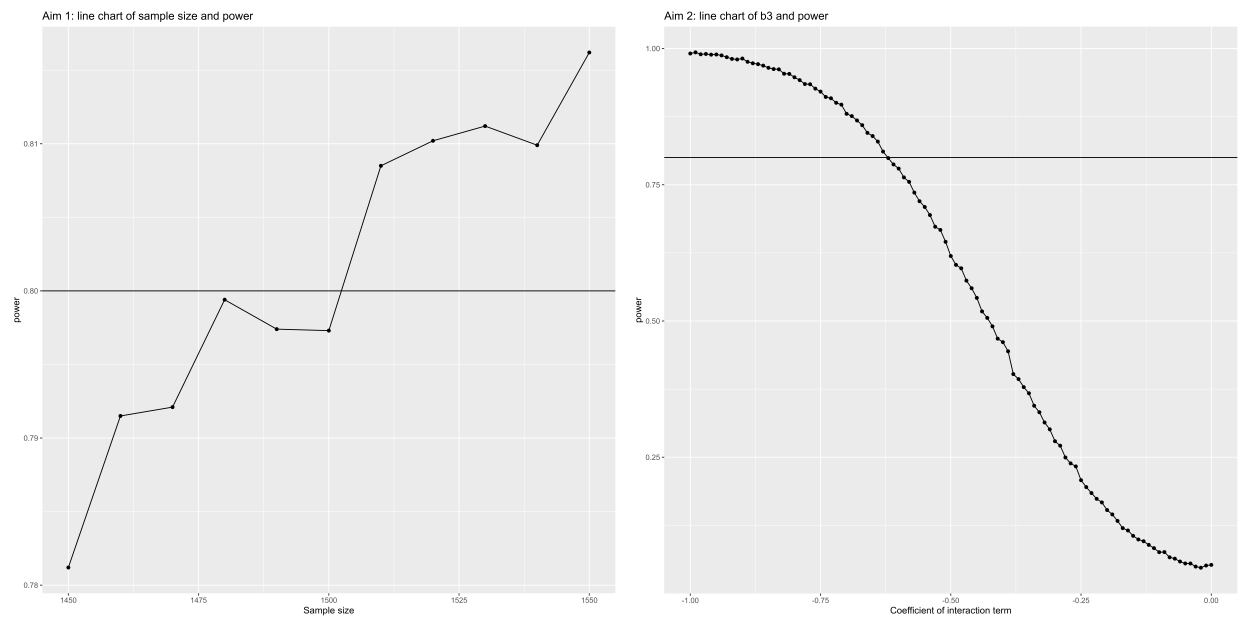
5 APPENDIX

All analysis was done by my own.

Figure 1: The figure for Aim 1 is a line chart illustrating the relationship between sample size and power, showing how power varies across different sample sizes. The figure for Aim 2 is a line chart displaying the power corresponding to different values of the interaction coefficient (b3) for gender and treatment group, highlighting the impact of varying interaction effects on power.