Final Class Project
BSTA 513/613, Spring 2021

Risk factors for fall

MrOS, a research study funded by the National Institutes of Health, began in 2000. Six clinical centers in the United States, located in Birmingham, AL; Minneapolis, MN; Palo Alto, CA; the Monogohela Valley near Pittsburgh, PA; Portland, OR; and San Diego, CA recruited 5,994 men at the baseline visit (2000 - 2002). The baseline examination included the assessment of risk factors for fractures and other conditions, including neuromuscular, visual and cognitive function tests; bone mineral density (BMD); x-rays of the spine; QCT scans of the hip and spine and the collection of biospecimens. Eric Orwoll et al. (2005) provides more information about the study.

In this exercise, we will look at a subset of the baseline characteristics (medical conditions, self-reported measures on health, body mass, bone mineral density (BMD) etc.) and evaluate how these baseline characteristics are associated with the risk of fall within a year.

Here is the list of variables:

id	Mr Os Id
site	Site
giage1	Verified Age at Enrollment (years)
mhdiab	Have you ever had diabetes? (1 = Yes; 0 = No)
mhstrk	Have you ever had a stroke? (1 = Yes; 0 = No)
mhpark	Have you ever had Parkinsons? (1 = Yes; 0 = No)
mhcopd	Have you ever had COPD? (1 = Yes; 0 = No)
mharth	Have you ever had arthritis or gout? (1 = Yes; 0 = No)
mhcancer	Have you ever had cancer? (1 = Yes; 0 = No)
pascore	PASE Score https://www.physio-pedia.com/Physical_Activity_Scale_for_the_Elderly_(PASE)
qlhealth	How would you rate your overall health? (1 = Excellent; 2 = Good; 3 = Fair; 4= Poor; 5 = Very Poor)
hwbmi	BODY MASS INDEX(KG/M**2)
b1tbfkg	Total body fat mass (kg)
b1tblkg	Total body lean mass (kg)
gsgrpavg	Avg OF right/left grip strength, kg
nfwlkspd	Walk speed in m/s using both times(calc)
b1fnd	Corrected Femoral Neck BMD (g/cm2)
b1thd	Corrected Total Hip BMD (g/cm2)
mhfallv2	During the past 12 months, have you fallen? (1 = Yes, 0 = No) -- measured 12 months after baseline assessment
mhfalln2	How many times have you fallen during the past 12 months? -- measured 12 months after baseline assessment

The data are provided in *MrOS_Baseline_Falls_Project.xlsx*.

Question 1: Association between baseline characteristics and the risk of fall within a year (*outcome variable = mhfallv2)*.

1. Build what you think is a good model to assess the association between baseline characteristics and the risk of fall within a year (***outcome variable = mhfallv2)***. Make sure to present your rationales for the steps of variable selection and the choice of your final model, assess the overall fit of the model and residuals. Find out the discriminative ability of your model in identifying a faller within 12 months. Also note that in the process of model building, you need to make sure you have adequate number of events for each level of your categorical independent variable and there is no co-linearity problem.

2. Among the 5,994 participants, the outcome variable is missing in 770 participants. Compare the baseline characteristics between those with vs. those missing outcome at 12 months, and evaluate whether the participants missing outcome information are systematically different from those having outcome information. Do you think the missing data here will introduce biases to your model?

Question 2: Association between baseline characteristics and the risk of having more than one fall within a year.

Variable ***mhfalln2*** reported the number of falls within the past 12 months. In this question, instead of using mhfallv2 as the outcome variable, use having more than one fall as the outcome variable (Yes if ***mhfalln2*** > 1; Otherwise No). Otherwise, do the same exercises as in Question 1.

- In this case, find out the model's discriminative ability of your model in identifying a faller with more than one fall within 12 months.

Question 3: Association between baseline characteristics and the rate of fall.

Variable ***mhfalln2*** reported the number of falls within the past 12 months. In this question, use ***mhfalln2*** as the outcome variable, and evaluate the association between baseline characteristics and the rate of fall within a year

- You may use the same purposeful selection process to do variable section and model building.
- In this case, no need to assess the model's discriminative ability.
- Otherwise, do the same exercises as in Question 1.

For each problem,

1. Write up a summary of your findings of up to **FIVE** pages. Use the following section headers: (1) Study Background and Objective, (2) Methods, (3) Results, and (4) Discussion and Conclusion.
    a. In your results section, provide a table to show subject characteristics using the appropriate descriptive statistics of the variables in the dataset. – Whenever you do a real-life project, it is almost always that you would have a table like this, usually as your table 1.
    b. You may present any statistics or graphs that you feel help your summary. However, be sure to display the results of your final model in a table, including odds ratios (or rate ratios for Q3), the associated 95% confidence intervals and the corresponding P-values in the table (No need to present coefficients if you present ORs (or RRs)). Make sure to clearly present your referent group for categorical variables and provide interpretation of model coefficients.
    c. Include one plot each in the summary write-up to show your assessment for the scale of continuous variables and residuals.
    d. In the conclusion, also communicate your results in a way that could be understood by the general public with limited statistical knowledge.
    e. Note that the page limit is strictly enforced and you may attach other materials and codes as appendices.
2. Create an 18-20 minute oral presentation to present your results to the class. (Think about what are the most important results/message to present.) One slide should provide a description of role of each group member in completing the class project.
3. Submit your report and presentation slides by midnight, June 9th.

Group assignments

It is up to you to find your group members and pick a research question to work on. I will send out a separate email to record your group assignments. Each research question could be selected **up to 3 times**. Q1 and Q2 will need you to fit a logistic regression model; and Q3 will need you to fit a Poisson regression model. Each group may choose Q1 or Q2; however, if some groups want to take some extra challenges, consider to choose Q3. ☺☺ Poisson regression will be covered on 6/2; however, the majority of statistical techniques for logistic regression model (purpose selection, overall test of model fit based on deviance and Pearson's residuals, model diagnostic, check for collinearity, etc.) are similarly applicable to Poisson regression.