BIOSTATISTICS 515/518, Winter 2020
DATA ANALYSIS PROJECT

Inflammation and Heart Size

The size and function of the heart are related to cardiovascular disease (CVD). A thickening of the muscular tissue of the heart (the myocardium) of the heart's left ventricle is called left ventricular (LV) hypertropy (enlargement). LV hypertrophy is a potent risk factor for cardiovascular disease (CVD), which includes ischemic heart disease, chronic heart failure, and cardiovascular death. Stress on the heart, including hypertension and aortic stenosis, can contribute to myocardium thickening. LV ejection fraction is the proportion of blood pumped out of the left ventricle with each heartbeat. Low values indicate heart problems. LV hypertrophy can reduce LV ejection fraction, and reduced LV ejection fraction is a sign of impaired cardiac function.

LV mass differs by sex and, in addition, larger people are expected to have higher LV mass.

LV mass is a size measurement and LV ejection fraction is a measure of heart function. Echocardiolgraphy (ECG) or magnetic resonance imaging (MRI) can be used to measure LV mass and LV ejection fraction.

Several biological pathways have been observed to be associated with LV hypertrophy. Inflammation is believed to damage the inner lining of the arteries and make having a heart attack more likely. C-reactive protein (CRP) is a biomarker of general levels of inflammation in the body. However, the connection between high CRP levels and CVD risk is not well-understood.

Data

The Multi-Ethnic Study of Atherosclerosis (MESA) is one of the largest studies of a racially diverse population of older Americans. When participants enter the MESA study they are asymptomatic of CVD (no clinical disease), although they may have subclinical CVD. Extensive data collection in MESA includes inflammatory and hemostatic biomarkers. MESA also collected MRI data, including LV mass and LV ejection fraction. MESA captured technically acceptable MRI images for 73% of participants (5004 out of 6814 MESA participants). Most missing MRI data occurred because the participant was ineligible for MRI due to a metallic fragment, implant, or device in the body (7% of 6814), or because the participant was claustrophobic and unable to complete the MRI. (14% of 6814).

In the dataset you will find information on LV mass (lvmass), c-reactive protein (crp1), and other variables. Although MESA is a prospective cohort study, the data available for this project are cross-sectional.

The **Aims** of your investigation are:

1. Characterize the association between C-reactive protein and LV mass in the population.

2. Investigate the evidence that inflammation, as measured by C-reactive protein, contributes to LV hypertrophy.

Comments

1. Investigators graciously provided us with these data for teaching purposes only. The datafiles must not be shared with anyone outside of the class. You should delete the datafiles at the end of the project. You have signed an agreement that you will honor these terms.
2. The two project aims may seem similar at first glance. However, it may be appropriate to adjust for some covariates for one aim and not the other aim.
3. It is likely you are not familiar with C-reactive protein. Consider its distribution and remember you have options for how to include a variable in a regression model (e.g., you can choose to transform it).
4. Methods for addressing missing data have not been covered in this class and are *outside* the scope of this project -- it is anticipated that you will do complete case analyses. It is *within* the scope of the project to consider the role of missing data and possible impact on your results.
5. Extensive literature review is outside the scope of this project and should not be done. Any literature review should be limited to helping you "learn the basics" about LV hypertrophy and/or inflammation. Reviewing the literature on inflammation as a mechanism contributing to LV hypertrophy should *not* be done.
6. Although groups will collaborate on the project, only one person per group should submit the final paper on behalf of the group.
7. Label your submitted file according to this convention: Group03.docx for group 3; Group11.docx for group 11; etc. Word format is preferred but PDF is also acceptable.
8. Your paper should have a title. Do *not* write names in the paper – they will be graded blinded. You can write your group number as the authorship.
9. The most important part of your paper is its **abstract.** A abstract that is clear and both complete and concise is very important. In order to be complete, your abstract for this project might be longer than a typical abstract in the literature.
10. Typical organization of a paper is: **Introduction, Methods, Results, Discussion.**
11. **5 page maximum, double-spaced.** The abstract can be single-spaced.
12. Your job is to perform appropriate data analyses and describe and justify your analytic choices. Since most of us are not experts in this subject area, you will be graded on whether analytic choices are appropriate given some sensible reasoning, not whether you made the same choices that an expert in cardiovascular disease would have made. Of course, your reasoning has to meet a standard of common sense. For example, a claim of a causal effect of smoking on race would not be reasonable.
    a. Your paper should resemble a paper in the scientific literature, except perhaps for a longer abstract and/or greater detail on the rationale for covariates.
13. Analytic methods not covered in Biostat 514/515/517/518 are outside the scope of this project. For example, we have not covered methods for handling missing data (see comment 4).

14. Direct questions to the course instructor. Use the course discussion board to ask questions if appropriate. If you are unsure, email your question to katiek@uw.edu rather than using the Discussion board. You should not consult with students in the class outside your group, and you should not consult outside sources.
15. A grading rubric will be available within ~1 week of the due date.
16. Some additional guidelines
    a. Do not plagiarize, including from this document
    b. Take care when using language that sounds causal.
    c. It is best to reserve the word “significant” for either statistical significance or clinical significance. If you really mean “large” or “substantial” or “important” then use one of those words instead of “significant.”
    d. “Gender” and “sex” are not synonyms. You should use these terms correctly, even when others do not. There is a variable in the dataset labeled gender, but it seems clear this is actually recording subjects' biological sex rather than gender or gender-identity.
    e. Proofread your reports for spelling and grammar.

You are assigned into groups of 4-5 students for this project. Your group should work together to write a paper reporting your analysis. Group members should contact each other and develop a plan to proceed. You may only discuss the project with members of your group. Direct all questions to the course instructor.

A literature review is outside the scope of this project and should *not* be done. Few or none of you is an expert in cardiovascular disease (nor am I).

NOTE: Investigators graciously provided us with these data for teaching purposes only. The datafiles must not be shared with anyone outside of the class. You should delete the datafiles at the end of the project.

Each group should work together to produce a single paper that addresses the study goals. Please follow the following guidelines

- Although groups will collaborate on the project, only one person per group should submit the final paper on behalf of the group.
- Label your submitted file according to this convention: Group03.docx for group 3; Group11.docx for group 11; etc.
- Word format is preferred but PDF is also acceptable.
- Label your paper with your **group number** but **no names**
- Your paper must include an **abstract.** A clear abstract that is both complete and concise is very important.
- Typical organization of a paper is: **Introduction, Methods, Results, Discussion.**
- **5 page maximum, double-spaced.**
- Your job is to perform appropriate data analyses and describe and justify your analytic choices. Analytic methods not covered in Biostat 514/515/517/518 are outside the scope of this project. For example, we have not covered methods for handling missing data, so these methods are outside the scope of this project. (That does not mean you should completely ignore missing data.)
- Direct questions to the course instructor. Use the course discussion board to ask questions if appropriate. If you are unsure, email your question to katiek@uw.edu rather than using the Discussion board. You should not consult with students in the class outside your group, and you should not consult outside sources.
- There is a grading rubric available.
- A primary way your report will be evaluated is whether you provide reasonable rationale for analytic choices.
- Here are some additional guidelines:
    - Do not plagiarize, including from this document
    - Take care when using language that sounds “causal”.
    - It is best to reserve the word “significant” for its technical meaning in statistics. Etymologically, “significant” means “signifying something”. If you really mean

"large" or "substantial" or "important" or "major" then use that word instead of "significant."

- "Gender" and "sex" are not synonyms. You should use these terms correctly, even when others do not. See http://www.who.int/gender-equity-rights/knowledge/glossary/en/
- Some students mistakenly think they should identify a single analysis that answers all questions. This is very rarely appropriate. Different questions require different analyses.
- Proofread your reports for spelling and grammar.