Comments on the paper authored by Group 04 as refereed by Group 03

1. Summary
    a. Do you intend to state more than a single primary goal? I don't think you need to, and expect that "goals" was just a typo
    b. Including actual estimates/p-value/conf. interval will help strengthen your statements here
    c. Are there any limitations that could be briefly mentioned?

2. Background
    a. These are great descriptions of the biomarkers, but I'd like to know if there is any reason why we are suspicious that they could be associated with cardiovascular disease.
    b. If you look at the project assignment description, identifying the "overall goal" in terms of public health impact of disease would help us understand the importance of the study

3. Questions of Interest
    a. You might try to rephrase the questions rather than those listed on the dataset description. For example, when mention the "predictive value", you might instead state what kind of summary measure you chose.
    b. Again, reference the project description as Scott wanted the "client's" questions as well as "the questions you answered" – somehow distinguishing these would be helpful.

4. Description of the Data

    a. During discussion sections, Scott has talked about re-organizing your list of variables into some broader categories. It would help your reader to know we have some demographic characteristics, some behavioral characteristics, some functional (test) characteristics, and some past medical history.

5. Statistical Methods
    a. Please state the software used
    b. I like how your paragraphs are ordered to match the questions of interest.
    c. The project assignment description asks for two levels of analysis. The first should be a technical description fit for publication, which is what you have provided – it's very clear and well written. However, the assignment also calls for a lay explanation of your methods.
        i. For example, I'm the reader without any stats background – what are your Kaplan Meier curves telling me?
        ii. Another example: does Missing at Random mean the missing data are OK? Help the reader understand that you're not worried about any of the missing data patterns.

iii. You might also explain how your regression models will help you identify effect modification and/or confounding.
    1. I find the last paragraph a little confusing. It sounds like you are using the regression for identifying confounders. Scott has mentioned that we want to rely mainly on descriptives for identifying confounders/precision variables: http://emersonstatistics.com/courses/formal/b517_2006/qa1.txt

d. In statistical method part, when you transfer the data to log scale. We would like to have the reason why you do so. (Any scientific reason cause you to model CRP as a continuous log-transformed variable? Is the log transformed data leads to a better model?) Please explain this in the paper. You explain why it's transformed well in the results, but a brief explanation here e.g., "because it's skewed…" here could help clarify this section.

6. Results
    a. The explanation for transforming CRP is very good here. Maybe in the methods you could mention that you will transform and explain the rationale in the results.
    b. "It appears from the table that being black, having prior AD, diabetes, higher BMI, and higher cholesterol are associated with higher fibrinogen levels" – should estrogen be included in this sentence?
    c. In the descriptive statistics part, we suggest more detailed words on your table 2, rather than just a conclusion: you find the association.
    d. In the "Differences in association between biomarkers and mortality, defined by survival status at three years" part, we are not very clear how you combine the standard error. We think it would be helpful (in the method part or in the result part) to define the way you calculate the combined standard error. This is the same as other effect modification variables.
    e. In the associations sections, the language is very good for another statistician – would be great for our homework assignments – but each paragraph could use a little "translation" to explain the findings in lay terms.
        i. The sections that would benefit the most from this would be those on effect modification from 3 year survival and sex. How exactly do the findings from your tests after stratification prove to the reader that effect modification exists? Your methods look sound, just difficult to interpret for a non-statistician
    f. The most pressing need is further explanation for your choice of covariates in the final model (e.g., race, smoking status, estrogen use, and diagnosed diabetes). You've shown that these variables are associated with the POI, but have yet to show that they are associated with mortality. That might be something you do up in the descriptive statistics section, but regardless of where you put it, you can then explain that you think these are confounders (or precision variables) and thus included for adjustment. One thought would be to show the KM curves or KM-outcome tables for those variables.

7. Discussion
    a. For the limitation of the study, we prefer more detailed description.

b. Future research plan might be mentioned at the end. Did any findings in the study seem suspicious or interesting – what else would you study?

8. Figures/Tables
    a. For the format of the tables, we prefer the title of the graph/table stay in the same page as the graph/table. Some tables are broke up between pages, you might need to fix that.