Comments on the paper authored by Group 3 as Refereed by Group 2

Summary:

- What confounders and precision variables did you adjust for?
- In general, the summary could use more statistical inference (i.e. estimates, confidence intervals, p-values). Right now it seems to contain a lot of words but not a lot of information (in particular, your questions of interest).

Background:

- The background part may be too similar to that in the data description document
- For ". *It has become an important tool in the diagnostic and prognostic repertoire of most modern medical institutions as it provides high resolution visualization of structures not easily captured on computed tomography (CT) or x-ray imaging"*

I might rephrase it as "It has become an important diagnostic and prognostic tool and provides high resolution visualization…"

Questions of Interest:

- The way you write this part may not be appropriate for an analysis report aimed at a common person. It seems that your reader is only Dr. Emerson.
- I am not sure whether it is appropriate to merge the 4 questions in the document into the two questions you proposed. To me, each of the four questions in the documents need to be answered separately.
- Re: Q1, what do you mean by "lifespan"? Survival?
- Re: Q2, I might reword "…association true after…" to something like, "What is the association after adjusting for…" Using "true" seems like too strong of a word, especially for a presumably non-stats audience.

Source of the Data:

- It was helpful that you discussed censored data

Statistical Methods:

Descriptive Statistics: Technical

- Your sentence: "In Table 1, descriptive statistics for these variables…." Your table 1 should identify potential confounders, describe the sample, reflect trends, etc., but it should not detect associations, as that is beyond the scope of descriptive statistics.
- Your sentence: "The categories of atrophy have no scientific relevance, but…." Is there an inherent bias to your method of categorization? I believe categories should be either based on scientific inference or the range of variable, but not necessarily in order to create equal subsample sizes. It might help if you just say what those categories were.

Descriptive Statistics: Theoretical

- Your sentence: “Risk factors that are confounders will likely be associated with…” They, by definition, MUST be.
- On a more subjective basis, I found your organization of methods a little redundant. I think it’d be preferable to identify your technical methods and justify them with theory (or note the conclusions that may be drawn from the method) all at once, rather than having the two different sections.

Inferential Statistics: Technical

- What is your justification for using Huber White sandwich estimation?
- Second paragraph, your sentence: “These covariates will be added in a stepwise, model-building approach; covariates that do not impact the model will be excluded.” How did you determine if a covariate impacted the model?
- Why do you make atrophy score a categorical variable? Why not keep it continuous?

Inferential Statistics: Theoretical

- First paragraph, your sentence: “This model requires Cox proportional hazards, which is the…” That’s not true. Cox PH is very common, but I don’t think it is the ONLY way.
- Second paragraph, your sentence “…we then have to ask how these variables behave: are they linear? Quadratic? Log-scale?” If we are talking about behavior, multiplication, additive, etc. is more meaningful to me than quadratic, log-scale, etc. And more importantly, I want to know which one you chose.
- Next sentence: “Any variables that, on scientific grounds, may not be linear were modeled with either dummy variables or linear splines.” I do not understand what you mean. Are you talking about continuous variables? Even if a variable is known to be nonlinear, it is sometimes still better to model it as such. (Assuming it quantitative.)
- A couple sentences later: “The associated confidence intervals tell us that with 95% confidence, the true…” This is not entirely true. To be explicit, it says that if we do this same experiment over and over again, 95% of the time…
- Last couple sentences of the last paragraph: What is DSST? I don’t understand your justification for exclusion.

<u>Results:</u>

- The paper might flow better if you place all the figures and tables at the end. Additionally, the tables were hard to read (font size, orientation, etc.) and the graphs were too small to make sense of.
- Third paragraph: It would be more digestable if you phrased differences in context, such as instead of “*larger weight and height*” *and* “*higher creatinine*”

  Say “heavier,” “taller,” or “have worse renal function or evidence of kidney dysfunction” which is what creatinine is a surrogate measure for. Hmm, it might be extrapolating too much to make the leap with creatinine, but overall I recommend translating what a higher or lower value of a variable means.

- Fourth paragraph (sentence): “Most of the data appears to be normally distributed. However, platelet count may be mildly right-skewed.” Can you explain this further?

- Since atrophy score is your POI, I think you should really highlight that KM plot.
- You keep mentioning that your categories have no scientific significance. Is this really true? Maybe the thresholds are not of particular importance, but you noted in your background section that cerebral atrophy is a measure of the overall decrease in brain mass. Is the atrophy score not a quantification of that decrease? So categories with higher scores can be compared to those of lower scores in some sort of scientific meaning?
- Any thoughts on running the model using atrophy score as a continuous variable?

Discussion:

- From your first paragraph, I cannot tell if you think atrophy scores are clinically important. Maybe expand a little on its unadjusted predictive value? (I.e. your question 1)
- Maybe you could further discuss what covariates are the major confounders of the association of interest, what are the potential modifiers and what are precision variables.
- In the second paragraph, you mentioned that brain atrophy might be a clinical sign for cardiovascular disease in nature. To me, this is counter-intuitive and hard to believe. Maybe there is an association between the two because they are both associated with another risk factor such as age.
- Overall, what did you learn? You stated that you failed to reject the null hypothesis, but did note some impressive potential associations from the KM curves and descriptive statistics. What explains the discrepancy?
- Limitations and strengths of the analyses plan may also be discussed.

General Comments:

- It seems like it might be more work to provide both explanations – the lay version and the technical version. Scott could probably shed some light and save you some work