Comments on the paper authored by Group 11 as Refereed by Group 10

Comments by Section:

Summary:

1. "We find that in …. Physical activity, and stroke" Define what you mean by stroke.
2. Add confidence intervals and P-values if you claim the results are significant
3. "Exploratory analysis suggests that atrophy levels reflect underlying risk factors separate from other known diseases (cardiovascular disease and stroke)."- What does this sentence mean?

Background:

1. "For several years now researchers have found higher rates of atrophy in older brains,…" should this be older individuals?
2. "…but the scientific community has not fully understood the implication of a brain with atrophy"- this sentence is not clear.
3. When you talked about the 3 different beliefs in the scientific community, the categorization was not clear.
4. You defined atrophy score but what is atrophy? Briefly define it.

Questions of interest:

1. "…relationships between atrophy and other common measures. "- Common measures of what?
2. "Other studies have shown that several variables available have been found to be predictive of mortality, so we perform a secondary analysis to examine whether the additional information gained by knowing the atrophy score is relevant if we already know a variety of other health measures. "- This sentence was not clear. What is the main idea you wish to express here?
3. The third paragraph does not clearly explain what your third question of interest is.
4. "Note that the validity of this analysis depends on the measure of cardiovascular disease being complete so as to not confound atrophy with other diseases." Need to specify that you are talking about stroke and cardio vascular disease.

Source of Data:

1. "Our dataset contains 735 generally healthy patients and 30 variables that were recorded for each patient over an eleven year timeframe"- need to state that variables were measured at baseline. This sounds like they were measures periodically over 11 years.
2. "…with the intention of observing the incidence of cardiovascular disease and cerebrovascular disease." If this was the original study then you need to explain why it is okay for you to use this data for predicting death by atrophy.
3. ". Subclinical risk factors were also included, such as systolic blood pressure, LDL (usually considered "bad") cholesterol levels, FEV (maximum amount of air forcibly expelled in 1 second), albumin (a liver function indicator), creatinine (a kidney function indicator), platelets (indicates ability to blood clot), and DSST (a test of cognitive ability)." For platelets, I believe you mean platelet count? Secondly, write out what DSST stands for at-least once in the beginning.

Statistical methods:

1. "…we will create a table of descriptive statistics" specify table number here.
2. "We also present univariate regression analyses as exploratory in this question, since we will not take into considerations issues regarding multiple comparisons as we are simply trying to get an overall picture." What kind of regression? Mean, geometric mean etc? Secondly, how do you get your standard error, robust or not? Wald statistics or something else?
3. "We first stratify atrophy scores into three groups (0-33 for "low", 34-66 for "medium", and 67-100 for "high"), "- Repetition of stratification not needed.
4. "To obtain statistical inference, we perform a Cox Proportional Hazards Regression of the variable death on atrophy."- Again, you need to specify robust SE or not.
5. You spent a large part of this section explaining the theory of cox proportional hazard model. We feel this is not needed or at least you should make it more concise.
6. "After discussion with our team, we adjust for the variables atrophy, age, weight, pack years (smoking),…" –explain what the variable pack years means.
7. "…years after quitting smoking, alcohol intake, physical activity, platelets, coronary heart failure,"- For platelets: again specify if this is platelet count or what this measures. Similarly, for coronary heart failure I think you mean history of coronary heart failure.
8. "We decided to not include the disease variable coronary heart disease because we decided that the variable coronary heart failure is a stronger indicator of heart issues." - Explain this sentence? Not clear why the variable was excluded?
9. "We did include stroke, which we modeled as a dummy variable since it is an unordered categorical variable."- No need to mention this or improve sentence: This sentence sounds like you included stroke BECAUSE it was a categorical variable.

Results:

1. "Measures for participants with medium levels of atrophy tended to be between low and high levels, so they are omitted here, since no major non-linarites were found."- Did you perform a test of non-linearity or was this just an empirical observation.
2. "…history of congestive heart failure and stroke will be discussed more in the following"- what does this sentence mean?
3. "The participants' weight and platelets level are also considered for scientific reasons:…" what does it mean to be considered? Include in the model or do you study them separately?
4. "Thus, we estimate mean atrophy score 0.698 higher for a group one year older; 6.182 points higher in males than females; and 6.622 higher in a group of people with stroke prior to baseline than those who do not have any prior cerebrovascular disease." -Need to specify if this was a risk difference or risk ratio?
5. "Years since quitting smoking is a complicated variable, since its interpretation is related to the pack years smoked, so we did not consider it in this part of the analysis. "- This sentence is not clear.
6. "Table 3 and Figure 1 present these results."- Table 3 is empty.
7. "We performed a secondary exploratory analysis of associations between atrophy and non-MRI covariates…"- What are non-MRI covariates? Variables other than atrophy?
8. Instead of always referring to the methods section it might be helpful to briefly say by linear regression or by cox proportional hazards regression etc.

9. Generally results section was similar to assignment solutions. Please make sure this is not too similar to Scott's solution keys and perhaps you may want to make it less repetitive and present different results more concisely.

10. "The final line in Table 4 presents the results of the Cox Proportional Hazards Regression model adjusting for the variables coronary heart failure and stroke"- This line should also mention diabetes as in Table 4 or perhaps table 4 should not mention diabetes.

Discussion:

1. "The final line in Table 4 presents the results of the Cox Proportional Hazards Regression model adjusting for the variables coronary heart failure and stroke"- This sentence is not clear.

2. Perhaps for the discussion section you could include recommendations or possible limitations of study?