Comments on the paper authored by Group 01 As Refereed by Group 14

Summary

The role of risk factor is not clear in the second sentence. Is this a variable you are assessing? Or a group of variables defined as risk factors?

Is the risk of death an absolute or a relative increase? Specify what this difference is measuring.

The definition for the risk factor is confusing. The questions of interest are not clear in the summary part. In the conclusion, it seems that atrophy score is response variable in the linear model, but in the 4th line, atrophy score looks like a covariate.

The estimated risk of death was estimated by proportional hazard regression, not Kaplan-Meier methods.

Questions of Interest

It may be easier to read if the questions were limited in more of a list.

It is unnecessary to state methods used in this portion. Just state the main questions you want to answer and reference these questions in the stats methods to explain what was used to answer which question.

Question 4: Still not clear what risk factors you are exactly looking at here or what these known disease processes are.

Also, make sure your primary question of interest is clear and distinct from secondary analyses.

The second question will be addressed by using proportional hazard regression examining atrophy and survival.

> In the statistical methods, Kaplan-Meier curves was implemented for survival rate among atrophy groups.

Source of Data

Observation time and death are censored data.

Statistical methods

Paragraph 1: how will “different across cerebral atrophy score groups” be determined? What values will lead to a conclusion of difference?

Paragraph 2: it is unclear whether the risk factors are individually regressed with cerebral atrophy or if they are all individually modeled.

Paragraph3: Why do you choose 5.75 year restricted mean?

Results

The description of missing data is tedious and could be more easily summarized by a statement such as: "Data was > XX% complete". Specific missing data can be given in the table.

Paragraph 1: are smoking and alcohol history the only behavioral variables used? You should either mention all of them here or somehow mark them in the table to convey which ones you though had an association.

Table1: There is no indication for missing data for each variable? From table 1 it seems like all variables are balanced. Sample sizes within atrophy score strata have huge difference. I think the categorization should be adjusted. Or is there scientific evidence to categorize atrophy score like this?

Nice analysis to missing data pattern. But I think this part should be reflected in the Table 1.
Table2: Which variables are confounders in each adjusted regression model?
And I think interpretation to a pvalue should also be included in the result part.

Table 2 & 4: need to interpret the regression coefficients in a table legend. I.e. "for each 1 unit change in the predictor..." and then list the units clearly on each line of the table.

Figure 1, Table 1 & Methods: How did you determine the cutpoints for the atrophy score? Are these based on previously reported cutpoints? Tertiles of the data? How clinically relevant are these cutpoints? i.e. what do they mean?

Cerebral Atrophy and Survival Paragraph 1: Your tables do not match up with your claim “The lowest atrophy score group averages 4.4776 years while the highest group averages 5.5274 years.” This statement does not match the rest of the claims you are making. It appears the lower score group has the higher year value.

Paragraph 2: State which risk factors are statistically associated with survival or make it obvious in your table which ones you are stating are significant.

What does it indicate when some potential confounders in the model are not significant? Give some interpretation.

Cerebral Atrophy Section: Are the percent risks or death differences relative or absolute? Make sure this is clearly stated.

Discussion

Paragraph 1: I think the reason not to choose height, physical activity and AAI as potential confounders is that these variables may not be associated with atrophy score in the sample.

Paragraph 2: Is a 30 unit change in atrophy score actually more clinically relevant or does it simply demonstrate a more noticeable change in risk of death? 30 units can still be used but if labeled as clinically relevant it should have some true background music

Paragraph 2: Are your Kaplan-Meier estimates derived from a proportional hazards regression? The current wording makes it sound like it is from the Kaplan Meier plot which is not appropriate given the claim you appear to be making.