

## Comments on the paper authored by Group 12 as Refereed by Group 11

### Overall comments

- Please paraphrase for questions. Not just copy and paste from inflamm. Doc
- Why are the inflammatory biomarkers C-reactive protein and fibrinogen important? What are these measuring in the body (clinically speaking, foregoing any associations to death in the data)?
- Justify (or at least explain) how you came to the decision to split C-reactive protein up to 4 levels.
- It's unclear what short and long term is. Did you dichotomize time into short (1-3years) and long (3-6years) or something like that
- How were risk factors identified? Is it just common knowledge? Are confounders chosen by descriptive statistics? Why did you choose those specific covariates? Did you omit any available ones? If so, which why?

### Summary

- Please provide numbers for **not significant evidence** of effect modification by sex.
- When mentioned about similar trends in CVD-specific mortality but the estimates were more **extreme**, we are wondering what about p-values in CVD specific mortality; whether the adjusted estimate was the only non-significant one as well.

### Background

- What do you mean by "longevity"?
- Alternative suggestion is great! But when mentioned CVD risk factors are **protective in these diseases**, could it be overpowering or hiding these other diseases?

## Source of Data

- Please provide the name of the four regions of the United States the data comes from. I'm not convinced this won't effect interpretation of the final results, even when randomly sampled.
- When you mentioned the third point of unavailable data, it might be worth explaining why you would want this data more specifically

## Statistical Methods

- You compare characteristics between cases with missing data and cases with complete data, why not just throw out all data entries with missing points? Explain why you didn't or won't do this if you don't like that idea.
- When mentioned estimate the probability of survival for 3 years and 5 years, one suggestion is to explain briefly how and why Kaplan-Meier curves work. We are assuming an audience that is not very knowledgeable, so it would be hard to make the assumption that they know what Kaplan-Meier is. We don't have to teach them, but we can explain briefly how and why it works.
- What do you mean that the curves cannot be **estimated perfectly**? What about the curves is incorrect?
- Why aren't the lines - the estimates of the 10th, 50th (median), and 90th percentiles on the Kaplan-Meier plots? Just checked and didn't see them.
- Nice description for the part of the model for assessing whether elevated levels of these two inflammatory biomarkers are associated with earlier death.
- To deal with the log transformation problem when C-reactive protein levels equal to zero, is this standard to do **replace “zero” values** with half of the minimum observed non-zero value? If so, please mention that. Is this the only way? It seems really messy and confusing.
- One suggestion is to explain hazard ratio briefly, or describe it in terms of "instantaneous risk of death" which is easier to understand.
- You mentioned the estimated hazard ratio would be **significantly** different from 1.0

if effect modification exists; I think you are right, but the wording might be strong for the method used to reach this conclusion. Many other factors go into this.

- Is any further evidence to say whether the analyses should be treated as **confirmation** of any interaction in the primary analysis?
- Briefly explaining why disease pathways may be different between sex.
- Since you are comparing many variables across a few groups multiple times, do you think a multiple comparisons problem exists? If so, how do you account for that? Is it negligible? Is it mitigated in some way?

## **Results**

### Descriptive Statistics

- The only missing values in the dataset were in the biomarkers?
- Maybe you can restructure the document to have the relevant tables close to your descriptions...
- Table 1 & 2 : range (min- max) would be better to provide
- Figure 1: it's better to re-label the y-axis for the Kaplan-Meier curves. Perhaps cut out the bottom half of the chart and have y go from 0.5 to 1.

### Analysis of Time to Death

#### *Primary (Unadjusted) Analysis*

- Beautiful explanation for hazard ratios.

#### *Comparison of Short-Term and Long-Term Associations*

- Check the math on “remaining **4571** subjects”. This is 4899 subjects who were included minus 330 observed to die which should be 4569 remaining instead of 4571 right? Where did those other 2 people go?

- Maybe you can state more clearly that inflammatory biomarkers are better indicators of short-term survival **than** what?

#### *Adjusting for Sex and Known Risk Factors*

- In the last sentence, it might make it clearer to show explicitly in the same sentence that number "show" this. (Rather than “This shows...”)

#### **Discussion**

- The discussion seems like a suggestion, but the discussion should be more about the implications of the results, not a summary.
- Perhaps it's good idea to give an explanation on why not to over-interpret the results; the wording “over-interpreting” should be check as well because "over-interpreting" any results is not good.
- Are there additional reasons to explain the association is not mitigated by the inclusion of other risk factors?