Hello Jane and Amin,

Sorry please discard my previous email, it is not yet finished. Here is the full email:

Thanks for the good work in building this web application, it looks so great! I have written the function for prediction and attached the code here. In addition, I made a few revision to the other codes, just for your information (I am not sure whether they are the right codes),  because I have some concerns and comments, could you please advise? Thanks!

- for the regression code: I like the current version, it is what we want, I didn't change anything.

 - for the prediction code:

   a)  I predicted 20-year changes in two scenarios: quitted (or never smoke) and continue to smoke (use current daily-cigarettes to calculate cumulative smoke packyears over time).

   Please see comment 4 for the calculation of smoke packyears.

   b)  I simply used the results from the regression code. But can you change the input variable's names, make them the same as the variable names I have in data\_rf4, so that the prediction function can work? Otherwise it doesnt work.

Comments:

1. For Jane's second question, regarding data\_rf4 is not the right data - Actually, it is the right data. Please note that our prediction model is centered and scaled, data\_rf4 is the final dataset with centered variables that are continuous and time-fixed (including fev1 itself).

    -  To answer your question, why we can't replicate Table 2's coefficients, it is because of the same reason - the coefficients from data\_rf4 needs to be back-transformed to generate Table 2's coefficients, because the former is for transformed data, and the latter is on original scales.

2. Related to comment 1, I am wondering, what are your functions "FEV" and "FEV\_RC" for? If you want to calculate rate of decline, similarly, the predictors need to be centered and scaled using data\_rf4's mean and SD parameters, and then the result needs to be back-transformed.

   - Don't worry about the transformation of predictor variables, in my prediction function, I did the transformation on final dataset (full\_file\_name right?), so your FEV and FEV\_RC function can go following it.

3. Current data input did not include baseline FEV1, but we do need it to do the prediction. If the patient has no clue, perhaps we can give 3 options: 1.25L, 2.25L, 3.25L, just like Zafar did in Mohsen's another web app. Please check what I added to the "FEV\_sidebarPanel" codes

4. Some of the predictors, such as alcohol index and cumulative smoke pack-years, are DERIVED variables, and should not be directly asked.

    - alcohol index is calculated by weekly intake of beer, wine and cocktail (highballs in specific). Please see what I have revised in the "FEV\_sidebarPanel" codes. Could you please make similar changes?

    - cumulative smoke years is derived by years of active smoking and daily cigarettes. I used "daily cigarettes" to calculate future smoke years if the participant does not quite smoking.

5. Please add an option that, for age>=65 years, abandon age\_category in the prediction. Framingham data do not have patients aged>=65 years, so if we include the age category, it wont predict for people aged>=65 years. Let's chat.

Please let me know if you have any questions, my comments and edits are in the attached codes

Good job, really really appreciate it!!!!!!!!!!

Wenjia

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

On Thu, Nov 23, 2017 at 3:16 PM, Jane Krot <[janesnow179@gmail.com](mailto:janesnow179@gmail.com" \t "_blank)> wrote:

Hello Wenjia,

Thank you for your email!

Below, I have provided answers to your questions.

In total, there are three things that we require for the web application - function that will take custom inputs and generate an lmer function, the correct dataset and make\_predictions() function.

Please view the attachment, which contains a table describing the three functions in detail.

*What would you like me to program?*

**Function #1:** Could you please provide a function that will take custom inputs and generate a correct lmer function? In my code (attached to this email), I do have a function which generates the lmer function based on the inputs. You may find it useful - FEV\_calculate\_lmer\_fn() function, located in the attached FEV\_functions.R file.

**Function #2:** At the moment, the dataset (i.e. data\_rf4) does not generate the estimates which correspond to the covariates in Table 2 of the manuscript. Therefore, could you please provide the correct dataset or the formula that will produce the correct dataset?

*Please note***:** At the current moment, when generating the lmer summaries, I am using data\_rf4. Therefore, the values of the estimates are incorrect.

**Function #3:***The predictions right?*

Yes, could you please write a function make\_predictions().

*Mohsen provided the following structure:*

make\_predictions<-function(lmer\_object, predictors)

{

                #Take the predictors and make predictions using the lmer object. The results will be processed by us to generate figures and tables.

#Includes prediction interval, or confidence interval

}

Thank you for your time and help!

If you have questions about anything, please feel free to contact me any time.

Respectfully yours,

Jane Krot

P.S.

**Appendix (optional - might be helpful for understanding the structure of my code)**

The code for the web application consists of four files**.**

**Description of files:**

1. **FEV\_7\_4.R**: In order for the web application to work, the main file that needs to be run is the FEV\_7\_4.R. You will probably need to update the setwd() within the FEV\_7\_4.R file.

2. **FEV\_functions.R**: File with list of functions used in the FEV program.

*Please note***:** At the current moment, some of these functions are not being used to generate the lmer summary. At the beginning of developing the Shiny web application, I generated a linear regression and plot that corresponded to the manuscript (i.e. when all 17 inputs are entered into the web application by the clinician).

3. **FEV\_na\_inputs\_check.R**: Function that checks for NA in inputs and returns NULL if input is not available. Inhibits errors being displayed in place of  reactive outputs when inputs have not been updated initially, or when inputs are intentionally left blank.

File will be used when we have reactive outputs, i.e. a plot that is dependent only on user inputs. At the current moment, this file is not playing any role in the application.

4. **FEV\_sidebarPanel.R**: A sidebar panel containing input controls.