# Overview

Title of project:

The Relationship of Diagnostic Markers with the Ebb and Flow of Pneumocystis Colonization

Author of project:

Emily Rayens

# Instructions

Write your comments and feedback below for each section/compontent of the project. The goal should be to help the author improve their project. Make comments as constructive and actionable as possible. You can provide both criticism and praise.

For each component, pick one summary statement by deleting the ones that do not apply and keeping only the one that you think most closely summarizes a given component.

Delete any sections/text of this template that are not part of your final review document.

# Specific project content evaluation

Evaluate the different parts of the project by filling in the sections below.

## Background, Context and Motivation

How well is the context of the project described? Is a comprehensive background, including summary of previous/related work given? Is the project well placed into the context of existing work (including proper referencing of existing work). Is it clear why the project was undertaken and what new information it hopes to provide?

### Feedback and Comments

There are enough descriptions about the object studied. The author well explained what is limitation in the field. It is clear why the project was conducted. However, it is not a comprehensive background and no citation.

### Summary assessment

* some contextualization and motivation

## Question description

How well and clear are the question(s)/hypotheses the project aims to address described? Is it clear how the questions relate to the data?

### Feedback and Comments

The question is clear and related to data.

### Summary assessment

* question/hypotheses fully clear

## Data description

How well is the data overall described? Is the source provided? Is a codebook or other meta-information available that makes it clear what the data is?

### Feedback and Comments

Data was explained good enough. Each variables were explained what is the meaning. However, the author forgot to mention about time which is included in Survival curve.

### Summary assessment

* source and overall structure of data somewhat explained

## Data wrangling and exploratory analysis

How well is the data cleaned/processed and explored? Are all steps reasonable and well explained? Are alternatives discussed and considered? Are meaningful exploratory results shown (e.g. in the supplementary materials)?

### Feedback and Comments

I do not know how the author cleaned and processed data. There is no code for this process.

### Summary assessment

* major weaknesses in wrangling and exploratory component

## Appropriateness of Analysis

Were the analysis methods appropriate for the data? Was the analysis done properly? Were different components of the analysis (e.g. performance measure, variable selection, data pre-processing, model evaluation) done in the best way possible and explained well?

### Feedback and Comments

There is no description about methods which were used to analyze the data. There is no variable selection, model evaluation, and multivariate analysis. It seems to me that the author analyze the data somewhat properly. The author should describe the methods for analysis. For example, using cox proportional hazard to analyse time to event data. Which variables you use to adjusted for or include in the final model, what is crude and adjusted hazard ratio.

### Summary assessment

* defensible but not optimal analysis

## Presentation

How well are results presented? Are tables and figures easy to read and understand? Are the main figures/tables publication level quality?

### Feedback and Comments

The tables are not met the publication level of quality. The author did not include the captions for some figures and pictures. The author should not use the variables’ name to present the results in manuscript.Overall, the project is understandable.

### Summary assessment

* results are poorly presented, hard to understand, poor quality

## Discussion/Conclusions

Are the study findings properly discussed? Are strengths and limitations acknowledged? Are findings interpreted properly?

### Feedback and Comments

Dr. Handel template does not requires about the study findings discussed (Why he ask here?). The strengths and limitations is mentioned. Findings are interpreted properly.

### Summary assessment

* strong, complete and clear discussion

## Further comments

*Add any other comments regarding the different aspects of the project here. Write anything you think can help your classmate improve their project.*

Some analysis you do not do in R. Do all your analysis in R. When you mention the results (for example p-value), you should use R code to make it reproducible.

# Overall project content evaluation

Evaluate overall features of the project by filling in the sections below.

## Structure

Is the project well structured? Are files in well labeled folders? Do files have reasonable names? Are all “junk” files not needed for analysis/reproduction removed? By just looking at files and folders, can you get an idea of how things fit together?

### Feedback and Comments

The projects are well structured. Files are well in label. Files have reasonable names. The manuscript is well structured too. One figure in results folder is not used and should delete it. The authors have some raw datasets. Some are not used. These data should be deleted too.

### Summary assessment

* well structured

## Documentation

How well is the project documented? Are you able to understand each step of the whole analysis, each decision that was made, and each line of code? Is enough information provided as comments in code or as part of Rmd files?

### Feedback and Comments

The results and analysis are understandable. The code is well documented and undestandable. However, the author did something outside ouf R (some analysis and cleaning process). If you can do whole analysis in R and write a reproducible manuscript, the project will be great.

### Summary assessment

* decently documented with some gaps

## Reproducibility

Are all results fully reproducible? Is documentation provided which clearly explains how to reproduce things, and does it work without the need for any manual intervention?

### Feedback and Comments

The manuscript document is reproducible. The numbers in manuscrip is not reproducible. However the analysis and cleaning are not reproducible. Some analysis did not do in R.

### Summary assessment

* major parts not reproducible

## Thoroughness

How thorough was the overall study? Were alternatives (e.g. different ways of processing the data or different models) considered? Were alternatives discussed? Were the questions/hypotheses fully and thoroughly addressed?

### Feedback and Comments

The author should include the multivariate in the analysis. Also, you should have better label for y axis of figure and table. For example what is the PCP percent. I am not clear what is PCP. The last table is reported p-value? The question are fully addressed. The work flow is understandable and reasonable.

### Summary assessment

* strong level of thorougness

## Further comments

*Add any other comments regarding the overall project here. Write anything you think can help your classmate improve their project.*

* Report 95% CI for RR rather than p-value
* Report HR for the time-to-event analysis rather than p-value
* Have table one to describe the chacteristics of your population. This tables can be divided by vaccination status and may include p-value to compare two groups (it will be better than your second and your last tables). You can look at my code and manuscript.
* present how you calculate percent PCP, and percent colonization.