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

Emily Rayens

2019-10-11

# Summary/Abstract

*Pneumocystis jirovecii* (Pc) is an opportunistic fungal infection, primarily associated with an AIDS diagnosis. However, constant monitoring shows healthy and HIV+ indviduals lacking other diagnostic criteria are also transiently colonized with Pc, and demonstrate capability of clearing this infection before it develops into an intersitial pneumonia. In this analysis, the diagnosis of Pc through PCR was compared to the other diagnostic criteria to more accurately capture the relationship of these criteria to the transient colonization of Pc. Hopefully we find something.

# Illustrating setup

*This section is only there to show how to insert results from other places in the project and how to cite figures and other references. Delete this whole section at some point.*

# Introduction (required for part 1)

## Background

*Pneumocystis jirovecii* (Pc) is the causative agent of an interstitial pneumonia in immunocompromised populations, inclulding those with HIV, transplant recipients, and cancer patients undergoing chemotherapy or radiation treatments. Pc presents a number of clinical issues including a lack of FDA-approved vaccines and limited treatment options.\_

Among the issues associated with Pc infection, it is difficult to determine the incidence of this infection because of a lack of clear diagnostics. Unlike most bacterial and fungal infections, Pc cannot be cultured for infection confirmation. Instead, diganosis is reliant on a combination of parameters including sampling of bronchoalveolar lavage fluid (BALF) for PCR, cell differential, and smear. Additionally, analysis of criteria of the immunecompromised status, such as CD4+ T cell count in HIV+ individuals or transplant recipients, may provide additional information.

The story is further complicated by the fact that healthy and HIV+/non-AIDS patients can be transiently colonized and can appropriately clear the infection. In this analysis, the relationship of all Pc diagnostic criteria will be examined in relationship to the most reliable diagnostic criteria (PCR) to capture the immune system response to colonization and fulminant pneumonia.

## Description of data and data source

This data set has been generated from a number of studies in a non-human primate model of HIV and Pc co-infection (n=30). Data was collected prior to and following infection with Simian Immunodeficiency Virus (SIV) through ~40 weeks post infection.

The data collected includes:

### 1. *Pneumocystis* diagnosis from bronchoalveolar lavage fluid (BALF)

A. PCR status:

i. First round positive (deemed equivalent to pneumonia)  
ii. Second round positive (equivalent to colonization)  
iii. Negative

B. BALF differential for neutrophil count (marker for infection)

C. BALF Smear

i. Positive (evidence of cysts or trophs)  
ii. Inflammation (foamy exudate)  
iii. Clear

### 2. Immune System Status

A. Lymphocyte Count

B. CD4+ T cell count

C. Viral Load

D. Vaccine Status

library(readxl)  
Pc <- read\_excel("data/raw\_data/Pc.xlsx")

str(Pc)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 421 obs. of 9 variables:  
## $ Identifier : num 6717 6717 6717 6717 6717 ...  
## $ Timepoint : num 0 4 6 10 16 20 24 28 32 36 ...  
## $ PcStatus : num 3 3 3 3 3 3 2 3 2 2 ...  
## $ Neutrophils: num 1 1 1 0 7 0 0 0 2 1 ...  
## $ BALSmear : num 3 3 3 3 3 3 3 3 3 1 ...  
## $ Lymphocytes: num 1800 2110 1610 2310 4160 3420 NA 3430 1990 2200 ...  
## $ CD4\_PBL : num 567 1076 887 1265 856 ...  
## $ Viral\_Load : num NA NA NA 1328971 1167963 ...  
## $ Vaccinated : num 1 1 1 1 1 1 1 1 1 1 ...

summary(Pc)

## Identifier Timepoint PcStatus Neutrophils   
## Min. :2116 Min. : 0.00 Min. :1.000 Min. : 0.000   
## 1st Qu.:6917 1st Qu.:10.00 1st Qu.:3.000 1st Qu.: 0.000   
## Median :7316 Median :26.00 Median :3.000 Median : 1.000   
## Mean :7166 Mean :25.57 Mean :2.731 Mean : 1.314   
## 3rd Qu.:7715 3rd Qu.:40.00 3rd Qu.:3.000 3rd Qu.: 2.000   
## Max. :8017 Max. :50.00 Max. :3.000 Max. :16.000   
## NA's :1 NA's :1 NA's :104 NA's :75   
## BALSmear Lymphocytes CD4\_PBL Viral\_Load   
## Min. :1.000 Min. : 920 Min. : 40.25 Min. : 1   
## 1st Qu.:3.000 1st Qu.: 2340 1st Qu.: 390.03 1st Qu.: 249537   
## Median :3.000 Median : 3250 Median : 701.67 Median : 704992   
## Mean :2.932 Mean : 3471 Mean : 825.15 Mean : 1927529   
## 3rd Qu.:3.000 3rd Qu.: 4150 3rd Qu.:1098.11 3rd Qu.: 1600364   
## Max. :3.000 Max. :14400 Max. :3860.38 Max. :37326865   
## NA's :229 NA's :30 NA's :41 NA's :177   
## Vaccinated   
## Min. :1.000   
## 1st Qu.:1.000   
## Median :2.000   
## Mean :1.512   
## 3rd Qu.:2.000   
## Max. :2.000   
## NA's :18

## Questions to be addressed

1. Does experimental vaccination reduce the incidence of Pc in this cohort?
2. Are there predictive diagnostic criteria in transient vs. progressive *Pneumocystis* infection?
3. Are there immune changes of note related to Pc colonization or infection?

# Methods and Results

*In most research papers, results and methods are separate. You can combine them here if you find it easier. You are also welcome to structure things such that those are separate sections.*

## Data aquisition

*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*

## Data import and cleaning

*Write code that reads in the file and cleans it so it’s ready for analysis. Since this will be fairly long code for most datasets, it might be a good idea to have it in one or several R scripts. If that is the case, explain here briefly what each file does. The files themselves should be commented well so everyone can follow along.*

## Univariate analysis

## Bivariate analysis

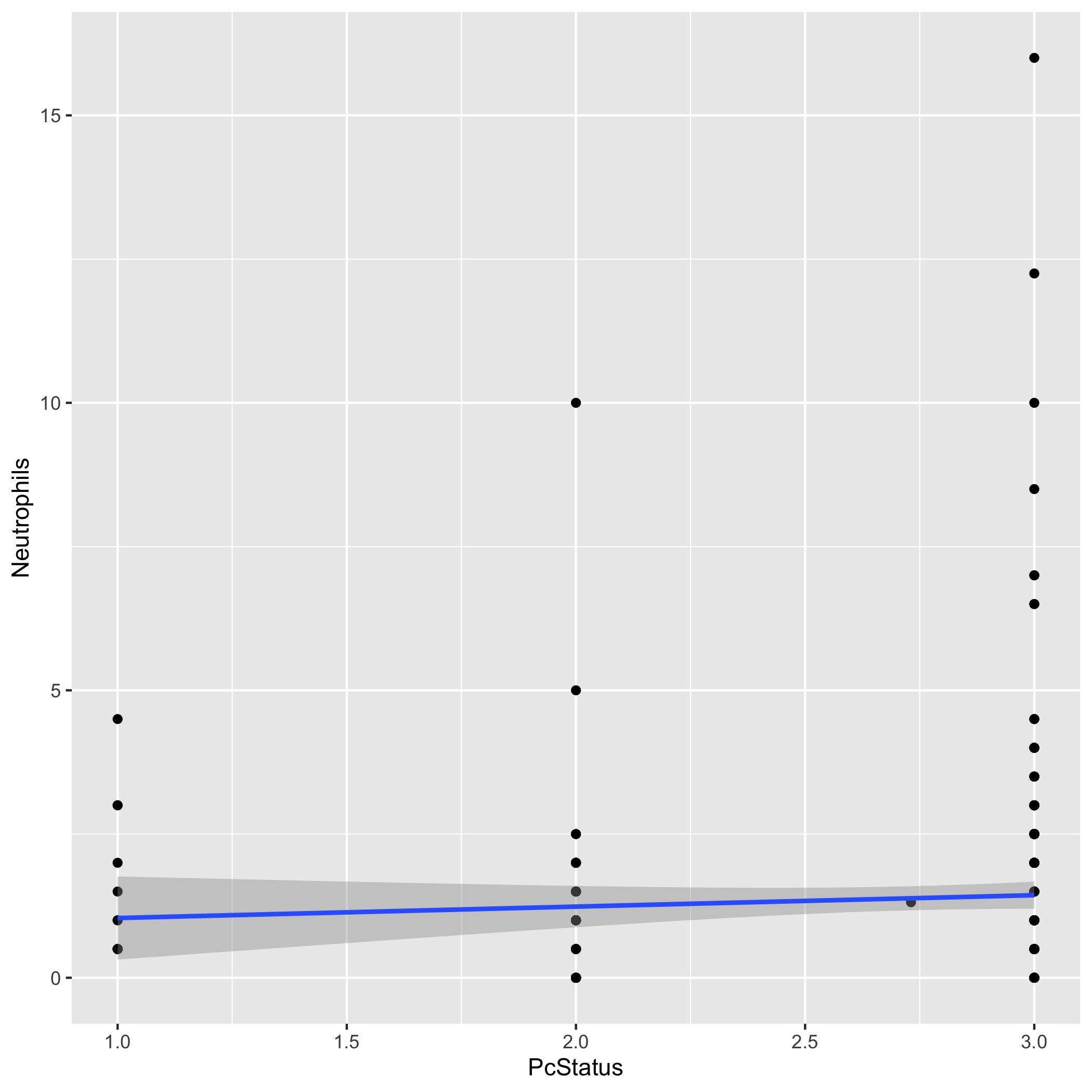


Figure 1: Analysis figure.

## Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You can then load the results produced by this code*

# Discussion

## Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

# References