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 :1.533   
## 3rd Qu.:2.000   
## Max. :2.000   
##

## 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

### Exploring the Pc Diagnostic Variables by Identifier

This is a summary of the values of the Pc Diagnostic Variables by Identifier. I am struggling to display the table I generated in the analysisscript and what I can see within the RMarkdown document when I run the chunk of code. Any assistance on this is appreciated.

Table 1: Summary of Pc Characteristics by Individual

|  |
| --- |
| V1 |

|

Table 1

Group.1

PcStatus

Neutrophils

BALSmear

Lymphocytes

CD4\_PBL

Viral\_Load

Vaccinated

2116

2.785714

1.4583333

3.000000

2545.385

672.8244

1062977.04

1

6615

2.777778

1.4090909

NaN

3935.714

730.3514

873158.53

1

6715

3.000000

1.2000000

NaN

4009.167

1021.3023

4848734.98

2

6717

2.428571

1.1923077

2.571429

2381.538

725.8650

769540.42

1

6815

3.000000

3.0000000

NaN

3589.286

784.8579

127616.88

1

6817

2.928571

1.1538462

3.000000

4026.923

812.3254

429642.20

1

6915

3.000000

1.3750000

NaN

4460.000

928.6896

5113149.09

1

6917

2.785714

0.9615385

3.000000

5719.286

1950.5067

108432.11

1

7015

2.888889

1.5416667

NaN

3514.286

552.3512

1698627.14

2

7017

2.714286

0.7307692

3.000000

7247.857

1837.4735

395514.28

1

7115

2.400000

1.3888889

NaN

2234.000

397.9443

20565531.31

2

7117

2.785714

0.6818182

2.857143

3950.000

1293.0650

1527131.08

1

7215

3.000000

0.5833333

NaN

1933.571

427.7707

1018647.93

1

7217

2.428571

0.4615385

3.000000

3580.714

587.2368

2231267.06

1

7315

3.000000

1.7083333

NaN

2967.143

453.0355

1210564.14

2

7317

2.384615

0.3076923

3.000000

2182.143

540.5180

118039.24

2

7415

2.714286

2.5555556

NaN

2619.167

401.9947

2518083.26

1

7417

2.769231

2.1875000

3.000000

2090.714

353.8679

756294.12

2

7515

3.000000

0.8333333

NaN

3575.000

1179.3794

21863.53

2

7517

2.375000

1.7222222

2.875000

3547.500

546.0115

7735580.70

2

7615

3.000000

1.2727273

NaN

4835.714

1242.4463

134638.74

2

7617

2.307692

1.8461538

2.846154

2392.143

783.3108

696261.11

2

7715

2.875000

1.1111111

NaN

2198.000

299.9739

1522775.32

2

7717

2.538461

1.1923077

2.846154

3988.571

1077.7516

2357744.29

2

7815

3.000000

1.6000000

NaN

3134.286

730.5493

631339.94

2

7817

2.600000

0.8000000

3.000000

2832.000

525.5889

621162.84

2

7915

3.000000

0.5416667

NaN

3515.714

585.5310

766442.78

1

7917

2.200000

1.5500000

3.000000

2328.000

722.6838

364821.67

2

8015

3.000000

0.8636364

NaN

3348.462

615.1221

3669352.11

1

8017

2.769231

2.5769231

3.000000

4091.429

1276.5329

28550.92

2

### Exploring the Pc Diagnostic Variables by Vaccine Status

This is to get first impressions of differences in the vaccinated and unvaccinated populations in a first attempt to answer question 1.

Table 2: Summary of Charateristics by Vaccine Status

|  |
| --- |
| V1 |

|

Table 2

Group.1

PcStatus

Neutrophils

BALSmear

Lymphocytes

CD4\_PBL

Viral\_Load

Vaccinated

1.000000

2.784810

1.182635

2.918367

3802.158

901.7518

1440438

1.000000

1.533333

2.731013

1.313768

2.931937

3470.821

825.1482

1927529

1.533333

2.000000

2.677215

1.436798

2.946237

3156.050

754.3774

2394978

2.000000

## Bivariate analysis

This is a clumsy attempt to analyze the neturophil count (major marker of infection) by vaccine status. We can actually see here that our vaccinated population (on the left) has a lower number of neutrophils counted over the course of the study.

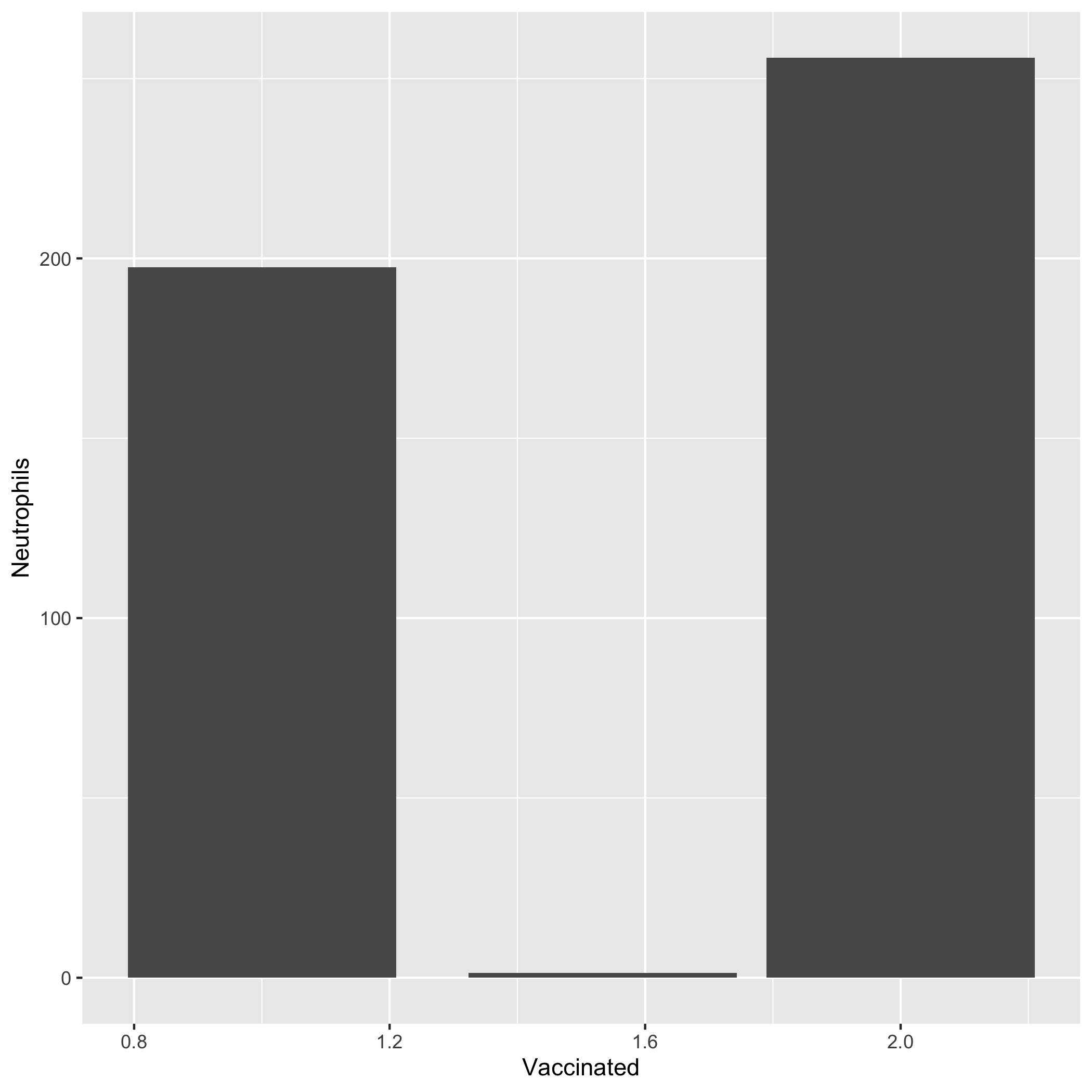
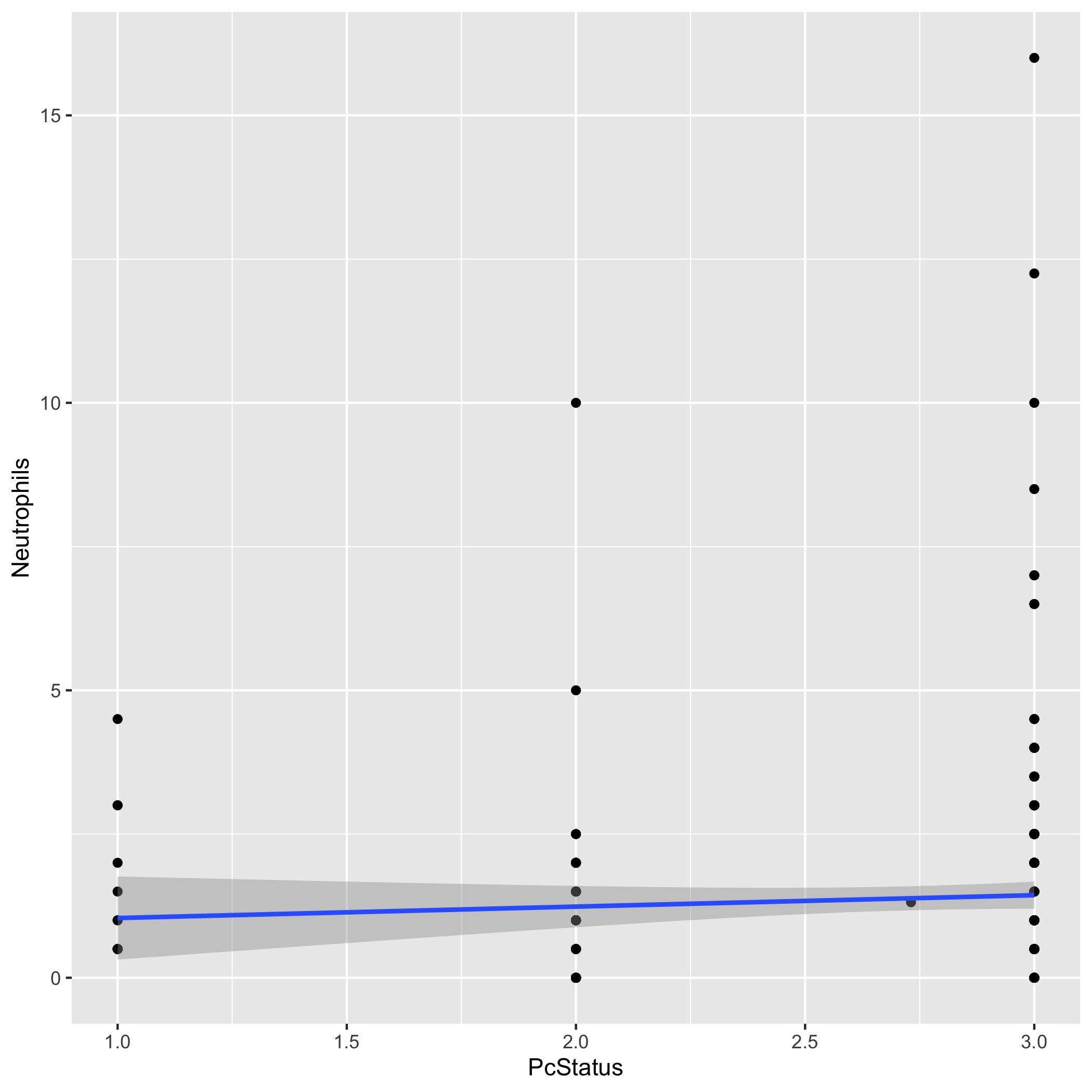


Figure 1: Analysis figure.

This is a brief display of the relationship between Pc Status and Neutrophil Count. We can see that as Pc status gets categorized more highly (greater disease), measured neutrophil count increases. 

## 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

# Ongoing Issues:

Displaying a simple table in the univariate analysis section. The tables look as I expect in RStudio but it does not appear that way in the output.