Remy Lagrois

Stats II – 402

Regression Analysis

02/15/2016

Interferon Gamma as a Treatment for Chronic Granulomatous Disease

1. **Introduction**

Chronic granulomatous disease (CGD) is an inherited genetic disorder which leads to an immunodeficiency. Males are particularly susceptible to the disease as the defective gene can be present on the X chromosome. However females can also have CGD in its autosomal recessive form (at least four genes, including the X linked gene, can be responsible for the disease). [1] The defective genes encode subunits of an oxidase enzyme. [2] White blood cells (WBCs) normally use this enzyme to produce hydrogen peroxide and other activated O2 compounds which destroy ingested bacteria and fungus.[1] Without this enzyme, however, WBCs have greatly reduced ability to control bacterial and fungal infections. These infections can become very severe leading to abscesses and lesions in the lungs, lymph nodes, and on the liver (among other organs). The disease is treatable though does have a rather high fatality rate. During a CGD study published in 2009 a total of 20% of patients looked at died while the data was being collected.[2]

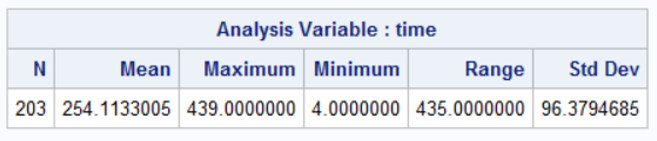
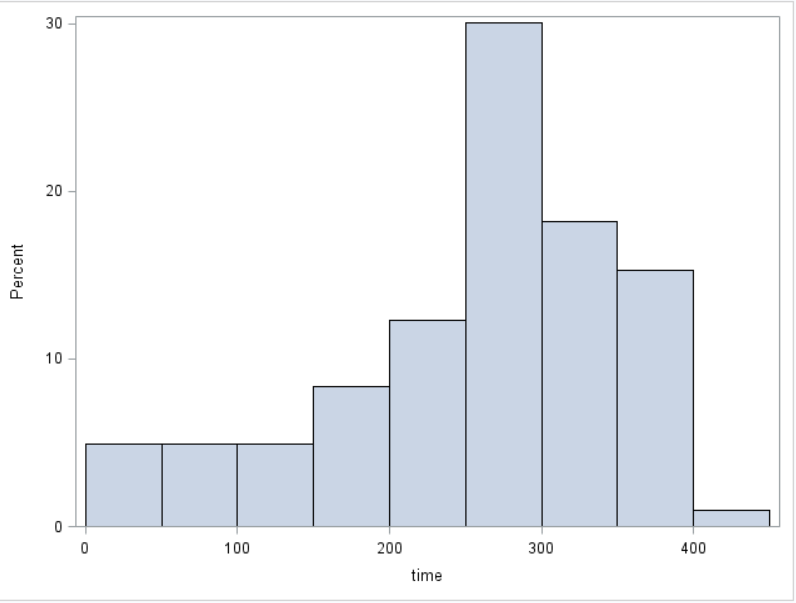
In 1989 a clinical trial was conducted to assess the efficacy of Interferon Gamma (IFNγ) in the treatment of CGD.[3] IFNγ is a class of molecules called cytokines which are involved in cell to cell signaling within the body. Specifically, IFNγ is an important part of a functioning immune system being involved in the proliferation, growth, and activation of various immune cells.[4]  In fact, the interferon class of cytokines is so named due to their ability to ‘interfere’ with viral replication. The authors of the study had found evidence that IFNγ may restore activity of the oxidase which should prevent sever infections in the first place.[3] Patients were enrolled into the study on a rolling basis and were randomly assigned to either a placebo or treatment with IFNγ. Other factors, which will be expanded on later, were also collected. Treatment continued until a serious infection developed and the number of days between entry into the study and the onset of infection were used to compare the placebo and IFNγ.

1. **Factors and Basic Statistics**

A number of categories of information were gathered about each participant that were not directly of interest in our analysis; such as entry date, infection date, and region (note: region was briefly investigated but held no statistical value). The factors which were thought to be of possible significance are listed as follows:

* **Treatment:** IFNγ (d\_treat 0) or placebo (d\_treat 1)
* **Inheritance type:** X linked or autosomal recessive (d\_inherit 0 or d\_inherit 1)
* **Age:** In years
* **Height:** In centimeters
* **Weight:** In kilograms
* **Corticosteroids:** In use at time of entry (d\_steroid 0) or not in use (d\_steroid 1)
* **Prophylactic Antibiotics:** In use at time of entry (d\_anti 0) or not in use (d\_anti 1)
* **Elapsed Time:** Number of days between entry and diagnosis of infection (time)

Time is our independent variables and most of our independent variables of interest are categorical (with the exception of age, height, and weight). The data from all participants (totaling 203) was used in the analysis.



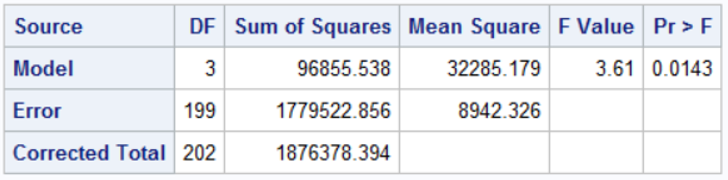
**Fig 1** The sample distribution of time as well as population mean, max, min, range, and standard deviation for time

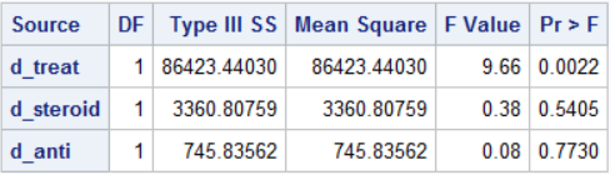
The above figure (fig 1) shows the sample statistics for the dependent variable time. There is some left skew to the population but otherwise shows no evidence against a normally distributed population. Additionally we can invoke the central limit theorem due to the sample size. Even with the apparent left skew all data points fall within 2-3 standard deviations.

The main analysis performed was multiple regression to see if these factors had a connection with the length of time before infection and if so what the nature of that connection was. Of primary interest was the treatment type. Secondarily the effects of steroid or antibiotic use were of interest as well as their possible interaction with the type of treatment. The sex of each participant was also thought to possibly have an effect as the X linked form of the disease tends to be more severe and almost exclusively affects males (ipso facto males primarily have the X linked form). Of tertiary interest was how the age, weight, and height may play a role.

1. **Analysis**

The first test run was a regression analysis using all the variables bulleted in section II. All the parameters were insignificant except for the treatment type (data not shown). In order to refine the model a second regression was done with only the variables thought to be most likely to have an effect on the number of days before infection; steroid use, antibiotic use, IFNγ treatment.

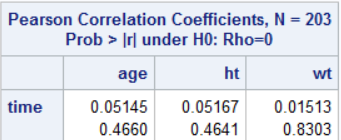




**Fig 2** The overall F-test and Type III sum of squares for the model using treatment, steroid, and antibiotic use.

Figure 2 shows that while the overall test was significant (p-value of 0.0143), neither the steroid use nor the antibiotic use were (note the antibiotics/steroids were not taken during the study). The treatment type, with a p-value of 0.0022 did have statistically significant effect on time.

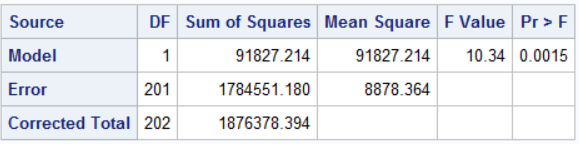
Other combinations of the factors were also tested along with and separate the treatment using regression but produced no statistically significant results (see appendix A for tables). The only other variable that showed promise for possibly having an effect was weight.

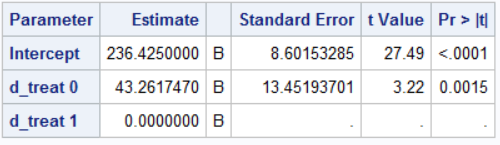


**Fig 3** Correlation test for age, height (ht), and weight (wt)

Weight did have a statistically significant positive correlation with time (figure 3, p-value 0.015 and r of 0.83). However when weight was added to the model using treatment (and the interaction of weight and treatment) none of the parameters were significantly different from zero (Appendix A).

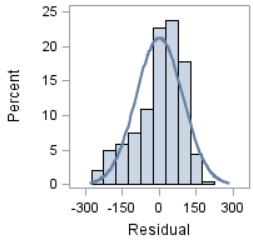
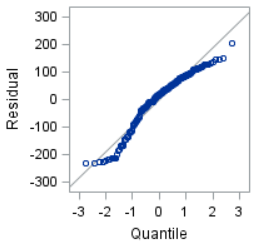
The final model was simply time = β0 + β1(treatment), the results of which were statistically significant.





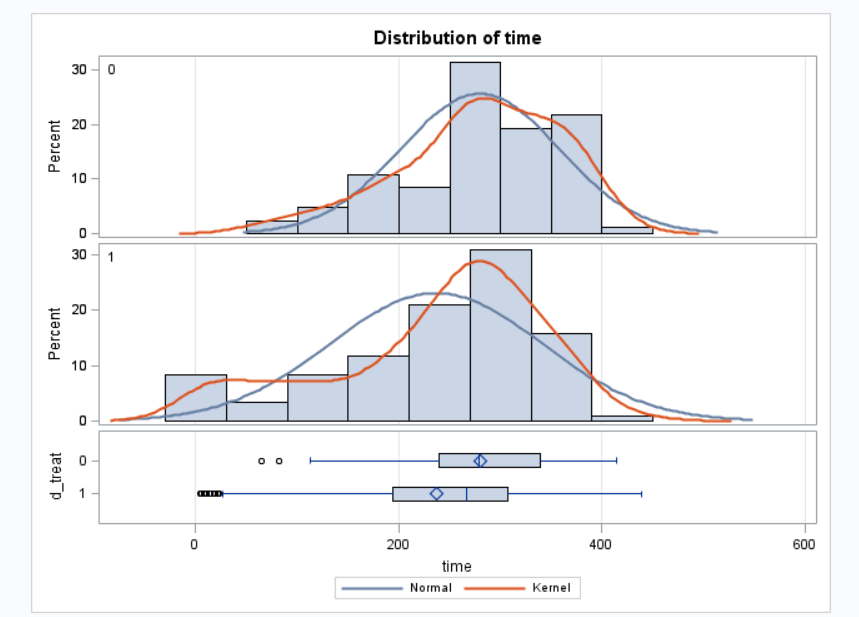
**Fig 4** The overall F-test and parameters for the final model

The intercept (fig 4) represents the mean number of days before infection for those who were on the placebo. The row ‘d\_treat 0’ shows β1 with the placebo as reference. Treatment with IFNγ added an average of 43 days before a severe infection was diagnosed (95% confidence interval of 16.7 to 69.8 days). However the R2  was very low at 0.05, clearly there is more at work than this study could cover.



**Fig 5** Residual QQ plot and histogram

While the QQ plot wasn’t perfectly normal neither it nor a histogram of the residuals offer extreme evidence against normality (figure 5). Additionally the histogram of time for both the placebo and IFNγ groups show equal variance and no evidence against normality (figure 6). A log transformation was tried on the time variable but did not improve the distribution at all. The assumption of independence was also not violated as each subject had no connection to the others and when one became infected had no influence over when the others did. There were also several high leverage data points according to Cook’s D, however there was no reason to believe they were due to nothing more than natural variation and their exclusion had no helpful effect on the model. Both the treatment and the subjects were randomized so these results can be applied to the broader population of CGD patients.



**Fig 6** Histogram of time for IFNγ (top) and placebo (bottom)

1. **Conclusion**

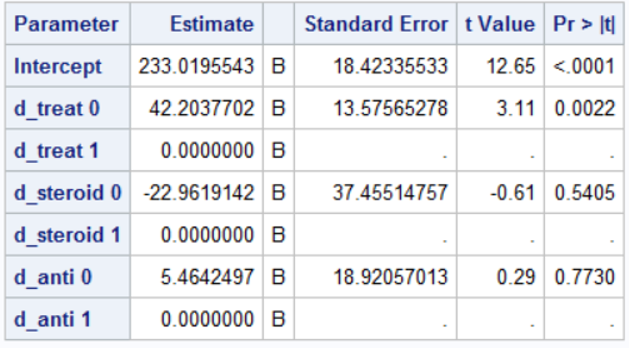
The use of IFNγ as a treatment for CGD clearly does have a positive effect on patients. However the practical significance of it is debatable. On the low end it may only delay infection by as little as 16 days with the longest most likely up to nearly 70. While the patient may appreciate the extra days not spent in a hospital it probably isn’t enough in the long run. Each infection could lead to death or lasting damage caused by the lesions and abscesses. Since the disease is incurable surviving one infection doesn’t mean it’s over and a delay of only up to just over 2 months still means the patient is likely to have at least one severe infection every year. Infection prevention needs to be on the order of years to have a significant impact on CGD patients’ quality and length of life. Given that only about 5% of the difference in time before infection can be explained by IFNγ treatment there clearly are many other factors at play which this study could not measure. Perhaps now with whole genome sequencing the differences in CGD patients’ immune systems can be found which explain why some become infected more frequently than others. Sequencing may also allow for more personalized treatments to be developed. The use of IFNγ to treat CGD will need to be combined with other concurrent treatments. A new study should be done on IFNγ alongside antibiotics and steroids but for now IFNγ is not sufficient on its own.

References

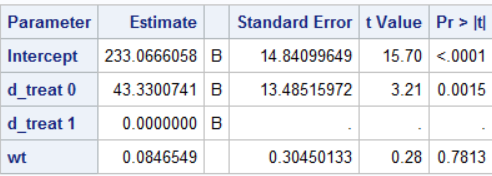
1. Fernandez , J. Chronic Granulomatous Disease. Merck. <http://www.merckmanuals.com/professional/immunology-allergic-disorders/immunodeficiency-disorders/chronic-granulomatous-disease>
2. Van den Berg, J. M., van Koppen, E., Åhlin, A., Belohradsky, B. H., Bernatowska, E., Corbeel, L., … Kuijpers, T. W. (2009). Chronic Granulomatous Disease: The European Experience. *PLoS ONE*, *4*(4), e5234. <http://doi.org/10.1371/journal.pone.0005234>
3. UMass Statistics Department. <http://www.umass.edu/statdata/statdata/data/cgd.txt>
4. Davidson College Biology Department. Interferon Gamma. <http://www.bio.davidson.edu/courses/immunology/students/spring2006/v_alvarez/ifn-gamma.html>

**Appendix A**

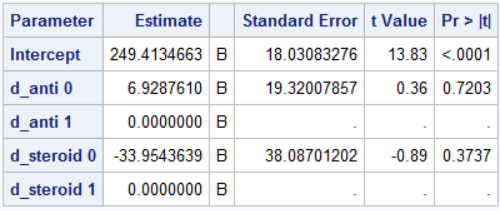
Model using treatment, steroid and antibiotic use (prior to study)



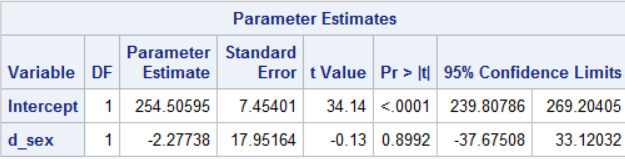
Treatment and weight



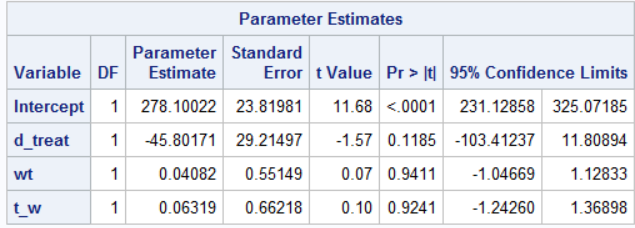
Steroid/antibiotics on their own



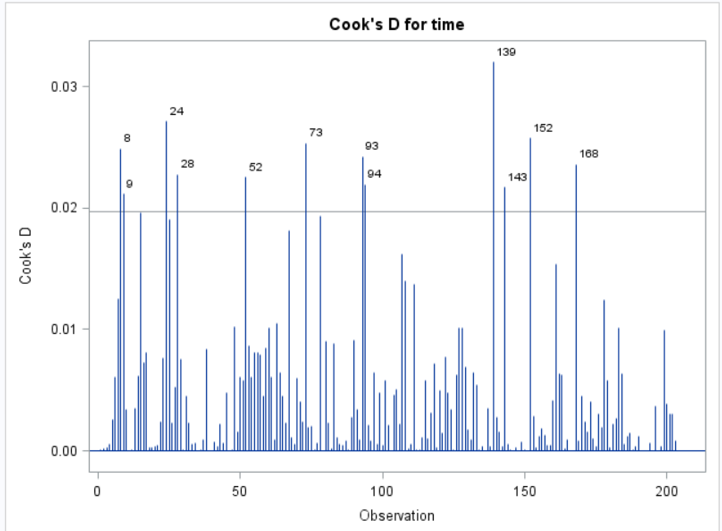
Sex alone



Treatment, weight, and interaction



Cook’s D for final model



**Appendix B**

(note that some SAS code was reused so the number of procedures seen does not reflect the actual number run, see file pharynx.csv attached to submission for raw data)

**data** cgd;

infile '\\Client\D$\SMU DS\Stats II\Pro 1\pharynx.csv' firstobs = **2** dlm = ',';

input ID $ ent\_date $ inf\_date $ treatment inherit age ht wt steroid antibio sex region time;

**data** cgd2;

set cgd;

d\_treat = treatment - **1**;

d\_inherit = inherit - **1**;

d\_steroid = steroid - **1**;

d\_sex = sex - **1**;

d\_anti = antibio - **1**;

l\_time = log(time);

s\_i = d\_sex \* d\_inherit;

t\_sa = d\_treat \* d\_steroid;

t\_a = d\_treat \* d\_anti;

l\_treat = log(d\_treat);

l\_steroid = log(d\_steroid);

l\_anti = log(d\_anti);

t\_w = d\_treat \* wt;

**run**;

**proc** **print** data=cgd2;

**run**;

/\*final model?\*/

**proc** **glm** data=cgd2 PLOTS=(DIAGNOSTICS RESIDUALS);

class d\_treat;

model time = d\_treat / solution ss3;

**run**;

**proc** **reg** data=cgd2 plots = cooksd(LABEL);

var d\_treat;

model time = d\_treat wt t\_w / CLB;

**run**;

**proc** **means** data=cgd2 n mean max min range std;

var time;

**run**;

**proc** **sgscatter** data=cgd2;

matrix d\_treat time;

**run**;

**proc** **glm** data=cgd2 PLOTS=(DIAGNOSTICS RESIDUALS);

class d\_anti d\_steroid;

model time = d\_anti d\_steroid / solution ss3;

**run**;

**proc** **ttest** data=cgd2;

class d\_treat;

var time;

**run**;

**proc** **corr** data=cgd2;

var age ht wt;

with time;

**run**;

**proc** **sgplot** data=cgd2;

histogram time;

**run**;