BIOS 622 - Project

Student: Nguyen Doan
Professor: Sijin Wen
Class: BIOS 622-Time to Event Analysis

December 12, 2018

Abstract

Chronic Granulomatous Disease (CGD) is a group of inherited rare disorders of the immune function characterized by recurrent pyogenic infections which usually result in death during early childhood. Research has suggested that gamma interferon, which is an important macrophage activating factor can reduce the rate of serious infections of patients. A study was conducted to test the effectiveness of the gamma interferon between October 1988 and March 1989 with 203 eligible patients. The patients were randomized to placebo vs. gamma interferon. This paper considers two approaches, proportional hazard model and parametric model to analyze the affects of different covariates on survival rate of patients. The proportional hazard model and parametric model exhibit similar results in term of survival rate and median follow-up time. The two approaches also provides similar significance of included variables.

Contents

1 Data Description

The data consisted of 203 patients who joined the study from July 15, 1989 to September 22, 1989. The gamma interferon and placebo were randomly given to each patient at the first scheduled visit and patients were monitored throughout the study.
Categorical variables are coded as dummy variable. A list showing details for each variable below is helpful to understand further analysis.

- Treatment: 1 = Gamma Interferon, 2 = placebo
- Pattern of inheritance: 1 = X-linked, 2 = autosomal recessive
- Using corticosteroids at time of study entry: 1 = yes, 2 = no
- Using prophylactic antibiotics at time of study entry, 1 = yes, 2 = no
- Gender: 1 = male, 2 = female
- Hospital category: 1 = US, 2 = Europe
- Censored: 0 = Censored, 1 = Non-censored

The time to the initial infection from study entry until the first scheduled visit was recorded. Other important variables such as pattern of inheritance, age, height, weight, whether a patient was using corticosteroids or prophylactic at time of study entry, gender, and hospital category were also recorded. Table 1 and 2 include descriptive information of all variables with censored data.

Table 1: Descriptive Table for Categorical Variables

Variables	Description	Yes	No
Treatment	Yes=Gamma Interferon, No=Placebo	83	120
Pattern of Inheritance	Yes=X-linked, No=Autosomal recessive	131	72
Using corticosteroids		7	196
Using prophylactic antibiotics		172	31
Gender	Yes=male, No=female	168	31
Hospital	Yes=US, No=Europe	149	54
Censored		127	76

Table 2: Descriptive Table for Quantitative Variables

Variables	Description	Min	Max	Mean	Standard Deviation
Age	In years	1.0	44.0	13.7	9.3
Height	In cm	76.3	189.0	138.1	31.4
Weight	In kg	10.40	101.50	39.34	21.8
Time to disease	In days	4.0	439.0	254.1	96.4

2 Initial Assessment

First, because CGD mainly results in death in early childhood, we would expect to see that age has a negative correlation with the diseases. CGD is also a group of inherited disorders, we would expect to see the variable pattern of inheritance to be significant. Assuming that hospitals in the US and Europe randomly assigning treatments to patients, we would not expect this variable to be significant in predicting hazard ratio. In addition, if gamma interferon is in fact an important macrophage activating factor which could restore superoxide anion production and bacterial killing by phagocytes in CGD patients, a survival rate for patients with gamma interferon treatment should be higher than those with placebo.

3 Proportional Hazard Model

3.1 Proposed Model

Cox proportional hazard model is used for this analysis due to its usefulness with both categorical and quantitative data. First, I consider a model that include all variables that were mentioned above. I perform a step-wise selection method based on AIC to achieve a final model that includes:

- Using corticosteroids at time of study entry
- Weight
- Age
- Treatment

Treatment, age and using corticosteroids appear to be significant at 5% level. Although weight is very close to 5% significant level, including it gives us a model with lowest AIC of 713.58. The likelihood ratio test also show a significant result of 22.64 on 4 degrees of freedom and p-value of 0.0004.

The positive slope of treatment means that patients who were given gamma interferon have a lesser chance of getting CGD. Those who were given placebo have more than twice the chance of getting CGD. High age appears to be negatively correlated with risk of getting CGD, i.e: older patients have lower chance of getting CGD. Similarly, Using corticosteroids at time of study entry is negatively correlated with risk of getting CGD. Using corticosteroids reduces the hazard by a factor of 0.30. People with higher weight have higher risk of getting CGD.

Table 3: Cox Proportional Hazard Model

Variable	Coef	exp(Coef)	SE(Coef)	z	Pr($>\|z\|$)	95%CI
treatment2	0.97052	2.63931	0.26755	3.627	0.000286	(1.5622, 4.4589)
age	-0.08393	0.91950	0.03439	-2.441	0.014654	(0.8596, 0.9836)
weight	0.02434	1.02464	0.01289	1.889	0.058931	(0.9991, 1.0509)
cortico2	-1.20004	0.30118	0.56205	-2.135	0.032751	(0.1001, 0.9062)

3.2 Diagnostic

There are several conditions that we need to check in order to confirm that the proposed proportional hazard model is good. However, the proportional hazard assumption is the most important condition. This assumption means that hazard ratio between two treatments has to

be constant over time.
The function cox.zph() in R correlates the corresponding set of scaled Schoenfeld residuals with time, to test for independence between residuals and time. Addtionally, it performs a global test for the model as a whole. The hypothesis test would be:

- H_0: hazard ratio is constant
- H_a: hazard ratio is not constant.

In table 4, no variables show significant p-value. The global test for proportional hazard also shows no significant p-value. Hence, we can safely say that the assumption holds for our model. In addition, stratification is not necessary since we satisfy the proportional hazard assumption.

Table 4: Proportional Hazard Test

Variable	Rho	chisq	p-value
treatment2	0.01786	0.02573	0.873
age	-0.04482	0.11282	0.737
weight	0.05370	0.16997	0.680
cortico2	-0.00649	0.00282	0.958
GLOBAL	NA	0.19407	0.996

Performing a graphical diagnostics based on the scaled Schoenfeld residuals also shows similar result. In figure 1, Each plot does not show random or non-constant pattern over time.

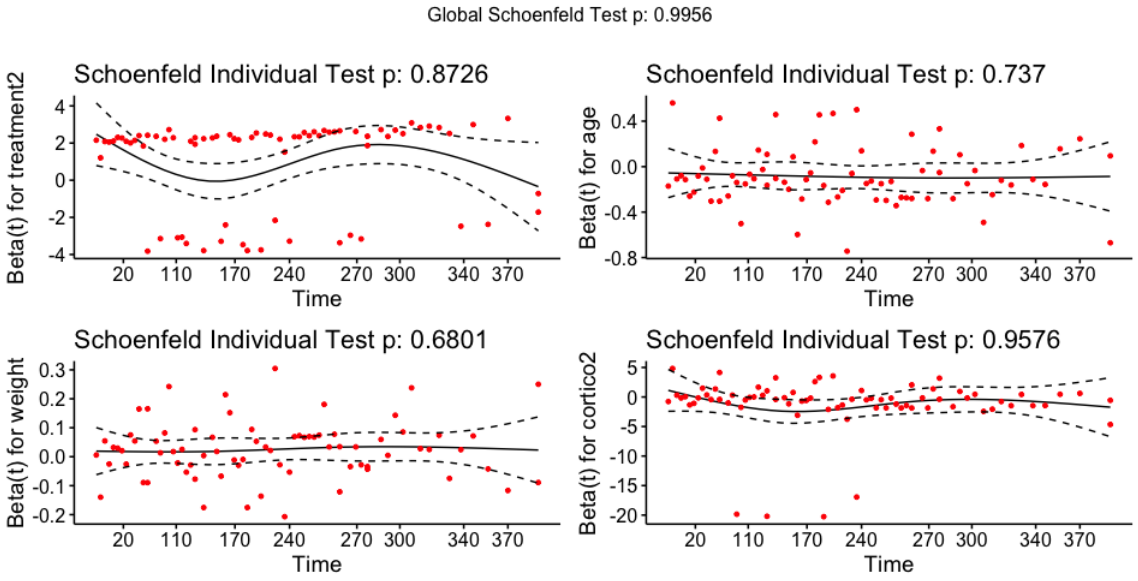

Figure 1: Proportional Hazard Test

We can also check for influential observations or outliers by using deviance residuals. The deviance residual is a normalized transform of the martingale residual. These residuals should

be roughly symmetrically distributed about zero to indicate no influentials observations. Figure 2 shows a symmetrical distribution of residuals about zero. Hence, the data does not have any outliers or influential observations.

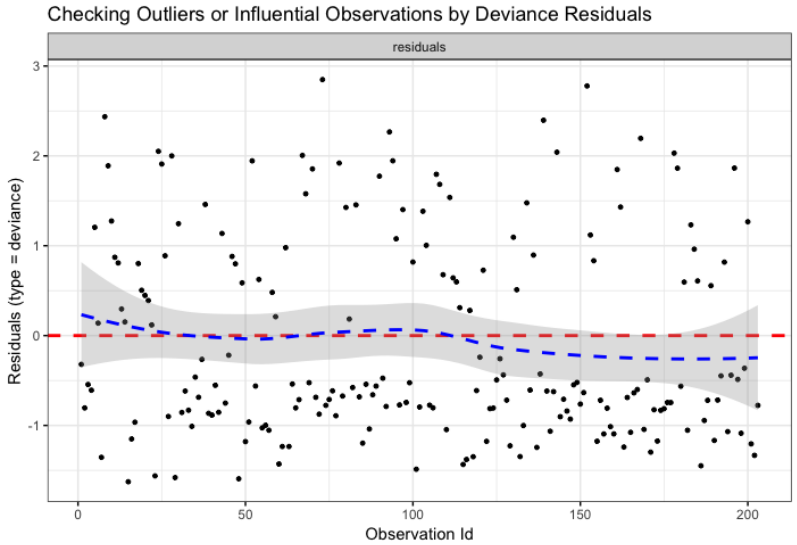

Figure 2: Checking outliers and influential observations

Although we often assume that continuous variables have a linear form, it is better that we check this assumption. Plotting the martingale residuals against each continuous variable can assess the functional form of the variable. Nonlinearity is not an issue for categorical variables, hence we only need to examine age and weight in the model. Figure 3 only shows a slight non-linear pattern in age and weight against martingale residuals. Hence, we can safely assume that linearity assumption holds for our model.

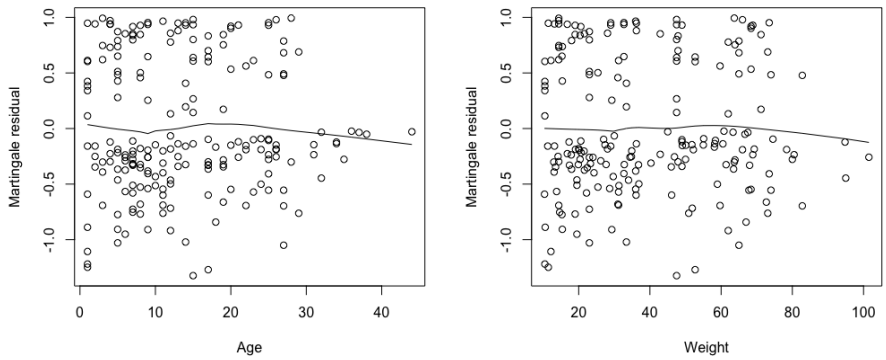

Figure 3: Checking linearity assumption

3.3 Survival Curve

Survival curve shows that patients that received gamma interferon treatment have higher survival rate compared to those who received placebo. The median follow-up time for placebo treatment is 307 days while gamma interferon treatment has median follow-up time beyond the time frame of this study. This also means that patients with gamma interferon have higher median follow-up time.

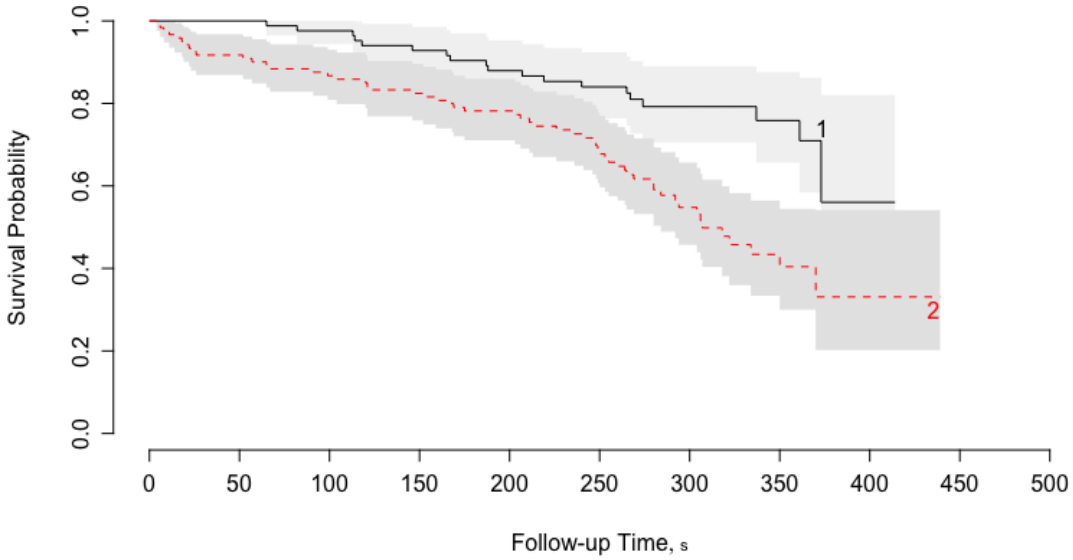

Figure 4: Survival Curve between 2 Treatments

4 Parametric Model

In this section, we want to assume that our survival function follows Weibull distribution. We can fit a parametric model and compare it with proportional hazard model. We also use the same variables that were included in PH model. The log likelihood ratio test of this model has chisquare = 20.48 on 4 degrees of freedom and p-value = 0.0004. All variables also shows significance or close to significance.

Table 5: Weibull Distribution Model

Variable	Coef	exp(Coef)	SE(Coef)	z	Pr(> |z|)	95%CI
intercept	5.85289		0.44759	13.08	<2e-16	
treatment2	0.8859	2.4252	0.2634	-3.23	0.0012	(1.4471, 4.0643)
age	-0.0853	0.9181	0.0343	2.43	0.0153	(0.8584, 0.9820)
weight	0.0259	1.0263	0.0128	-1.99	0.0467	(1.0008, 1.0524)
cortico2	-1.1020	0.3322	0.5627	1.94	0.0524	(0.1103, 1.0009)
log(scale)	-0.31880		0.10625	-3.00	0.0027	

Figure 5 shows survive curves for 2 treatments. Similar to Cox model, gamma interferon still has higher survival probability than placebo. The median follow-up time for gamma interferon is 618.21 days and for placebo is 328.98 days.

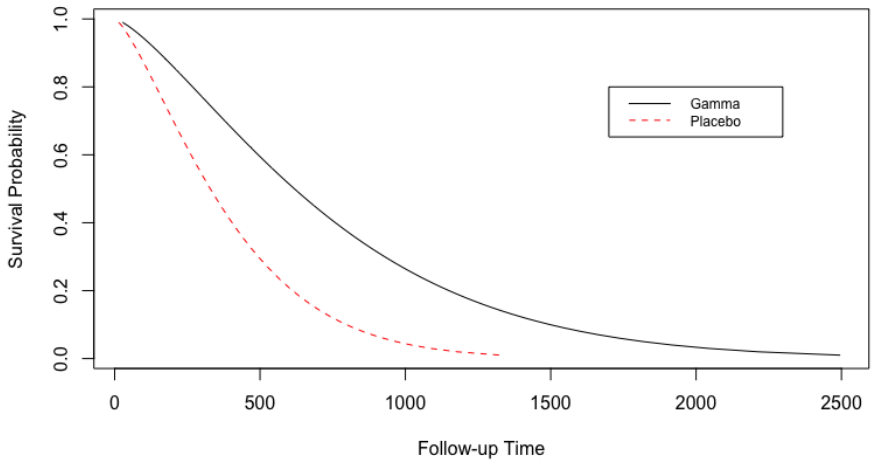

Figure 5: Survival Curve between 2 Treatments for Weibull Model

5 Comparison between PH and Parametric Model

In section 3, we fit a proportional hazard model that indicates gamma interferon treatment increases survival probability for patients. We also saw that age and using corticosteroid at time of entry have negative correlation to hazard ratio while weight has a positive correlation. In section 4, all variables in Weibull model shows similar result. Patients with gamma interferon treatment still perform better. However, the parametric model allows us to calculate median follow-up time easier.

6 Conclusion

Both approaches show similar result in term of variables that directly affect CGD. Although CGD is estimated to occur in 1 in 200,000 to 250,000 people worldwide, it is a very fatal disease, considering that it mainly affect early age baby and children (1). They have very weak immune system in their young age, hence, the disease can have an exponential affect on reducing their immune system. This study did not find any significance in pattern of inheritance. However, further analysis should explore more on this issue.

7 References

1. "Chronic Granulomatous Disease - Genetics Home Reference - NIH." U.S. National Library of Medicine, National Institutes of Health, ghr.nlm.nih.gov/condition/chronic-granulomatous-disease-statistics.