Homework 4: BST 665: Survival Analysis

Devin Koehl

Spring Semester 2019

1. An investigator is studying infections in patients receiving dialysis. He believes that the method used to access the blood during dialysis may be associated with the risk of infection.

A. Using the time to infection as the survival time, fit a Cox proportional hazards model using diabetes status and access method as covariates. Use the Prentice-Williams-Peterson conditional probability (PWP-CP) model. Create a table reporting the estimated hazard ratios (and 95% confidence intervals) for each of the effects in this model. Provide your interpretations of these results.

Table 1: Infections for Patients Receiving Dialysis

Covariate	Hazard Ratio	Lower 95% CI	Upper 95% CI	p-value
Diabetes	2.93	2.03	4.25	<0.0001*
Method				
Fistula	0.56	0.37	0.86	0.008*
Graft	0.65	0.44	0.96	0.03*

**p<0.05; CI = Confidence Interval*

The PWP-CP model is similiar to the AG model except the baseline hazard function varies by event. Therefore in the "strata" statement we will include the "infection" variable.

From the results, we can see that Diabetes has an increase in risk of 193% (p<0.0001). The reference group for method was catheter, while the reference group for Diabetes was "No". The fistula method compared to the catheter method corresponds to a 46% decrease in risk (p=0.008), while the graft method compared to the catheter method corresponds to a 35% (p=0.03).

B. Repeat Part A, this time using the Prentice-Williams-Peterson gap time (PWP-GT) model.

Table 2: Infections for Patients Receiving Dialysis

Covariate	Hazard Ratio	Lower 95% CI	Upper 95% CI	p-value
Diabetes	4.24	2.90	6.20	<0.0001*
Method				
Fistula	0.68	0.40	1.15	0.12
Graft	0.72	0.44	1.12	0.19

**p<0.05; CI = Confidence Interval*

From the results, we can see that Diabetes has an increase in risk of 324% ($p<0.0001$). The fistula method compared to the catheter method corresponds to a 32% decrease in risk, while the graft method compared to the catheter method corresponds to a 28%. Neither of these methods reached statistical significance.

C. The investigator would like you to explain the difference between the results from Parts A and B. Write a short paragraph explaining the PWP-CP and PWP-GT models. (This explanation should be written so that it can be understood by the investigator, who has not taken a survival analysis course).

In many patient populations, the event of interest can occur multiple times in a patient. These events are termed recurrent events. Many statistical analyses focus on time to first event, such as time to CVA or device malfunction. Time to first event ignores any subsequent events. The underlying theory behind recurrent event models is that events are ordered and a participant is only at risk for one event at a time. The Prentice, Williams, and Peterson Conditional Probability model (PWP-CP) orders multiple events by stratification. All participants are at risk in the first strata. For the patients that experienced an event in the first stratum, they are now at risk for another event in the second stratum. In other words, subjects are not at risk for the k^{th} event until the $(k-1)^{th}$ has occurred. The shape of the hazard function is allowed to vary with the event number. For the PWP-CP model, the time scale we are using is the total time from beginning to of study.

Another form of the Prentice, Williams, and Peterson model uses the "gap time" as the time scale. The gap time is the amount of time between events. When we use the PWP-GT model, we are essentially resetting the time to zero after the recurrence of the event. Analyzing gap time may be useful when we have events that are spread out over a population.

In summary, while both models are similar the time scale changes from the CP to the GT model. In the CP model we are analyzing the total time from the beginning of the study, while the GT model is using the "gap time" which is the amount of time between events.

D. The investigator is preparing a manuscript on this study and has asked for your help writing the methods and results sections. Write 1-2 paragraphs for the investigator's paper. Be sure to include your interpretation of any results you include and to take the stated goal of the study into consideration.

An investigator sought to study infections in patients receiving dialysis. A Cox Proportional Hazards model was fit using diabetes status and access method as covariates. SAS 9.4 was used to generate outcomes and a p-value of <0.05

was statistically significant. A recurrent event model is one where the event of interest can happen multiple times to a patient. The recurrent event is this analyze is infection. A Prentice-Williams-Peterson conditional probability model (PWP-CP) was fit to analyze diabetes status and access method as a covariate. Likewise, the Prentice-Williams-Peterson Gap Time (PWP-GT) model was also used to assess the effect on the covariates given the gap time. With each model, the shape of the hazard function is allowed to vary with the event number. Subjects are not at risk for the k^{th} event until the $(k-1)^{th}$ has occurred. From the data, we can see there 100 instances of a first infection, 72 of a second infection, 52 of a third infection, 34 of a 4th infection, and 33 of a 5th infection.

From the results of the PWP-CP model, having diabetes has an increase in risk of 193% (p<0.0001). The fistula method compared to the catheter method corresponds to a 46% decrease in risk (p=0.008), while the graft method compared to the catheter method corresponds to a 35% (p=0.03). Overall, this model showed that diabetes and method both has significant effects on the model (p<0.0001 and 0.03 respectively). Analyzing the results of the PWP-GT model, we are now interested in the time between events. From these results, we can see that Diabetes has an increase in risk of 324% (p<0.0001). The fistula method compared to the catheter method corresponds to a 32% decrease in risk, while the graft method compared to the catheter method corresponds to a 28%. Neither of these methods reached statistical significance. Diabetes had a significant effect on the model (p<0.0001) while method did not (p=0.34). Overall, diabetes and the type of method significantly has an effect on the risk of recurring infection.

2. You are the statistician for a large clinical trial that is being designed to evaluate a new drug for preventing stroke in high-risk patients. The outcome of interest for this study will be the time to an occurrence of stroke. The planned study will last for a total of ten years, with patients being enrolled for the first four years of the study and then followed for an additional six years. Half of the subjects will be randomized to receive the new drug; the other half will receive the traditional drug.

A. The trial's principal investigator believes that patients taking the new drug will have a 30% reduction in their risk of stroke as compared to those receiving the standard treatment. How many strokes will the PI need to observe in order to have 90% power to detect this effect at significance level $\alpha = 0.05$?

$m = \frac{(Z_{\frac{\alpha}{2}} + Z_{\beta})^2}{\theta^2 * \pi * (1-\pi)}$

$m = \frac{(-1.96-1.28)^2}{ln(0.7)^2 * 0.5 * (1-0.5)} = 330$ events

B. Results from a prior study suggest that 12% of patients taking the traditional drug will have a stroke within the first year of being diagnosed as high-risk. Use this finding to estimate the probability that a subject enrolled in the new study will have a stroke.

We are given that 12% of patients taking the traditional drug will have a stroke within the first year of being diagnosed as high-risk.

$S_1(t) = e^{\frac{-t}{\mu}}$

$0.88 = e^{\frac{-1}{\mu}}$

Solve for μ

$\frac{-1}{\mu} = ln(0.88) = 7.822683$

Given that we know μ we can solve for the other survival times. (6,8,10) given by f, $\frac{a}{2} + f$ and a+f.

$S_1(6) = e^{-6/7.822683} = 0.4644$

$S_1(8) = e^{-8/7.822683} = 0.3596$

$S_1(10) = e^{-10/7.822683} = 0.27850$

Now to get S_2

$S_2(t) = S_1(t)^{\beta * x}$

$S_2(6) = 0.4644^{0.7} = 0.58455$

$S_2(8) = 0.3596^{0.7} = 0.48876$

$S_2(10) = 0.27850^{0.7} = 0.40867$

$\bar{S}(t) = \pi S_1(t) + (1 - \pi) S_2(t)$

$\bar{S}(6) = 0.5 * 0.4644 + 0.5 * 0.58455 = 0.524475;$

$\bar{S}(8) = 0.5 * 0.3596 + 0.5 * 0.48876 = 0.42418;$

$\bar{S}(10) = 0.5 * 0.27850 + 0.5 * 0.40867 = 0.343585;$

Now given these times we can calculate the probability of an event.

$\text{P(event)} = 1 - \frac{1}{6}[\bar{S}(f) + 4\bar{S}(\frac{a}{2} + f) + \bar{S}(a + f)]$

$\text{P(event)} = 1 - \frac{1}{6}[0.524475 + 4 * 0.48876 + 0.343585] = 0.529$

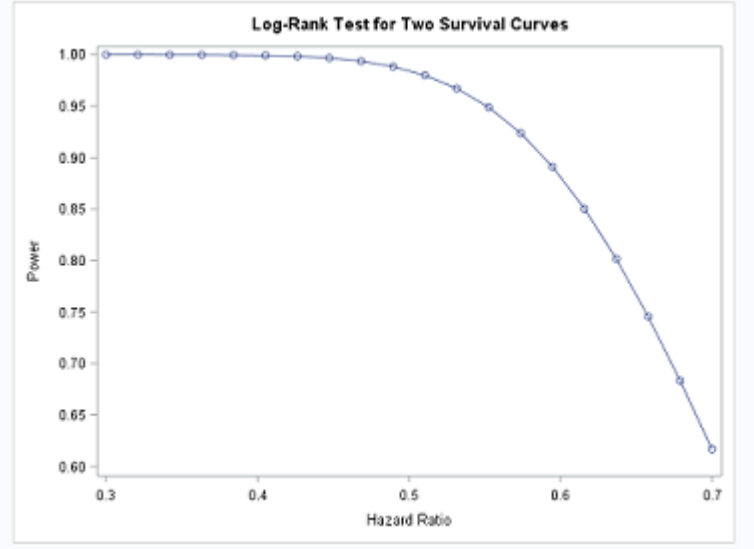
Now that we are given the P(event) and we know the number of events, we can solve for the sample size.

$n = \frac{m}{P(event)}$

$n = \frac{330}{0.529} = 623.8 = 624$ subjects. We will nee 312 subjects in each group.

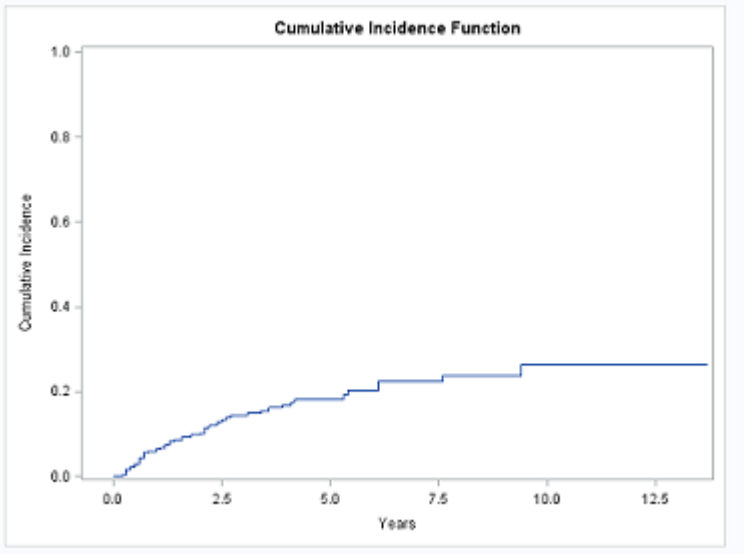
C. The principal investigator needs to include a sample size justification in her grant application for this trial. Propose a sample size for this trial and write a brief paragraph explaining your choice. Include a figure showing the study power at a range of hazard ratios.

We are seeking to examine a new drug to administer to patients with a high-risk stroke. The outcome of interest for this study will be the time to an occurrence of stroke. The study will last for a total of 10 years, with patients being enrolled for the first four years of the study and then followed for an additional six years. Half of the subjects will be randomized to receive the new drug; the other half will receive the traditional drug. In order for the new drug to have a 30% decrease in risk with 90% power with an $\alpha = 0.05$, we would need to observe a total of 330 events. Given that 12% of patients taking the traditional drug have a stroke in the first year, we hypothesize that the probability of an event occuring for a subject enrolled in the new study is 0.529. Using this knowledge, we anticipate a sample size of 624 subjects, 312 in each group is needed. Below a graph is shown for power at a range of hazard ratios.



1. A nephrologist is studying the time to disease progression following nephrectomy for patients with renal cell carcinoma. While the nephrologist is primarily interested in progression, he knows that he must also take into account the fact that death is a competing risk for this population.

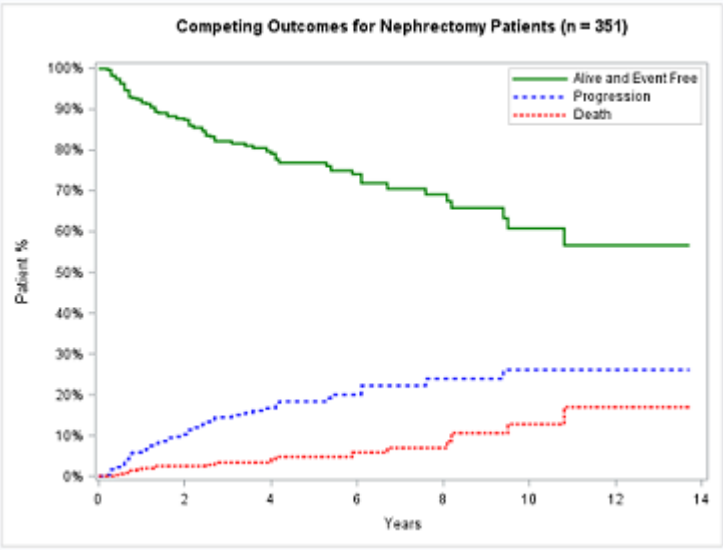
A. Plot the cumulative incidence function for progression. What proportion of patients will have experienced disease progression at two years after nephrectomy? Provide a 95% confidence interval for this proportion.



The cumulative incidence given is SAS is 0.1019 with 95% confidence intervals (0.0700,0.1408). If we turn this into a percentage we get

10.2% experienced disease progression at two years with 95% confidence intervals (7.0,14.1)%

B. Plot the cumulative incidence functions for progression and death on the same plot. Include the Kaplan-Meier estimate of survival in the plot. What conclusions can we draw from this figure?.



From the graph, we can see that the incidence of progression is much higher than the incidence of death for this population. The alive curve is the Kaplan-Meier is alive free from death or progression. At two years we have 10.2% of patients experiencing progression as compared to 2.4% for death.

C. Does the cumulative incidence of progression vary by cancer stage? Conduct an appropriate hypothesis test to answer this question. (Make sure you state the null and alternative hypotheses, level of significance used, test statistic, p-value and conclusion in terms of the problem).

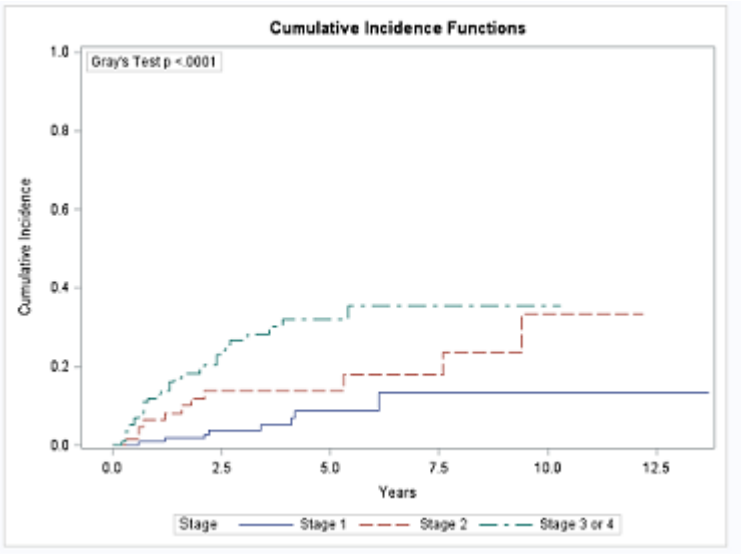
To conduct this test we will use Gray's test for the CIF. The null and alternature hypothesis are as follows:

$H_0 : F_{11}(t) = F_{12}(t) = F_{13}(t)$

The CIF for event 1 for Stage 1, event 1 for Stage 2, and event 1 for Stage 3 or 4 are the same

H_A : At least two of them are different

The Chi-Square test statistic is 20.0498. A p-value of <0.05 was considered statistically significant. The p-value is <0.0001. Therefore, we reject the null hypothesis and conclude there is a difference in at least two of the CIF functions.


Cumulative Incidence Functions
Gray's Test p <.0001
Cumulative Incidence
1.0
0.8
0.6
0.4
0.2
0.0
0.0
2.5
5.0
7.5
10.0
12.5
Years
Stage
Stage 1
Stage 2
Stage 3 or 4