Survival Analysis Models in SAS

This project was made to showcase different statistical models and tests in survival analysis, including the Cox Proportional Hazards model, how to treat competing events, and determining minimum sample size to attain a good statistical power. Graphs were created using SAS's basic ODS graphics system.

1. *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 six years, with patients being enrolled for the first two years of the study and then followed for an additional four 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 25% 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 α = 0.05?*

$$\boldsymbol{m = \frac{(z_{\alpha/2} - z_{\beta})^2}{\theta^2 \pi(1-\pi)}}$$

$$= \frac{(-1.96 - 1.28)^2}{\log(0.75)^2\, 0.5(1-0.5)} = 507.37$$

Rounding up to the nearest integer, we need 508 strokes.

B. *The PI has provided you with the results from a published trial that examined the risk of stroke in patients taking the traditional drug. Use these results to estimate the probability that a subject enrolled in the new study will have a stroke.*

$$P(event\ occurs) = 1 - \frac{1}{6}[\hat{S}(f) + 4 * \hat{S}\left(\frac{\alpha}{2} + f\right) + \hat{S}(\alpha + f)]$$

$$= 1 - \frac{1}{6}[\hat{S}(4) + 4 * \hat{S}\left(\frac{2}{2} + 4\right) + \hat{S}(2 + 4)]$$

$$= 1 - \frac{1}{6}[0.63 + 4 * 0.51 + 0.47] = 0.4767$$

Subjects in the study have a 47.67% risk of stroke.

C. *A colleague of the PI suggests that the trial results used in Part B may not be accurate and suggests you use another study to estimate the rate of strokes in patients receiving the traditional drug. The suggested study found that 9% 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.*

$$S_1(1) = 0.91 = \exp\left(\frac{-1}{\mu}\right)$$

$$\mu = \frac{-1}{\log(0.91)} = 10.6$$

$$S_1(4) = \exp\left(\frac{-4}{10.6}\right) = 0.686$$

$$S_1(5) = 0.624 \; and \; S_1(6) = 0.568$$

Therefore,

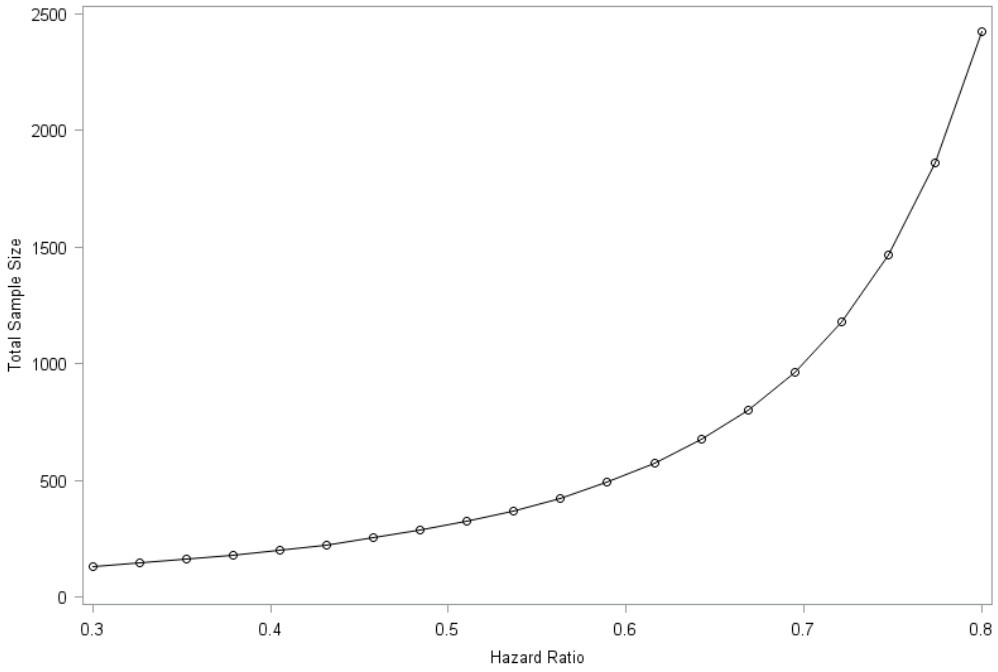
$$P(event\ occurs) = 1 - \frac{1}{6}[0.686 + 4 * 0.624 + 0.568] = 0.375$$

D. *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.*

$$n = \frac{508}{0.375} = 1354.67$$

Rounding up, we need 1355 samples.

Sample size justification. 1491 subjects will be enrolled and randomized (1:1) to receive either the new or traditional drug. We assume up to 10% will be lost to follow-up, leaving at least 1355 subjects available for analysis. Assuming a significance level of 5%, this six-year study will have therefore have approximately 90% power to detect a 25% decrease (or a hazard ratio of 0.75) in the risk of stroke for the new drug compared to the traditional drug.



2. *A plastic surgeon working with burn patients has developed a new treatment that she believes will help prevent infections. She has collected data on time to infection following admission to the hospital for 116 patients.*

Hazard Ratio (new vs. old) = 0.501

95% Confidence Interval: (0.254, 0.987)

The new treatment results in a 50% decrease in risk of event with Pr > ChiSq = 0.0458.

A. *Using the length of the follow-up period as the survival time and the infection status at last follow-up as the censoring variable, fit a Cox proportional hazards model to assess the effect of treatment assignment on time to infection. [Treatment should be the only effect in the model]. Report a hazard ratio and 95% confidence interval for the effect of the new treatment relative to the standard treatment. Provide your interpretation of this hazard ratio.*

$$\log(h(t|\beta, x) = \log\big(h_0(t)\big) + \beta_T x_T + \beta_{TA} x_{TA}$$

$$x_T = \begin{cases} 1\ for\ New\ treatment \\ 0\ for\ Standard\ treatment \end{cases}$$

$$x_{TA} = interaction\ variable\ indicating\ age\ and\ treatment\ type$$

B. *Define the variables in this model.*

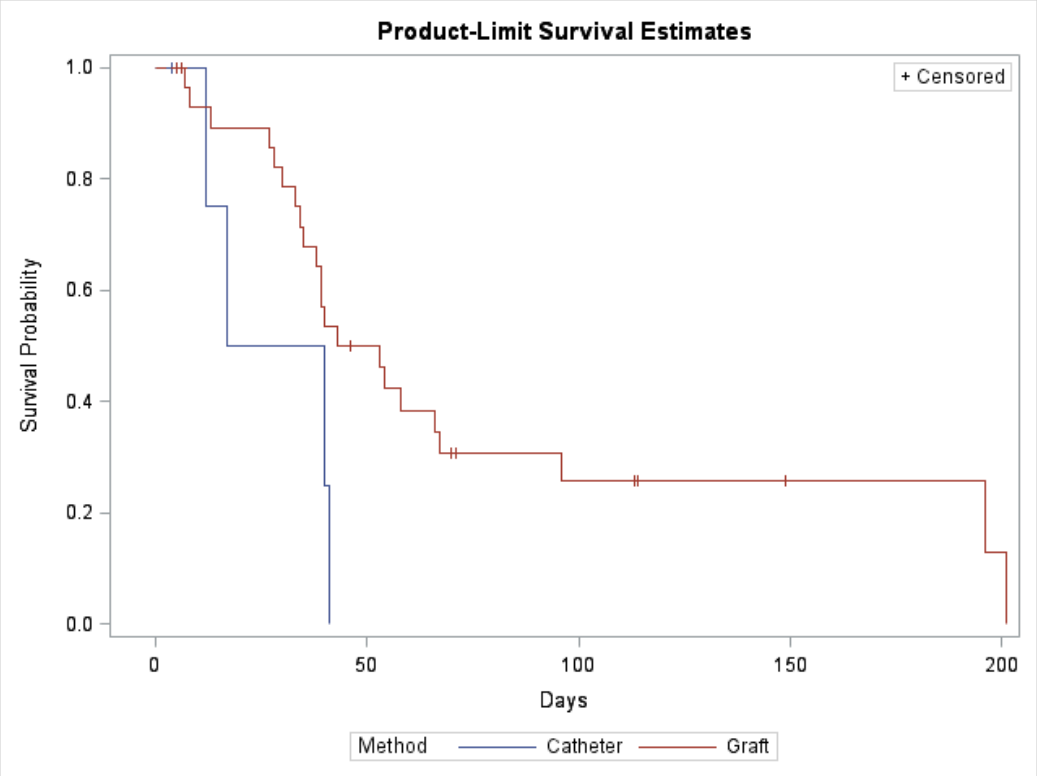
See above.

C. *Fit the model you proposed in Part B. Is there evidence that the effectiveness of the new treatment depends on how much of the patient's body is burned? Provide justification for your answer.*

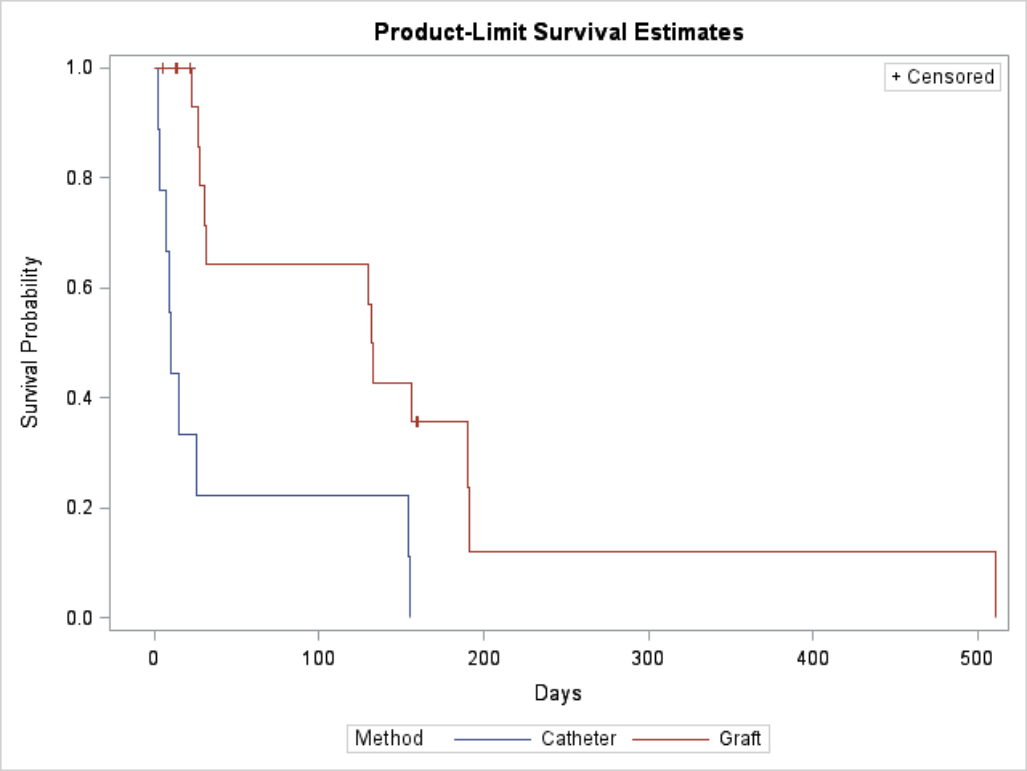
The Wald test statistic is 0.0003 and the p-value is 0.9851. We therefore fail to reject H_0 and conclude that the effectiveness of the new treatment does not depend on how much of the patient's body was burned.

3. ***A****. For subjects with acute nephritis, plot the Kaplan-Meier estimate for each dialysis method on the same graph. Create similar graphs for each of the other three disease subtypes. [You should have four graphs (one for each disease subtype), where each graph has two survival curves (one for each dialysis method)]. Comment on any differences you see.*

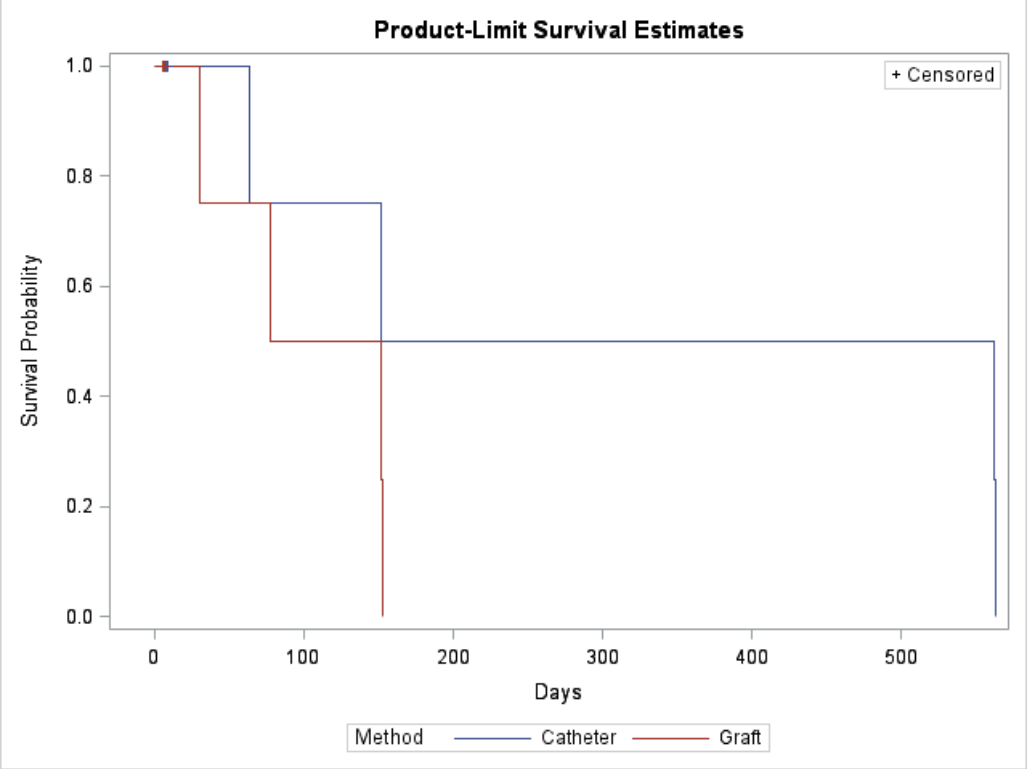
Acute*:*



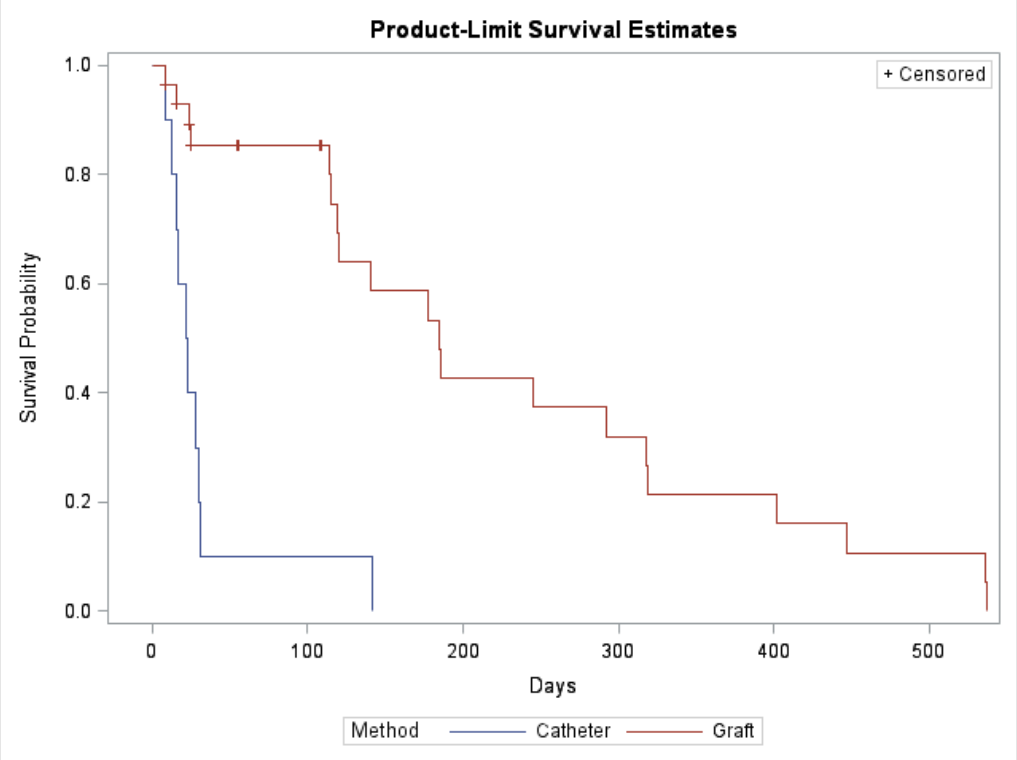
Glomerulo*:*



Polycystic*:*



Other:



Disease Subtype	Catheter	Graft	Log-rank p-value
Acute Nephritis	28.5	48	0.0554
Glomerulo Nephritis	10	132.5	0. 0012
Polycystic Kidney	357	115	0.2119
Other	22.5	185	<.0001

Comment: I noticed the Graft seems to help significantly with survival time for every group besides the Polycystic Kidney group, but for that group the results aren't significant.

B. *Stratifying on disease subtype, fit a stratified Cox proportional hazards model using the length of the follow-up period as the survival time and the status at last follow-up as the censoring variable. Include effects for age and dialysis method and assume the effects of these covariates are the same for each disease type. [Age and dialysis method should be the only effects in your model]. Write the model using mathematical notation (i.e., using β coefficients). Be sure to define any variables used in the model, including any dummy variables.*

$$\log(h_D(t|\beta, x) = \log\big(h_{Do}(t)\big) + \beta_{Age}x_{Age} + \beta_M x_M$$

D indicates disease subtype and can be *A* for Acute Nephritis, *G* for Glomerulo, *P* for Polycystic Kidney Disease or *O* for "other".

x_{Age} indicates patient age and $x_M = \begin{cases} 1\ if\ method = catheter \\ 0\ if\ method = graft \end{cases}$

***C**. Report estimated hazard ratios and 95% confidence intervals for each of the effects of the model you fit in Part B. Provide your interpretations of these hazard ratios.*

Hazard Ratio (age): 1.002	LCL: 0.983	UCL: 1.022
Hazard Ratio (method): 3.601	LCL: 2.063	UCL: 6.286

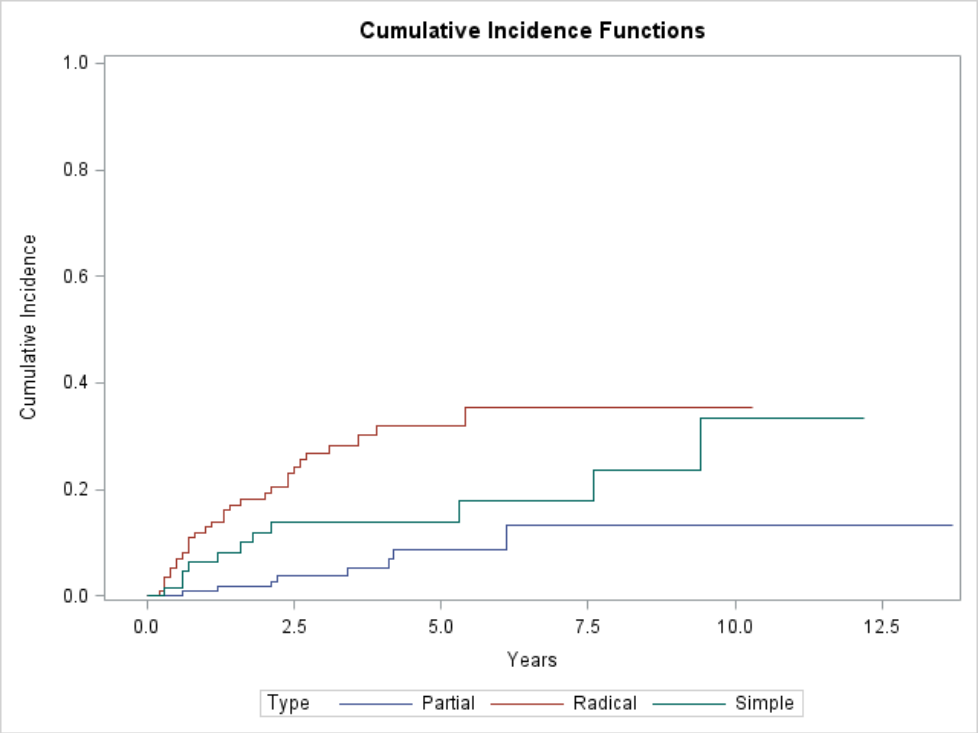
A 1-unit increase in age results in a 0.2% increase in risk of event, while the choice of a catheter over graft increase probability of event by 360%.

***D.** Refit the model from Part B, this time allowing the effect of dialysis method to vary by disease subtype. (As before, assume the effect of age does not differ by disease subtype). Write this revised model using mathematical notation (i.e., using β coefficients). Be sure to define any variables used in the model, including any dummy variables. Report the estimates of the hazard ratios and 95% confidence intervals for effect of dialysis method for each disease subtype. Provide your interpretations of these hazard ratios.*

Catheter vs. Graft at Subtype	Hazard Ratio	LCL	UCL
Acute nephritis	4.092	1.436	11.660
Glomerulo	3.797	1.826	7.896
Other	5.674	2.621	12.28
Polycystic Kidney Disease	1.0	1.0	1.0

4. *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 for each type of nephrectomy on the same plot. For each nephrectomy type, what proportion of patients will have experienced disease progression at two years post-nephrectomy? Provide a 95% confidence interval for each proportion.*



At two years, from the output we see 2.72% of Partial, 19.32% of Radical Nephrectomy patients, and 13.9% of Simple Nephrectomy patients experienced the events.

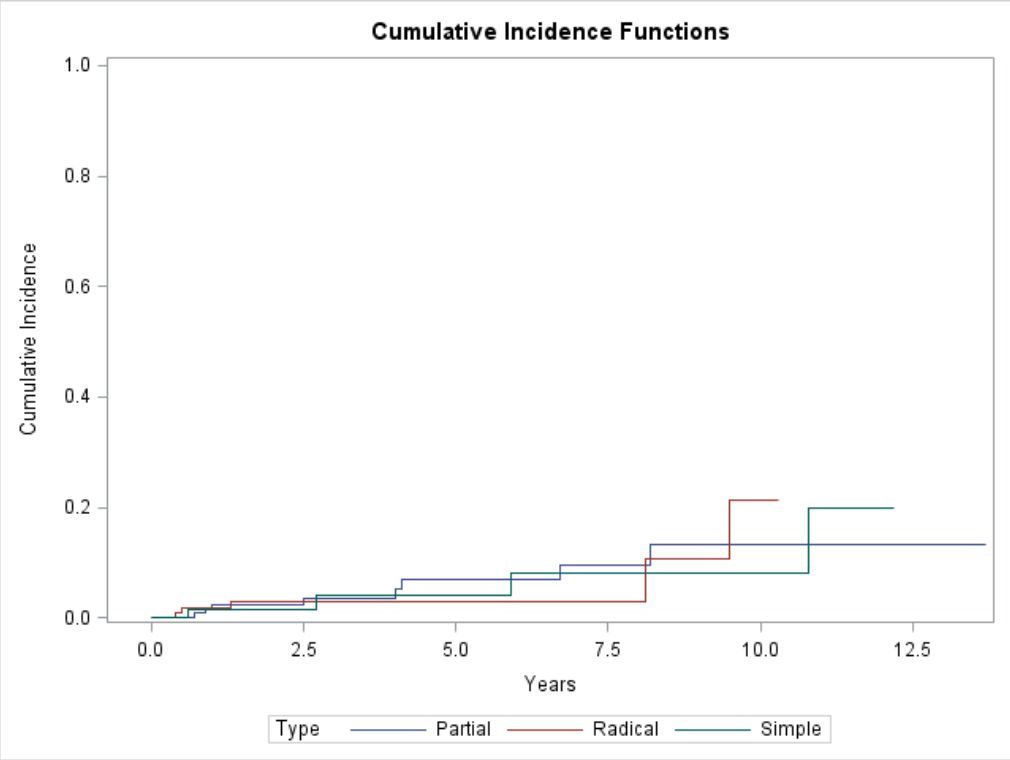
Simple	13.9%	6.41%	24.37%
Partial	2.72%	0.719%	0.718%
Radical	19.32%	12.29%	27.54%

B. *Does the cumulative incidence of progression vary by nephrectomy type? Conduct an appropriate hypothesis test to answer this question.*

The CIF does vary by nephrectomy type. Gray's test indicates significant difference between cumulative incidence functions.

Chi-Square with 2 DF: 20.0498
p-value: <.0001

C. *Plot the cumulative incidence function for death for each type of nephrectomy on the same plot. For each nephrectomy type, what proportion of patients will have died at five years post-nephrectomy? Provide a 95% confidence interval for each proportion.*



D. *Does the cumulative incidence of death vary by nephrectomy type? Conduct an appropriate hypothesis test to answer this question.*

It does not. Gray's test does not indicate significant difference among CIFs. Chi-Square with 2 df: 0.0319 and p-value 0.9842