Case Study: The impact of tumor location on survival outcomes for patients diagnosed with pancreatic cancer and treated with surgery in Louisiana between 2004 and 2016

James Henegan
Department of Data Science
University of Mississippi Medical Center
BDS 725 Final Project, Fall 2019

Abstract

Background. The impact of tumor location on survival outcomes for patients diagnosed with pancreatic cancer and treated with surgery has been investigated in the past. For example, Artinyan et. al found that patients who had cancer in the body or tail of the pancreas had survival outcomes that were worse than those who had cancer in the head of the pancreas.

Methods. For this observational study, we obtained data from the Surveillance, Epidemiology, and End Point Surveillance (SEER) program and investigated survival outcomes for $N = 988$ patients diagnosed with pancreatic cancer and treated with surgery in Louisiana between 2004 and 2016. The event of interest was all-cause mortality, and the time origin for each patient was time of diagnosis for the cancer. The main objective was to determine the extent to which tumor location influenced survival probabilities. Specifically, we compared the outcomes of patients who had a tumor in the head of the pancreas with the outcomes of patients who had a tumor in the body or tail of the pancreas.

Results. Of the $N = 988$ subjects in our sample, 68.4% of them ($N = 676$) had a recorded death. The median survival time was 18 months (95% CI: 17,20). Patients with tumors in the head (H) of the pancreas had a median survival time of 17 months (16,19) while patients with tumors in the body or tail (B/T) of the pancreas had a median survival time of 26 months (20,38). Kaplan-Meier curves indicated that overall survival outcomes were better for patients in the B/T group, as compared to the H group. The log-rank test established that the difference between these two curves was statistically significant ($p < 0.0001$). Using the B/T group as the reference group, an unadjusted Cox model (Model 1) for tumor location estimated that the Hazard Ratio between these groups was 1.58 (95% CI: $1.32, 1.90$). Another unadjusted Cox model (Model 2) based on year of diagnosis indicated that patients diagnosed during or after 2011 had outcomes that were better than patients diagnosed before 2011 (HR: 0.66, 95% CI: 0.57,0.77). Multivariate models indicated that when variables such as cancer grade, cancer stage, and age at diagnosis were considered in addition to tumor location, the impact of tumor location was negligible [Tumor Location HR of 1.11, with a 95% CI of (0.91, 1.35)].

Conclusions. With Model 1, we found that the survival outcomes for patients with cancer in the head of the pancreas were worse than the survival outcomes patients who had cancer in the body or tail of the pancreas. In this sense, our results ran opposite to the ones obtained by Artinyan et. al. However, their study was based on data collected between 1998 and 2004, while our study was based on data collected between 2004 and 2016. This suggests that survival outcomes may be influenced by year of diagnosis. This claim appears to be supported by Model 2, as mentioned above. This aligns with the idea that the quality of clinical treatment (e.g., surgical procedures and medicine) may be improving over time. Also, Artinyan's result was based data from all of the registries in the SEER database, while we only used data from the Louisiana registry.

Finally, our results indicate that the impact of tumor location disappears when variables such as cancer grade and stage are considered.

Contents

1 Introduction

The American Cancer Society estimates that 55,770 people in the United States will be diagnosed with pancreatic cancer in 2019, and that 45,750 will die from the disease [1]. For the sake of comparison, the same organization estimates that approximately 96,480 people in the United States will be diagnosed with melanoma skin cancer, and that 7,230 will die from it. These estimates indicate that, while not as common as skin cancer, pancreatic cancer is much more fatal. Based on data from 2009-2015, the estimated percentage of patients surviving five years after receiving a diagnosis for pancreatic cancer is 9.3% [2].

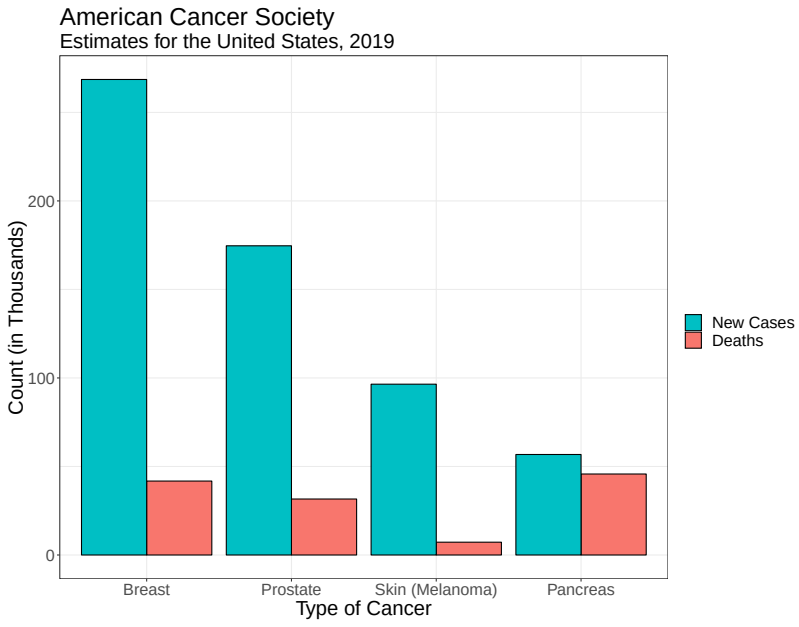

Figure 1. American Cancer Society Estimates.

If a patient has been diagnosed with pancreatic cancer, then, in many cases, surgery is not an option [3]. This follows from the fact that the disease is difficult to detect when it is in its early stages [4]. However, in certain cases, surgery may be performed in an effort to cure the disease [5].

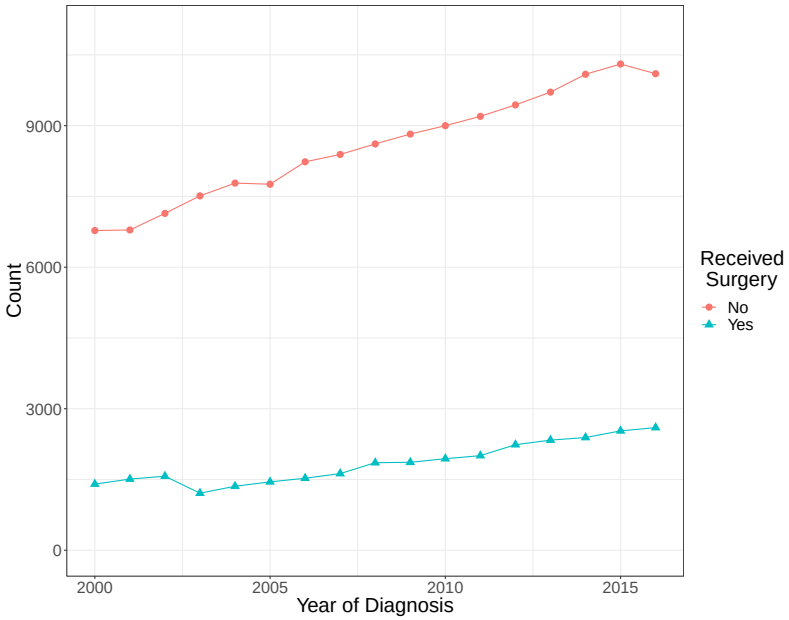

Figure 2. Of all the patients in the SEER database diagnosed with pancreatic cancer between 2000 and 2016, how many received surgery? How many did not?

The impact of tumor location on survival outcomes for patients diagnosed with pancreatic cancer and treated with surgery has been investigated in the past. For example, Artinyan et. al [6] found that patients who had cancer in the body or tail of the pancreas had outcomes that were worse than those who had cancer in the head of the pancreas. The purpose of this study is to continue this investigation between the relationship between tumor location and survival probability for patients diagnosed with pancreatic cancer and treated with surgery. We restricted our attention to patients from Louisiana who had been diagnosed with the disease between the years of 2004 and 2016, inclusive. The decision to focus on patients from Louisiana was partially based on the idea that survival outcomes may be influenced by geographic location. For example, the quality of treatment in one part of the United States may be better or worse than the quality of treatment in another. By focusing on patients from Louisiana, we hoped to eliminate some potential variability in survival outcomes that may be due to geography.

2 Methods

In this section, we provide a summary of the methods used in our investigation. We also provide details on how we collected the data used in our project.

2.1 Summary of Methods

Data Collection. Using data obtained with permission from the Surveillance, Epidemiology, and End Point Surveillance (SEER) program [7], we investigated survival outcomes for $N = 988$ patients diagnosed with pancreatic cancer and treated with surgery in Louisiana between 2004 and 2016. The data available from the SEER database on patients diagnosed with pancreatic cancer contains information on the location of the patient's tumor. This variable assumes one of three values: `Head`, `Body/Tail`, or `Other`. The purpose of our study was to compare the outcomes of patients with a `Head` location to the outcomes of patients with a `Body/Tail` location.

Definition of Survival Data. The event of interest was all-cause mortality. The initial starting time for each patient was the time of diagnosis for the cancer. The data for our study was subject to right censoring.

Descriptive Statistics. To describe certain non-survival aspects of our data, we created a characteristics table for our sample, grouped by tumor location (i.e., `Head` against `Body/Tail`). We utilized data visualization techniques to gain a better understanding of our sample with respect to age distribution and year of diagnosis. Data visualization was also used to explore how tumor location, tumor grade, and cancer stage manifested themselves within our sample.

To describe our survival data, we computed the median survival time for our sample, along with a 95% confidence interval for the true median survival time. We constructed an overall Kaplan-Meier curve, an overall Nelson-Aalen cumulative hazard curve, and an overall smoothed hazard curve. Furthermore, we plotted smoothed hazard estimates by tumor location, marginalized by tumor grade and cancer stage, while adjusting for age of diagnosis and year of diagnosis.

Hypothesis Testing. Kaplan-Meier curves were constructed to visually inspect the differences in survival outcomes for groups defined by the following variables: (i) tumor location, (ii) cancer stage, (iii) tumor grade, (iv) race, (v) sex, (vi) age at diagnosis, and (vii) year of diagnosis. Log-rank and Wilcoxen tests were conducted to assess the statistical significance of these differences.

Single-variable (unadjusted) Cox proportional hazards models were built to estimate differences between the hazard rates for these groups. The validity of each univariate Cox model was checked by computing Schoenfeld residuals. We also constructed a Cox model with time-varying variables to model the impact of tumor location on survival outcomes.

Additional multivariate analysis was performed. We constructed a stratified Cox model involving tumor location, age at diagnosis, and year of diagnosis. This model was stratified by cancer stage and tumor grade. After diagnosing the stratified model with Schoenfeld residuals, we looked for influential points with the help of DFBETAs. Step-wise model selection and the use of random forests assisted in the multi-variate modeling process.

2.2 Data Collection

The following diagram provides details about how we used the SEER program to obtain the data used for our analysis.

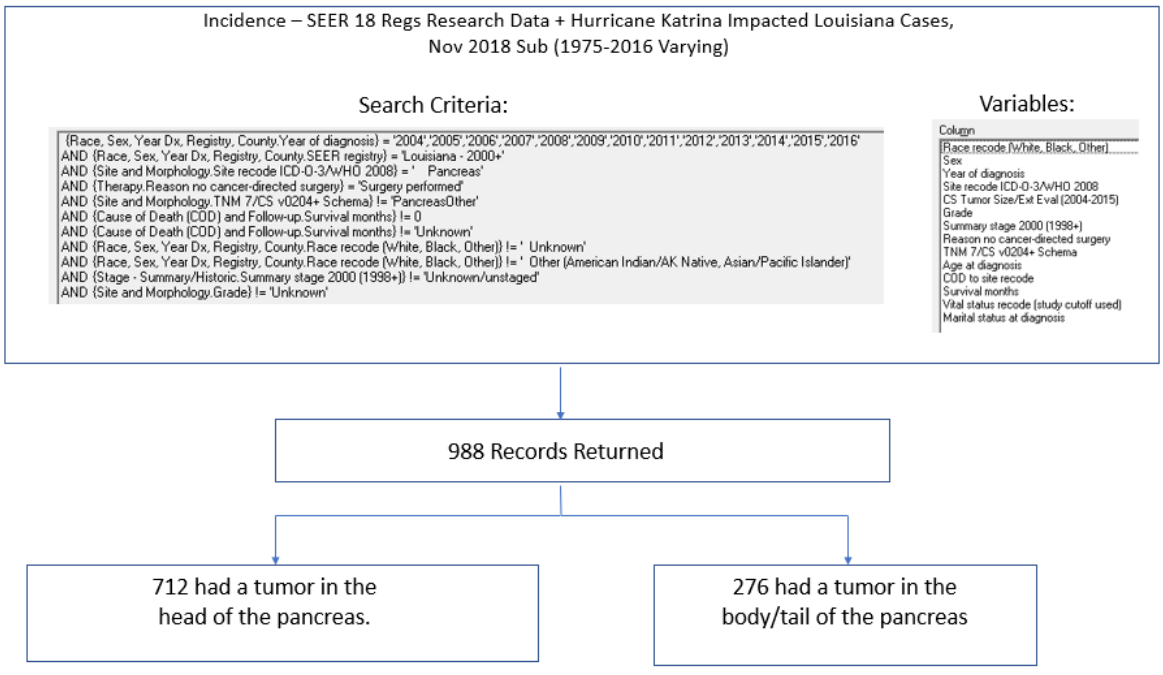

Figure 3. Flow Diagram.

The essential search criteria may be summarized as follows. We looked for patients who were diagnosed with and received surgery for pancreatic cancer in Louisiana between 2004 and 2016, inclusive. We excluded patients who had a survival time of zero months, patients who had a race of "Other" (i.e., other than "White" or "Black"), and patients who what a tumor location of "PancreasOther" (i.e., other than the head or body/tail of the pancreas). Furthermore, we excluded patients with an unknown measurement for survival months, race, stage, or grade value.

Upon receiving the results from our query, we used the R library `tidyverse` to perform most of our data cleaning [8].

We remark that the variable `TNM 7/CS v0204+ Schema` was used to obtain information about tumor location. As mentioned above, this variable assumed one of three values: `Head`, `Body/Tail`, or `Other`. More information about this schema may be found at the website for the Collaborative Stage Data Collection System [9]. We also mention that the variable `Summary stage 2000 (1998+)` attained the values of `Local, Regional`, and `Distant`. Meanwhile, the variable `Grade` attained values of `I`, `II`, `III`, and `IV`. The decision to include both `Stage` and `Grade` in studies involving prognosis of pancreatic cancer has been been a position argued for in the past [10]. Finally, we also remark that we did not make use of the variable `CS Tumor Size/Exit Eval(2004-2015)`, even though it was downloaded. The decision not to use this variable was based upon the fact that, out of the 988 returned results in our data set, 837 (85%) of them had a value of `3`, while 87 (9%) had a missing value. Moreover 6 observations had a value of zero, despite the fact that these were patients who received surgery for pancreatic cancer. These issues caused us to question the integrity of the data for this variable.

3 Results

In this section, we describe the results of our investigation. We begin by providing some descriptive statistics. Then we give the results of our hypothesis testing. We finish by presenting a Stratified Cox Model.

3.1 Descriptive Statistics

We begin with a characteristics table, summarizing various properties of the $N = 988$ members of our sample. Then we summarize the overall survival data for our sample. We conclude the section by summarizing the survival data by tumor location.

3.1.1 Summary of Characteristics

Of the 988 members of our sample, 712 had a tumor located in the head of the pancreas, while 276 had a tumor located in the body or tail of the pancreas. A characteristics table is provided below.

	Tumor Location: Head of Pancreas ($N = 712$)	Tumor Location: Body/Tail of Pancreas ($N = 276$)
Age at Diagnosis (years)	64.89 ± 10.77	63.05 ± 12.22
Year of Diagnosis	2010.84 ± 3.55	2011.60 ± 3.38
Sex		
Female	354 (50%)	151 (55%)
Male	358 (50%)	125 (45%)
Race		
Black	146 (21%)	60 (22%)
White	566 (79%)	216 (78%)
Grade		
I	109 (15%)	105 (38%)
II	329 (46%)	90 (33%)
III	256 (36%)	73 (26%)
IV	18 (3%)	8 (3%)
Stage		
Local	92 (13%)	95 (34%)
Regional	573 (80%)	130 (47%)
Distant	47 (7%)	51 (18%)

Table 1. Sample Characteristics.

Below, we provide visualizations related to the data presented above. We note that grades III and IV were combined for the *Characteristics II* plot, due to the relatively small sample size for patients with Grade IV cancer.

Figure 4. Sample Characteristics.

Generally speaking, we noted that there were more patients in our sample with tumors located in the head of the pancreas than there were patients with tumors in the body or tail of the pancreas. As pointed out by the American Cancer Society in [5], curative surgery is usually performed to treat cancers in the head of the pancreas. We note that our sample is similar to the one obtained by Artinyan, et. all in [6] in this respect. Meanwhile, grade II tumor were the most common for members of the head group, while grade I tumors were most common for members of the body/tail group.

3.1.2 Summary of Overall Survival Data

In this section, we present a summary of the overall survival data. Of the $N = 988$ subjects in our sample, there were a total of 676 deaths. The median survival time for all participants was 18 months (95% CI: 17,20). The median follow up time was 49 months (95% CI: 44, 56).

The mean survival time was 38.671 months, with a standard error of 1.798. A plot of the Kaplan-Meier survival estimate is shown in Figure 5, along with dashed lines indicated the median survival time of 18 months.

	Total ($N = 988$)
Dead	676 (68%)
Median Survival Time (Months)	18 (95% CI: 17,20)
Median Follow Up Time (Months)	49 (95% CI: 44,56)

Table 2. Overall Survival Data.

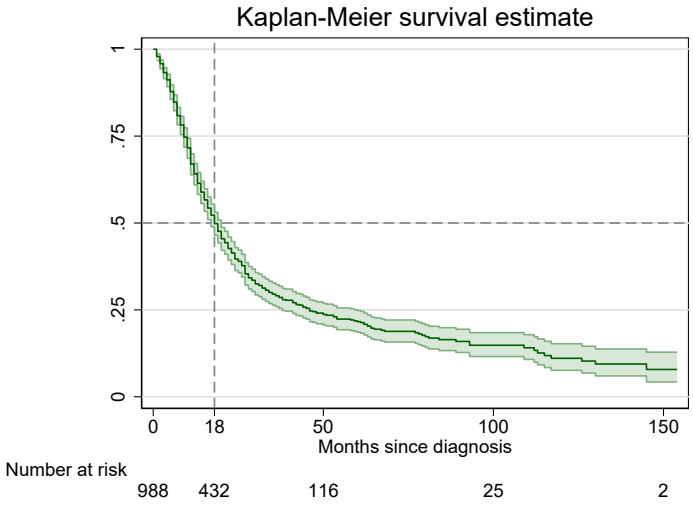

Figure 5. Overall Kaplan-Meier Curve.

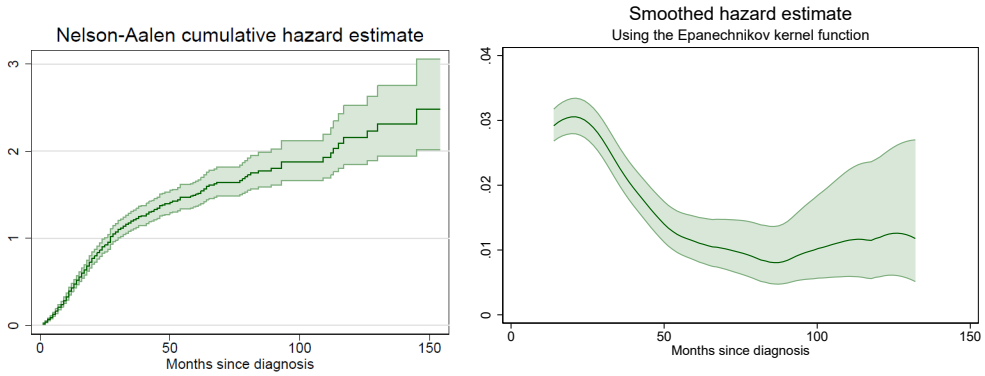

Figure 6. Nelson-Aalen cumulative hazard function and smoothed hazard estimate.

The Nelson-Aalen cumulative hazard estimate is shown in Figure 6, along with a visualization of the smoothed hazard estimate, using the Epanechnikov kernel function.

3.1.3 Summary of Survival Data, by Tumor Location

The primary research question for our project was to determine the extent to which tumor location impacted survival outcomes. The following results are directly related to this question. In general, these results suggest that patients with tumors in the body or tail of the pancreas had outcomes that were better than patients with tumors in the head of the pancreas.

	Tumor Location: Head ($N = 712$)	Tumor Location: Body/Tail ($N = 276$)
Dead	529 (74.30%)	147 (53.26%)
Median Survival Time in months (w/ 95% CI)	17 (16,19)	26 (20,38)
Median Follow-Up Time in months (w/ 95% CI)	53 (44,65)	46 (40, 52)

Table 3. Survival Data, by Tumor Location.

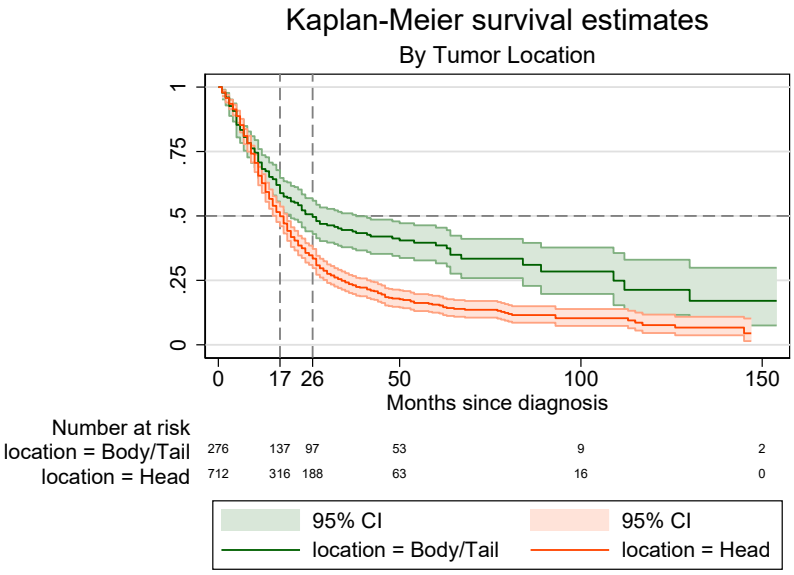

Figure 8. Kaplan-Meier Survival Estimates by Tumor Location.

The following hazard estimates (by tumor location) were obtained by adjusting for age at diagnosis and year of diagnosis, and then by marginalizing by either grade or stage. For example, the following command was used to obtain the graph in the upper left hand corner. Here, `aad` stands for "Age at Diagnosis," and `yod` stands for "Year of Diagnosis."

```
sts graph if grade == 1, haz by(location) adjust(aad yod) title("Grade = 1")
```

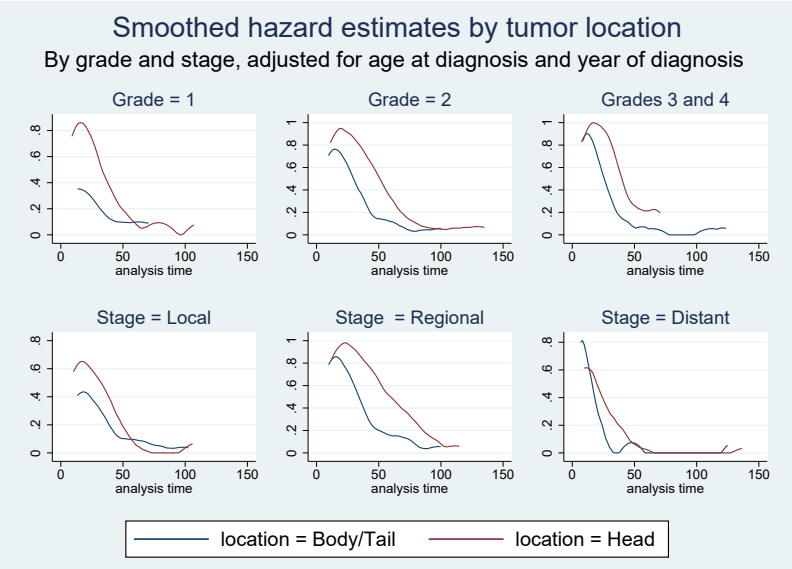

Figure 9. Smoothed Hazard Estimates by tumor location. Marginalized by tumor grade and cancer stage, adjusted for age at diagnosis and year of diagnosis.

These estimates indicate that, in general, the hazard functions for patients with a tumor located in the head of the pancreas were greater than the hazard functions for patients with a tumor located in the body or tail of the pancreas.

3.2 Hypothesis Tests for Univariate Models

In this section we report results from hypothesis tests based on univariate models. We provide Kaplan-Meier curves, results from log-rank and Wilcoxon tests, and results from unadjusted Cox models.

3.2.1 By Tumor Location

The Kaplan-Meier curve by tumor location presented in section 3.1.3 suggests that, on the whole, patients with a tumor in the Head of the pancreas had outcomes that were worse than those who had a tumor in the Body/Tail of the pancreas. The log-rank test indicated that the difference between these curves was statistically significant. Using the Body/Tail group as the baseline group, an unadjusted Cox proportional hazards model estimates that the hazard rate between these two groups is 1.58 (95% CI: 1.32, 1.90). A Schoenfeld p-value of 0.07 indicates that we have insufficient evidence to reject the proportional hazards assumption at an $\alpha = 0.05$ level of significance.

Variable	Unadj. HR	Unadj. HR 95% CI	Schoenfeld p-value	Log-rank p-value	Wilcoxon p-value
Tumor Location				< 0.0001	0.0022
Head	1.58	(1.32, 1.90)	0.07		
Body/Tail	–				

Table 4. Results from the unadjusted Cox model for Tumor Location.

However, when we constructed a log-log plot by tumor location, we saw that the proportional hazards assumption was clearly violated, especially for the first eight months after time of diagnosis. Indeed, note that the two curves in the plot below cross before Log(Time) = 2.

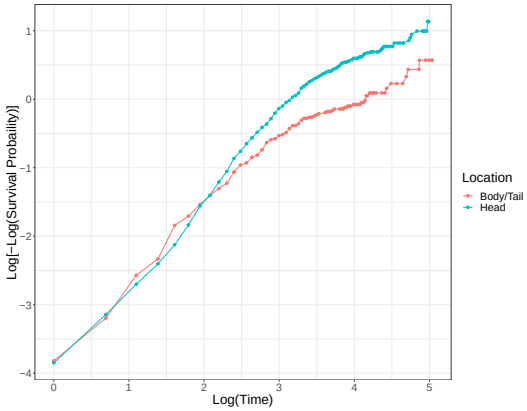

Figure 10. Log-log plot by tumor Location.

This prompted us to consider a model with two time-dependent covariates:

$$Z_1(t) = \begin{cases} 1, & \text{Tumor in the head of the pancreas and } t \leq \tau \\ 0, & \text{otherwise} \end{cases}$$

$$Z_2(t) = \begin{cases} 1, & \text{Tumor in the head of the pancreas and } t > \tau \\ 0, & \text{otherwise} \end{cases}$$

We found that the value of τ which maximizes the partial likelihood for this model was $\tau = 5$ months. This caused us to fit the model

$$h(t, \boldsymbol{X}) = h_0(t) e^{x_1 Z_1 + x_2 Z_2},$$

where

$$Z_1(t) = \begin{cases} 1, & \text{Tumor in the head of the pancreas and } t \leq 5 \\ 0, & \text{otherwise} \end{cases}$$

$$Z_2(t) = \begin{cases} 1, & \text{Tumor in the head of the pancreas and } t > 5 \\ 0, & \text{otherwise} \end{cases}.$$

The estimates obtained by this model are described below.

Variable Name	Hazard Ratio	95% CI for HR
Z_1	0.77	(0.52, 1.13)
Z_2	1.89	(1.53, 2.33)

Table 5. Results from the Cox model for Tumor Location with time-dependent variables.

These results suggest that the impact of tumor location on survival outcomes was not statistically significant for the first five months after diagnosis. For times after that, however, the instantaneous risk of death for patients with a tumor in the head of the pancreas is 89% higher than the instantaneous risk of death for patients with a tumor in the body or tail of the pancreas.

3.2.2 By Stage, Age, Year of Diagnosis, and Sex

Although tumor location was our primary variable of interest, we also considered the impact of other variables on survival outcomes. In this section, we consider Cancer Stage, patient age at diagnosis, year of diagnosis, and sex. The corresponding Kaplan-Meier curves are presented below.

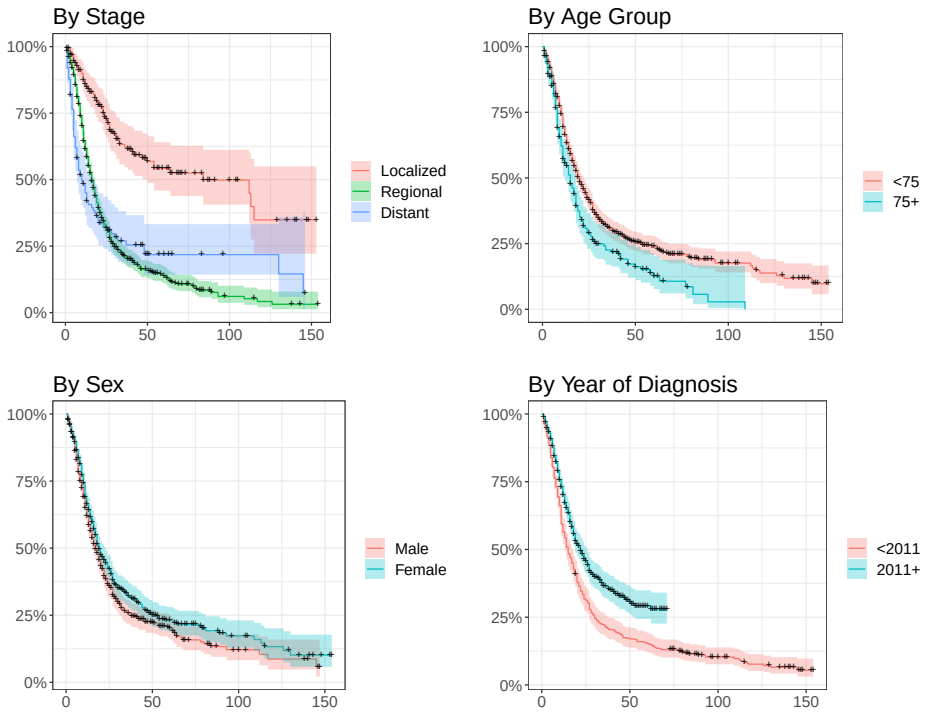

Figure 11. Kaplan-Meier estimates by Cancer Stage, Age Group, Sex, and Year of Diagnosis.

Below we provide the results for the corresponding hypothesis tests and univariate Cox models.

Variable	Unadj. HR	Unadj. HR 95% CI	Schoenfeld p-value	Log-rank p-value	Wilcoxon p-value
Cancer Stage				< 0.0001	< 0.0001
Distant	3.45	(2.47, 4.81)	0.0012		
Regional	3.27	(2.53, 4.22)	0.9877		
Localized	–				
Age at Diagnosis				0.0001	0.0003
≥ 75	1.43	(1.19, 1.70)	0.2675		
< 75	–				
Sex				0.0447	0.0518
Male	1.16	(1.00, 1.35)	0.9059		
Female	–				
Year of Diagnosis				< 0.0001	< 0.0001
≥ 2011	0.66	(0.57, 0.77)	0.9809		
< 2011	–				

Table 6. Results from the unadjusted Cox models for Cancer Stage, Age at Diagnosis, Sex, and Year of Diagnosis.

In summary, we see that the outcomes for regional and distant stages were much worse than outcomes for patients with a localized stage. Meanwhile, patients who were diagnosed at an age younger than 75 years had outcomes that were better than those who were diagnosed at an age greater than or equal to 75. Also, the outcomes for females were slightly better than those for males. Finally, we saw that patients who were diagnosed during or after 2011 had survival outcomes better than those who were diagnosed before 2011.

We remark that the "split year" of 2011 was obtained by constructing univariate Cox models based on a range of potential "split years" between 2004 and 2016. By comparing the resulting log-likelihoods from these models, it was determined that 2011 the optimal year on which to split the data.

3.2.3 By Tumor Grade

We investigated the impact of tumor grade on survival outcomes. We note that the number of patients with a grade of 4 is small, relative to the number of patients with other grades.

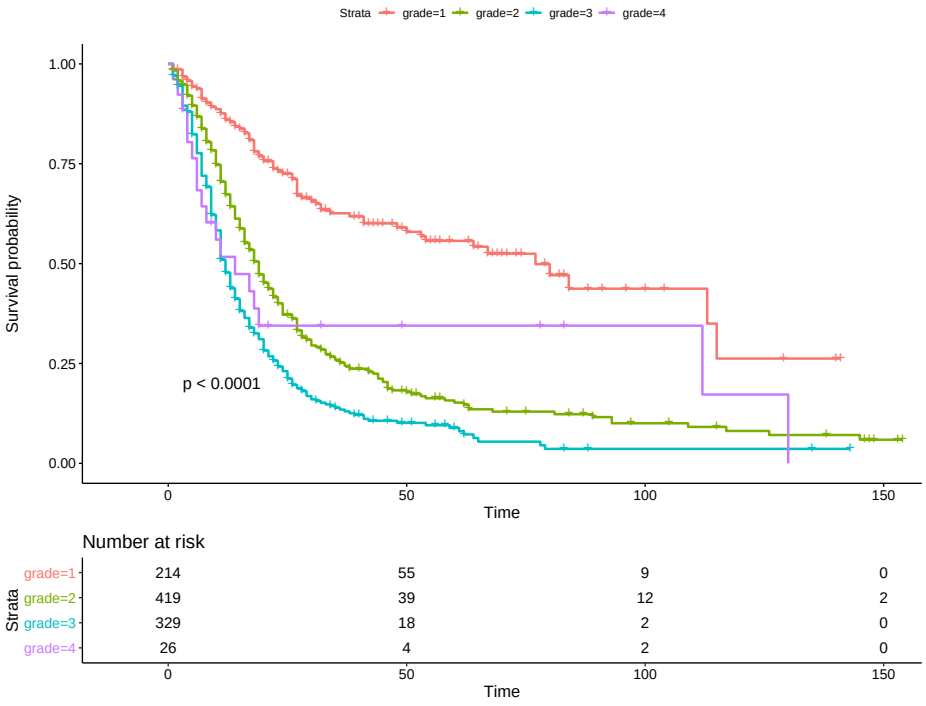

Figure 12. Kaplan-Meier estimates by Tumor Grade.

Variable	Unadj. HR	Unadj. HR 95% CI	Schoenfeld p-value	Log-rank p-value	Wilcoxon p-value
Tumor Grade				< 0.0001	< 0.0001
I	–				
II	2.75	(2.14, 3.52)	0.47		
III	4.25	(3.30, 5.47)	0.18		
IV	2.76	(1.65, 4.61)	0.18		

Table 7. Results from the unadjusted Cox models for Tumor Grade.

We remark that the hazard ratio between grades I and III is the highest hazard ratio we have seen in the unadjusted models.

3.3 Stratified Cox Model

We constructed a Cox Model, stratified by cancer stage and grade, where tumor location was the main effect. We adjusted for age at diagnosis and year of diagnosis. The results are given below

Cox Model, Stratified by Cancer Stage and Tumor Grade

Variable	HR	HR 95% CI	Schoenfeld p-value
Tumor Location			0.1977
Head	1.11	(0.91, 1.35)	
Body/Tail	–		
$\frac{1}{10}\times$(Age at Diagnosis, years)	1.15	(1.07, 1.24)	0.1843
Year of Diagnosis			0.7893
≥ 2011	0.75	(0.64, 0.88)	
< 2011	–		
			0.3465 (Global)

Table 8. Results from the stratified Cox Model.

These results indicate that when cancer grade and stage are considered, the impact of tumor location fails to be statistically significant. We constructed a Cox-Snell residual plot and determined that the overall fit was acceptable.

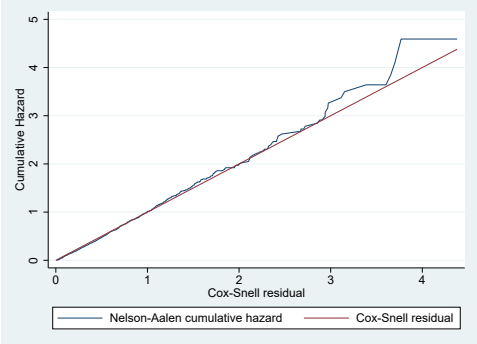

Figure 13. Cox-Snell residual plot for the Stratified Cox Model.

When we examined the DFBETA plots [12] for the stratified model and looked for influential points, we noticed that the residuals associated with each variable tended to be high for patients who survived over 50 months with a regionalized stage and a grade II cancer. Other than this, we were unable to discern any pattern among the residuals. Recall that patients with values of `Regional` and `Stage 2` were common in our sample.

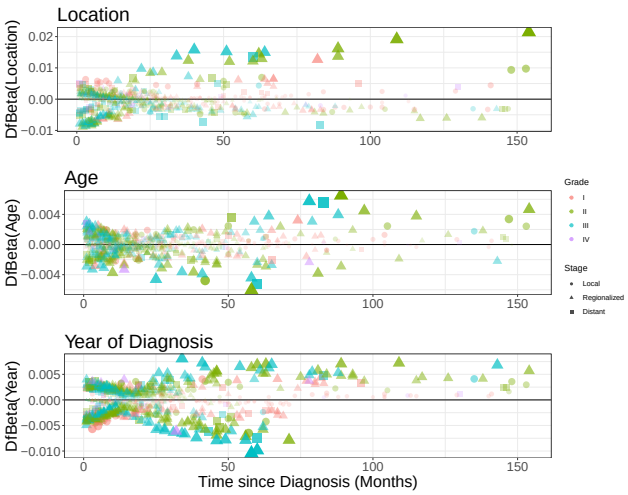

Figure 14. DFBETA plots for the Stratified Cox Model.

Overall, we concluded that the stratified cox model was acceptable.

4 Discussion

4.1 Variable Importance

The results of the Stratified Cox Models suggest that the significance of tumor location vanishes when additional variables are considered. Other approaches to multivariate modeling confirmed this notion. For example, when we used a step wise variable selection technique to build a multivariate Cox model in SAS, we saw that cancer grade, cancer stage, age at diagnosis, and year of diagnosis were included in the model. However, tumor location was not included.

Step	Entered	p-value
1	Grade	< 0.0001
2	Stage	< 0.0001
3	Age at Diagnosis	0.0003
4	Year of Diagnosis	0.0011

Table 9. Results from the step wise variable selection process, in an attempt to build a Cox model using all of the recorded variables.

Similarly, when we used random forests to discern variable importance in R, tumor location ended up being the least important variable considered.

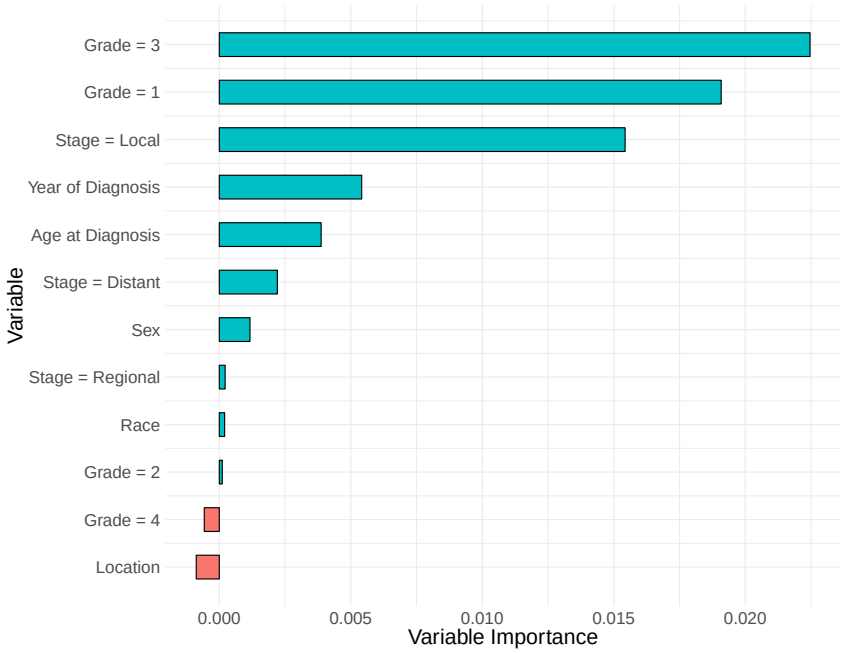

Figure 15. Variable Importance results from the Random Forest Model.

When these results are combined with the results of our stratified Cox model, it appears as though tumor location does not have a significant impact on the survival outcomes, provided that information on cancer grade and stage is considered.

4.2 Competing Risks

As mentioned previously, the event of interest in our study was all-cause mortality. In this section, we use Kaplan-Meier estimation with competing risks to gain some intuition about the relationship between the cause-specific hazard of death from pancreatic cancer and the hazard of death from any other cause. The plot below indicates that most of the patients in our sample who died within the first 25 months after diagnosis died from pancreatic cancer. Analysis was performed following methods provided in [11].

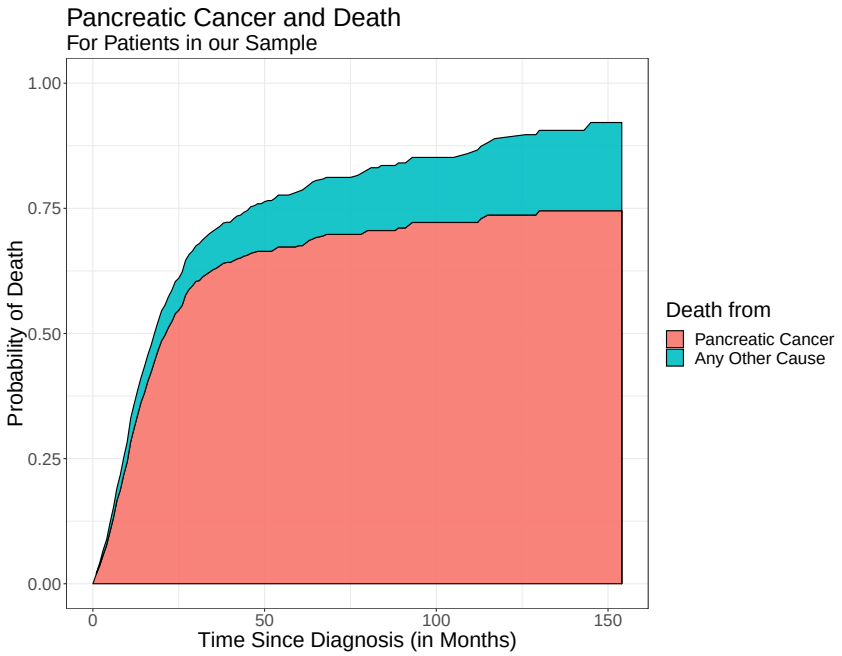

Figure 16. Pancreatic Cancer and Death
for patients in our sample, viewed through the lens of competing risks.

For the sake of comparison, we obtained some additional data from the SEER program to gain some intuition about the relationship between the cause-specific hazard of death from melanoma skin cancer and the hazard of death from any other cause. Specifically, we collected data on patients in Louisiana who were diagnosed with melanoma skin cancer between 2004 and 2016.

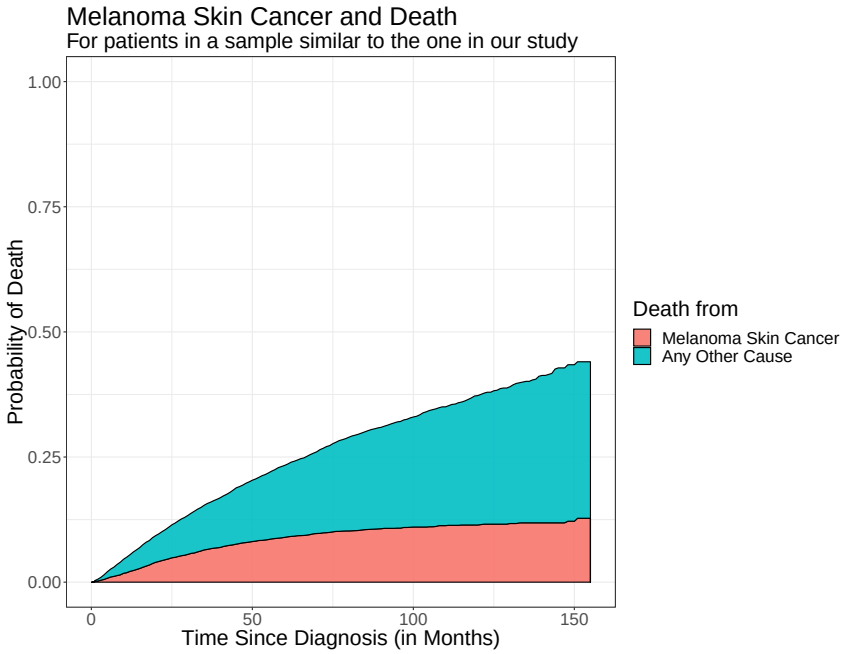

Figure 17. Melanoma Skin Cancer and Death, viewed through the lens of competing risks.

4.3 Future Research

In this section, we briefly describe some ideas for some possible future research related to this project.

4.3.1 Geographic Location and Survival Outcomes

As noted in the introduction, our decision to focus on patients from Louisiana was partially based on the idea that survival outcomes may be influenced by geographic location. This hypothesis may be investigated in the future.

4.3.2 Competing Risk Analysis

The plots provided in Section 4.2 point towards the idea that a disease such as pancreatic cancer is much more lethal than a disease such as melanoma skin cancer. We hypothesize that competing risk analysis can be used to somehow quantify the extent to which a particular disease, such as pancreatic cancer, is "more fatal" than a disease such as melanoma skin cancer. For example, if a patient diagnosed with pancreatic cancer died during the fist 10 months of diagnosis, then it was very probable that their cause of death was pancreatic cancer. On the other hand, if a patient diagnosed with melanoma skin cancer died within the first 10 months of diagnosis, then they very well may have died from some disease other than melanoma skin cancer. We hope to be able to articulate these ideas and observations more precisely with future research.

4.3.3 The Relationship Between Surgery, Tumor Location, and Survival Outcomes

As we noted in section 3.1.1, there were more patients in our sample with tumors located in the head of the pancreas than there were patients with tumors located in the body or tail of the pancreas. This was expected based on statements made by the American Cancer Society. In particular, they say that most curative surgery procedures are performed when the tumor is in the head of the pancreas [5]. In the future, we would like to explore the relationship between surgery, tumor location, and survival outcomes. For example, by examining some data obtained from SEER, it was clear that patients who received surgery had outcomes that were better than patients who did not. In other words, it appears though surgery influences survival outcomes. Meanwhile, it appears as though tumor location plays a role in determining whether or not a patient will have surgery. Meanwhile, the purpose of the present surgery was to determine the extent to which tumor location influences survival outcomes.

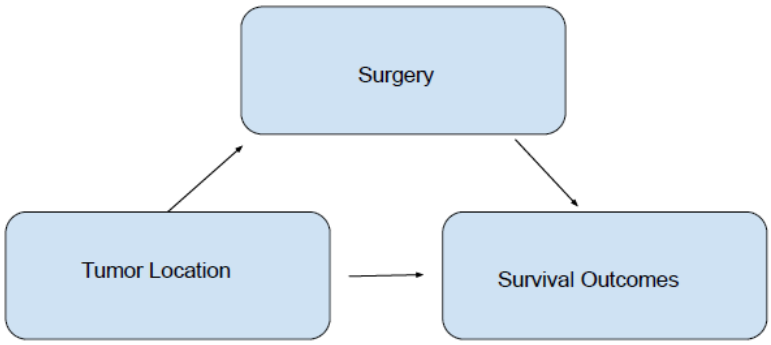

Figure 18. On the Relationship Between Surgery, Tumor Location, and Survival Outcomes.

In the future, we would like to obtain a better understanding of how these variables are related.

4.4 Conclusion

When no other variables were considered, we found that the survival outcomes for patients with cancer in the head of the pancreas were worse than the survival outcomes patients who had cancer in the body or tail of the pancreas. In this sense, our results ran opposite to the ones obtained by Artinyan et. al. However, their study was based on data collected between 1998 and 2004, while our study was based on data collected between 2004 and 2016. This suggests that survival outcomes may be influenced by year of diagnosis. This claim appears to be supported by Model 2, as mentioned

above. This aligns with the idea that the quality of clinical treatment (e.g., surgical procedures and medicine) may be improving over time. Also, Artinyan's result used data from all of the registries in the SEER database, while we only used data from the Louisiana registry.

Finally, our results indicate that the impact of tumor location disappears when variables such as tumor grade and cancer stage are considered.

References

[1] *American Cancer Society* website. *Key Statistics for Pancreatic Cancer.* Last Revised: February 11, 2019. Direct link:
https://www.cancer.org/cancer/pancreatic-cancer/about/key-statistics.html

[2] *Pancreatic Cancer - Cancer Stat Facts* website. Direct Link:
https://seer.cancer.gov/statfacts/html/pancreas.html

[3] Griffin J, Prouk K, Wolfgang C. *Pancreatic cancer surgery: past, present, and future.* Chin J Cancer Res. 2015 Aug; 27(4): 332-348.

[4] *The Sol Goldman Pancreatic Cancer Research Center* website. Direct Link:
https://pathology.jhu.edu/pancreas/TreatmentSurgery.php?area=tr

[5] *American Cancer Society* website. *Surgery for Pancreatic Cancer.* Last Revised: February 11, 2019. Direct Link:
https://www.cancer.org/cancer/pancreatic-cancer/treating/surgery.html

[6] Artinyan A, Soriano P, Prendergast C, Low T, Ellenhorn J, Kim J. A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. *The anatomic location of pancreatic cancer is a prognostic factor for survival.* HPB, 2008; 10: 371-376.

[7] *Surveillance, Epidemiology, and End Results (SEER) Program Research Data (1973-2014).* National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch.

[8] Wickham H, Averick M, Bryan J, Chang W, McGowan L, et. al. *Welcome to the Tidyverse.* Direct Link:
https://tidyverse.tidyverse.org/articles/paper.html

[9] *Collaborative Stage Data Collection System* website. Direct Link:
https://cancerstaging.org/cstage/schema/Pages/version0205.aspx

[10] Wasif N, Ko Cl, Farrell J, Wainberg Z, Hines O, Reber H, Tomlinson J. *Impact of Tumor Grade on Prognosis in Pancreatic Cancer: Should We Include Grade in AJCC Staging?* Annals of Surgical Oncology. 2010 Sep; 17(9): 2312-2320

[11] Moore D. *Applied Survival Analysis Using R.* Springer, 2016.

[12] Cleves M, Gutierrez R, Gould W, Marchenko Y. *An Introduction to Survival Analysis Using Stata.* Second Edition. Stata Press, 2002.

[13] Klein J, Moeschberger M. *Survival Analysis: Techniques for Censored and Truncated Data.* Springer, 1997.

[14] Kleinbaum D, Klein M. *Survival Analysis: A Self-Learning Text.* Second Edition. Springer, 2005.