12-2020

SURVIVAL ANALYSIS OF COLORECTAL CANCER PATIENTS WITH LIVER METASTASIS

BRANDON O'GRADY

SURVIVAL ANALYSIS OF COLORECTAL CANCER PATIENTS WITH LIVER METASTASIS

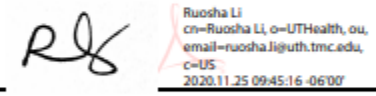
by

BRANDON O'GRADY, BS

APPROVED:

11/24/2020

ASHRAF YASEEN, PHD



RUOSHA LI, PHD

2020-11-29

MICHAEL CANNELL, PHD MPH

SURVIAL ANALYSIS OF COLORECTAL CANCER PATIENTS WITH LIVER METASTASIS

by

BRANDON O'GRADY
BS, Texas A&M University, 2017

Presented to the Faculty of The University of Texas

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH
Houston, Texas
December 2020

ACKNOWLEDGEMENTS

I first like to thank my parents who have supported me my entire life. Without them, none of my life goals could have been accomplished. I would like to thank all of my professors at the University of Texas School of Public Health for giving me the proper education and all the guidance they have given me through the years. I personally want to thank Dr. Yun Shin Chun for given me the opportunity to work with her and help me create the dataset for this project. I also personally want to thank Dr. Ruosha Li for helping me through the entire project. None of this would have been possible without the help of Dr. Chun and Dr. Li. I would like to thank Dr. Ashraf Yaseen for his excellent guidance as my academic advisor throughout the years.

SURVIAL ANALYSIS OF COLORECTAL CANCER PATIENTS WITH LIVER METASTASIS

Brandon O'Grady, BS, MS
The University of Texas
School of Public Health, 2020

Thesis Chair: Ruosha Li, PHD

Background- Colorectal cancer is the third most common cancer in the world. I investigated the survival rates among colorectal cancer patients diagnosed with hepatic metastasis to see if any variables are associated colorectal risk and survival.

Methods- Patients were diagnosed from 2000-2019 and collected through MD Anderson's database. A descriptive analysis, univariate analysis, Kaplan-Meier with Mantel log-rank test, Cox proportion hazard regression and a Stratified Cox Model was performed to investigate death. A competing risk regression was implemented to investigate liver recurrence.

Results- There was a clear difference in the survival outcome between liver surgery patients and non-liver surgery patients with a 99.1% two-year survival rate for the surgery group and a 47.1% two-year survival rate for the non-liver surgery group. Though the survival rate is higher for the 220 liver surgery patients, liver recurrence did occur out of 161 patients and 36 of them has died by end of follow-up. Age of liver diagnosis, extrahepatic metastasis, size path, synchronous, right colon primary, bilateral metastasis, and the number of liver

metastasis were significantly associated with worse survival. Liver surgery, primary surgery, and BMI were significantly significant with a greater overall survival in univariate analyses. The results competing risk regression showed that the log number of liver metastasis (SHR: 1.30, 95% CI: 1.07-1.35) and node positive (SHR: 1.42, 95% CI: 1.08-1.86) were significantly associated with a poorer result for liver recurrence.

Conclusions- Resection of the liver and primary cancer is an optimal way to treat patient with colorectal cancer with colorectal cancer with liver metastasis. It is important to note that even if a patient elects to go through surgery, there is a strong chance that recurrence will happen. There is also a significant difference in the survival outcome between the patients who diagnosed with adenocarcinoma of the right colon.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

BACKGROUND

Literature Review

Colorectal cancer is a cancer that starts in the colon or the rectum. The different parts of the colon include the cecum, ascending, transverse, descending, and the sigmoid colon. The large intestine is an important organ that of the digestive system. The rectum's job is to receive stool from the colon and transport it to the anus. The colon's job is to absorb water and salt from the remaining food from the small intestine and passes the remains into the rectum (American Cancer Society, n.d.). According to the World Cancer Research Fund, colorectal cancer is the second most common tumor among women and third most common tumor among men making it the third most common cancer worldwide (2019). In the United States it is the second leading cause of cancer death in women and the third for men, and it is estimated that about 1 in 23 women and 1 in 21 men in the United States will be diagnosed during their lifetime (Jemal et al., 2011). The liver is the most common site for metastasis to occur from the colorectal cancer (Kow, C. Wei, A., 2019).

Adenocarcinoma

There are different types of cancer in the colon and rectum. The most common type of cancer is adenocarcinomas, and it makes up about 96% of colorectal cancers (American Cancer Society). The dataset in this thesis work will only include patients that have been diagnosed with adenocarcinomas. According to the National Cancer Institute (NIH), adenocarcinoma is a cancer that starts in the glandular cells that are found in tissue that lines the internal organs that makes and releases substances in the body.

Symptoms

There are multiple symptoms that colorectal cancer can cause. Unexplained weight loss, blood in stool, rectal bleeding, abdomen pain, lump in the abdomen, and changes in bowel habits are all common symptoms of colorectal cancer (Brazer, 2018). Doctor visit is recommended if any of these symptoms continue for four weeks or more.

Risk factors

It is still unknown what exactly causes a person to develop colorectal cancer, but many risk factors are strongly linked to the disease (Cancer Treatment Centers of American, 2020). Older age may increase the risk of colorectal cancer. The majority of the cases are older than 50, and the rates of colon cancer in people younger than 50 are increasing (Mayo Clinic, 2019). Around 9 out of 10 people diagnosed with colorectal cancer are aged 50 or older (Columbia University Herbert Irving Comprehensive Cancer Center, 2019). Inherited syndromes, gene mutations, are another factor that can increase colon cancer risk and family history of colon cancer increases your risk. About 5 to 10 percent of people with inherited gene mutations develop colorectal cancer and four out of ten mutations are in the KRAS gene (Dinu et al., 2014). Lifestyle is a risk factor as well. Obesity, smoking, diet, Type II diabetes, lack of physical activity, and high alcohol consumptions are major risk factors (Brazer, 2018). The African American race and Ashkenazi Jews has the greatest risk of colon cancer than other races do (Columbia University, 2019).

Treatment

Chemotherapy

Treatment can depend on multiple factors. The most common types of treatment options include radiation, chemotherapy, and surgery. Systemic chemotherapy drugs are injected into a vein or by mouth to kill the cancer cells (Kow, C. Wei, A., 2019). This type of treatment is usually given for cancer that has metastasis. Neoadjuvant therapy is a chemotherapy treatment given before surgery to help shrink the cancer before surgery and adjuvant therapy is chemotherapy given after the surgery to help the can cancer from reoccurring (Columbia University, 2019). The chemotherapy drugs that are commonly used in in combination to treat: 5-Fluorouracil (5-FU), Capecitabine (Xeloda), Irinotecan, Oxaliplatin (Columbia University, 2019). These drugs are also used in combination with VEGF targeted drugs, Bevacizumab (Avastin), and EGFR targeted drugs, Cetuximab (Erbitux) and panitumumab (Vectibix), to help combat more advanced cancers. The targeted therapies drugs are given by an infusion and is given one to three weeks at a time depending on the doctor's treatment plan (Columbia University. 2019).

Radiation

Radiation is a unique treatment that targets the tumor while reducing the radiation around the healthy tissues. The treatment can be daily, weekly, every other day, or one single treatment. The radiation can be given by an external beam radiation, 3D Conformal radiotherapy, and Intensity modulated radiation therapy (Columbia University. 2019).

Surgery

Colectomy, laparoscopic, and liver resection are all forms of surgery that combat colon cancer. Colectomy, or colon resection, is the most popular form of surgery when a person develops colorectal cancer (Columbia University, 2019). It is a procedure were the

surgeon removes the part of the colon or rectum that is infected with the cancer and can remove up to 12 inches of the organ.

A laparoscopic surgery is the standard care for most of the colorectal surgeries. This is surgery is less invasive and can be conducted with small abdominal incisions (Columbia University 2019).

Liver resection, or hepatectomy, is a surgery that removes all or a part of your liver. The liver is divided into two main parts that can be further divided into eight segments. The liver function is to metabolize drugs and toxins, removes ammonia and bilirubin from the blood, and synthesis proteins and enzymes (Hyperarts). When surgery of the liver is required, the most common cancer of the liver that is removed is malignant neoplasms that arise from the colorectal metastasis (Hyperarts). According to the American Cancer Society, the best option to cure liver cancer is to perform a hepatectomy. One can live with thirty percent of liver resected and the liver will grow back to full size over time (Christiano, 2018).

Stages

There are 5 different stages of cancer from stage 0 to stage 4. Stage 0 is the earliest stage of cancer. This is when the can is still within the inner layer of the colon or rectum. Stage 1 is where the cancer has grown through the inner layer of the colon or rectum but has not spread beyond the wall of the organ. Stage 2 is where the cancer has spread through the wall but has not reach the lymph nodes. Stage 3 is where the cancer has invaded the lymph nodes but has yet to spread to the other parts of the body. Stage 4 is latest stage of cancer and it is where the cancer has spread to the other parts of the body (Brazer, 2018).

Public Health Significance

Stage IV cancer features a low life expectancy and investigating the significant impact on public health could help the overall quality of life. According to the American Cancer Society, the 5-year relative survival rate for patients with stage 4 colon cancer that has metastasis is 14%. Colorectal cancer is a major public health problem, and anyone could agree with that.

The cost for one year of treatment for a patient with late stage cancer in the colon are as high as $310,000 with an annual cost nationwide of $14 billion (Karen, 2015). It is important to research and implement the most effective treatment to help minimize cost. The quality of life takes a toll at the population level through the economic burden by the cost of treatment and the long-term effects of cancer (Cancer Net, 2020).

The American Society of Clinical Oncology (ASCO) provides a resource-stratified guideline to provide expert suggestion to clinicians for treatment options through different scenarios. The ASCO strongly suggests that upfront surgery of hepatic metastases and moderately suggests selective internal radiation therapy with systemic chemotherapy (Chiorean, et al., 2020). Even though this is the ASCO suggestion, they concluded that there were some limitations to liver metastases-directed therapies. This is why continuous research is important to improve the quality of life in cancer patients.

Stage IV colorectal cancer survival rates are low and finding any patterns to increase the overall quality of life is worth it. With multiple risk factors, treatments, and different stages of cancer; investigating the treatment outcomes will help determine the best treatment options to increase the best overall quality of life.

Specific Aims

Aim 1. Conduct descriptive analyses and K-sample comparisons to examine the patient characteristics by subgroups. The subgroups are defined by the primary cancer site, gender and whether the patient had liver surgery or not.

Aim 2. Conduct univariate analysis and K-sample comparisons to investigate death and liver recurrence, overall and by whether the patient received liver surgery.

Aim 3. Conduct a semi-parametric multivariable regression model to detect prognostic factors of liver recurrence and death. A refined Cox model (e.g., stratified proportional hazards model) will be conducted if the proportional hazards assumption is violated.

METHODS

Table 1: List of Variables

Variables	Description
Dependent: Surv(time, event)	The dependent variable that will be created in R that will test a survival object. Time is for right censor data and event is the status indicator.
Dependent: Surv(Futime, Status)	The dependent variable that will be created to test survival analysis on survival outcome on the entire data set.
Dependent: Surv(LiverFutime, LiverRecur)	The dependent variable that will be created to test survival analysis on liver recurrence for liver resection patients
Futime	number of days between liver diagnosis and the earlier of death and censoring
Liver Futime	number of days between liver surgery and the date of liver recurrence
Follow up	Patient's most recent follow-up date
DxAge	Age of diagnosis of colorectal cancer
DxAgeLiver	Age of diagnosis of liver mets
Gender	Male/Female
EhMets	Mets area outside the colorectal, lymph nodes, and liver region during the liver resection.
DateDx	The date the patient was diagnosis with colorectal cancer
DateDxLiver	The date the patient was diagnosis with liver metastasis.
MSI	If the patient had microsatellite instability
KRAS	If the patient had a KRAS gene Mutation
BMI	Weight is collected in grams and height collected in centimeters to calculate Body Mass Index
Synchronous	Disease-free interval <6 months from diagnosis of primary tumor to discovery of liver metastasis
Status	Patient is alive or not
Primary	If the primary tumor is in the colon or rectum
Primary tumor	Specific site of the primary tumor

Primary Surgery	If the patient had resection on the primary tumor
Liver Surgery	If the patient had liver resection
Liver surgery date	Date of liver surgery
Date liver Recur	Date of liver recurrence
Bilateral	If the patent had mets in both lobes of the liver
Size path	Size of largest tumor in liver in centimeters
NoMets	Number of mets in the liver
Node Positive	If the cancer has spread to the lymph nodes
HTN	If the patient had hypertension
Diabetes	If the patient had diabetes
LiverRecur	If liver recurrence happens after treatment/surgery

Study Subjects

There will be 470 patients from MD Anderson, with 220 that had liver surgery and 250 that did not have liver resection. The data is entered by me, doublechecked by Dr. Chun, M.D. and is under the oversight of an Institutional Review Board at MD Anderson. The variables are extracted from MD Anderson's electronic medical record and stored in Excel and REDCap. The variables range from basic demographics (sex, date of birth, bmi, and etc.) to cancer information (number of liver metastasis, size of tumor, location of primary and metastasis). These data will be de-identified. All patients are diagnosed with a primary colorectal tumor with metastasis to the liver (stage IV cancer). Eligible patients must have at least a two-year follow up, unless there is a record that the patient has pass away in that given time frame.

Data Analysis Plan

The study design of my thesis is to use survival analysis approaches to answer questions about the dataset. The analysis will be conducted in R version 3.6.0 (R Foundation

for Statistical Computing, Vienna, Austria). The Type-I error rate will be set at 0.05 for all planned analyses.

The survival (time to event) outcomes are the time in days since liver metastasis diagnosis to the time of death. The event indicator will be coded as 1 if the patient died and 0 otherwise. The other outcome is the time in days since liver metastasis diagnosis to liver recurrence, which is subject to the competing risks censoring by death. We will code the event indicator as 1 if the patient experience liver recurrence, 2 if the patient dies without liver recurrence, and 0 if the patient is censored due to other reasons. The death outcome will be analyzed using classical survival analysis methods in R package survival (Therneau T, 2020), and the recurrence outcome will be analyzed using competing risks methods in R package cmprsk (Gray, 2020).

Descriptive Analysis

Descriptive analysis will be used first to summarize the data and to find any patterns in the data. Patient characteristics will be compared across the subgroups, using the Wilcoxon rank-sum test for continuous variables and the Chi-square test for categorical ones.

Univariate Methods and K-sample Comparisons

Kaplan-Meier curve is an estimate of survival probability at each point in time. This curve is a purely descriptive method, and it helps to obtain quartiles, medians, and 95% confidence limits of time to death. It helps to estimate the population survival curve from the data set and helps compute the fraction surviving at each time. The Kaplan-Meier method will be used to estimate the survival curve of all the 470 patients with the variable “Status” equaling 0 as the censored event. For the liver recurrence outcome, I will estimate the

cumulative incidence probabilities while treating death as a competing event, using the cuminc function in R package cmprsk.

Next, we will plot the estimated Kaplan-Meier and cumulative incidence curves by liver surgery group (Yes/No). The Mantel log-rank test is a comparison of the Kaplan-Meier curve for more than one group. The log-rank test will be performed to compare the survival probabilities between patients that had liver resection and those that did not receive surgery. Similarly, I will apply the Gray's K-sample test to compare the tumor recurrence outcome between the two groups.

Multivariable Regression Methods

Cox proportional hazards regression is the model that is more useful than the non-parametric methods that uses multivariate approaches that controls the covariates. This method is commonly used in medical research for investigating the association between the survival time of patients and potential risk factors. I will construct a multivariable Cox proportional hazards model to investigate the association of patient survival time and the predictors. A full model will be constructed containing all available variables in their optimal functional form, and an Akaike information criterion (AIC)-based stepwise selection model will be performed to build model with only the key factors. The liver surgery (Yes/No) variable will be forced into the model due to its clinical importance. Results will be summarized by the hazard ratio (HR) and the corresponding 95% confidence intervals.

Next, for the liver recurrence outcome, I will implement competing risks regression method, using the crr function in R package cmprsk. I will summarize the results in terms of the HR of the subdistributional hazard and 95% confidence intervals.

Model Checking and Refinements

After testing for proportional hazards, I will stratify the model for any variables with non-proportional hazards. I will verify that all hazards were proportional in the final model. The final model will be evaluated by several diagnostic methods. A deviance residual plot could be used to identify outliers in the data and a plot of transformed score residuals could also be used to check for influential points on the plot.

Human Subjects, Animal Subjects, or Safety Considerations

This thesis project will use de-identified, existing data from the project approved by the IRBs at Texas MD Anderson Cancer Center and the University of Texas Health Science Center at Houston. My thesis project has been determined to qualify for exempt status 45 CFR 46.101(b).

RESULTS

Aim 1

Table 2 shows the results for the descriptive analysis between patients that had liver resection vs. the patients that did not have liver resection. Interesting findings is that the median follow up time in days is 2228 for the surgical group compared to 685 days for the non-surgical group. The baseline number of liver metastasis is increased in just the non-surgical group. The median is 6 liver metastases for the non-surgical group and 2 metastases for the surgical group. The median for the biggest size path for the non-surgical group was larger at 4.75 centimeters compared to 2 centimeters for the liver surgery groups. These variables could indicate why a patient was able to receive surgery or not.

Table 2: Descriptive Statistics of Liver Surgery Groups

Characteristic	Liver Surgery Group	Non-Surgical Group
Age of Liver Diagnosis		
Mean	53.2	54.86
Median	54	55
Standard Deviation	11.06	11.7
Min-Max	22-78	26-82
Survival Time (Days)		
Mean	2260	774.36
Median	2228	685
Standard Deviation	886.25	450.09
Min-Max	763-5142	76-3589
Liver Recurrence Time (Days)		
Mean	119.83	485
Median	873.50	329
Standard Deviation	1082.8	436.58
Min-Max	35-4901	56-3589
BMI		
Mean	29.22	28.17
Median	28.17	27.20
Standard Deviation	6.06	6.12

Min-Max	18.18-47.97	15.28-55.15
Size Path (cm)		
Mean	2.83	5.49
Median	2.00	4.75
Standard Deviation	2.26	3.94
Min-Max	0.3-15.0	0.5-20.30
Number of Mets		
Mean	3.35	11.36
Median	2	6
Standard Deviation	3.35	12.9
Min-Max	1-21	1-80

After running the descriptive statistics, I ran Wilcoxon rank-sum test, Pearson Correlation test, and Chi-square test to see if there are any clear patterns in the dataset. I ran a Wilcoxon rank-sum test by the liver surgery groups against the continuous variables and I rejected the null hypothesis against BMI, size path, and the number of metastasis. I also ran the Wilcoxon rank-sum test by gender and the only variables that were significant was the age at diagnosis of colorectal cancer and age at diagnosis of liver metastasis. The median age for women being diagnosis with colorectal cancer is 52 and the median age for men is 55.5. The rank-sum test was used on the primary site of the cancer (colon/rectum). When running the analysis, the test was significant with size of the liver metastasis and number of metastasis.

The Pearson Correlation test was conducted to correlate the continuous variables in the dataset. The only two variables that had a significant correlation was size path and number of metastasis in the liver. It had a p-value of <0.001 and a weak positive correlation of 0.25.

The Chi-square test was performed to compared categorical variables by subgroup to see if there is an association between them. After running the test between the liver surgery groups against gender, KRAS, synchronous, hypertension, diabetes, bilateral metastasis, and node positive, the only variables that tested significantly at a 0.05 significance level were the presence of synchronous and bilateral metastasis. There were 260 patients that had bilateral metastasis only 85 had surgery. The Chi-square test was performed again with the primary tumor site (right colon, left colon, rectum) and the variables that tested <.05 were gender, KRAS mutation, and synchronous. It is important to note that 78 out of 234 (33.3%) had a mutation in the left colon, 53 out of 106 (50%) had a mutation in the Rectum, and 88 out of 130 (66.7%) had a mutation in the right colon.

Aim 2

Aim 2 consists of running a univariate analysis, Kaplan-Meier curves with Mantel Log-Rank Tests, and Gray's K-sample test to compare the survival and tumor recurrence outcome. Table 3 shows the results of the univariate Cox regression analysis for the survival outcome. Gender, KRAS mutation, Microsatellite instability, hypertension, diabetes, primary tumor between colon and rectum, and node positive are not associated with mortality at a 0.05 significance level. An older age of liver diagnosis was associated with a 2% increase in hazard of mortality per year. Liver surgery was associated with an 86% decrease in hazard of

mortality compared to patients that did not have any type of liver resection. The presence of extrahepatic metastasis was associated with an 87% increase in hazard of mortality per year compared to the absence of extrahepatic metastasis. An increase in BMI was associated with a 3% decrease in hazard of mortality per year. The increase the largest size of liver metastasis measured in centimeters was associated with an 11% increase in hazard of mortality per year. The presence of synchronous was associated with a 74% increase of hazard of mortality compared to the absence of synchronous. Compared to left colon, primary tumor in the rectum was not associated with mortality but the presence of a primary tumor in the right colon was associated with a 37% increase in hazard of mortality relative to left colon (Figure 1). Primary surgery was associated with a 72% decrease in hazard of mortality compared to patients that did not have colorectal surgery. The presence of bilateral metastasis was associated with a 67% increase in hazard of mortality compared to the absence of bilateral of bilateral metastasis. An increase in the number of liver metastasis was associated with a 4% increase in hazard of mortality.

Table 3: Univariate analysis of each variable.

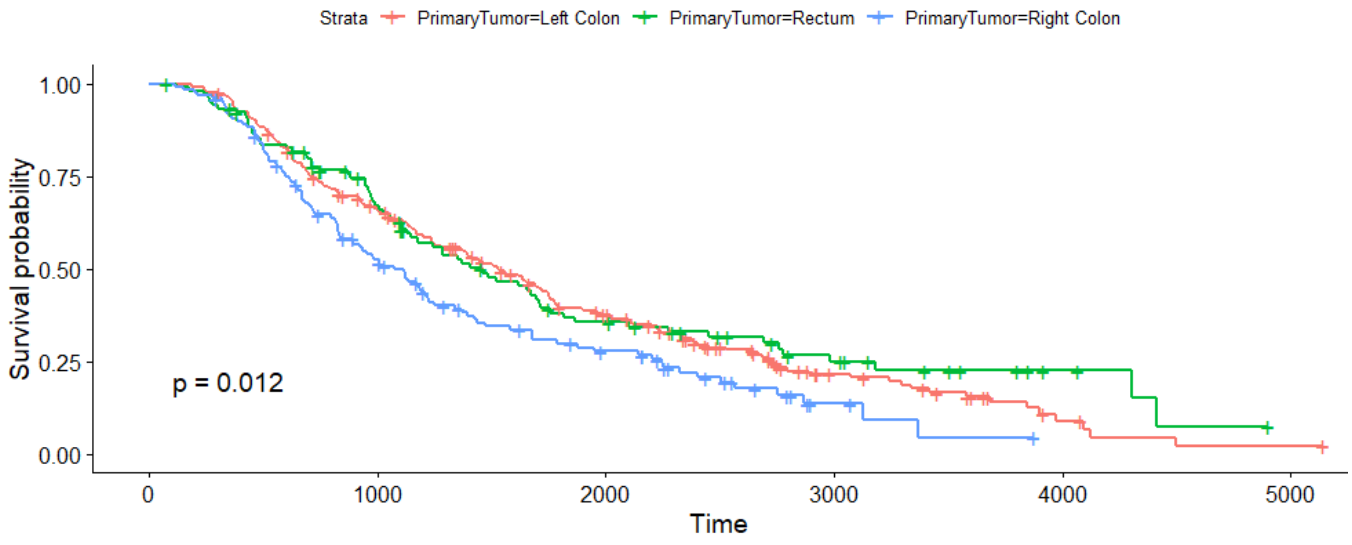
Dependent: Surv(Futime, Status)		all	Hazard Ratio (95% CI, p-value)
Gender	Female	196 (100.0)	-
	Male	274 (100.0)	1.07 (0.87-1.33, p=0.534)

Age of Liver Diagnosis	Mean (SD)	54.1 (11.4)	1.02 (1.00-1.02, p=0.003)
Liver Surgery	No	250 (100.0)	-
	Yes	220 (100.0)	0.14 (0.11-.18, p<0.001)
Extrahepatic Mets	No	332 (100.0)	-
	Yes	138 (100.0)	1.87 (1.49-2.35, p<0.001)
KRAS Mutant	No	251 (100.0)	-
	Yes	219 (100.0)	1.23 (1.00-1.52, p=0.055)
BMI	Mean (SD)	28.6 (6.1)	0.97 (0.95-0.99, p=0.002)
Size Path	Mean (SD)	4.3 (3.5)	1.11 (1.08-1.14, p<0.001)
Microsatellite Instability	No	345 (100.0)	-
	Yes	46 (100.0)	1.16 (0.82-1.65, p=0.42)
	N/A	79	

Synchronous	No	79 (100)	
	Yes	392 (100)	1.74 (1.30-2.33, p<0.001)
Primary	Colon	352 (100)	-
	Rectum	118 (100)	0.78 (0.61-1.01, p=0.057)
Primary Tumor	Left Colon	234 (100)	
	Rectum	106 (100)	0.90 (0.68-1.90, p=0.463)
	Right Colon	130 (100)	1.37 (1.07-1.76, p=0.012)
Primary Surgery	No	161 (100)	
	Yes	309 (100)	0.28 (0.23-0.36, p<0.001)
HTN	No	255 (100)	
	Yes	215 (100)	0.91 (0.73-1.12, p=0.353)
Diabetes	No	399 (100)	

	Yes	71 (100)	1.22 (0.91-1.64, p=0.192)
Bilateral	No	210 (100)	
	Yes	260 (100)	1.67 (1.34-2.06, p<0.001)
Number of Liver Mets	Mean (SD)	7.6 (10.4)	1.04 (1.03-1.05, p<0.001)
Node Positive	No	113 (100)	
	Yes	357 (100)	1.20 (0.93-1.53, p=.161)

Figure 1: Kaplan-Meier Curve for the Survival Outcome by Primary Cancer Site



The second part of Aim 2 was the creation of Kaplan-Meier curves with Mantel Log-Rank Tests. Figure 1 shows that the primary cancer in the right colon was statistically different compared the left colon. Figure 2 is the Kaplan-Meier curve of the entire dataset and Figure 3 is Kaplan-Meier curves for the two surgery groups with the Mantel Log-Rank test <.0001 indicating that there is a clear different between the surgical groups. The analysis showed that 99.1% two-year survival rate (730 days) for the liver surgery group and 47.1% two-year survival rate for the non-liver surgery group. Figure 4 is the plot of the cumulative incidence curve. A Kaplan-Meier curve was also conducted between patients who had bilateral metastasis and those who did not (Figure 5). The Log-Rank test indicates that the two groups are statistically different. The analysis showed an estimated 64.5% two-year survival rate for patients that had bilateral metastasis and an estimated 81.4% two-year survival rate for patients that did not have bilateral metastasis. Extrahepatic metastasis, primary surgery, and synchronous were also significant when running the Mantel Log-Rank Test (Figure 6,7,8).

Figure 2: Kaplan-Meier Curve for the Overall Survival

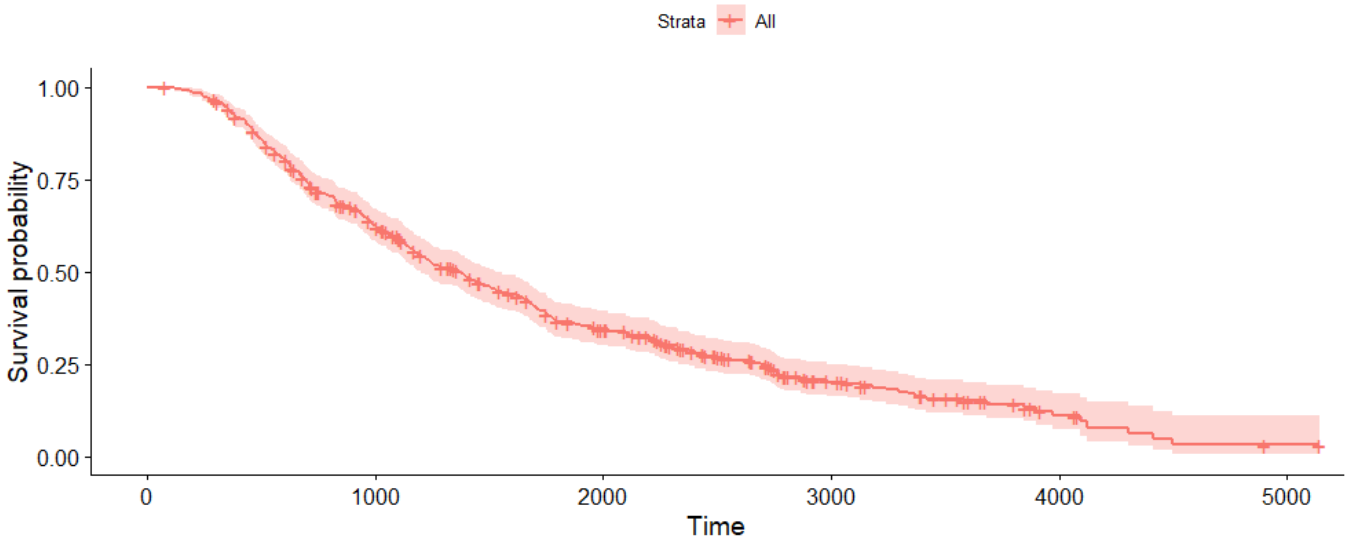


Figure 3: Kaplan-Meier Curve for the Survival Outcome by Liver Surgery Groups

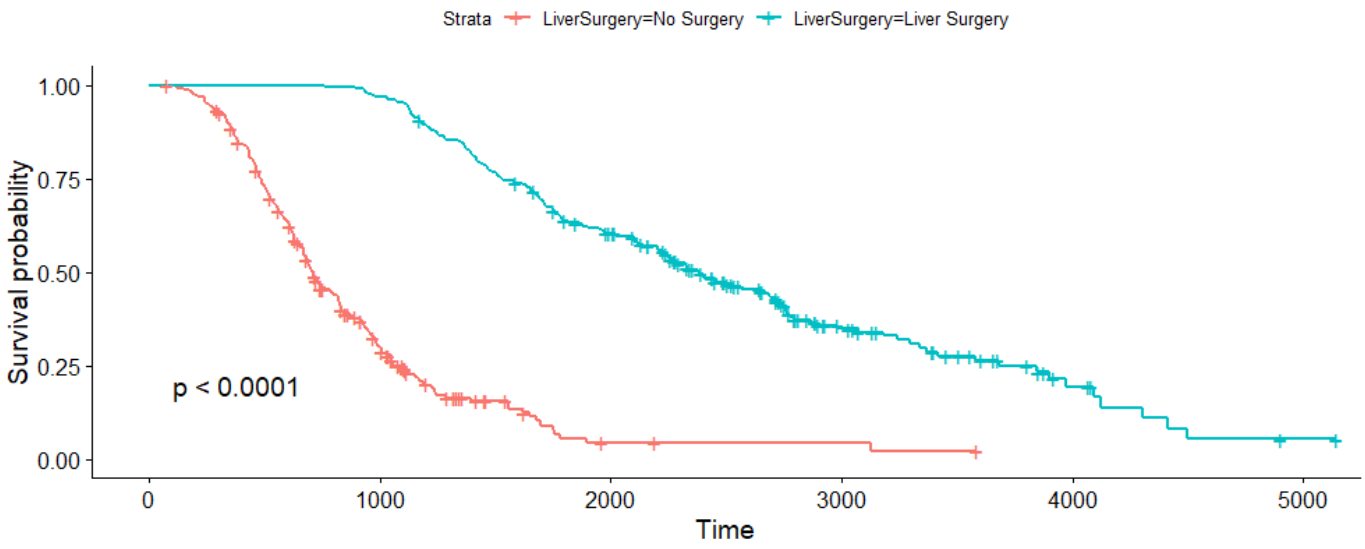


Figure 4: Cumulative Incidence Curves of Liver Surgery Groups for Death

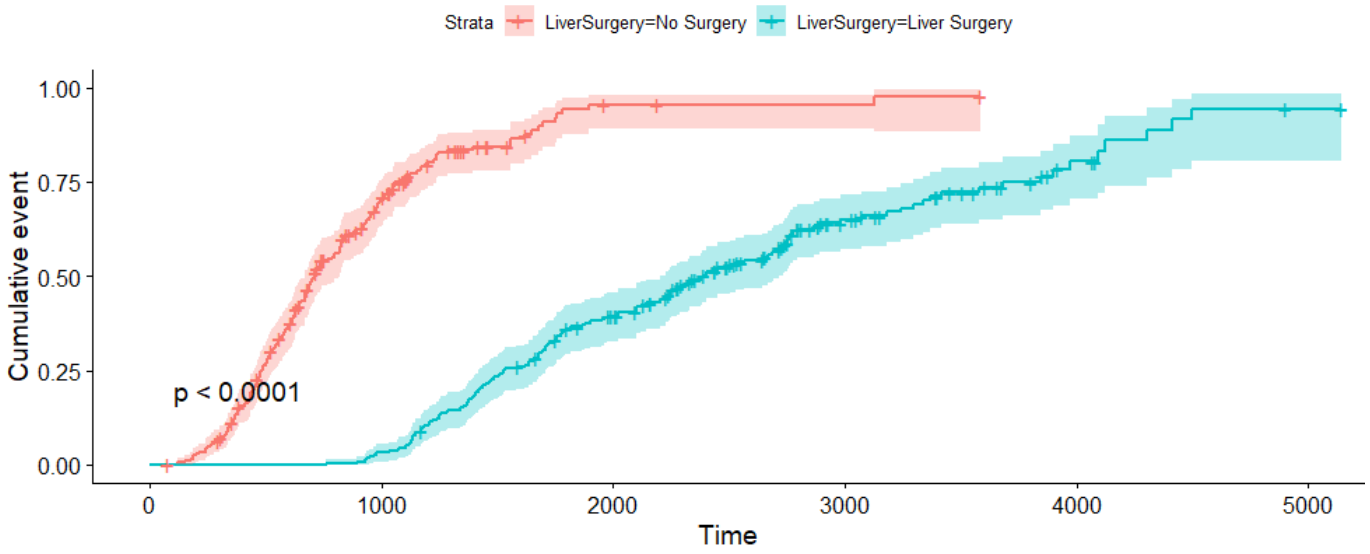


Figure 5: Kaplan-Meier Curve for the Survival Outcome by Bilateral Metastasis

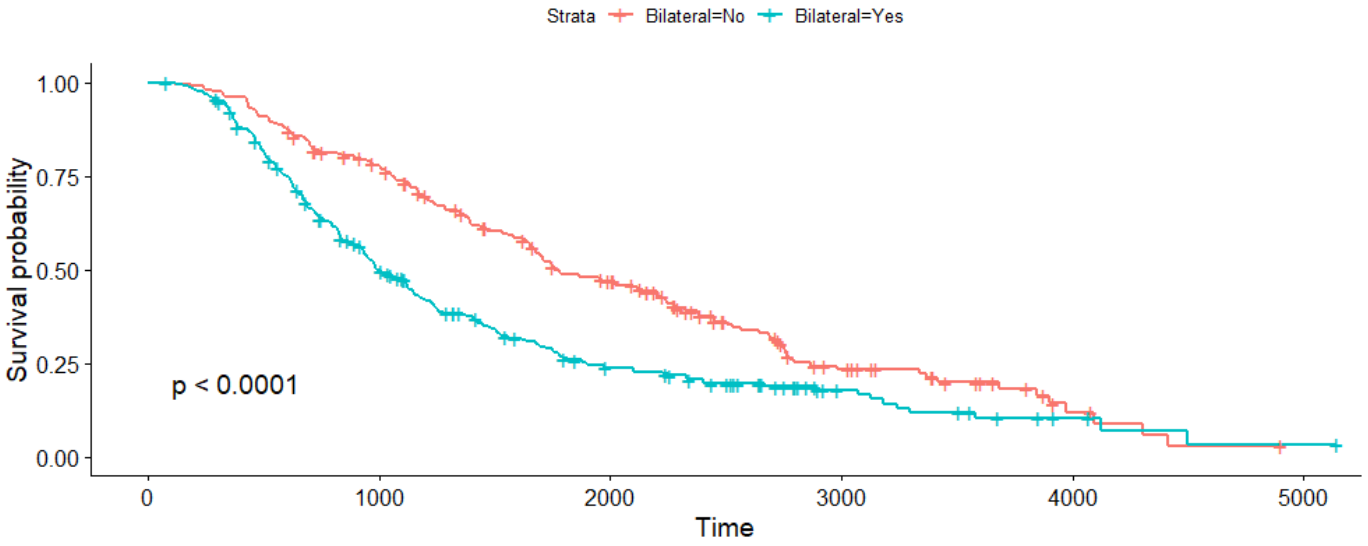


Figure 6: Kaplan-Meier Curve for the Survival Outcome by Extrahepatic Metastasis

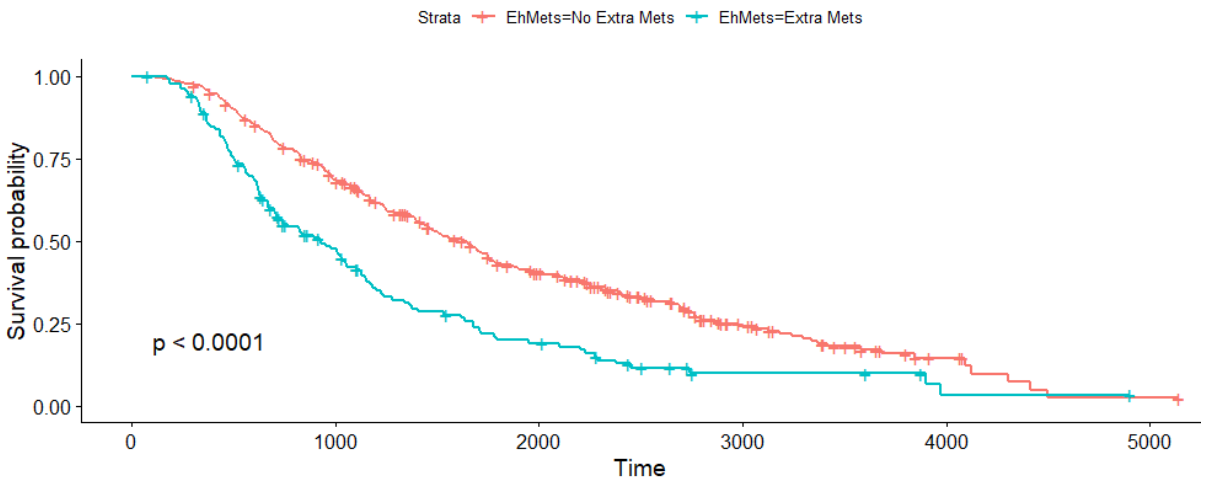


Figure 7: Kaplan-Meier Curve for the Survival Outcome by Primary Surgery

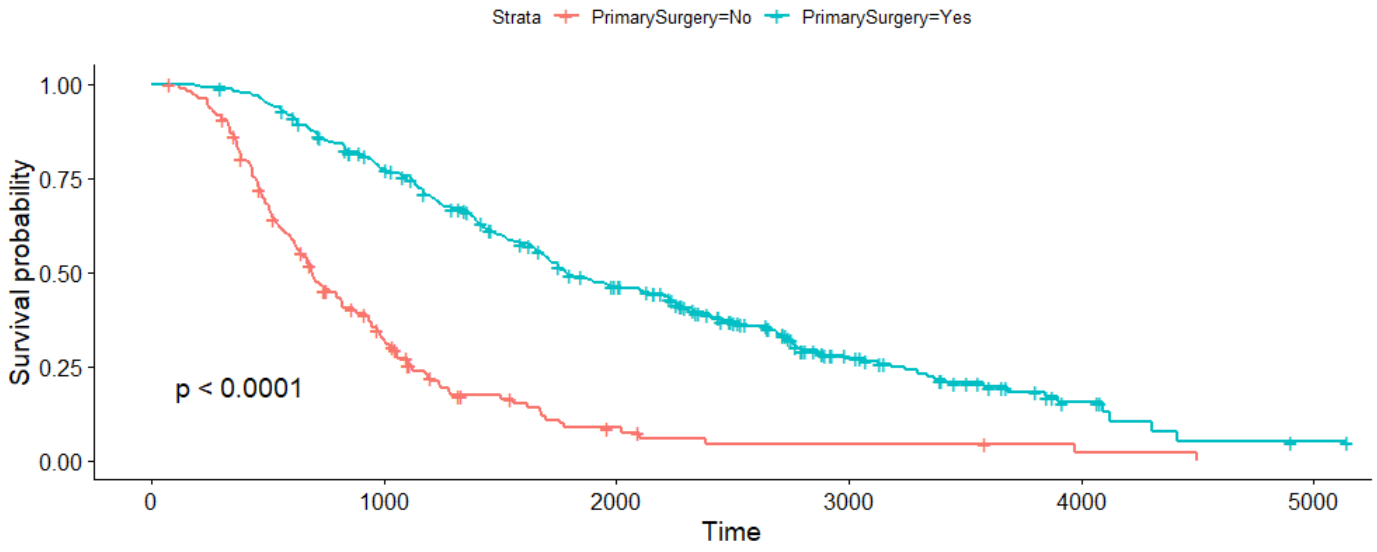
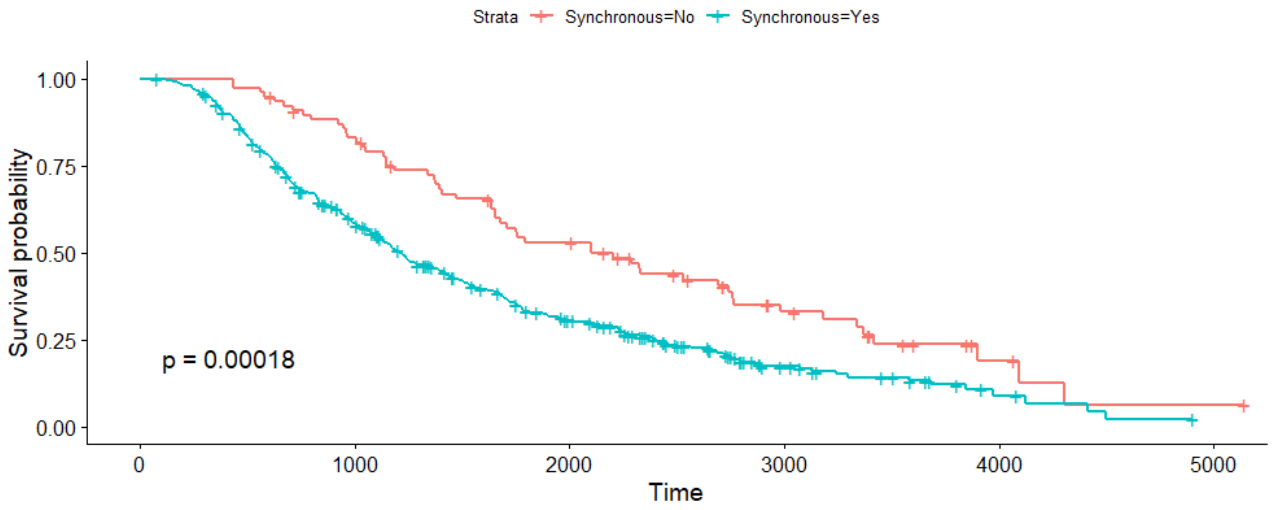


Figure 8: Kaplan-Meier Curve for the Survival Outcome by Synchronous



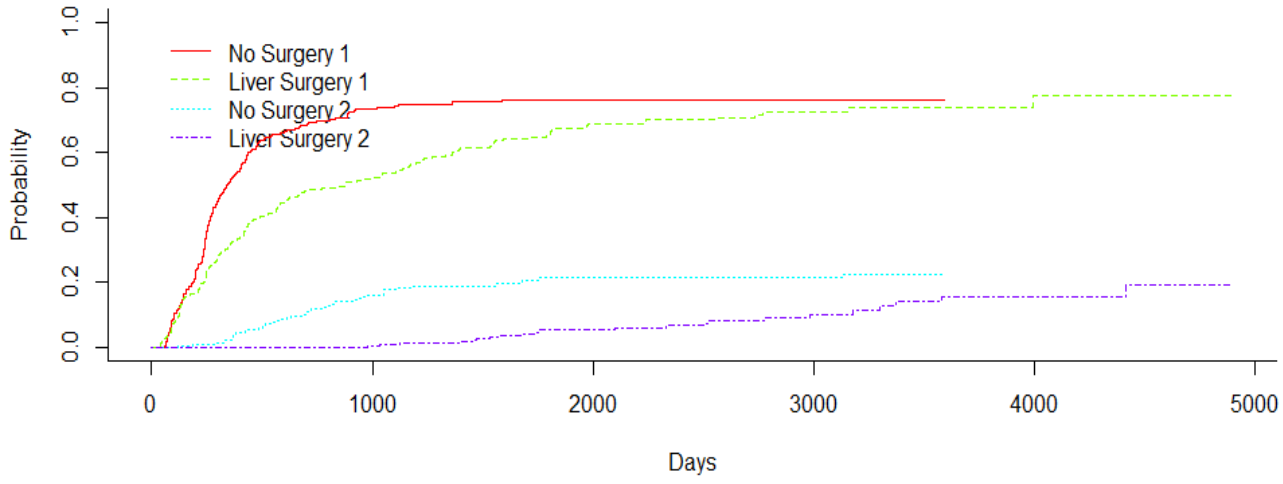
The third part of Aim 2 was to investigate liver recurrence. Shown in Table 4, we can see liver recurrence happens often after treatment. A total of 343 of 470 (73%) patients have had liver recurrence/progression after their treatment, 12% of patients never had recurrence, and 15% of patients died from other cause without having liver recurrence. Running a Gray's test for equality of for the competing risk data to investigate liver recurrence, we see that the liver surgery groups are statistically different for liver recurrence at a p-value of <0.001 and statistically different for death without liver recurrence at a p-value of <0.001 (Figure 9). With liver recurrence being coded as 1 and death without liver recurrence being coded as 2, there was a 73.2% chance that liver recurrence will happen at 1000 days after treatment for the non-liver surgery group and 51.8% chance that liver recurrence will happen at 1000 days after treatment for the liver surgery group. There was a 16.2% chance that death without liver recurrence will happen at 1000 days after treatment for the non-liver surgery group and a

0.45 percent chance that death without liver recurrence will happen at 1000 days after treat for the liver surgery group.

Table 4: Liver Recurrence Outcome

	Non-Liver Surgery	Liver Surgery Group
Censored	18	39
Liver Recurrence	185	158
Death without Liver Recurrence	47	23

Figure 9: Cumulative Incidence Function by Surgery Groups for Liver Recurrence, where 1 Corresponds to Liver Recurrence and 2 Corresponds to Death without Recurrence



Aim 3

The start of Aim 3 is the construction of a multivariable Cox proportional hazard model to investigate the association of patient survival time and the predictors. An AIC-

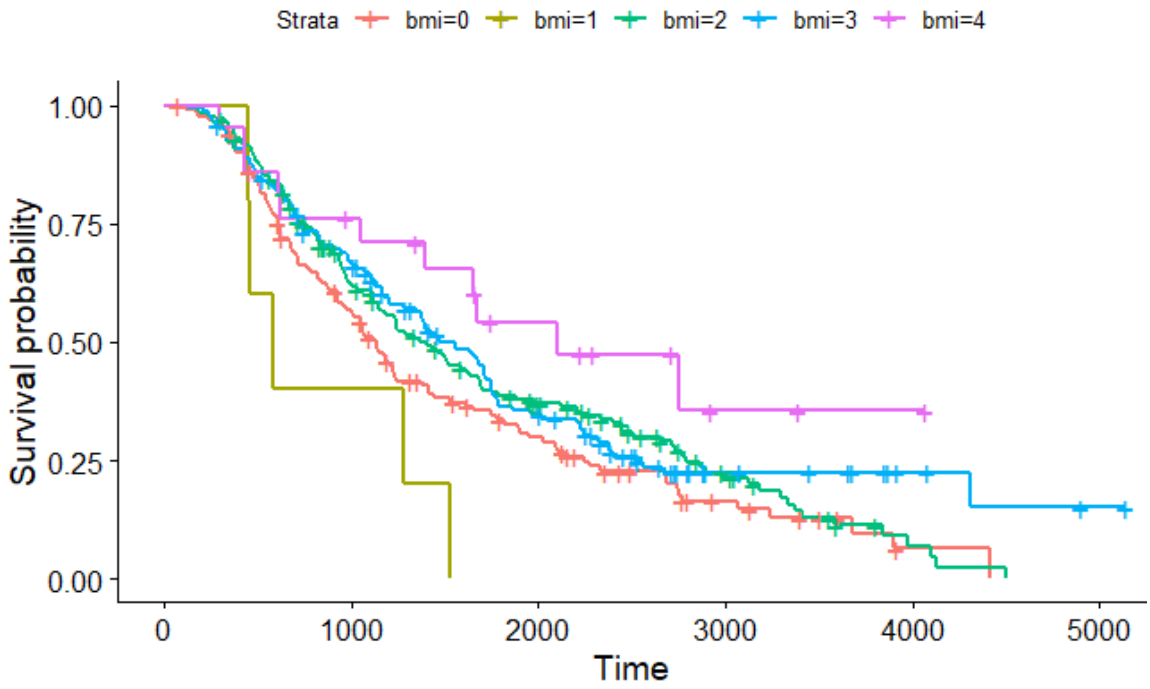
based stepwise selection model was performed with the log-transformation for ‘BMI’, ‘Size Path’, and ‘NoMets’. After building the model, it was found that age of liver diagnosis and liver surgery predictors had non-proportional hazards. With the ‘DxAgeLiver’ being a continuous variable, a new variable called ‘age’ was created to categorize age into three separate groups to help with the stratification. The three age groups where greater than or equal to 65, in between the ages of 50 and 65, and less than the age of 50 coded as 0, 1, 2, respectively. The model was stratified on the variable ‘LiverSurgery’ and ‘age’ that resulted in proportional hazard for the global model. Diagnostic plots showed several outliers with deviance residuals value just over |2| but none over |3|, and dfbeta plot showed one extreme outlier on the variable ‘Bilateral’ over |.04| and several extreme outliers on the log-transformation variable ‘BMI’ over |.04|.

It was noted that the presence of bilateral metastasis was associated with a 67% increase in the hazard of mortality compared to the absence of bilateral metastasis. When included in the final model, it was no longer statistically significant and when controlling for the other covariates I proceeded to drop this variable from the final model.

I decided to split the BMI data into ‘Normal’, Under Weight’, ‘Overweight’, ‘Obese’, and ‘Extremely Obese’ (coded 0-4) with the data being compared to normal weight. After running a univariate analysis, it was concluded that obese and extremely obese patients where statistically different, but there was a 26% decrease in the hazard for obese patients and a 50% decrease in hazard for the extremely obese patients (Figure 10). There were 5 underweight, 131 normal weight, 169 overweight, 144 obese, and 21 extremely obese patients in the dataset. Unexpected weight loss is a symptom of colorectal cancer and it could

be noted that it could be important to help add weight to a patient before they start their treatment. According to Jonathan Korcarnik, weight loss was significantly associated with decreased long-term survivors (Kocarnik et al.). The Center for Disease Control and Prevention has concluded that 40 percent of cancer diagnosis are associated with being overweight and obesity (CDC). The dataset includes 334 patients being overweight or higher, which supports that most of the patients have a larger BMI are diagnosis with cancer. It is important to note that doctors should help regulated the patient's weight throughout treatment and help add weight if the patient is considered underweight. BMI was only collected once right before treatment. I believe the variable should be dropped from the model and observed furthermore in future studies by recording if the patient had weight changes throughout treatment.

Figure 10: Kaplan-Meier Curves for the Survival Outcome by BMI Groups



It was concluded that bilateral metastasis and BMI had complications and should be excluded from the final model. After re-running the model with age and liver surgery still stratified, diagnostic plots (Figure 11, 12) showed several outliers with deviance residuals value over |2|, and few outliers on the variables extrahepatic metastasis, node positive, and KRAS mutant over |.02| and one extreme outlier on node positive over |.03| for dfbeta.

Figure 11: Deviance Residuals

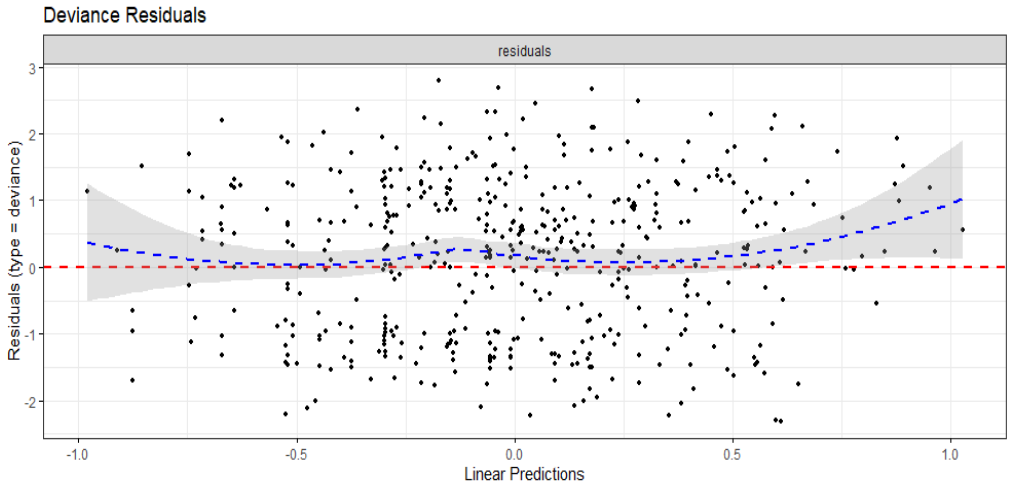
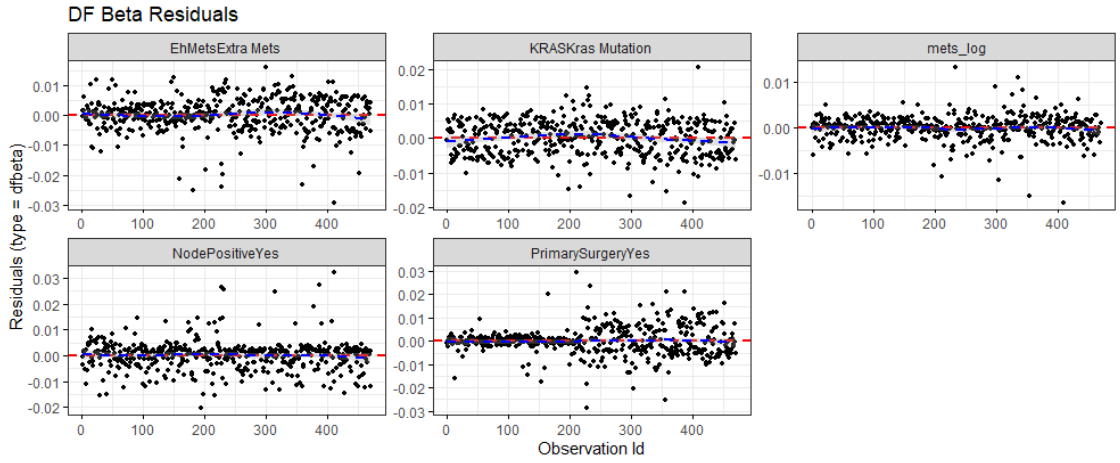


Figure 12: DF Beta Residuals



A global likelihood ratio test, Wald test, and Score test all returned p-values <0.001, indicating the significance of the model and that no more covariates need to be removed or added to the model.

Table 5 details the final model. The variables that were statistically significant associated with survival time were the log transformation of the number of liver metastasis and primary surgery of the colorectal cancer. Controlling for the other covariates, I assess the increase or decrease in hazard for mortality associated with each factor. The hazard was increase by 23.7% for one log increase in liver metastasis. The hazard will decrease by 36.5% for having primary surgery. The hazard will increase by 54.9% for having the presence of extrahepatic metastasis before treatment. The hazard will increase by 27.1% for the presence of having a KRAS mutation and the hazard will increase by 25.3% for the being node positive. Even though 'Node Positive' was not statistically significant, I felt that it was an important variable to add for the investigations for further studies. The AIC of this model is 2469.63.

Table 5: Final Cox PH Model

Dependent: Surv(Futime, Status)		all	HR (multivariable)
Age of Liver Diagnosis**	>= 65	96 (100)	-
	50-64	213 (100)	-
	< 50	161 (100)	-
Extrahepatic Mets	No	332 (100.0)	-
	Yes	138 (100.0)	1.55 (1.22-1.97, p<0.001)
KRAS Mutation	No	251 (100.0)	-
	Yes	219 (100.0)	1.27 (1.02-1.58, p=0.032)
Number of Liver Mets*	Mean (SD)	7.6 (10.4)	1.24 (1.10-1.39, p<0.001)
Node Positive	No	113 (100.0)	-
	Yes	357 (100.0)	1.25 (0.97-1.62, p=0.083)

Primary Surgery	No	161 (100.0)	-
	Yes	309 (100.0)	0.63 (0.49-0.83, p<0.001)
Liver Surgery**	No	250 (100.0)	-
	Yes	220 (100.0)	-

*Log transformation, **Stratified

The last part of Aim 3 was implementing a competing risk regression method. The results are show in Table 6. With forcing liver surgery and age of liver diagnosis into the model, the variables of interest that were statistically significant were node positive and the log transformation of metastasis. With controlling all the other covariates, the presence of node positive was associated with a 42.1% increase in hazard of liver recurrence and an increase in the log number of liver metastasis was associated with a 19.8% increase in hazard of liver recurrence. It is interesting to note, even though was not statistically significant, that the extrahepatic metastasis is associated with a 19.8% decrease in the hazard of liver recurrence. This could be due to the fact that it is known that cancer is in other locations of the body and the current treatment is helping from liver recurrence from happening.

Table 6: Competing Risk Regression using the Fine and Gray model, where effects are summarized as the subdistributional hazard ratio (SHR)

Dependent: crr(LiverFutime, LiverRecure, cbind(x))		all	SHR
Age of Liver Diagnosis	>= 65	96 (100)	-
	50-65	213 (100)	-
	< 50	161 (100)	0.99 (0.85-1.15, p=0.880)
Extrahepatic Mets	No	332 (100.0)	-
	Yes	138 (100.0)	0.78 (0.61-1.01, p=0.063)
KRAS Mutation	No	251 (100.0)	-
	Yes	219 (100.0)	1.19 (0.96-1.47, p=0.120)
Number of Liver Mets*	Mean (SD)	7.6 (10.4)	1.20 (1.07-1.35, p=0.002)
Node Positive	No	113 (100.0)	-
	Yes	357 (100.0)	1.42 (1.08-1.86, p=.011)

Primary Surgery	No	161 (100.0)	-
	Yes	309 (100.0)	1.00 (0.75-1.33, p=0.990)
Liver Surgery	No	250 (100.0)	-
	Yes	220 (100.0)	0.84 (0.62-1.12, p=0.230)

*Log transformation

FUTURE WORK

To improve the stratified Cox PH model, this study could increase the number of attributed variables that are significant predictors of survival time. More types of comorbidities could be added or just having the presence of a condition could be added to the dataset to help understand the health behaviors associated with colorectal patients. Other variables to add could be the other types of gene mutations (BRAF, TP53), when and how many chemo treatments a patient received, family history, race, smoking and alcohol status. A different direction could focus on BMI more to help understand on whether or not there is a difference in survival rates between groups. They could focus on whether the patients with higher BMI is associated with cancer because of their weight and the patients with a lower BMI is associated with cancer because gene mutations or another cause. It would be important to track their BMI throughout the treatment process if there is risk when a patient loses weight during treatment and if it is more common for normal to underweight patients to lose weight. The other future work could be added more patients to the liver and non-liver surgery groups, splitting the data between groups, running two separate models, and observing any clear difference the groups may have.

REFERENCES

American Cancer Society (n.d.). *Survival Rates for Colorectal Cancer*. https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html

American Cancer Society (n.d.). *What Is Colorectal Cancer?* Retrieved July 27, 2020, from https://www.cancer.org/cancer/colon-rectal-cancer/about/what-is-colorectal-cancer.html

Brazer, Y. (2018, January 9). *Colorectal cancer: What you need to know.* Medical News Today. https://www.medicalnewstoday.com/articles/155598

Cancer Net (2020, March 10). *Colorectal Cancer - Statistics*. https://www.cancer.net/cancer-types/colorectal-cancer/statistics

Cancer Treatment Centers of America (2020, July 17). *Colorectal cancer risk factors.* https://www.cancercenter.com/cancer-types/colorectal-cancer/risk-factors

CDC (n.d.). *Cancer Associated with Overweight and Obesity Make up 40 percent of Cancers Diagnosed in the United States.* https://www.cdc.gov/media/releases/2017/p1003-vs-cancer-obesity.html

Chiorean, E. G., Nandakumar, G., Fadelu, T., Temin, S., Alarcon-Rozas, A. E., Bejarano, S., . . . Chamberlin, M. D. (2020). Treatment of Patients With Late-Stage Colorectal

Cancer: ASCO Resource-Stratified Guideline. *JCO Global Oncology*, (6), 414-438. doi:10.1200/jgo.19.00367

Christiano, D. (2018, November 30). *Can You Live Without a Liver?* Health Line https://www.healthline.com/health/can-you-live-without-liver

Columbia University Herbert Irving Comprehensive Cancer Center (2019, February 19). *Colorectal Cancer: Treatments.* https://cancer.columbia.edu/colorectal-cancer-treatments

Dinu, D., Dobre, M., Panaitescu, E., Bîrlă, R., Iosif, C., Hoara, P., . . . Ardeleanu, C. (2014). *Prognostic significance of KRAS gene mutations in colorectal cancer--preliminary study.* Journal of Medicine and Life. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4316144/

Gray B. (2020). *Subdistribution Analysis of Competing Risk*. R package version 2.2-10, https://cran.r-project.org/web/packages/cmprsk/cmprsk.pdf

Hyperarts, R. (n.d.). *Liver Resection*. University of California of San Francisco Department of Surgery. https://surgery.ucsf.edu/conditions--procedures/liver-resection.aspx

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. *CA: A Cancer Journal for Clinicians, 61*(2), 69-90. doi:10.3322/caac.20107

Karen, K. (2015, March 13). *Eliminating colorectal cancer as a major public health problem.* Science Life Administrator. https://sciencelife.uchospitals.edu/2015/03/02/eliminating-colorectal-cancer-as-a-major-public-health-problem/

Kocarnik, J. M., Hua, X., Hardikar, S., Robinson, J., Lindor, N. M., Win, A. K., . . . Newcomb, P. A. (2017). Long-term weight loss after colorectal cancer diagnosis is associated with lower survival: The Colon Cancer Family Registry. *Cancer, 123*(23), 4701-4708. doi:10.1002/cncr.30932

Kow, C. Wei, A. (2019). Hepatic metastasis from colorectal cancer. *Journal of Gastrointestinal Oncology*, *10*(6), 1274-1298. doi:10.21037/jgo.2019.08.06

Mayo Clinic. (2019, October 08). *Colon cancer.* https://www.mayoclinic.org/diseases-conditions/colon-cancer/symptoms-causes/syc-20353669

National Cancer Institute (n.d.) *NCI Dictionary of Cancer Terms*. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/adenocarcinoma

Therneau T. (2020). *A Package for Survival Analysis in R*. R package version 3.2-7, https://cran.r-project.org/web/packages/survival/index.html.

World Cancer Research Fund (2019, July 17). *Colorectal cancer statistics.* https://www.wcrf.org/dietandcancer/cancer-trends/colorectal-cancer-statistics

APPENDICES

Appendix A:

```
## Brandon O'Grady - Thesis Project
## 9/4/2020
## Survival Analysis of Colorectal Cancer with Liver Metastasis
########## Libraries ####
library(survival)
library(survminer)
library(tidyverse)
library(dplyr)
library(tidyr)
library(psych)
library(ggplot2)
library(readxl)
library(pastecs)
library(finalfit)
library(cmprsk)
library(kSamples)

########## Load Data Set ####
CRLM <- read_excel("~/Thesis.xlsx")
######### Format Data #####
crlm <- CRLM %>%
  mutate(
    Gender = factor(Gender, labels = c("Female","Male")),
    EhMets = factor(EhMets, labels = c("No Extra Mets", "Extra Mets")),
    KRAS = factor(KRAS, labels = c("Non-Kras Mutation", "Kras Mutation")),
    Synchronous = factor(Synchronous, labels = c("No", "Yes")),
    Primary = factor(Primary, labels = c("Colon", "Rectum")),
    PrimarySurgery = factor(PrimarySurgery, labels = c("No", "Yes")),
    LiverSurgery = factor(LiverSurgery, labels = c("No Surgery", "Liver Surgery")),
    Bilateral = factor(Bilateral, labels = c("No", "Yes")),
    NodePositive = factor(NodePositive, labels = c("No", "Yes")),
    HTN = factor(HTN, labels = c("Normal", "Hypertension")),
    MSI = factor(MSI, labels = c("Microstatellite Stability", "Microstatellite Instability")),
    Diabetes = factor(Diabetes, labels = c("No", "Yes")),
    PrimaryTumor = factor(PrimaryTumor))

###### Aim 1 #####
##Discriptive Analysis##
#summary of data set#
summary(crlm)
```

```
sd(crlm$DxAgeLiver)
#Descriptive Statistics for whole dataset
stat.desc(crlm)

#Descriptive statistics by Surgical vs. Non-Surgical Groups#
describe.by(crlm,
        crlm$LiverSurgery, na.rm = T)

#Descriptive statistics by Gender#
describe.by(crlm,
        crlm$Gender, na.rm = T)

#Descriptive Statistics by Primary Tumor
describe.by(crlm,
        crlm$PrimaryTumor, na.rm = T)

#Descriptive Statistics by Colon or Rectum
describe.by(crlm,
        crlm$Primary, na.rm = T)

### Wilcoxon Rank-Sum test ####
#Liver Surgery vs. Non-Liver Surgery Groups
wilcox.test(BMI ~ LiverSurgery, data = crlm) #Reject the Null
wilcox.test(SizePath ~ LiverSurgery, data = crlm) #Reject the Null
wilcox.test(NoMets ~ LiverSurgery, data = crlm) #Reject the Null
wilcox.test(DxAgeLiver ~ LiverSurgery, data = crlm) #Fail to reject the null

#Gender
wilcox.test(BMI ~ Gender, data = crlm) #Fail to reject the null
wilcox.test(SizePath ~ Gender, data = crlm) #Fail to reject the null
wilcox.test(NoMets ~ Gender, data = crlm) #Fail to reject the null
wilcox.test(DxAgeLiver ~ Gender, data = crlm) #Reject the Null
wilcox.test(DxAge ~ Gender, data = crlm) #Reject the Null

#Rectum or Colon
wilcox.test(BMI ~ Primary, data = crlm) #Fail to reject the null
wilcox.test(SizePath ~ Primary, data = crlm) #Reject the Null
wilcox.test(NoMets ~ Primary, data = crlm) #Reject the Null
wilcox.test(BMI ~ Primary, data = crlm) #Fail to reject the null

### Chi-Square ###
#Liver Surgery vs. Non-Liver Surgery Groups
chisq.test(crlm$Gender, crlm$LiverSurgery) #Fail
```

```
chisq.test(crlm$MSI, crlm$LiverSurgery) #Fail
chisq.test(crlm$KRAS, crlm$LiverSurgery) #Fail
chisq.test(crlm$Synchronous, crlm$LiverSurgery) #Reject the null
chisq.test(crlm$HTN, crlm$LiverSurgery) #Fail
chisq.test(crlm$Diabetes, crlm$LiverSurgery) #Fail
chisq.test(crlm$Bilateral, crlm$LiverSurgery) #Reject the null
chisq.test(crlm$NodePositive, crlm$LiverSurgery) #Fail

table(crlm$LiverSurgery, crlm$Bilateral)

#Gender
chisq.test(crlm$MSI, crlm$Gender) #Fail
chisq.test(crlm$KRAS, crlm$Gender) #Fail
chisq.test(crlm$Synchronous, crlm$Gender) #Fail
chisq.test(crlm$HTN, crlm$Gender) #Reject the null
chisq.test(crlm$Diabetes, crlm$Gender) #Fail
chisq.test(crlm$Bilateral, crlm$Gender) #Fail
chisq.test(crlm$NodePositive, crlm$Gender) #Fail

#Primary Tumor Site
chisq.test(crlm$Gender, crlm$PrimaryTumor) #Reject the null
chisq.test(crlm$MSI, crlm$PrimaryTumor) #Reject the null
chisq.test(crlm$KRAS, crlm$PrimaryTumor) #Reject the null
chisq.test(crlm$Synchronous, crlm$PrimaryTumor) #Reject the null
chisq.test(crlm$HTN, crlm$PrimaryTumor) #Fail
chisq.test(crlm$Diabetes, crlm$PrimaryTumor) #Fail
chisq.test(crlm$Bilateral, crlm$PrimaryTumor) #Fail
chisq.test(crlm$NodePositive, crlm$PrimaryTumor) #Fail

table(crlm$KRAS, crlm$PrimaryTumor)
##### continuous variables ########
cor.test(crlm$DxAgeLiver, crlm$BMI, method="pearson") #not correlated
cor.test(crlm$SizePath, crlm$BMI, method="pearson") #not correlated
cor.test(crlm$SizePath, crlm$NoMets, method="pearson") #correlated at .25
cor.test(crlm$SizePath, crlm$DxAgeLiver, method="pearson") #not correlated
cor.test(crlm$NoMets, crlm$BMI, method="pearson") #not correlated

###########AIM 2##########

###### Univariate Cox-Regression Analysis of Each Variable######
summary(coxph(Surv(Futime, Status) ~ Gender, data = crlm))
summary(coxph(Surv(Futime, Status) ~ DxAgeLiver, data = crlm))
summary(coxph(Surv(Futime, Status) ~ LiverSurgery, data = crlm))
```

```
summary(coxph(Surv(Futime, Status) ~ EhMets, data = crlm))
summary(coxph(Surv(Futime, Status) ~ KRAS, data = crlm))
summary(coxph(Surv(Futime, Status) ~ MSI, data = crlm))
summary(coxph(Surv(Futime, Status) ~ BMI, data = crlm))
summary(coxph(Surv(Futime, Status) ~ Synchronous, data = crlm))
summary(coxph(Surv(Futime, Status) ~ Primary, data = crlm))
summary(coxph(Surv(Futime, Status) ~ PrimaryTumor, data = crlm))
summary(coxph(Surv(Futime, Status) ~ PrimarySurgery, data = crlm))
summary(coxph(Surv(Futime, Status) ~ HTN, data = crlm))
summary(coxph(Surv(Futime, Status) ~ Diabetes, data = crlm))
summary(coxph(Surv(Futime, Status) ~ Bilateral, data = crlm))
summary(coxph(Surv(Futime, Status) ~ SizePath, data = crlm))
summary(coxph(Surv(Futime, Status) ~ NoMets, data = crlm))
summary(coxph(Surv(Futime, Status) ~ NodePositive, data = crlm))
summary(coxph(Surv(Futime, Status) ~ CEA, data = crlm))

### Kaplan-Meier non-parametric analysis
kmsurvival <- survfit(Surv(Futime, Status) ~ 1, data = crlm)
summary(kmsurvival)
ggsurvplot(kmsurvival, data=crlm)

### Kaplan-Meier non-parametric analysis by Liver Surgery group
km_surgery <- survfit(Surv(Futime, Status) ~ LiverSurgery, data = crlm)
summary(km_surgery)
print(km_surgery)
ggsurvplot(km_surgery, data=crlm, pval =T)
#cumulative hazard
ggsurvplot(km_surgery, pval = T, fun = "cumhaz")
#cumulative event by liver surgery group
ggsurvplot(km_surgery, pval = T, conf.int = .95, fun = "event")

### Kaplan-Meier non-parametric analysis by Sex
km_sex <- surv_fit(Surv(Futime, Status) ~ Gender, data = crlm)
summary(km_sex)
ggsurvplot(km_sex, data=crlm, pval = T)

### Kaplan-Meier non-parametric analysis by Rectum and Colon
km_rectum <- surv_fit(Surv(Futime, Status) ~ Primary, data = crlm)
summary(km_rectum)
ggsurvplot(km_rectum, data=crlm, pval = T)

### Kaplan-Meier non-parametric analysis by Site
km_site <- surv_fit(Surv(Futime, Status) ~ PrimaryTumor, data = crlm)
```

```
summary(km_site)
ggsurvplot(km_site, data = crlm, pval = T)

## Kaplan-Meier non-parametric analysis by Bilateral
km_bilateral <- surv_fit(Surv(Futime, Status) ~ Bilateral, data = crlm)
summary(km_bilateral)
ggsurvplot(km_bilateral, data = crlm, pval = T)
print(km_bilateral)

## Kaplan-Meier non-parametric analysis by Synchronous
km_synchronous <- surv_fit(Surv(Futime, Status) ~ Synchronous, data = crlm)
summary(km_synchronous)
ggsurvplot(km_synchronous, data=crlm, pval = T)

## Kaplan-Meier non-parametric analysis by Extrahepatic Mets
km_ehmets <- surv_fit(Surv(Futime, Status) ~ EhMets, data = crlm)
summary(km_ehmets)
ggsurvplot(km_ehmets, data=crlm, pval = T)

## Kaplan-Meier non-parametric analysis by Primary Surgery
km_primarysurgery <- surv_fit(Surv(Futime, Status) ~ PrimarySurgery, data = crlm)
summary(km_primarysurgery)
ggsurvplot(km_primarysurgery, data=crlm, pval = T)

##### Liver recurrence outcome ####
liver_recur <- factor(crlm$LiverRecur, levels = c(0,1,2),
                      labels = c("Censored", "Liver Recurrence", "Death w/o Liver Recurrence"))
table(liver_recur, crlm$LiverSurgery)

summary(liver_recur)
##Equality of Cumulative Incidence Fuction among Liver Surgery Groups as the grouping
variable
fit <- cuminc(crlm$LiverFutime, crlm$LiverRecur, crlm$LiverSurgery, cencode = 0)
print.cuminc(fit)
plot.cuminc(fit, color = rainbow(4), xlab = "Days")

################ AIM 3 ##############
######## Functional Forms #####
ggcoxfunctional(Surv(Futime, Status)~DxAgeLiver, crlm)
ggcoxfunctional(Surv(Futime, Status)~SizePath, crlm)
ggcoxfunctional(Surv(Futime, Status)~BMI, crlm)
ggcoxfunctional(Surv(Futime, Status)~NoMets, crlm)
```

```
####Normality Check####
##Histograms of Continuous Variables##

#Age of Liver Met Histogram#
ggplot(crlm) +
  aes(x = DxAgeLiver) +
  geom_histogram()

#BMI Histogram#
ggplot(crlm) +
  aes(x = BMI) +
  geom_histogram()

#Liver Met Size#
ggplot(crlm.t) +
  aes(x = SizePath) +
  geom_histogram()

#Number of Liver Mets in Liver#
ggplot(crlm) +
  aes(x = NoMets) +
  geom_histogram
###Normal Q-Q Plots###
ggqqplot(crlm$DxAgeLiver)
ggqqplot(crlm$BMI)
ggqqplot(crlm$SizePath)
ggqqplot(crlm$NoMets)

###Shapiro-Wilk Test###
shapiro.test(crlm$DxAgeLiver) #can assume normality
shapiro.test(crlm$BMI) #cannot assume normality
shapiro.test(crlm$SizePath) #cannot assume normality
shapiro.test(crlm$NoMets) #cannot assume normality

####Mutate Non-normal variables#####
#log had the best transformations
crlm.t <- crlm %>%
  mutate(
    age = case_when(
      DxAgeLiver >= 65 ~ 0,
      DxAgeLiver >= 50 & DxAgeLiver < 65 ~ 1,
      DxAgeLiver < 50 ~ 2),
    bmi_log = log(BMI),
```

```
    size_log = log(SizePath),
    size.t = case_when(
      size_log >= 1.75 ~ 0,
      size_log >= .5 & size_log < 1.75 ~ 1,
      size_log < .5 ~ 2),
    mets_log = log(NoMets))
crlm.t <- crlm.t %>%
  mutate(
    size.t = factor(size.t),
    age = factor(age))

## Fuctional Form with Transformation
ggcoxfunctional(Surv(Futime, Status)~size_log, crlm.t)
ggcoxfunctional(Surv(Futime, Status)~bmi_log, crlm.t)
ggcoxfunctional(Surv(Futime, Status)~mets_log, crlm.t)


#### Build Model with Improved Functional Forms #####
### Droping MSI due to missing data

cox.mod <- coxph(Surv(Futime,
Status)~age+EhMets+Synchronous+size_log+mets_log+KRAS+

PrimaryTumor+LiverSurgery+HTN+Bilateral+NodePositive+bmi_log+Diabetes+PrimarySu
rgery, data= crlm.t)
summary(cox.mod)
#### AIC-Based Stepwise Model Reduction ####

step.mod <- step(cox.mod, direction = "both") #size_log and size.t were ran and are both
excluded from the model
summary(step.mod)

#### Test Proportional Hazards #####
step.mod %>% cox.zph() %>% ggcoxzph()
cox.zph(step.mod)

#Age of Liver Diagnosis and Liver Surgery will be strata
strat.mod <- coxph(Surv(Futime, Status) ~  strata(LiverSurgery, age) + EhMets + mets_log +
KRAS +
                     Bilateral + NodePositive + bmi_log + PrimarySurgery, data = crlm.t)
strat.mod
extractAIC(strat.mod)
## Re-check proportional hazards
```

```
strat.mod %>% cox.zph() %>% ggcoxzph() ## All Hazard Proportional, including the global
model
cox.zph(strat.mod)

##### Final model diagnostics ####

## Check dfbeta residuals for model

ggcoxdiagnostics(strat.mod, "dfbeta", title = "DF Beta Residuals")

## Check deviance residuals for model

ggcoxdiagnostics(strat.mod, "deviance", title = "Deviance Residuals")

#Final Report Model
summary(strat.mod)

### oberserving bmi in groups

crlm.t <- crlm.t %>%
  mutate(
    bmi = factor(bmi, labels = c("Normal", "UnderWeight", "Overweight", "Obese",
"Extremely Obese" )))

summary(coxph(Surv(Futime, Status) ~ bmi, data = crlm.t))
table(crlm.t$bmi)
fisher.test(crlm.t$bmi, crlm.t$LiverSurgery)

km_bmi <- survfit(Surv(Futime, Status) ~ bmi, data = crlm)
summary(km_bmi)
ggsurvplot(km_bmi, data=crlm.t)


####Dropping bmi_log, Bilateral due to complications and significance
#run final model

strat.final <- coxph(Surv(Futime, Status) ~  strata(LiverSurgery, age) + EhMets + KRAS +
mets_log  +
                  NodePositive  + PrimarySurgery, data = crlm.t)
strat.final
## Re-check proportional hazards
```

```
strat.final %>% cox.zph() %>% ggcoxzph() ## All Hazard Proportional, including the global
model
cox.zph(strat.final)

##### Final model diagnostics ####

## Check dfbeta residuals for model

ggcoxdiagnostics(strat.final, "dfbeta", title = "DF Beta Residuals")

## Check deviance residuals for model

ggcoxdiagnostics(strat.final, "deviance", title = "Deviance Residuals")

#Final Report Model
summary(strat.final)
extractAIC(strat.final)

### Competing risk regression with the variables of interest###
x <- cbind(crlm.t$age, crlm.t$EhMets, crlm.t$KRAS,
        crlm.t$mets_log, crlm.t$NodePositive,
        crlm.t$PrimarySurgery, crlm.t$LiverSurgery)

x.cr <- crr(crlm$LiverFutime, crlm$LiverRecur, x)
summary(x.cr)
table(crlm.t$age)
```