University of Essex
Department of Mathematical Sciences

MA981: DISSERTATION

Understanding the Impact of COVID-19 on Liver Cancer: A Survival Analysis

Dhanya Pezhumkattil Jayaprakash

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

One of the most recent pandemics, COVID-19, has had a wide range of effects on individuals all around the world. Studies are still in progress; thus, it is unclear how COVID-19 has affected things. Hospital operations and their activities were significantly interrupted during the pandemic. Patients with fatal conditions are impacted in a variety of ways. The health of an individual is significantly impacted by both COVID-19 and liver cancer. According to several research, COVID-19 may accelerate the spread of liver cancer. Furthermore, COVID-19-infected liver cancer patients are more prone to have severe COVID-19 symptoms and need hospitalization. There are various factors that could make COVID-19 worsen the course of liver cancer. One of the most frequent causes is COVID-19, which can harm the liver and make it more challenging for the body to fight off the virus. Because they frequently have underlying medical issues and are older, patients with liver cancer may be more sensitive to the effects of COVID-19. Chemotherapy and radiation therapy are two liver cancer treatments that might compromise immune function and increase the risk of infection in patients.

This study aims to analyze the dynamic interaction between the COVID-19 pandemic and the survival outcomes of newly diagnosed liver cancer patients by evaluating how these outcomes have been influenced by the pandemic and the multiple patient characteristics. By comparing the survival rates between the pre-pandemic and pandemic periods, the study seeks to completely grasp how the pandemic has affected the long-term prognosis for individuals with liver cancer. Additionally, it analyzes various patient characteristics and tries to reveal useful data on the intricate factors that affect survival trends throughout this distinct pandemic and pre-pandemic time period. It will give some valuable insights to handle liver cancer patients when unprecedented global challenges come.

There are various methods for resolving this issue. One approach to resolving this issue is using survival analysis. Quantitative survival analysis tools evaluate how COVID-19, together with other patient traits and conditions, influences the outcomes for patients with liver cancer. A specialist statistical method called survival analysis is used to examine how long it will be before a particular event, like failure or death, occurs. The fact that the data set contains time-dependent data to investigate and that it is particularly effective at handling censored data, which means some persons may be lost to follow-up or still living at the end of the study, are two of the many reasons to use this approach. Survival analysis can efficiently handle censored data, offering a mechanism to account for these people.

1.1 Problem Statement

The objective of this study is to examine COVID-19's impact on patients with newly discovered liver cancer. This study compares the outcomes from the pre-pandemic period with those from the pandemic period to ascertain the long-term impact of COVID on patients with liver cancer. Additionally, it looked at how different patient characteristics and circumstances linked to these patients' survival results during the pandemic and pre-pandemic periods.

This study first looks at how people with liver cancer have been impacted by the COVID-19 pandemic. To achieve this, the study seeks to comprehend the variations in outcomes between two distinct time periods: the "pre-pandemic period" and the "pandemic period," which are both referred to throughout the study. By contrasting these two times, the study seeks to ascertain any potential long-term effects of COVID-19 on patients with liver cancer. In addition, the study will take into account a number of variables that could affect whether someone died or survived during these times. These variables include characteristics of the patients themselves, such as age and gender, as well as their health problems, including the size of their tumors, and the type of therapy they get. By examining these elements, the study seeks to understand how different patient traits and environmental factors may have influenced the fatality rates seen throughout the pre-pandemic and pandemic periods.

CHAPTER 2

Literature Review

The numerous effects and causes of COVID-19 on individuals with chronic liver disorders (CLD) are covered in this research[1]. The SARS-CoV-2 virus, treatment side effects, hypoxia, immunological stress, and inflammation are a few of the causes that cause liver damage in COVID-19 patients. Furthermore, research indicates that most COVID-19 patients have minor, temporary increases in AST and ALT liver enzymes, which harm the liver. This article also discusses cirrhosis, non-alcoholic fatty liver disease, and hepatocellular carcinoma (HCC). The findings demonstrate that there is no connection between COVID-19 infection, which can result in serious illness or death, and autoimmune liver disease (AILD). The effect of COVID-19 on people who have had liver transplants was also covered in this essay. It demonstrates that fewer liver transplant cases have occurred during the pandemic. To lower the risk of COVID-19 infections, the report also covered the significance of immunization in patients with chronic liver disease or in individuals who have had liver transplants. Individuals with acute chronic liver syndrome must be followed up on regularly, especially elderly individuals.

COVID-19 has equally affected patients and healthcare professionals and had a substantial impact on the hepatology community [2]. Patients with chronic liver disease, hepatobiliary malignancy, and previous liver transplantation are also given instructions and guidance. The development of the omicron form has added further complexity; however, the COVID-19 vaccine and therapies have improved the clinical prognosis. EASL has tried to offer current information to researchers and doctors due to the dynamic nature of the COVID-19 virus.

Studies demonstrate that chronic liver damage is frequent during the pandemic, with raised liver enzyme levels being prevalent while severe liver dysfunction is uncommon. Systemic inflammation, cytokines, ischemia, and medication effects are a few of the complex aspects. In addition, the virus itself can infect hepatocytes directly, causing liver disease. The likelihood of liver damage varies with different COVID-19 formulations and relates to the overall course of the disease. The journal also covers secondary sclerosing cholangitis (SSC), a condition that is frequent in people with severe COVID-19. After COVID-19, it is typically marked by elevated bilirubin and ALP values, particularly in patients who already have chronic liver disease.

The topic of liver damage in COVID-19 patients is caused by viruses, drugs, or underlying disorders[3]. It investigates how COVID-19 affects people with chronic liver disease (CLD), who are more vulnerable since their immune and metabolic systems are weaker. According to studies, COVID-19 results may be worsened by antecedent non-alcoholic fatty liver disease (NAFLD) and viral hepatitis. According to this study, there are some limitations on the causes of liver damage resulting from COVID-19 advancement because there is a dearth of information regarding the effects of COVID-19 infection on people with chronic liver diseases (CLD). Patients with non-alcoholic fatty liver disease (NAFLD) frequently have comorbid metabolic disorders such as diabetes and hypertension, which individually raise the likelihood of a COVID-19 outcome. The understanding of this link is further complicated by a few uncertainties and a lack of complete information regarding the underlying hepatic diseases.

There are some studies showing the impact of COVID-19 on liver cancer management. Around 76 countries participated in this research, including Europe, South America, North America, Asia, and Africa[4]. This study highlights several impacts and changes, including modification or cancellation of screening programs (80.9%), Cancellation of curative and/or palliative treatments (50%), Modification of the liver transplantation program (41.7%), Reduced access to diagnostic procedures (40.8%), and Increased use of telemedicine (51.4%). The research demonstrates that COVID-19 has had a detrimental effect on the pandemic's treatment of liver cancer. It presents unexpected difficulties for the global healthcare system. The likelihood of patients with liver cancer during the pandemic having severe illnesses is high, and the risk is also high. Receiving treatment and regular reviews is also delayed because of the lockdown and other factors. This increased the severity of diseases. The authors contend that additional investigation is necessary to determine COVID-19's long-term effects.

To evaluate and determine the prognostic factors affecting long-term survival in patients with hepatocellular carcinoma (HCC) who had hepatic resection, a retrospective analysis is proposed [5]. Between 1989 and 1995, 204 patients at a single hospital were the subject of this investigation. According to the findings, the median overall survival (OS) period for patients is almost 35 months. The median OS during a five-year period is 38.7%. Poor OS is influenced by a number of variables, including tumor size, vascular invasion, the number of tumor nodules, preoperative liver function, and the retention value of indocyanine green at 15 minutes. The pTNM classification system was employed as the approach in this study, and it has a strong relationship with postoperative survival. This study discovered that the long-term prognoses of individuals with HCC were significantly influenced by the preoperative assessment of hepatic function.

Some of the processes that lead to direct viral entry into hepatocytes and cholangiocytes, immune-mediated hepatitis, hypoxia, and drug-related hepatotoxicity[6]. During the pandemic, cancer patients often exhibited worse COVID-19 outcomes and an increased risk of infection, particularly those who had recently completed cancer treatment. For the treatment of HCC, a decision should be made based on the accessibility of medical resources, the level of infection risk from COVID-19, and the patient's particular risk-benefit ratio.

To analyze the survival time of patients after diagnosis of COVID-19 deep learning and Cox regression are used [7]. The features are extracted from the clinical data of patients using a deep learning model and the Cox regression model is used to assess the chance of mortality. These methods help to identify those patients who are at high risk and develop tailored interventions to improve their prognosis. The dataset for analysis includes 1085 patient records. The proposed model predicted a death rate of 89%. This study found that a variety of variables raise the risk of diabetes. This includes concomitant diseases like diabetes, hypertension, and heart disease, as well as age and male gender. a protracted hospital stay and serious breathing issues. The technique could be used to identify individuals who are most at risk of dying and develop targeted medications to improve their prognoses.

Some of the theoretical investigation's main focus is on the Cox regression model's asymptotic properties [8]. We begin by introducing Cox regression models and their underlying assumptions. The asymptotic distribution of the model parameters is also generated, along with the maximum likelihood estimators. This derivation is important because it gives the researchers the confidence to make judgments about the model parameters. The asymptotic effectiveness of the estimators was then evaluated, as was the model's resistance to assumptions being broken. Selecting the best estimator for a given data set is equally important for researchers. Even if the assumptions are not totally accurate, we may still assess the reliability of the findings by assessing how robust the model is to assumptions being broken. The primary focus of this theoretical inquiry is on the asymptotic properties of the Cox regression model. First, the Cox regression models, and their underlying assumptions are presented. It is equally important for researchers to select the best estimator for their unique data collection. Looking at how resilient the model is to assumptions being broken allows us to assess the validity of the results even if the assumptions are not totally true.

A statistical strategy for assessing data on the interval before an event of interest happens is called survival analysis [9]. The likelihood that a person will not encounter an event of interest at a specific moment is known as the survivor function. The hazard function is the instantaneous rate at which the event of interest occurs. A statistical model called the Cox proportional hazards regression model, which assumes that hazard ratios remain constant across time, can be used to analyze the relationship between variables and the hazard function. A variety of statistical software tools can be used to fit the Cox regression model. The results are interpreted in terms of hazard ratios.

When comparing the rates of liver cancer that were respectable before and after the COVID-19 pandemic[10]. Patients who had their livers examined for cancer between January 2019 and June 2021 were included in the study, which was carried out at the Johns Hopkins Multidisciplinary Liver Cancer Clinic.. The COVID-19 epidemic has affected healthcare services in all medical specialties. To evaluate the effect, a retrospective analysis of patients with hepatocellular carcinoma (HCC) or biliary tract cancer (BTC) that was either suspected or confirmed was done. The data showed that during the first several months of the COVID-19 epidemic, there was a considerable decline in the surgical respectability rate for liver cancer. During the pandemic, several circumstances contributed to the treatment of liver cancer being delayed. This includes giving COVID-19 patients priority care. Additionally, the closure of some healthcare institutions, the disruption of cancer screening and diagnostic procedures, and the worry of catching COVID-19 in hospital settings The author makes certain recommendations, such as giving cancer patients priority in receiving care and making sure that cancer screening and diagnostic procedures are accessible during the pandemic.

Recent Studies demonstrate that individuals with liver cancer who had a SARS-CoV-2 infection had a greater probability of dying than patients who did not have the infection [11]. Patients with advanced liver cancer faced a heightened chance of death. The patients' deaths were caused by several health-related circumstances. In addition to elderly age, these conditions also include chronic lung, heart, and kidney diseases, diabetes, and advanced liver cancer. The study led the scientists to the conclusion that SARS-CoV-2-infected liver cancer patients have a higher probability of dying. Patients with liver cancer should take precautions to prevent contracting SARS-CoV-2, including vaccination, mask use, and social seclusion. It is based on the information gathered from patients with liver cancer at 29 hospitals worldwide. The patients were monitored for a full year. The 30-day mortality rate and its 95% confidence intervals (95% CI) were calculated using the Kaplan-Meier method. To calculate the sub-distribution hazard ratios (HR) and their 95% confidence intervals, Cox regression models were also used. The studies reported a variety of liver abnormalities in patients with COVID-19, including elevated liver enzymes, steatosis, hepatitis, and acute liver failure[14]. The patients with COVID-19 were more likely to develop liver abnormalities if they were older, had underlying liver disease, or were critically ill. A COVID-19 patient has several ways to damage the liver through the immune response to the virus.

The instantaneous rate at which the relevant event occurs is the hazard function [12]. Given that the individual has not yet experienced the event, it is defined as the likelihood that the individual will do so within a short period of time. The probability that a person won't experience the relevant event at a particular moment is known as the survival function. The form of the hazard function is not assumed in the non-parametric Kaplan-Meier method.

The semi-parametric Cox proportional hazards model takes the hazard ratios to be constant across time. It also examined the log-rank test, which compares several groups statistically. When doing a survival analysis, it's crucial to take note of the beginning and end of the survival period as well as any censored observations that occur when a subject is not tracked throughout the entire study. It can be used to determine prognostic factors and evaluate the effectiveness of treatments in clinical trials.

The association between risk factors and time to occurrence is investigated using the statistical model known as Cox regression [13]. The hazard ratio for a risk factor is constant in the model. However, a covariate's effect frequently changes over time. In this study, a technique for fitting time-invariant variables into the Cox regression model is presented. Cancer patient lung datasets are used for the implementation and analysis. The suggested approach is effective and looks promising. The time-varying coefficient is handled in this study using a step function. With this approach, the analysis time is divided into several intervals, and the Cox proportional model is stratified for each of these intervals. Another method that has been put out describes the time-varying coefficient using a continuous function. According to the findings, the suggested approach can increase prediction accuracy when compared to the conventional Cox regression model using time-invariant covariates.

The Cox proportional hazards model, the Weibull model, and the accelerated failure time model are just a few examples of the several models that can be used for multivariate survival analysis that are covered by the authors [**?**]. There are various techniques for evaluating the fit of a survival model, including the log-rank test, the Cox-Snell residuals, and the Schoenfeld residuals. The residual from the survival model, the residual plot, and the Kaplan-Meier survival curve were also used to evaluate the suitability of a model. Additionally, proportional hazards are evaluated using a variety of methodologies; in the case studies, log survival plots are used.

The additive hazards model is more flexible and able to handle categorical and continuous exposures[16]. Each person's incident risk is determined by the model while considering their exposure status and other variables. The deviation from additivity, or the interaction between the exposures, is also estimated by the model. The paper begins by covering the many techniques to assess interaction in survival analysis. The most frequent strategy is to utilize the Cox proportional hazards model, which determines the proportionate risk of an occurrence linked with each exposure. However, the Cox model cannot directly examine additive interactions. There is an empirical example of the application of the additive hazards model to analyze the impact of smoking and education on lung cancer risk. Results from the Cox model and additive hazard models are compared, and data from the recently created Social Inequality in Cancer database is examined.

For evaluating the proportional hazard assumptions, it covers the Cox proportional hazard model, the most popular survival analysis model[17]. According to the proportional hazard hypothesis, the risk ratio for an occurrence stays constant across time. The Cox model might not be accurate if it is broken. Several techniques for evaluating the proportional risk hypotheses are examined and put into practice. These techniques include statistical tests like the Schoenfeld test as well as graphical techniques like the log-minus-log plot. To deal with the situation of broken hazard assumptions, some extended Cox models exist. The time-dependent Cox model, the robust Cox model, and the stratified Cox model are some examples of these models. All of these techniques are used to analyze the survival of cancer patients using clinical data. Additionally, it examined cardiac disease patients' survival rates and post-operative patient survival rates. Reviewed and investigated the effectiveness of the log-rank test for comparing two groups or trends.

CHAPTER 3

Methodology

In this study, the effect of COVID-19 on patients with newly detected liver cancer who were diagnosed during the pre-pandemic period was compared to those who were diagnosed during the pandemic period using the Cox proportional hazard model. It is both a multivariate statistical methodology and a survival analysis technique. It measures the time until an event occurs and is employed to analyze survival in relation to numerous factors at once. This study evaluates the effect of the pandemic on the likelihood of dying from liver cancer. This kind of research is a good fit for the crucial statistical technique known as the Cox proportional hazard model. It is an effective method for assessing how covid 19 affects patients with recently diagnosed liver cancer. This model is effective at estimating the impact of one or more characteristics on the likelihood of an event, like death. The occurrence in this instance is a death from liver cancer.

There are numerous methods of survival analysis accessible. It contains the frailty model, the proportional hazards regression model, the parametric survival model, and the Kaplan-Meier model. Because it makes no assumptions about the data, the Kaplan-Meier model is a viable option for studies with small sample sizes or when the data are not regularly distributed. For predicting the survival function for a single variable or a particular group of interest, use the Kaplan-Meier estimator. If the data are not normally distributed, a non-parametric approach is utilized. There are various distributional assumptions made by the parametric survival model. It is effective and used when data is normally distributed. When the hazard ratio varies over time, the proportional hazards regression model is utilized. This model makes various data assumptions. When the results indicate that patients' survival is not independent of one another, the frailty model is applied.

3.1 COX Proportional Hazard Model

Using the COX proportional hazard model has various benefits. It is a semi-parametric model that may be applied with a range of data formats and does not make any assumptions about the data. It is perfect for investigations with low resources because it is relatively simple to fit and interpret. It can be used to contrast patient survival rates with various parameters. The hazard ratio, or instantaneous rate of risk given that a person has so far lived, is estimated by the COX proportional hazard model. The risk factor assessed here is death.

It is crucial to define the time and event variables with all of the study's predictors in order to fit the COX proportional hazard model. "Survival from MDM" is the time variable in this study, while "Alive_Dead" is the status variable. Predictors can also incorporate additional factors like age, gender, size, treatment group, and other medical issues. Given the predictor variables, the Cox proportional hazard model calculates the hazard function, which indicates the instantaneous risk of encountering the event (in this example, death) at any given moment. The hazard ratio for each predictor is assumed by the model to remain constant across time. Mathematically the hazard function is written as

$$h(t) = h_0(t) \times \exp(b_1x_1 + b_2x_2 + \ldots + b_px_p)$$

where t is the survival time, h(t) is the hazard function determined by a set of p covariates (X1, X2...Xp), b1,b2...bp are coefficients that measure the impact of covariates, h0 is the baseline hazard. If the hazard ratio is greater than one, it indicates an increase in hazard, and if the hazard ratio is less than one it indicates a reduction in hazard. The hazard ratio is equal to zero indicating no effect. The results of the COX regression proportional hazard model are interpreted based on the hazard ratio, the coefficients, and p values. A coefficient greater than one indicates an increased risk of death for patients with covid 19 compared to those without. If the coefficient is less than one, it shows a lower risk of death due to covid 19. There are also upper and lower confidence intervals for the hazard ratio. Finally, the model also gives p values for all variables in the dataset. It is used to identify the significant variables in the model and assume that the null hypothesis is true. If the p-value is less than the selected significant level then the null hypothesis is rejected. This means there is a statistical relationship between covid 19 on survival time. Moreover, the results are also interpreted using graphical methods. The survival curves of the fitted model are used to interpret the results easily. In R programming survfit() function estimates the survival proportion.

3.2 COX Proportional Hazard model with interaction terms

It is a substitute for the COX proportional hazard model and is used to determine whether the interaction between two predictor variables and survival time affects the connection between the two. In contrast to the sum of the individual effects of the two variables, the interaction term evaluates the combined impact of the two factors on the hazard. The mathematical notation of the COX model with interaction term is written as,

$$h(t|x) = h_0(t) \cdot \exp(\beta_1 \cdot C + \beta_2 \cdot A + \beta_3 \cdot (C \cdot A))$$

Where, h(t | x) is the hazard function at a given time, h0(t) is the baseline hazard function,β_1, β_2, and β_3 are the coefficients associated with the predictor variable and the interaction term, C and A are binary or continuous predictor variables, and $C \cdot A$ is the interaction term between variable C and variable A.

The coefficient for the interaction term is interpreted as the change in the event's hazard ratio for a one-unit increase in one variable while holding the other variable constant. To test the effect modification, the interaction terms with the COX model are used. It occurs when the degree to which one variable affects the likelihood of an event fluctuates based on another one. The Cox proportional hazards model includes the interaction term to check for effect modification. If the interaction term's coefficient is statistically significant, shows the effect of the pandemic on the risk of death is different for patients with different stages of liver cancer. To improve model fit, this form of model is used.

The Cox proportional hazards model with interaction terms facilitates analysis of the relationships between patient variables and survival outcomes, such as death status, that have changed between the pandemic and pre-pandemic periods. The interaction terms explain how the relationship between patient characteristics and survival changes over time. We can establish whether specific patient characteristics have a more significant impact on survival during the epidemic than they did before it by using interaction terms.

CHAPTER 4

Design and Implementation

4.1 Process Diagram

For analysis, the COVID-19 Effect on Liver Data Set is collected from Kaggle. The process diagram is depicted in Figure 4.1. There are 451 observations and 27 attributes of raw data made up the dataset's initial state. Both category values and several missing values are present. In some columns, it also includes outliers. The dataset is handled for missing value management and categorical value encoding in the second step of data pre-processing. Additionally, prior to further processing, outliers are found and dealt with. The dataset's missing values are handled using the mean value imputation approach. Categorical values are transformed into a numerical format for further processing using the label encoding approach. Following data cleaning, the COX Proportional Hazard Model receives the cleaned data as input for model training. The output of the model is then assessed using the hazard ratio and P values. Other metrics for judging the model include the confidence interval and standard error term in the output. The results are finally explained based on the coefficients and hazard ratios in relation to the effects of COVID-19 on the liver cancer.

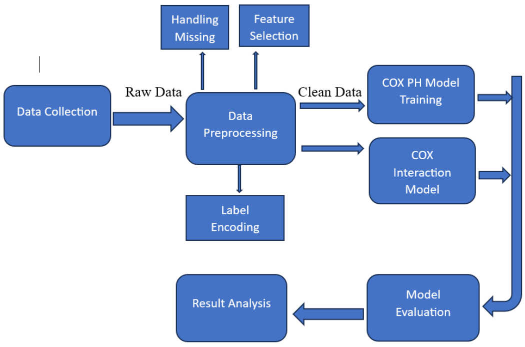


Figure 4.1: Process Diagram

4.2 Covid-19 Effect on Liver Data Set

One of the crucial datasets for examining how COVID-19 affects patients with newly diagnosed liver cancer is the COVID-19 Effect on Liver Cancer Prediction Data set. The exact effects of COVID-19 have not yet been thoroughly documented. Numerous studies demonstrate that it affects the management of high-risk diseases during the first wave of the pandemic (March 2020 to February 2021) and creates disruptions in hepatocellular carcinoma (HCC) services, a decrease in reported and newly detected cases, and an influence on these factors. The hepatopancreatobiliary multidisciplinary team (HPB MDT) at Newcastle-upon-Tyne NHS Foundation Trust (NUTH), where patients were sent throughout the first 12 months of the pandemic, is where the data is gathered. Details of patients with comparable conditions in the last 12 months are also included. All the information of patients with newly diagnosed hepatocellular carcinoma (HCC) or intrahepatic cholangiocarcinoma (ICC) confirmed radiologically or histologically based on international standards.

The dataset contains data on the clinical characteristics of the patients, including the following: cancer, year, month, bleed, mode of presentation, age, gender, etiology, cirrhosis, size, treatment groups, survival from MDM, alive/dead status, type of incidental findings, surveillance program, surveillance effectiveness, mode of surveillance detection, time diagnosis first Tx, date incident surveillance scan, PS the performance status, and surveillance time.

- Cancer: This is a binary variable that represents the presence or absence of cancer in the patient (Y or N).

- Year: The patient's year of cancer diagnosis is indicated by this category feature; the values include pandemic and pre-pandemic.

- Month: The patient's month of cancer diagnosis is indicated by this category feature.

- Bleed: This binary variable represents the patient's experience with spontaneous tumor hemorrhage (Y) or lack of experience (N). This categorical variable describes the manner in which the cancer was presented. The choices are symptomatic, incidental, or under surveillance.

- Age: The age of the patient at the time of diagnosis is indicated by this continuous variable.

- Gender: The gender of the sample is indicated by this category variable.

- Etiology: The underlying cause of the cancer is indicated by this category variable. The choices include hepatitis B virus, hepatitis C virus, hereditary hemochromatosis, primary biliary cholangitis, and autoimmune hepatitis, as well as CLD (chronic liver disease), ARLD (alcoholrelated liver disease), NAFLD (nonalcoholic fatty liver disease), and HCV (hepatitis C virus).

- Cirrhosis: This is a binary variable that represents the presence (Y) or absence (N) of underlying liver disease in the patient.

- Size: This continuous variable represents the tumor's diameter in millimeters.

- HCC TNM Stage: The stage of cancer as determined by the TNM staging system is indicated by this category variable. I, II, IIIA+IIIB, and IV are the available possibilities.

- HCC_ BCLC Stage: The Barcelona Clinic for Liver Cancer staging approach is used to determine the HCC BCLC Stage, which is a categorical variable. There are five choices: 0, A, B, C, and D.

- ICC_TNM stage: The intrahepatic cholangiocarcinoma (ICC) TNM stage is a categorical variable that describes the stage of the malignancy using the TNM staging technique. There are four choices: I, II, III, and IV.

- Treatment groups: The first-line treatment received is indicated by this category variable. The options include medical or supportive treatment, resection, ablation, TACE (transarterial chemoembolization), SIRT (selective internal radiation therapy), and OLTx (orthotopic liver transplantation).
- Survival from MDM: This continuous variable represents the length of time since the multidisciplinary meeting has taken place.
- Alive Dead: This is a categorical variable that indicates the status of whether the patient is alive (Alive) or dead (Dead).
- Type of incidental finding: This categorical variable describes the process used to make the incidental finding. Primary care routine, secondary care routine, primary care acute, and secondary care acute are the available alternatives.
- Surveillance program: Whether the patient was enrolled in a formal surveillance program is indicated by the binary variable "surveillance program" (Y) or not (N).
- Surveillance effectiveness: This category variable represents adherence to surveillance throughout the past year. Consistent, inconsistent, and missed are the available alternatives.
- Mode of surveillance detection: This category variable shows the mode of event surveillance. Mode of surveillance detection. Alpha-fetoprotein alone, ultrasound, and CT/MRI are the available alternatives.
- Time diagnosis 1st Tx: Time between diagnosis and initial treatment: This continuous variable represents the interval between a diagnosis and an initial medical intervention.
- PS: The patient's performance status is indicated by this categorical variable. There are four choices: 0, 1, 2, 3, and 4.
- Time MDM 1st treatment: This continuous variable represents the amount of time that has passed since the multidisciplinary meeting and the start of the first treatment.
- Time decision to treat 1st treatment: This continuous variable represents the passage of time between the decision to treat and the initial treatment.

- Prev known cirrhosis: A binary variable called "Prev. known cirrhosis" (Y or N) indicates if the patient had known cirrhosis previously.
- Months from last surveillance: This continuous variable represents how many months have passed since the last surveillance.

4.3 Exploratory Data Analysis

The dataset has 451 observations and 27 attributes. Both qualitative and numerical values are included. With the aid of various graphical tools to describe the data in depth, R software is used for all exploratory data studies. To prepare the data for analysis, several data preparation techniques are also used.

4.3.1 Missing Value Handling

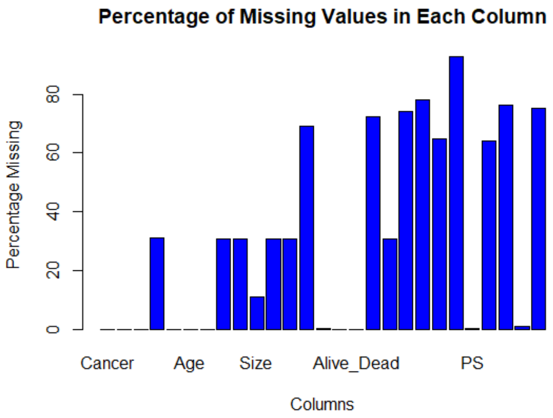


Figure 4.2: Percentage of Missing Values in Each Column

The study excludes attributes with missing values of more than 20%. Figure 4.2 gives a graphical explanation of the number of missing values in each column. The dataset only has 13 variables left after removing the columns with the highest percentage of missing values. The variables such as etiology, cirrhosis, HCC_TNM_Stage, HCC_BCLC_Stage, ICC_TNM_Stage, type of incidental finding, surveillance program, surveillance effectiveness, mode of surveillance detection, Time_diagnosis_1st_Tx, Date_incident_surveillance_scan, Time_MDM_1st_treatment, Time_decisiontotreat_1st_treatment, and Months_from_last_surveillance have 20% more missing values, so these columns dropped from the dataset. Size, Treatment_grps, PS, and Prev_known_cirrhosis are some of the columns that still have a very tiny number of missing values. To address missing data, the missing value imputation method is used in this column. When data is absent, the mean value imputation approach is applied. The data set is free of missing values following the successful imputation of the missing value with its mean. The data set contains the remaining 13 attributes after missing value imputation.

Mean Value Imputation Method

One of the most effective ways to deal with missing data is the mean value imputation approach. The mean value of the non-missing values in the same column is used in this method to fill in the gaps left by missing values. In this method, the algorithm first determines the mean value for each column and is configured to ignore missing data. After that, it repeats over each column, substituting the mean value for any missing data. This approach of missing value imputation is preferred for several reasons. First off, the variables' values or the values of any other variables in your dataset are unrelated to the missing data. When there is no correlation between the missing values and any other dataset variables, it performs well. The dataset in this case has only a minor amount of missing data.

4.3.2 Data Visualization

To uncover any underlying links and patterns in the data, numerous graphs and charts are plotted during this phase. To show the relationship between survival time and event status for different treatment groups, Figure 4.3 first plots a bar plot. The graph demonstrates the high overall survival rates across all therapy groups. The overall shape of the graph demonstrates that patients who got various therapies had variable survival rates. The greatest survival rate was shown in patients who underwent orthotopic liver transplantation (OLTx), followed by resection, ablation, TACE, SIRT, and medicinal treatment. The poorest survival rates were seen in patients who received supportive treatment.

Analyzing tumor size during and before the pandemic is crucial for determining the severity of COVID-19's effects. The violin plot in Figure 4.4 depicts the correlation between tumor size and various HCC/BCLC stages throughout both the pandemic and pre-pandemic

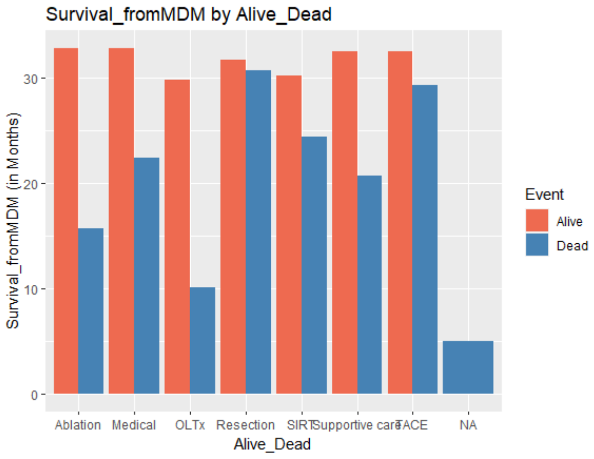


Figure 4.3: Relationship between survival time and Event by Different Treatment Groups

periods. The graph demonstrates that as compared to the pre-pandemic period, the average tumor size is higher during the pandemic. The patient population with HCC-BCLC stage tumors has the largest tumor diameter during the pandemic. The HCC-BCLC Stage A patient group, however, exhibits a smaller tumor size during the pandemic than during the pre-pandemic period.

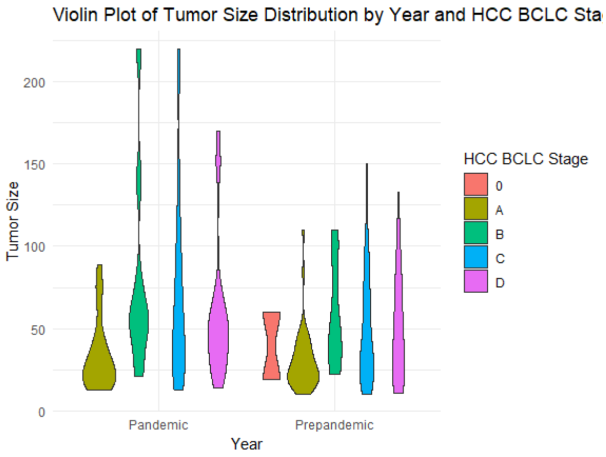


Figure 4.4: Tumor Size distribution by Year and HCC_BCLC Stage

The association between the liver cancer diagnosis method by month and the method of presentation is depicted in Figure 4.5 below. It specifies the month the ailment is discovered and the way it manifests. According to the graph, there were more newly recognized patients with symptomatic illnesses in the last month of December. A higher rate of liver cancer incidentally occurs in the months of May and August. In every month but January, the symptomatic manner of presentation predominates. The overall pattern of the graph shows that there is a peak in the number of diagnoses in the month of March.

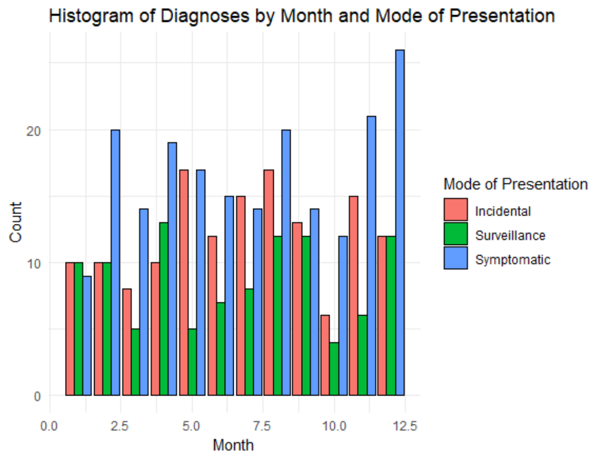


Figure 4.5: Mode of Diagnosis by Month and Mode of Presentation

The pie chart displays the distribution of event factors by year. The graph depicts how many people were alive and perished during the pandemic and pre-pandemic periods. The death rate increased somewhat during the pandemic compared to the pre-pandemic period, as seen by the graph in Figure 4.6. The COVID-19 pandemic is probably to blame for the discrepancy between the two pie charts. Worldwide, the number of fatalities has increased significantly as a result of the COVID-19 epidemic. There are many reasons for this higher level during the outbreak. Studies show that the COVID-19 infection and disturbance of the medical system both increase the rate. There are more people alive than dead, as evidenced by the pre-pandemic pie chart.

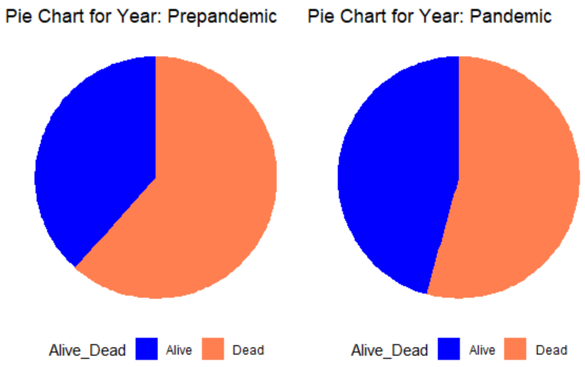


Figure 4.6: Alive/Death Status in Pandemic and Pre-pandemic time

According to the scatter plot in Figure 4.7, relative to their younger and middle-aged contemporaries, people in their 50s and 60s are more susceptible to developing cancer. There are more cancer patients with big tumors in the 65 to 80 age group. The graph also demonstrates that there is a difference in tumor size between cancer patients and non-cancer individuals. Cancer patients typically have larger tumors than non-cancer patients. This is probably because cancer cells develop more quickly than healthy cells. Additionally, some elderly patients exhibit modest tumor sizes. The relationship between age and size is favorable.

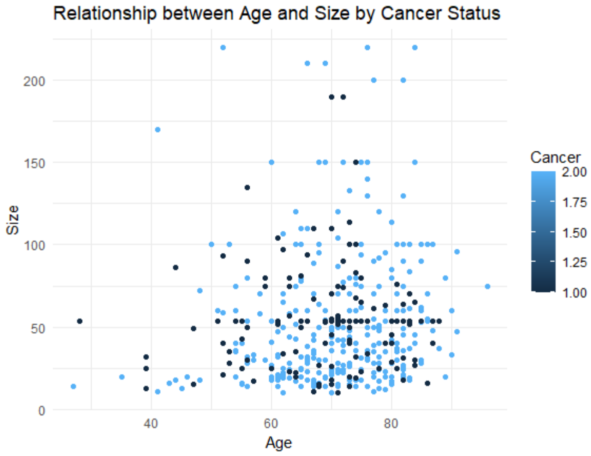


Figure 4.7: Relationship between Age and Size by Cancer Status

4.3.3 Outlier Detection

The data contains a few outliers, or data points that diverge noticeably from the rest of the data. In the dataset, outliers can be found using a variety of techniques. Figure 4.8 shows the box plot used in this study to identify outliers in the dataset. There are a few outliers in the data points for the month and size variables, as seen in the plot. We must remove the outliers from the size variable because it contains more of them. The outliers in this data set are handled according to the standard deviation values because the data is not normally distributed. Z-Score techniques are applied in this work to handle the outliers. This method views data points that differ from the mean by more than three SDs as outliers. After eliminating the mean from each data point, the z-score is calculated by first dividing the value by the standard deviation. If a data point is equal to the mean, it has a z-score of 0, and if it deviates one standard deviation from the mean, it has a z-score of 1. A data point that is one standard deviation from the mean has a z-score of -1. The data that has been

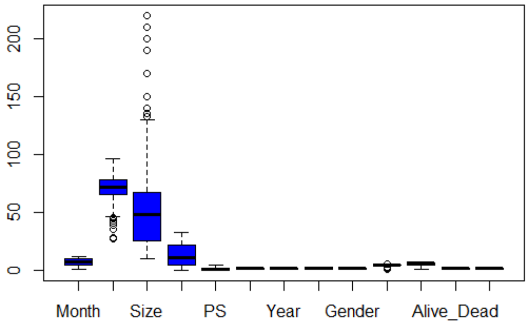


Figure 4.8: Box Plot for Outlier Detection

outlier-treated and deleted is then used for additional processing. Using Huber's method, the outlier-treated data was checked for outliers, and it is now outlier-free. This method calculates the dataset's median before calculating the median absolute deviation to account for data variability. To regulate the sensitivity of outliers, there is a tuning constant (k). The Z score is then determined using the method for each data point. The point is regarded as an outlier if the Z score is greater than the k value.

4.3.4 Categorical Value Encoding

There are only 13 features total in the data set after resolving missing values, of which 8 are categorical. Before entering categorical features into the COX proportional hazard model for model training, they must be converted into numerical format. The label encoding technique transforms category values into a numerical format for this purpose. of this method, each category of a categorical variable is given different interval values. Then it swaps out the category values in the dataset for their corresponding integer labels. The category variable is best handled by this method. The techniques used for label encoding rely on an ordinal relationship between the categories. Each category has a certain meaning or arrangement. It makes the data less dimensional than with one-hot encoding. In comparison to other encoding techniques, it requires less memory while processing huge datasets.

4.3.5 Feature Selection

The preprocessed dataset is used in the correlation analysis. It is a statistical method for determining the relationship between variables in a dataset and for comprehending patterns and dependencies. Pearson's correlation coefficient, which has a -1 to +1 range, is used to calculate how dependent the dataset's features are on one another. A perfect positive correlation is represented by a value of 1, a perfect negative correlation by a value of 1, and a perfect zero correlation by a value of 0. Using a correlation matrix, it is plotted in Figure 4.9. It is a table that offers a thorough understanding of all the connections between all the pairs of variables by showing the correlation coefficients between various variables in a dataset.

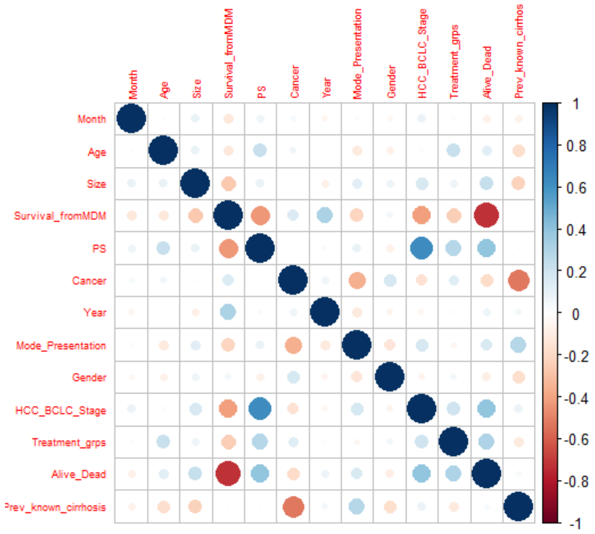


Figure 4.9: Correlation Matrix

4.3.6 Survival Analysis

KaplanMeier Survival Curves

The KaplanMeier Survival curve is used to visualize the survival probabilities over time for different features. This curve provides subsequent analysis and contributes to a more comprehensive understanding of the data's survival patterns before fitting the COX model. From this analysis, we can understand the certain impact of specific features. These curves are used to compare different groups and potential trends. For this initially, the time variable and event variable are defined as survival data. Here the event variable is Alive_Dead and the time variable is Survival_fromMDM.The KaplanMeier survival curves for two patient groups, one with liver cancer diagnosed during the pandemic (red line) and the other diagnosed during the pre-pandemic period (blue line) are depicted in the plot in Figure 4.10. The survival likelihood in both scenarios declines over time, according to the graphic. However, the group diagnosed during the prepandemic period has a better chance of surviving than the group diagnosed during the pandemic period. Accordingly, those who are given a liver cancer diagnosis before a pandemic have a better chance of surviving than those who receive it after one.

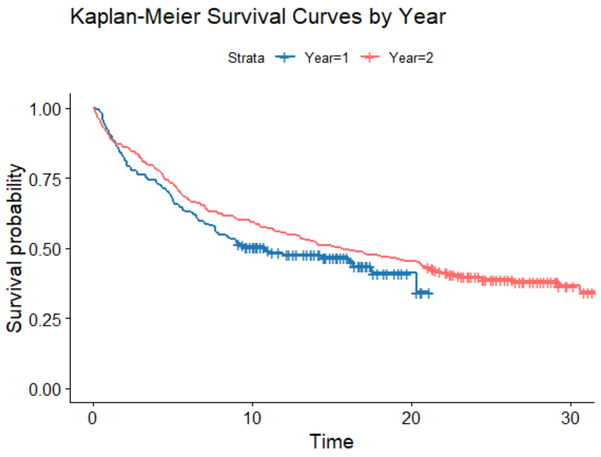


Figure 4.10: Kaplan-Meier Survival Curve by Year

The Kaplan-Meier survival curves for two patient groups with liver cancer diagnosed during the pandemic (red line) and the other diagnosed during the pre-pandemic period (blue line) are depicted in the plot in Figure 9. The survival likelihood in both scenarios declines over time, according to the graphic. However, the group diagnosed during the pre-pandemic period has a better chance of surviving than the group diagnosed during the pandemic period. Accordingly, those who are given a liver cancer diagnosis before a pandemic have a better chance of surviving than those who receive it after one.

The plot in Figure 4.11 shows the KaplanMeier survival curves for three groups of patients, one with the mode of presentation as incidental (red line), one with the mode of presentation symptomatic (blue line), and one with the mode of presentation surveillance (green line). The survival probability of all groups decreased over time. However, the survival probability for the group with the mode of presentation incidental is higher than the survival probability for the groups with the mode of presentation symptomatic and surveillance. This means that patients with a mode of presentation incidental have a better chance of survival than patients

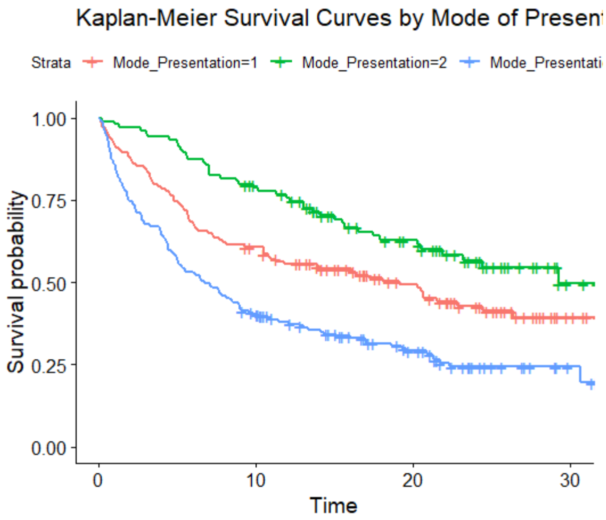


Figure 4.11: Kaplan-Meier Survival Curve by Mode of Presentation

with a mode of presentation symptomatic and surveillance due to several factors.

LogRank Test

Before fitting the COX proportional hazard model, the logrank test provides numerous benefits in the initial screening of characteristics and associations. The test results give a basic grasp of the traits and patterns of the data. It aids in identifying any additional survival disparities that demand attention. The log-rank test may identify substantial associations that merit additional investigation if it finds significant differences. It's employed to evaluate statistical significance. The logrank test is used in the current situation to examine the statistical significance of cancer, presentation mode, and treatment groups for the variable.

According to the logrank test results illustrated in Figure A.3, the log-rank test statistic for the cancer feature is 18.5, and it has 1 degree of freedom. The test's p-value is $p = 2 \times 10^{-5}$, which is substantially lower than the usually accepted significance level of 0.05. Therefore, the null hypothesis is strongly rejected by the evidence. The two groups are significantly different in terms of their chances of surviving, according to the "cancer" variable. It implies that the cancer variable affects a patient's chance of surviving a pandemic or pre-pandemic period. The logrank test statistic for the Mode of Presentation variable is 42.4 and has two degrees of freedom. The related p-value for the test is $p = 6 \times 10^{-10}$, which is below the cutoff. The "Mode_Presentation" variable affects survival outcomes, and there is sufficient evidence to reject the null hypothesis. The output is referred to in Figure A.3. The logrank test result on the variable "Treatment_grps" provides strong support for the conclusion that there are differences in the survival outcomes across the seven categories of the variable. It affects the outcome of survival, such as death.

CHAPTER 5

Result Analysis and Interpretation

5.1 Model Training and Evaluation

For training and evaluation of the model, the Cox proportional hazards regression model is fitted using the "Coxph" function from the survival package in R. The following Table 5.1 contains the formula and model summary of the fitted model. The given formula includes various predictor variables such as "Month", "Year", "Age", "Size", "PS"(Performance Status), "Cancer", "Mode_Presentation", "Gender", "HCC_BCLC_Stage" (Hepatocellular Carcinoma BCLC Stage), "Treatment_grps", and "Prev_known_cirrhosis", .These variables are used to predict the survival time represented by the surv_data variable, which is in the survival format, it includes the event and time variable created using the "Surv" function. For each predictor variable, there is a coefficient estimate of "coef", exponentials of the coefficients "exp(coef)", standard errors "se(coef)", z-scores, and corresponding p-values.

The exponentials of the coefficients are known as hazard ratios. A hazard ratio greater than 1 indicates an increased risk of the event occurring, while a hazard ratio less than 1 indicates a decreased risk. The detailed output is depicted in Figure A.1. Based on the results variables such as "Age", "Size", "PS", "Cancer", "HCC_BCLC_Stage" and "Treatment_grps" have an increased risk of event occurring. The statistical significance of the variable is associated with p values. Lower p-values typically < 0.05 suggest that the variable has a significant impact on survival. From the above-mentioned variables, expect the age

Feature Name	Hazard Ratio	P Values
Month	0.968171	0.09089
Year	0.939132	0.63886
Age	1.004819	0.47491
Size	1.008654	4.40e-08
PS	1.443704	3.53e-06
Cancer	0.437871	3.01e-06
Mode-Presentation	1.134153	0.12436
Gender	0.811393	0.13876
HCC-BCLC-Stage	1.484814	0.00102
Treatment-groups	1.311776	4.18e-08
Prev_known_cirrhosis	0.899909	0.54442

Table 5.1: The Cox proportional hazards regression model summary

variable to be statistically significant and it influences the survival outcome of the event of death. The "lower .95" and "upper .95" columns specify the lower and upper bounds of the 95% confidence interval for the hazard ratios. This interval provides a range within which the true hazard ratio lies. There is also a Concordance value, that specifies the measure of the model's predictive accuracy. A higher value closer to 1 indicates better performance accuracy. Here, the concordance value is 0.756.

The Cox proportional hazards regression model also includes some model evaluation test results such as the Likelihood ratio test, the Wald test, and the Score (log-rank) test. In all these tests the p-value is less than the 0.05 threshold level, which indicates that the model, and the individual variables, are highly statistically significant predictors of survival outcomes. Overall, the given model has a good fit and predictive accuracy, as indicated by the concordance index and test p-values. The predictor variables such as "Age", "Size", "PS", "Cancer", "HCC_BCLC_Stage", and "Treatment_grps" a significant impact on the hazard of the event of death occurring.

5.2 Further Analysis

To investigate the specific impact of each variable on the year variable, the Cox proportional hazards regression model with interaction terms is fitted. Analyzing how the specific characteristics affect when a person dies both during and before a pandemic using the COX model with interaction terms is helpful. The output and summary of the model are shown in Table 5.2. In this model, factors like the year of observation, age, tumor size, presence of cancer, disease stage, performance status, treatment groups, mode of presentation, previous cirrhosis, and gender are investigated with their association with the Year variable. These interactions imply that each predictor variable's impact on survival may vary over time, such as during pandemics and pre-pandemics. The model provides some values on coefficients, p-values, confidence intervals, and other pertinent information on the model's fit after being fitted.

The dataset for the provided model contains a total of 450 observations, of which 264 observations are death events that occurred. According to the model output given in Figure A.2, all variable-interaction terms for survival, apart from Year: Cancer, are not statistically significant. The interaction between "Year" and other factors except "Cancer" is not statistically significant. This suggests that the impact of the year on survival does not significantly differ based on whether an individual has previously known cirrhosis, tumor size, cancer stage, treatment group, gender, and age. Because of its statistical significance, the relationship between "year" and "cancer" has a Hazard Ratio of 3.249588 and a p-value of 0.00312. If p is a very small value, it means that the influence of year on survival is dependent on the presence of cancer. For every unit increase in a Year, the risk of the incident rises by a factor of about 3.25, especially for people with Cancer. The concordance value of 0.758 indicates the model performance in terms of the accuracy of prediction. In addition to this, the model output also gives model evaluation test results in terms of the Likelihood ratio test, the Wald test, and the Score log-rank test to check the performance of the predicted model in terms of accuracy. Overall, the lower p-value such as less than the threshold value of 0.05 indicates that the model is a good fit to the data.

Feature Name	Hazard Ratio	P Values
Year	0.160373	0.47491
Age	0.996353	0.87882
Size	1.011948	0.02397
Cancer	0.061586	5.87e-05
HCC-BCLC-Stage	2.131111	0.16381
PS	1.521602	0.12210
Treatment-groups	1.599947	0.01820
Mode-Presentation	1.064061	0.82285
Prev_known_cirrhosis	0.333428	0.12095
Gender	0.780180	0.62596
Year:Age	1.004687	0.74059
Year:Size	0.997993	0.53795
Year:Cancer	3.249588	0.00312
Year:HCC-BCLC-Stage	0.838623	0.55484
Year:PS	0.948717	0.74572
Year:Treatment-groups	0.888241	0.29112
Year:Mode-Presentation	1.050486	0.76898
Year:Prev_known_cirrhosis	1.785549	0.14533
Year:Gender	1.047212	0.87637

Table 5.2: The Cox proportional hazards regression model summary

5.3 Results Interpretation and Insights

5.3.1 Analysis of Covid19 Impact on Liver Cancer

The objectives of this study are to investigate the significant characteristics and effects of the pandemic and pre-pandemic time when the COVID-19 pandemic occurred on liver cancer based on the coefficients, hazard ratios, and statistical significance of each variable in the COX proportional hazard model, and to understand the variables, particularly the "year" variable, that affect the amount of time until an event (such as death). When interpreting and evaluating the model's findings, the coefficient of -0.062800 and p-value of 0.63886 for the variable "year" show that it has no considerable influence on the event's risk. The log hazard of the occurrence reduces by approximately 0.062800 units with each extra year, according to a negative coefficient like -0.062800. This indicates a decreasing trend of the hazard occurring.

The output of the model indicates that for the "month" variable, the coefficient of -0.032347 with a p-value of 0.09089 suggests that the hazard of the event lowers by around 3.2% for each unit increase in the "month" variable. But because the p-value is higher than 0.05, it's possible that the association is not statistically significant. "Age" has a coefficient of 0.004808. This suggests that the risk of the incident rises by roughly 0.5% for each additional year of age. The p-value 0.47491 is greater than 0.05, indicating that this effect is not statistically significant. It has a value of 0.008616 for the "Size" variable. This shows that the risk of the incident increases by roughly 0.9% for each unit increase in size. The p-value is lower than 0.05, which shows that this effect is statistically significant. The variable "PS" has a coefficient of 0.367212. This implies that the risk rises by roughly 36.7% for every unit that improves its performance status. "Cancer" has a coefficient of -0.825830. This suggests that the risk is much lowered when cancer is present. The risk is roughly 56.2% lower for individuals without cancer than for those with cancer. This effect is statistically quite important. It might be due to the extra care and meditation given to cancer patients than a normal person.

"Mode_Presentation" has a coefficient of 0.125886 with a p-value of 0.12436 and it is not statistically significant. The coefficient for "Gender" is -0.209002 and a p-value of 0.13876, this effect is not statistically significant, nonetheless. "HCC_BCLC_Stage" has a coefficient of 0.395289. The risk rises by roughly 39.5% for each unit increase in the Cancer stage. Statistics show that this effect is significant. "Treatment_grps" has a coefficient of 0.271382. This suggests that different treatment groups had different outcomes after accounting for other characteristics. Finally, when comes to the variable "previous_known_cirrhosis" has a coefficient of -0.105462. This suggests that having cirrhosis that has already been diagnosed is related to a marginally decreased risk, but it is not statistically significant with a p-value of 0.54442. Overall, several variables like "Size", "PS", "Cancer", and "HCC_BCLC_Stage" seem to be strong predictors of the event hazard, as they have statistically significant coefficients and lower p values. The results prove that these variables are the significant factors for the event occurring such as death.

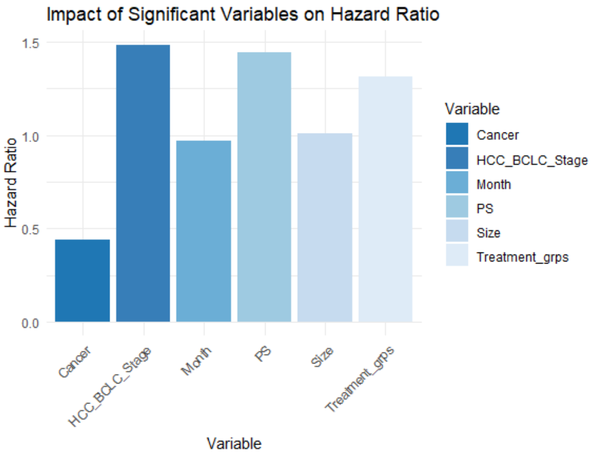


Figure 5.1: Impact of Significant Variables with Hazard Ratio

The above graph in Figure 5.1 describes the hazard ratio level for each significant variable based on model output. The graph shows that the hazard ratio for cancer is less than 0.5, which means that patients with liver cancer are 50% less likely to die than patients without liver cancer during the period of study. The hazard ratio for HCC-BCLC-Stage is 1.2, which means that patients with a more advanced stage of cancer are 20% more likely to die than patients with an earlier stage of cancer. The hazard ratio for the month is 0.8, which means that patients who have been diagnosed with liver cancer for a longer time are 20% less likely to die than patients who have been diagnosed for a shorter time. The hazard ratios for PS, size, and treatment_grps are not statistically significant, which means that these variables do not have a significant impact on the hazard ratio even if these variables have a higher hazard ratio with lower p values.

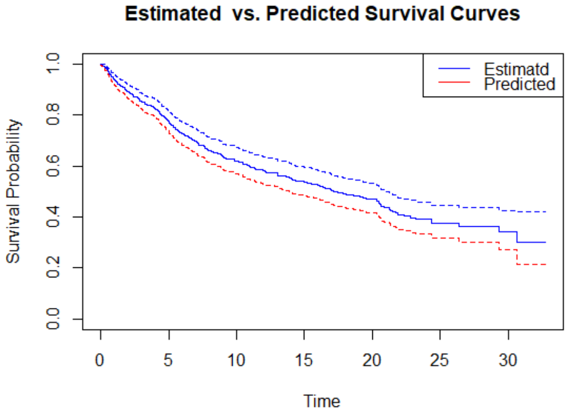


Figure 5.2: Survival curve for the COX model

The blue curve in this survival plot represents the estimated survival probability, whereas the red curve represents the predicted survival probability. According to the graph in Figure 5.2, the predicted survival curve is shorter than the one that was estimated. This indicates that the survival rate of the study's patients was lower than what the statistical model predicted. We can infer from this curve that the death rate somewhat increased during the relevant timeframe because of factors like "Age", "Size", "PS", "Cancer", "HCC_BCLC_Stage" and "Treatment_grps".At the end of the time frame, the predicted curve is steady and fluctuating compared with a gradual drop in the survival rate during the initial time frame. When comparing the first five months the average survival rate is 70%, however in the last 5-month time the average survival rate dropped to 20%.

5.3.2 Identification of Significant Features

The COX proportional hazard model with interaction term is used to examine individual features' effect on COVID-19 during the pandemic and pre-pandemic periods. The interactions between the Year variable and other variables in this model capture how the impacts of the other variables alter during this time. The p-value for "Year: Age" is 0.74059, while the coefficient for "Year: Age" is 0.004676. This interaction word describes how "year" and "age" together have an impact on the event's risk. According to the coefficient, the risk of the incident rises by around 0.004676 units for every additional year for every new unit of age. The p-value is more than 0.05, the interaction between "year" and "age" doesn't significantly affect the event's risk. "Year: Size" has a coefficient of -0.002009 and a p-value of 0.53795. This interaction term describes how the combined influence of "year" and "size" affects the event's risk. The coefficient indicates that for a one-unit increase in size, the hazard of occurrence lowers by around 0.009 units in pandemic and pre-pandemic time. The p-value is more than 0.05, it indicates that the interaction between "year" and "size" has no appreciable effect on the event's danger.

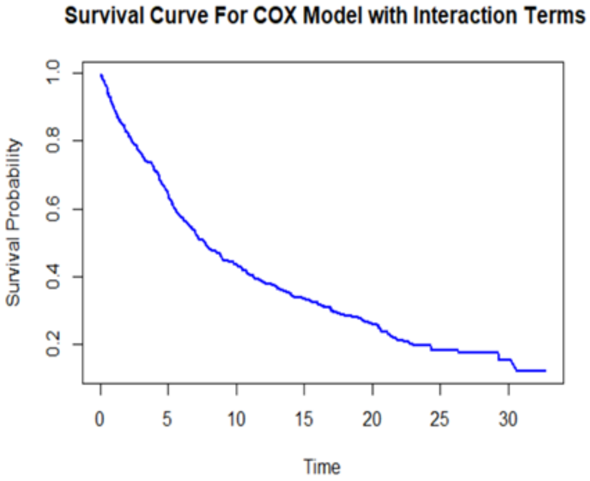


Figure 5.3: Survival Curve For COX Model with Interaction Terms

The coefficient for "Year: Cancer" is 1.178528 and the p-value is 0.00312. The positive coefficient shows that, specifically for people with cancer, the risk of occurrence rises by approximately 1.178528 units for every additional year. The p-value is less than 0.05, indicating that this rise is statistically significant. In other words, the relationship between the variables "year" and "cancer" significantly affects the risk of occurrence for those who have cancer. The p-value is 0.55484, while the coefficient for "Year: HCC_BCLC_Stage" is -0.175994. The coefficient indicates that with a one-unit increase in HCC_BCLC_Stage, the hazard of occurrence falls by approximately 0.175994 units for each additional year. The p-value is more than 0.05, hence this drop is not statistically significant. This indicates that the interaction between "year" and "HCC_BCLC_Stage" has no appreciable bearing on the event's hazard.

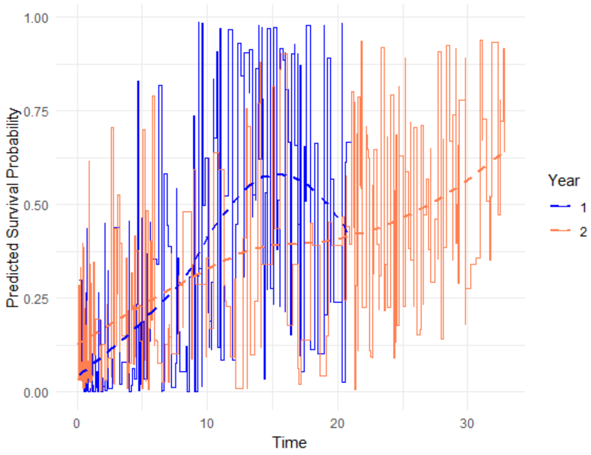


Figure 5.4: The COX Interaction Model survival probability plot for pandemic and pre-pandemic

The survival probability significantly decreased, as seen by the survival curve shown in Figure 5.3. A person's chance of surviving peaked at 10% near the end of the time. The likelihood of survival began to drastically decrease below 40% in the middle of the time. The model predicts survival probability for various years, including pandemic and pre-pandemic years, as well as how they change over time, as shown in the graph in Figure 5.4. The pandemic and pre-pandemic periods are represented by Years 1 and 2, respectively. The graph shows that from the pre-pandemic era, the likelihood of survival has grown. Although the likelihood of survival rose initially throughout the outbreak, it fell in the middle and eventually peaked at about 0.35. This indicates how COVID-19 affected the pandemic period, which explains why the overall trend is declining in contrast to an uptick in survival probability prior to the pandemic.

Figure 5.5 shows the dot plot of the p-values for the COX model with the interaction term. The graph shows that the variables cancer, year: cancer, treatment_grps, and size have a lower p-value of less than 0.05 and are statistically significant. This suggests that characteristics such as tumor size, liver cancer, and treatment group have a substantial impact on raising the event's risk. In comparison to others, those with cancer and larger tumors tend to pass away sooner.

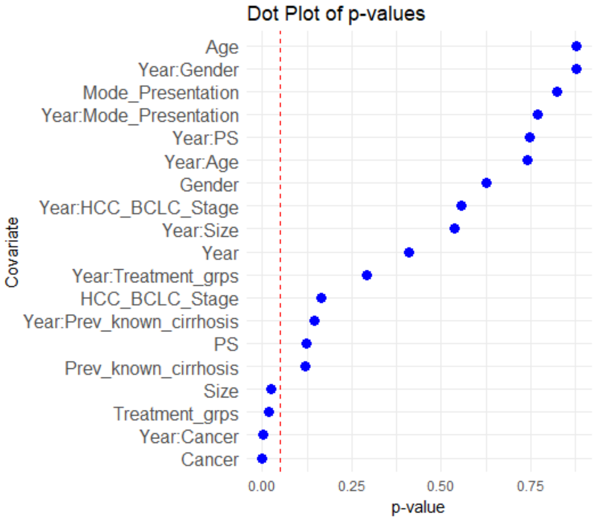


Figure 5.5: Dot Plot of p-Values for COX Interaction Model

CHAPTER 6

Conclusions

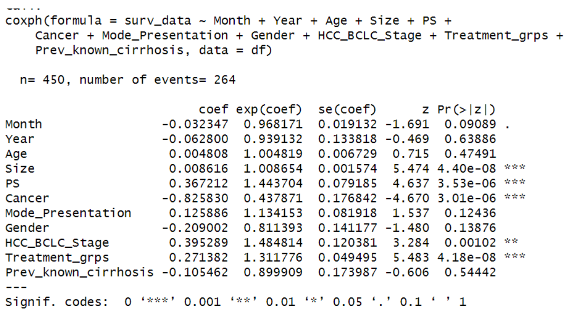
In conclusion, this study offers important new information about the relationship between COVID-19 and liver cancer survival. The study's main goal is to investigate how COVID-19 affects those who have just received a liver cancer diagnosis. The study also sought to examine how different patient traits and circumstances affected survival rates during the pandemic and pre-pandemic periods.

This study covers several significant insights. The analysis's first findings suggest that COVID-19 had no direct effect on patients with liver cancer's prognosis. This implies that patients with liver cancer did not have a worse prognosis due to the presence of COVID-19. However, there are some patient traits that have an impact on survival rates. It was discovered that characteristics such as the presence of cancer, the type of treatment used, and the size of the tumor had a substantial impact on the hazard ratio related to the event of death. These variables become significant risk factors, underscoring their significance in relation to patient outcomes. Additionally, the study discovered that the chance of survival was lower during the pandemic period when comparing the pre-pandemic and pandemic periods. The cumulative effect of different patient factors, including cancer, treatment group, and tumor size, can be used to explain this decrease. COVID-19 accepts that other traits and conditions played a more significant role in influencing survival outcomes throughout this exceptional period and that it has no direct effects on survival.

This thorough investigation is supported by information from a single hospital. Due to the complexity of the pandemic’s impact, it’s possible that the complete understanding of COVID-19’s consequences hasn’t been fully recorded; additional research is required to fully comprehend the effect. However, these discoveries advance knowledge of the complex interactions between internal and external elements that affect patient experiences during unanticipated pandemics in the future.

APPENDIX A

Model Output and Code

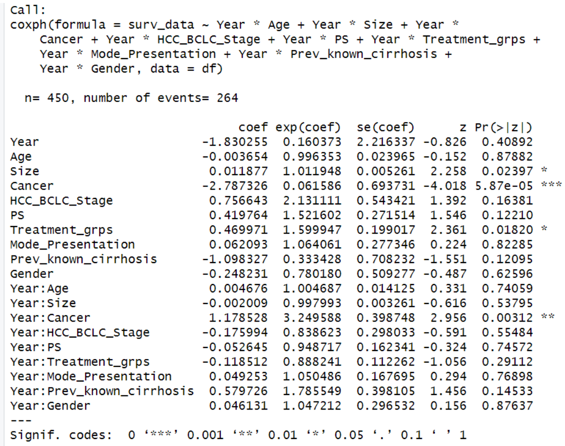


```
coxph(formula = surv_data ~ Month + Year + Age + Size + PS +
    Cancer + Mode_Presentation + Gender + HCC_BCLC_Stage + Treatment_grps +
    Prev_known_cirrhosis, data = df)

  n= 450, number of events= 264 

                          coef exp(coef)  se(coef)      z Pr(>|z|)    
Month                -0.032347  0.968171  0.019132 -1.691  0.09089 .  
Year                 -0.062800  0.939132  0.133818 -0.469  0.63886    
Age                   0.004808  1.004819  0.006729  0.715  0.47491    
Size                  0.008616  1.008654  0.001574  5.474 4.40e-08 ***
PS                    0.367212  1.443704  0.079185  4.637 3.53e-06 ***
Cancer               -0.825830  0.437871  0.176842 -4.670 3.01e-06 ***
Mode_Presentation     0.125886  1.134153  0.081918  1.537  0.12436    
Gender               -0.209002  0.811393  0.141177 -1.480  0.13876    
HCC_BCLC_Stage        0.395289  1.484814  0.120381  3.284  0.00102 ** 
Treatment_grps        0.271382  1.311776  0.049495  5.483 4.18e-08 ***
Prev_known_cirrhosis -0.105462  0.899909  0.173987 -0.606  0.54442    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.1: COX Proportional Hazard Model Output

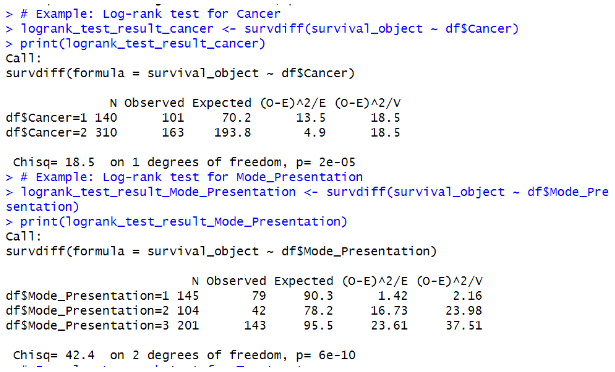


```
Call:
coxph(formula = surv_data ~ Year * Age + Year * Size + Year *
    Cancer + Year * HCC_BCLC_Stage + Year * PS + Year * Treatment_grps +
    Year * Mode_Presentation + Year * Prev_known_cirrhosis +
    Year * Gender, data = df)

  n= 450, number of events= 264 

                               coef exp(coef)  se(coef)      z Pr(>|z|)    
Year                      -1.830255  0.160373  2.216337 -0.826  0.40892    
Age                       -0.003654  0.996353  0.023965 -0.152  0.87882    
Size                       0.011877  1.011948  0.005261  2.258  0.02397 *  
Cancer                    -2.787326  0.061586  0.693731 -4.018 5.87e-05 ***
HCC_BCLC_Stage             0.756643  2.131111  0.543421  1.392  0.16381    
PS                         0.419764  1.521602  0.271514  1.546  0.12210    
Treatment_grps             0.469971  1.599947  0.199017  2.361  0.01820 *  
Mode_Presentation          0.062093  1.064061  0.277346  0.224  0.82285    
Prev_known_cirrhosis      -1.098327  0.333428  0.708232 -1.551  0.12095    
Gender                    -0.248231  0.780180  0.509277 -0.487  0.62596    
Year:Age                   0.004676  1.004687  0.014125  0.331  0.74059    
Year:Size                 -0.002009  0.997993  0.003261 -0.616  0.53795    
Year:Cancer                1.178528  3.249588  0.398748  2.956  0.00312 ** 
Year:HCC_BCLC_Stage       -0.175994  0.838623  0.298033 -0.591  0.55484    
Year:PS                   -0.052645  0.948717  0.162341 -0.324  0.74572    
Year:Treatment_grps       -0.118512  0.888241  0.112262 -1.056  0.29112    
Year:Mode_Presentation     0.049253  1.050486  0.167695  0.294  0.76898    
Year:Prev_known_cirrhosis  0.579726  1.785549  0.398105  1.456  0.14533    
Year:Gender                0.046131  1.047212  0.296532  0.156  0.87637    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure A.2: COX Proportional Hazard Model with Interaction Term Output



```
> # Example: Log-rank test for Cancer
> logrank_test_result_cancer <- survdiff(survival_object ~ df$Cancer)
> print(logrank_test_result_cancer)
Call:
survdiff(formula = survival_object ~ df$Cancer)

              N Observed Expected (O-E)^2/E (O-E)^2/V
df$Cancer=1 140      101     70.2      13.5      18.5
df$Cancer=2 310      163    193.8       4.9      18.5

 Chisq= 18.5  on 1 degrees of freedom, p= 2e-05 
> # Example: Log-rank test for Mode_Presentation
> logrank_test_result_Mode_Presentation <- survdiff(survival_object ~ df$Mode_Pre
sentation)
> print(logrank_test_result_Mode_Presentation)
Call:
survdiff(formula = survival_object ~ df$Mode_Presentation)

                          N Observed Expected (O-E)^2/E (O-E)^2/V
df$Mode_Presentation=1 145       79     90.3      1.42      2.16
df$Mode_Presentation=2 104       42     78.2     16.73     23.98
df$Mode_Presentation=3 201      143     95.5     23.61     37.51

 Chisq= 42.4  on 2 degrees of freedom, p= 6e-10 
```

Figure A.3: Log Rank Test Output

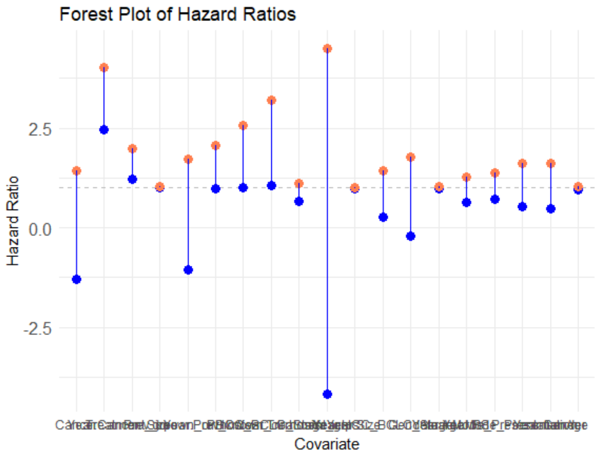


Figure A.4: Forest Plot of Hazard Ratios for COX Model with Interaction Term

Bibliography

[1] Li, P. Liu, Y. Cheng, Z. Yu, X., and Li, Y. COVID-19-associated liver injury: Clinical characteristics, pathophysiological mechanisms and treatment management *Biomed Pharmacother, Oct 2022.*

[2] *Marjot T, Eberhardt CS, Boettler T, Belli LS, Berenguer M, Buti M, Jalan R, Mondelli MU, Moreau R, Shouval D, Berg T, Cornberg M. Impact of COVID-19 on the liver and on the care of patients with chronic liver disease, hepatobiliary cancer, and liver transplantation: An updated EASL position paper* Epub 2022 Jul 20. PMID: 35868584; PMCID: PMC9296253.

[3] Hu X, Sun L, Guo Z, Wu C, Yu X, Li J. Management of COVID-19 patients with chronic liver diseases and liver transplants. *Ann Hepatol. 2022 Jan-Feb;27(1):100653. doi: 10.1016/j.aohep.2021.100653. Epub 2021 Dec 18. PMID: 34929350; PMCID: PMC8683212.*

[4] *Munoz Martinez, S. and Sapena, V. and Forner, A. and Nault, J.C. and Sapisochin, G. and Rimassa, L. Sangro, B. and Bruix, J. and Sanduzzi-Zamparelli, M. and El Kassas. Assessing the Impact of COVID-19 on Liver Cancer Management* Eupub 2021 Feb 23.

[5] Lau H, Fan ST, Ng IO, Wong J. Long-term prognosis after hepatectomy for hepatocellular carcinoma: a survival analysis of 204 consecutive patients. *Cancer. 1998 Dec 1;83(11):2302-11. PMID: 9840529.*

[6] *Chan SL, Kudo M. Impacts of COVID-19 on Liver Cancers: During and after the Pandemic.* Epub 2020 Sep 1. PMID: 33078127; PMCID: PMC7490489.

[7] Atlam M, Torkey H, El-Fishawy N, Salem H. Coronavirus disease 2019 (COVID-19): survival analysis using deep learning and Cox regression model. *Epub 2021 Feb 15. PMID: 33613099; PMCID: PMC7883884.*

[8] *Tsiatis, A.A. A Large Sample Study of COX Regression Model.* Annals of Statistics, 9, 93-108, 1981.

[9] Stel VS, Dekker FW, Tripepi G, Zoccali C, Jager KJ. Survival analysis II: Cox regression. *Nephron Clin Pract*, 2011;119(3):c255-60..

[10] Li D, Jia AY, Zorzi J, Griffith P, Kim AK, Dao D, Anders RA, Georgiades C, Liddell RP, Hong K, Azad NS, Ho WJ, Baretti M, Christenson E. Impact of the COVID-19 Pandemic on Liver Cancer Staging at a Multidisciplinary Liver Cancer Clinic. *Ann Surg Open*, 2022 Oct.

[11] Munoz-Martinez S, Sapena V, Forner A, et al. Outcome of liver cancer patients with SARS-CoV-2 infection: An International, Multicentre, Cohort Study. *Liver Int*, 2022;42(8):1891-1901.

[12] Bewick V, Cheek L, Ball J Statistics review 12: survival analysis. *Mind*, 59:433–460, 1950.

[13] Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CGM. Time-varying covariates and coefficients in Cox regression models. *Ann Transl Med.*, 2018 Apr;6(7):121.

[14] Kariyawasam JC, Jayarajah U, Abeysuriya V, Riza R, Seneviratne SL. Involvement of the Liver in COVID-19: A Systematic Review.

Am J Trop Med Hyg., 2022 Feb 24;106(4):1026â41.

[15] Bradburn MJ, Clark TG, Love SB, Altman DG. CSurvival analysis Part III: multivariate data analysis – choosing a model and assessing its adequacy and fit. *Crit Care.*, 2004;8(5):389-394.

[16] Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CGM. Time-varying covariates and coefficients in Cox regression models. *Ann Transl Med.*, 2018 Apr;6(7):121.

[17] In J, Lee DK. Survival analysis: part II - applied clinical data analysis. *Korean J Anesthesiol.*,2023 Feb.