Causal survival analysis with applications in R
Practical Day 1

In this practical we will explore the data we have and conduct a descriptive analysis of the data

Trial context

The dataset is a simulated dataset using the IMbrave150 trial. This was a Phase III, Open-Label, Randomized Study of Atezolizumab in Combination With Bevacizumab Compared With Sorafenib in Patients With Untreated Locally Advanced or Metastatic Hepatocellular Carcinoma.

Patients were originally randomized in a 2:1 ratio but we simulated a 1:1 randomization ratio for simplicity.

Trial population

Inclusion Criteria:

- Locally advanced or metastatic and/or unresectable Hepatocellular Carcinoma (HCC)
- No prior systemic therapy for HCC. Previous use of herbal therapies/traditional Chinese medicines with anti-cancer activity included in the label is allowed, provided that these medications are discontinued prior to randomization.
- At least one measurable untreated lesion
- ECOG Performance Status of 0 or 1
- Adequate hematologic and end-organ function
- For women of childbearing potential: agreement to remain abstinent
- For men: agreement to remain abstinent
- Child-Pugh class A

Exclusion Criteria:

- History of leptomeningeal disease
- Active or history of autoimmune disease or immune deficiency
- History of idiopathic pulmonary fibrosis, organizing pneumonia, drug-induced pneumonitis, or idiopathic pneumonitis, or evidence of active pneumonitis on screening chest computed tomography scan
- Known active tuberculosis
- History of malignancy other than HCC within 5 years prior to screening, with the exception of malignancies with a negligible risk of metastasis or death
- Pregnancy or breastfeeding, or intention of becoming pregnant during study treatment or within at least 5 months after the last dose of atezolizumab, 6 months after the last dose of bevacizumab, or 1 month after the last dose of sorafenib
- Known fibrolamellar HCC, sarcomatoid HCC, or mixed cholangiocarcinoma and HCC
- Untreated or incompletely treated esophageal and/or gastric varices with bleeding or high-risk for bleeding
- A prior bleeding event due to esophageal and/or gastric varices within 6 months prior to initiation of study treatment.
- Moderate or severe ascites
- History of hepatic encephalopathy
- Co-infection of HBV and HCV
- Symptomatic, untreated, or actively progressing central nervous system (CNS) metastases
- Uncontrolled tumor-related pain
- Uncontrolled pleural effusion, pericardial effusion, or ascites requiring recurrent drainage procedures
- Uncontrolled or symptomatic hypercalcemia
- Treatment with systemic immunostimulatory agents
- Inadequately controlled arterial hypertension
- Prior history of hypertensive crisis or hypertensive encephalopathy
- Evidence of bleeding diathesis or significant coagulopathy
- History of intestinal obstruction and/or clinical signs or symptoms of GI obstruction including sub-occlusive disease related to the underlying disease or requirement for routine parenteral hydration
- Serious, non-healing or dehiscing wound, active ulcer, or untreated bone fracture
- Metastatic disease that involves major airways or blood vessels, or centrally located mediastinal tumor masses
- Local therapy to liver within 28 days prior to initiation of study treatment or non-recovery from side effects of any such procedure
- Chronic daily treatment with a non-steroidal anti-inflammatory drug (NSAID)

Outcomes

The study used many outcomes including Progression Free Survival (PFS), Overall Survival (OS) and Objective Response Rate (ORR). However, our focus for the first course will be on the overall survival to aviod any issues with competing events. Overall Survival (OS) in the Global Population with a time frame: From randomization to death from any cause up to the clinical cut off date (CCOD) of 36 months. OS was defined as the time from randomization to death from any cause.

Data structure

Variable	Definition
id	Participant ID
Treatment	Randomized treatment
Age	Age at randomization (in years)
Sex	Sex (male; female)
ChildPugh	Baseline Child-Pugh Score (Classes A-C (most severe))
gregion	Geographic region
miexhep	Baseline macrovascular invasion or extrahepatic spread of disease
alpha_feto	Baseline alpha-fetoprotein level
ecog_perf	Eastern Cooperative Oncology Group (ECOG) Performance Status
hc_cause	Cause of hepatocellular carcinoma

Installing R

https://cran.r-project.org/bin/

Installing R studio

https://posit.co/download/rstudio-desktop/

- other less popular option https://code.visualstudio.com

R Packages: Installing and loading

```
install.packages(tidyverse)
install.packages(magrittr)
```

```
library(tidyverse)
library(magrittr)
```

```
# Loading the dataset
data <- read.csv("dataset.csv", stringsAsFactors =TRUE)
```

Exploring the dataset/ descriptive analysis

Now we will start to understand the dataset

```
str(data)

View(data)

# How many patients in total?
nrow(data)

# How many patients in each group?
table(data$Treatment)

# Age
tapply(data$Age, data$Treatment, summary)

# Categorize age to intervals
range(data$Age)

hist(data$Age)

data$age_group <- data %$% ifelse(Age <= 40 , "< 40",
                                  ifelse( Age <= 50, "41-50",
                                   ifelse(Age <= 60, "51-60",
                                    ifelse(Age <= 70, "61-70",
                                     ifelse(Age <= 80, "71-80",
```

```
                                ifelse(Age > 80, "Above 80", NA))))))


table(data$Treatment, data$age_group)
```

Now try to replicate this for the following variables:

- Sex
- Baseline Child-Pugh Score
- Geographic region
- Baseline macrovascular invasion or extrahepatic spread of disease
- Baseline alpha-fetoprotein level
- Eastern Cooperative Oncology Group (ECOG) Performance Status
- Cause of hepatocellular carcinoma

Table one package

Table one package is one of the most convenient packages to summarize baseline characteristics

```
install.packages(tableone)
install.packages(kableExtra)

library(tableone)
library(kableExtra)
```

```
vars <- c("Age", "age_group","Sex" ,"ChildPugh",
          "gregion", "miexhep","alpha_feto",
          "ecog_perf","hc_cause")

non_normal <- "Age"

#Create the table
baseline_tab <- CreateTableOne(vars = vars, strata= c("Treatment"),
                  data= data,factorVars= c("age_group","Sex",
                    "ChildPugh", "gregion", "miexhep","alpha_feto",
                    "ecog_perf","hc_cause"),
```

```
                      includeNA = TRUE,
                       addOverall = T)

# Print the table ####

kableone(baseline_tab,booktabs = TRUE, longtable = T,nonnormal = non_normal) %>%
  kable_styling(full_width = NULL,latex_options = "repeat_header")

# Save the table
tab_base <- print(baseline_tab, nonnormal = non_normal,
                  quote = FALSE, noSpaces = TRUE,
                  printToggle = FALSE, test = FALSE)
```