SECTION A. DATA EXPLORATION DESCRIPTION

step 1: load and summarize the dataset

```
# Load required packages
library(survival)
library(survminer)

## Loading required package: ggplot2

## Loading required package: ggpubr

##
## Attaching package: 'survminer'

## The following object is masked from 'package:survival':
##
##     myeloma

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(readr)

# Load the dataset
survival_data <-
read_csv("C:\\Users\\pc\\Downloads\\survival_data_group2.csv")

## Rows: 200 Columns: 7
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## dbl (7): id, time, status, age, sex, treatment, comorbidity
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

# Preview the data
head(survival_data)
```

```
## # A tibble: 6 × 7
##      id  time status   age   sex treatment comorbidity
##   <dbl> <dbl>  <dbl> <dbl> <dbl>     <dbl>       <dbl>
## 1     1   4.3      1    57     0         1           4
## 2     2   5.9      1    59     1         0           1
## 3     3   2.6      1    72     1         0           2
## 4     4   4.2      1    72     0         0           1
## 5     5   2.5      1    81     1         1           2
## 6     6   3        1    85     0         0           2
```

```
# Summary of variables
summary(survival_data)
```

```
##        id              time            status           age       
##  Min.   :  1.00   Min.   : 0.100   Min.   :0.000   Min.   :30.00  
##  1st Qu.: 50.75   1st Qu.: 1.100   1st Qu.:1.000   1st Qu.:56.00  
##  Median :100.50   Median : 3.000   Median :1.000   Median :65.00  
##  Mean   :100.50   Mean   : 5.194   Mean   :0.955   Mean   :65.19  
##  3rd Qu.:150.25   3rd Qu.: 6.525   3rd Qu.:1.000   3rd Qu.:74.00  
##  Max.   :200.00   Max.   :24.000   Max.   :1.000   Max.   :96.00  
##       sex          treatment      comorbidity   
##  Min.   :0.000   Min.   :0.000   Min.   :0.00  
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:1.00  
##  Median :1.000   Median :0.000   Median :2.00  
##  Mean   :0.545   Mean   :0.495   Mean   :1.75  
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:3.00  
##  Max.   :1.000   Max.   :1.000   Max.   :8.00
```

```
# Structure of dataset
str(survival_data)
```

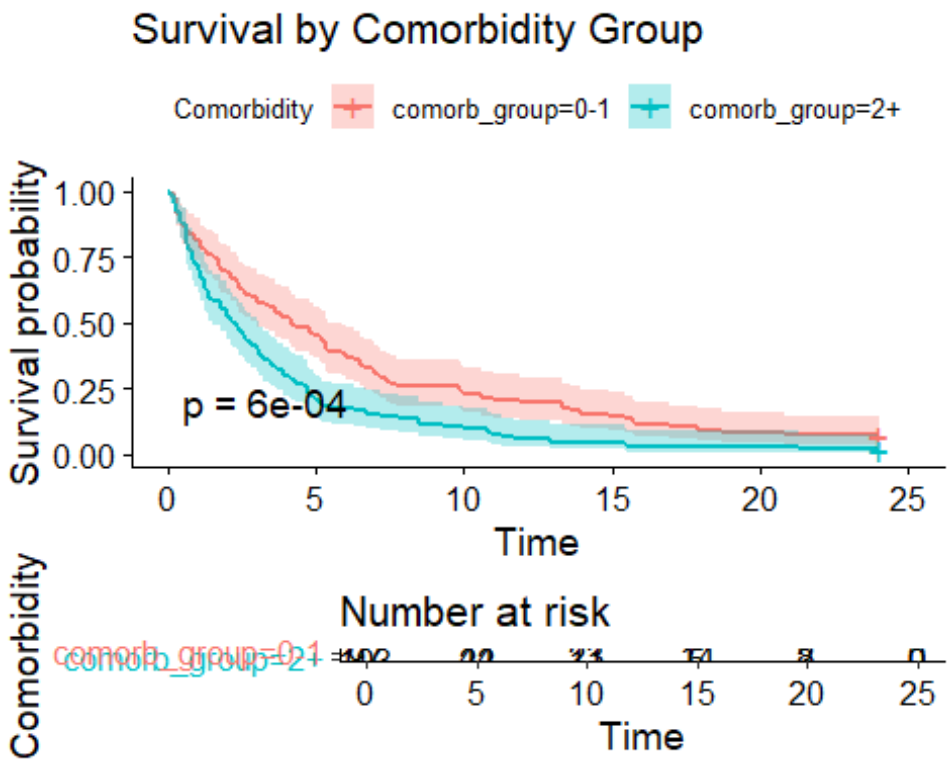
```
## spc_tbl_ [200 × 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ id         : num [1:200] 1 2 3 4 5 6 7 8 9 10 ...
##  $ time       : num [1:200] 4.3 5.9 2.6 4.2 2.5 3 0.4 2 3 1.4 ...
##  $ status     : num [1:200] 1 1 1 1 1 1 1 1 1 1 ...
##  $ age        : num [1:200] 57 59 72 72 81 85 69 70 61 61 ...
##  $ sex        : num [1:200] 0 1 1 0 1 0 0 1 0 0 ...
##  $ treatment  : num [1:200] 1 0 0 0 1 0 0 0 1 0 ...
##  $ comorbidity: num [1:200] 4 1 2 1 2 2 6 4 1 2 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   id = col_double(),
##   ..   time = col_double(),
##   ..   status = col_double(),
##   ..   age = col_double(),
##   ..   sex = col_double(),
##   ..   treatment = col_double(),
##   ..   comorbidity = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

step 2: median survival time by treatment

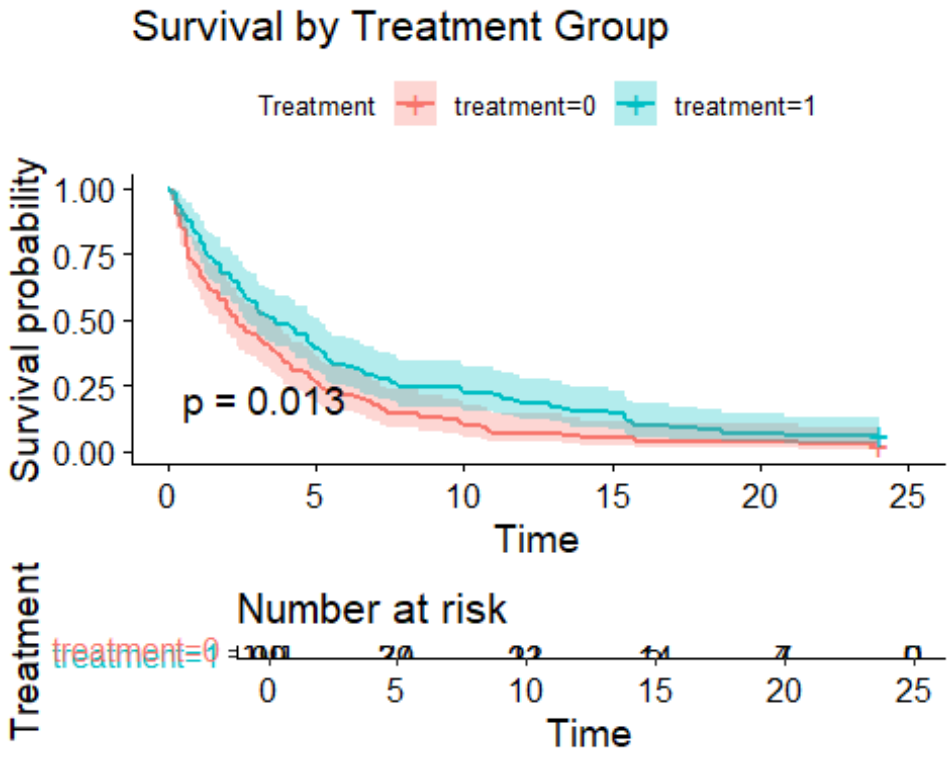
```
# Create survival object
surv_object <- Surv(time = survival_data$time, event = survival_data$status)

# Fit Kaplan-Meier by treatment
surv_fit_treatment <- survfit(surv_object ~ treatment, data = survival_data)

# Median survival time per group
summary(surv_fit_treatment)$table[, "median"]

## treatment=0 treatment=1
##         2.3         3.6
```

step 3: kaplan-meier curves by comorbidity and treatment

```
# Create comorbidity groups
survival_data <- survival_data %>%
  mutate(comorb_group = ifelse(comorbidity <= 1, "0-1", "2+"))

# KM by comorbidity
fit_comorb <- survfit(surv_object ~ comorb_group, data = survival_data)
ggsurvplot(fit_comorb, data = survival_data,
           pval = TRUE, conf.int = TRUE, risk.table = TRUE,
           title = "Survival by Comorbidity Group",
           legend.title = "Comorbidity")
```



```r
# KM by treatment
fit_treatment <- survfit(surv_object ~ treatment, data = survival_data)
ggsurvplot(fit_treatment, data = survival_data,
           pval = TRUE, conf.int = TRUE, risk.table = TRUE,
           title = "Survival by Treatment Group",
           legend.title = "Treatment")
```



SECTION B: SURVIVAL MODEL FITTING

step 1: fit cox model(age, sex, treatment)

```r
cox_model1 <- coxph(Surv(time, status) ~ age + sex + treatment, data =
survival_data)
summary(cox_model1)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ age + sex + treatment, data =
survival_data)
##
##   n= 200, number of events= 191
##
##                coef exp(coef)  se(coef)      z Pr(>|z|)
## age        0.016319  1.016453  0.005963  2.737  0.00620 **
## sex        0.150319  1.162205  0.150854  0.996  0.31903
## treatment -0.385089  0.680390  0.147018 -2.619  0.00881 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## age          1.0165     0.9838    1.0046    1.0284
## sex          1.1622     0.8604    0.8647    1.5620
## treatment    0.6804     1.4697    0.5101    0.9076
##
## Concordance= 0.59  (se = 0.022 )
## Likelihood ratio test= 16.57  on 3 df,   p=9e-04
## Wald test            = 16.15  on 3 df,   p=0.001
## Score (logrank) test = 16.3  on 3 df,   p=0.001
```

interpret the model output focusing on hazard ratios.

Age (HR = 1.016, p = 0.006)

#The hazard ratio greater than 1 indicates that increasing age is associated with an increased risk of readmission.

#Specifically, for every additional year of age, the hazard of readmission increases by 1.6%, holding sex and treatment constant.

#This effect is statistically significant, as the p-value is less than 0.05.

#Sex (HR = 1.162, p = 0.319)

#Males have a 16.2% higher hazard of readmission compared to females.

#However, this result is not statistically significant (p > 0.05), so we cannot conclude that sex has a meaningful effect on readmission in this model.

#Treatment (HR = 0.680, p = 0.009)

#Patients who received discharge planning treatment had a 32% lower hazard of readmission than those who did not.

#The hazard ratio less than 1 indicates a protective effect of the treatment.

#This effect is statistically significant, indicating that discharge planning is likely effective in reducing readmissions.

step 2: add comorbidity and compare models

```r
cox_model2 <- coxph(Surv(time, status) ~ age + sex + treatment + comorbidity,
data = survival_data)
anova(cox_model1, cox_model2, test = "LRT")
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(time, status)
##  Model 1: ~ age + sex + treatment
##  Model 2: ~ age + sex + treatment + comorbidity
##    loglik  Chisq Df Pr(>|Chi|)
## 1 -842.15
## 2 -833.21 17.875  1  2.359e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(cox_model2)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ age + sex + treatment +
##     comorbidity, data = survival_data)
##
##   n= 200, number of events= 191
##
##                  coef exp(coef)  se(coef)      z Pr(>|z|)
## age          0.017630  1.017786  0.006037  2.920   0.0035 **
## sex          0.268479  1.307974  0.154489  1.738   0.0822 .
## treatment   -0.334129  0.715961  0.148008 -2.258   0.0240 *
## comorbidity  0.235064  1.264989  0.053144  4.423 9.73e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             exp(coef) exp(-coef) lower .95 upper .95
## age             1.018     0.9825    1.0058    1.0299
## sex             1.308     0.7645    0.9663    1.7705
## treatment       0.716     1.3967    0.5357    0.9569
## comorbidity     1.265     0.7905    1.1399    1.4039
##
## Concordance= 0.625  (se = 0.022 )
## Likelihood ratio test= 34.44  on 4 df,   p=6e-07
## Wald test            = 34.58  on 4 df,   p=6e-07
## Score (logrank) test = 35.28  on 4 df,   p=4e-07
```

#A second Cox proportional hazards model was fitted by adding comorbidity to the initial model (which already included age, sex, and treatment).

#Model Fit Comparison (Likelihood Ratio Test):

#Hazard Ratio (HR) = 1.265,

#p-value < 0.001

#Interpretation: For each additional comorbid condition, the hazard of hospital readmission increases by 26.5%, holding other variables constant. This effect is statistically significant.

#model Fit Comparison (Likelihood Ratio Test): #Model without comorbidity: Log-likelihood = -842.15

#Model with comorbidity: Log-likelihood = -833.21

#Likelihood Ratio Chi-square = 17.875, p < 0.001

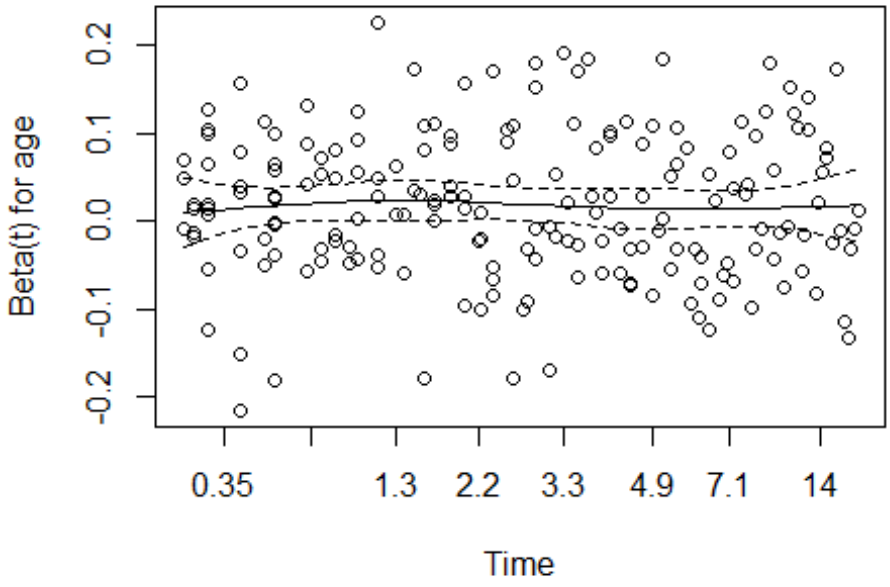
#Interpretation: #Adding the comorbidity variable significantly improves the model fit. The Likelihood Ratio Test shows a highly significant improvement (p < 0.001), indicating that comorbidity provides important explanatory power in predicting the time to readmission. Patients with more comorbidities are at a substantially higher risk of readmission.

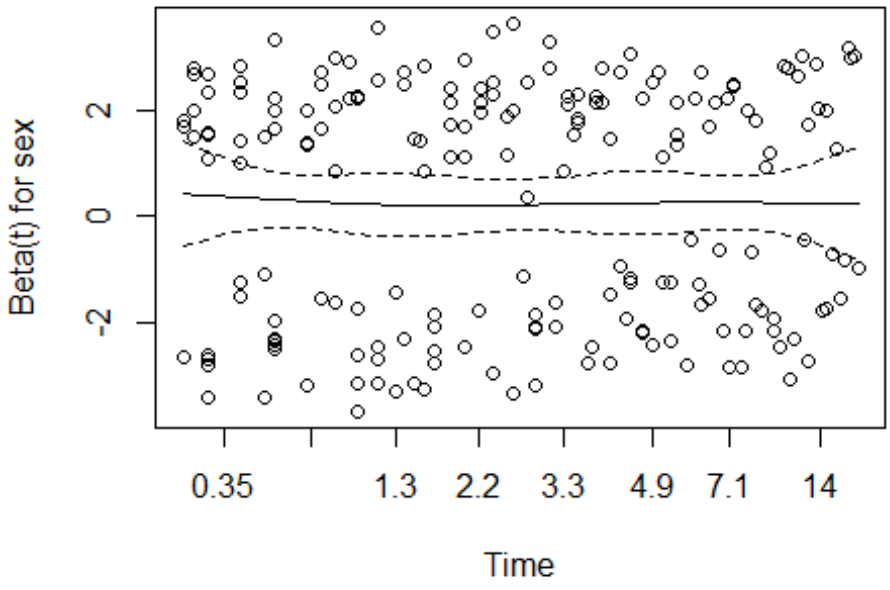
step 3: test proportional hazards assumption

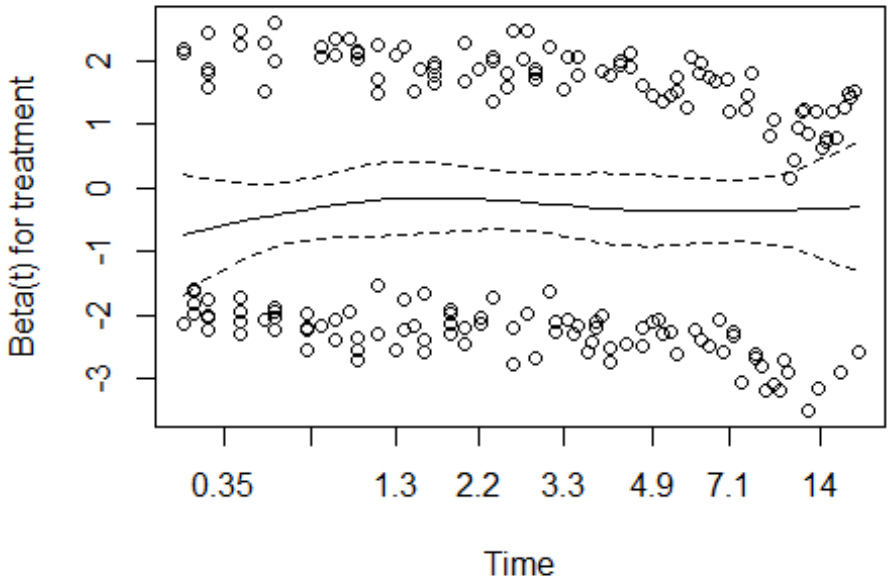
```
ph_test <- cox.zph(cox_model2)
print(ph_test)
```

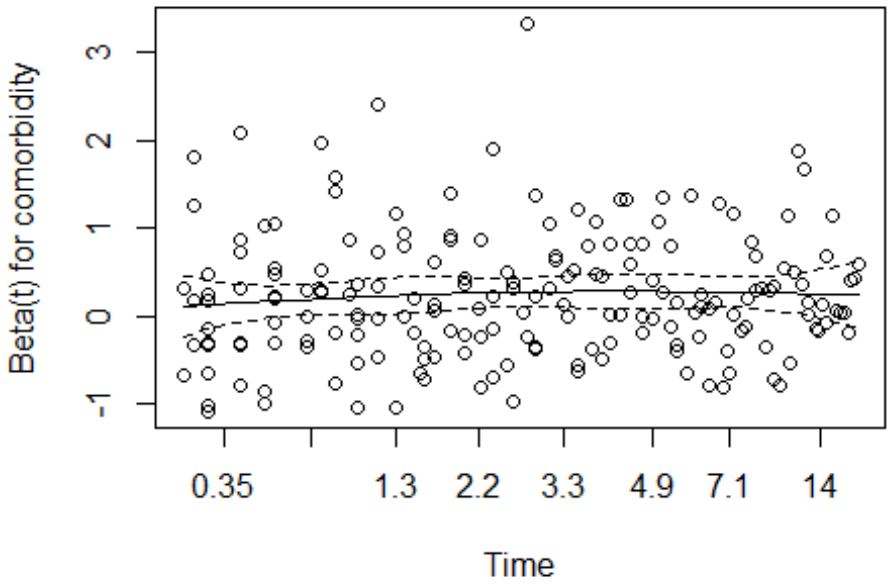
```
##              chisq df    p
## age         0.0665  1 0.80
## sex         0.1543  1 0.69
## treatment   0.0665  1 0.80
## comorbidity 0.7544  1 0.39
## GLOBAL      0.8904  4 0.93
```

```
plot(ph_test)
```


Beta(t) for age
0.2
0.1
0.0
-0.1
-0.2
0.35
1.3
2.2
3.3
4.9
7.1
14
Time


Beta(t) for sex
2
0
-2
0.35
1.3
2.2
3.3
4.9
7.1
14
Time





Interpretation:

#All covariates have p-values > 0.05, and the global test p-value is 0.93.

#This indicates that none of the covariates violate the proportional hazards assumption.

#The test confirms that the hazard ratios for all variables remain constant over time, which is a key assumption for the validity of the Cox proportional hazards model. #conclusion: The model satisfies the proportional hazards assumption. This validates the use of the Cox model and supports the reliability of its estimates.

step 4: check treatment-comorbidity interaction

```
cox_model3 <- coxph(Surv(time, status) ~ age + sex + treatment * comorbidity,
data = survival_data)
summary(cox_model3)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ age + sex + treatment *
##     comorbidity, data = survival_data)
##
##   n= 200, number of events= 191
##
##                             coef exp(coef)  se(coef)      z Pr(>|z|)
## age                      0.017635  1.017791  0.006037  2.921  0.00349 **
## sex                      0.267289  1.306418  0.155909  1.714  0.08646 .
## treatment               -0.323768  0.723418  0.236589 -1.368  0.17116
## comorbidity              0.238358  1.269163  0.079125  3.012  0.00259 **
## treatment:comorbidity -0.005782  0.994235  0.102970 -0.056  0.95522
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                         exp(coef) exp(-coef) lower .95 upper .95
## age                        1.0178     0.9825    1.0058     1.030
## sex                        1.3064     0.7655    0.9624     1.773
## treatment                  0.7234     1.3823    0.4550     1.150
## comorbidity                1.2692     0.7879    1.0868     1.482
## treatment:comorbidity      0.9942     1.0058    0.8125     1.217
##
## Concordance= 0.625  (se = 0.022 )
## Likelihood ratio test= 34.44  on 5 df,   p=2e-06
## Wald test            = 34.61  on 5 df,   p=2e-06
## Score (logrank) test = 35.45  on 5 df,   p=1e-06
```

#To examine whether the effect of discharge planning (treatment) depends on the level of comorbidity, an interaction term between treatment and comorbidity was added to the Cox proportional hazards model.

#Results: #Interaction term (treatment × comorbidity)

#Coefficient = −0.0058

#Hazard Ratio (HR) = 0.994

#p-value = 0.955

#Interpretation: #The p-value of 0.955 indicates that the interaction term is not statistically significant.

#This suggests that there is no meaningful evidence that the effect of treatment on the hazard of readmission differs based on the number of comorbidities a patient has.

#The hazard ratio near 1 (0.994) further implies that the joint effect of treatment and comorbidity is not synergistic or antagonistic, but rather independent and additive.

#Conclusion: There is no statistical interaction between treatment and comorbidity. The protective effect of discharge planning applies similarly across patients regardless of comorbidity burden.

SECTION C: APPLICATION AND INTERPRETATION

STEP 1: 6-month survival for hypothetical patients

```
# Final model
cox_model_final <- coxph(Surv(time, status) ~ age + sex + treatment +
comorbidity, data = survival_data)

# Patient data
patient1 <- data.frame(age = 55, sex = 0, treatment = 1, comorbidity = 0)
patient2 <- data.frame(age = 70, sex = 1, treatment = 0, comorbidity = 3)

# Predictions
surv_patient1 <- survfit(cox_model_final, newdata = patient1)
surv_patient2 <- survfit(cox_model_final, newdata = patient2)

# Survival at 6 months
summary(surv_patient1, times = 6)$surv

## [1] 0.577732

summary(surv_patient2, times = 6)$surv

## [1] 0.07114217
```

#Estimate and Interpret the 6-Month Survival Probabilities (5 marks) #Using the fitted Cox model with age, sex, treatment, and comorbidity, survival probabilities at 6 months were estimated for two hypothetical patients:

#Patient A: 55-year-old female, received discharge planning, with 0 comorbidities
Estimated 6-month survival probability: 57.8%

#Patient B: 70-year-old male, did not receive discharge planning, with 3 comorbidities
Estimated 6-month survival probability: 7.1%

#Interpretation: #Patient A has a markedly higher chance of avoiding hospital readmission within 6 months compared to Patient B. This large difference is explained by:

#Younger age

#Receipt of discharge planning

#Fewer comorbid conditions

#Female sex, though this factor was not statistically significant

#These results highlight the strong protective effect of discharge planning and the increased vulnerability of older patients with multiple comorbidities.

STEP 2: age group and treatment interaction

```
# Create age group
survival_data <- survival_data %>%
  mutate(age_group = ifelse(age < 60, "<60", "60+"))

# Cox model with interaction
cox_age_interact <- coxph(Surv(time, status) ~ age_group * treatment + sex +
comorbidity, data = survival_data)
summary(cox_age_interact)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ age_group * treatment +
##     sex + comorbidity, data = survival_data)
##
##   n= 200, number of events= 191
##
##                             coef exp(coef) se(coef)      z Pr(>|z|)
## age_group60+            0.70578   2.02542  0.23650  2.984  0.00284 **
## treatment              -0.15072   0.86009  0.26325 -0.573  0.56696
## sex                     0.22949   1.25796  0.15469  1.484  0.13793
## comorbidity             0.23581   1.26594  0.05424  4.348 1.38e-05 ***
## age_group60+:treatment -0.20685   0.81314  0.31730 -0.652  0.51446
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                        exp(coef) exp(-coef) lower .95 upper .95
## age_group60+              2.0254     0.4937    1.2741     3.220
## treatment                 0.8601     1.1627    0.5134     1.441
## sex                       1.2580     0.7949    0.9290     1.704
## comorbidity               1.2659     0.7899    1.1383     1.408
## age_group60+:treatment    0.8131     1.2298    0.4366     1.514
##
## Concordance= 0.641  (se = 0.023 )
## Likelihood ratio test= 40.31  on 5 df,   p=1e-07
```

```
## Wald test            = 40.67  on 5 df,   p=1e-07
## Score (logrank) test = 42.08  on 5 df,   p=6e-08
```

#Assess Whether Age Modifies the Effect of Treatment (5 marks) #To assess whether age modifies the effect of discharge planning, a new variable was created to categorize patients into two age groups: below 60 years (<60) and 60 years and above (60+). A Cox proportional hazards model was then fitted with an interaction term between age group and treatment, along with sex and comorbidity as covariates.

#The results showed that being in the 60+ age group significantly increased the hazard of hospital readmission, meaning older patients are at a higher baseline risk. However, the interaction between age group and treatment was not statistically significant, indicating that the effect of treatment did not vary significantly between the younger and older age groups.

#Interpretation: #There is no evidence that age modifies the effect of discharge planning on the risk of readmission. While older patients are generally at higher risk, the protective effect of treatment is consistent across age groups. Both younger and older patients benefit similarly from discharge planning.

Summary and Policy Implications (5 marks)

#Summary: #Age, comorbidity, and treatment significantly affect readmission risk. Older patients and those with more comorbidities are at higher risk. Discharge planning reduces the risk of readmission and is effective across all patient groups.

#Policy Implications: #Discharge planning should be standard for all patients, especially those who are older or have multiple conditions. Hospitals should strengthen discharge protocols and ensure proper follow-up care to reduce early readmissions.

#Section D: Conceptual Understanding (5 marks) #1. Explain what the hazard function represents in survival analysis. (3 marks) #The hazard function describes the instantaneous risk that an event (e.g., hospital readmission) will occur at a specific time, given that the individual has survived or not yet experienced the event up to that time. It helps in understanding how risk changes over time.

#2. What is left censoring and when does it occur? (2 marks) #Left censoring occurs when an individual has already experienced the event before entering the study, but the exact time of the event is unknown. For example, if a patient was readmitted before the study began but the date of readmission is not recorded, their data is left-censored.