Cox Proportional Hazard Model in Political Science

Shaka Y.J. Li David Akindoyin

2025-10-21

Table of contents

1 Introduction to Duration Analysis

What is “survival” in political science? In political science, survival analysis is often used to study the duration of political events or institutions. For example, we may want to know how long a regime survives before collapsing, how long a cabinet remains in office, or how long a particular leader stays in power. In this lecture note, we first explore the basic concept for duration model. Then, we introduce the idea behind Cox Proportional Hazard Model and explain how to model duration data based on this model. To make the concept more concrete, we then introduce two examples with `R` in political science and medical science to help the audience have better sense. We conclude by stating the importance of duration data and model in political science and future direction about this.

1.1 Basic Concept in Duration Analysis

1.1.1 Survival Function

The survival function $S(t)$ represents the probability that the event has **not yet occurred** by time t. Formally,

$$S(t) = P(T > t) = 1 - F(t), \tag{1}$$

where $F(t)$ is the cumulative distribution function (CDF), or the probability that the event **has occurred** by time t.

1.1.2 Hazard Function

It describes the **instantaneous rate** at which an event occurs, given that the individual has survived up to time t. Formally, it can be represented as:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \tag{2}$$

> ℹ Note
>
> In other words, hazard function answers the question: If a regime (or patient) has already survived past time t, what is the probability that it will suddenly collapse (or die) in the very next instant?

2 Cox Proportional Hazard Model

The hazard rate for the ith individual is

$$h_i(t) = h_0(t)exp(\beta' X), \tag{3}$$

where $h_0(t)$ is the baseline hazard function, and the $\beta' X$ is the covariates and regression parameters.

2.1 Key Features of the Cox Model

The Cox model seeks to answer the question: *Which variables (such as education, age, gender, etc.) affect the hazard rate?* In other words, it estimates how certain characteristics accelerate or decelerate the timing of an event's occurrence.

One key distinction between the Cox model and other duration models (such as Weibull or exponential models) is that the Cox model treats the baseline hazard function, $h_0(t)$, as **unknown and unparameterized**. That is, it models only the covariates through $\beta' X$, while leaving $h_0(t)$ unspecified. For this reason, the Cox model is often referred to as a **semi-parametric** model. Furthermore, because the baseline hazard is not explicitly modeled, Cox regression does **not** include an intercept term. To see why:

Proof. Consider a basic hazard rate for the ith individuals:

$$h_i(t) = h_0(t)exp(\beta' X), \tag{4}$$

Then we write the $\beta' X$ in a scalar form:

$$h_i(t) = h_0(t)exp(\beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki}) \tag{5}$$

Move $h_0(t)$ to LHS and take log on both sides:

$$log(\frac{h_i(t)}{h_0(t)}) = \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} \tag{6}$$

See!!! There's no β_0 on the RHS, which means we don't have the intercept term under the Cox model. □

2.2 Partial Likelihood of Cox Model

In parametric models, we often apply Maximum Likelihood Estimation (MLE)[1] to maximize the likelihood function and obtain the parameter estimates that best fit the data. However, in Cox model, we can only maximize part of the parameters, as we leave $h_0(t)$ is unspecified. Therefore, the likelihood function of Cox model is often called **partial** likelihood.

[1] We usually pronounce it "Emily"—we're not sure who started it either.

2.2.0.1 Step 1: Conditional Probability

To see the partial likelihood of Cox model, we first consider a conditional probability:

$$Pr(t_j = T_i \mid R(t_i)) = \frac{e^{\beta' \mathbf{x}_i}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j}} \tag{7}$$

At time t_j, we know exactly one event will occur. We also know which individuals are still at risk—that is, thos who could possibly experience the event at that time (this group is called the risk set, $R(t_i)$). Given this, what is the probability that individual i is the one who actually experiences the event at time t_j? In other words, we can interpret the numerator, $e^{\beta' \mathbf{x} i}$, as the **hazard [2] of the individual who actually experienced the event at time** t_i. Meanwhile, the denominator, $\sum j \in R(t_i) e^{\beta' \mathbf{x}_j}$, represents the **total hazard of all individuals still at risk at that time (i.e., the risk set).**

2.2.0.2 Step 2: Product of Conditional Probability

Then, we take the product [3] of the conditional probabilities. Define δ_i as an indicator that takes the value 0 if the observation is right-censored—meaning the event has not occurred by the end of the observation period—and 1 if the observation is uncensored, meaning the event has occurred. The result yields the **partial likelihood function**:

$$\mathcal{L}_p(\beta) = \prod_{i=1}^{K} \left(\frac{e^{\beta' \mathbf{x}_i}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j}} \right)^{\delta_i} \tag{8}$$

$$= \left(\frac{e^{\beta' \mathbf{x}_1}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j}} \right)^{1} \cdot \left(\frac{e^{\beta' \mathbf{x}_2}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j}} \right)^{0} \cdots \left(\frac{e^{\beta' \mathbf{x}_K}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j}} \right)^{1} \tag{9}$$

2.2.0.3 Step 3: Take log for the Product We Derived

We take log with respect to equation (8), and this yield the **partial log-likelihood function**.

[2] You can also interpret the "hazard" as a kind of probability—specifically, the instantaneous risk that an event (like death or regime collapse) happens at time t_i. In the terminology of survival analysis, this is often referred to as the "probability of failure" at time t_i. However, it's important to note that "failure" in this context simply means that the event of interest has occurred. It does not carry any normative or negative connotation — it could mean anything from a patient dying to a regime peacefully transitioning power.

[3] Each t_i represents an independently observed conditional probability of an event occurring at that time. To obtain the overall joint probability for the entire sample, we multiply all these conditional probabilities together according to the multiplication rule of probability.

$$\log \mathcal{L}_p(\beta) = \sum_{i=1}^{K} \delta_i \left[\log\left(e^{\beta' \mathbf{x}_i}\right) - \log\left(\sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j} \right) \right] \tag{10}$$

$$= \sum_{i=1}^{K} \delta_i \left[\beta' \mathbf{x}_i - \log\left(\sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j} \right) \right]. \tag{11}$$

By maximizing this log-likelihood function, the estimates of β are obtained! We call this Maximum Partial Likelihood Estimation (MPLE). The MPLE has same properties as MLE. So the parameter estimates from partial likelihood are asymptotically normal, efficient, consistent, and invariant.

2.3 Handling Tied data

We will not cover this part in the workshop. Essentially, these are methods designed to handle tied data. What we have covered so far assumes that at each time point, exactly one event occurs. However, in some cases, multiple events may occur at the same time point. Therefore, we need additional methods to address this issue. We list some of these methods below, but please refer to Box-Steffensmeier and Jones (2004) for more details if you are interested.

1. Breslow Method
2. The Efron Method
3. Average Likelihood
4. The Exact Discrete Method

3 Example 1: Duration in Authoritarian Regime

In this section, we demonstrate how to model duration data in authoritarian regimes. We obtain the data from **Autocratic Breakdown and Regime Transition Dataset** (Geddes et al., 2014), which documents regime transition information for 280 authoritarian regimes from 1946 to 2010. This *dataset* identifies how regimes exit power, what the subsequent regime types are after transition, and how much violence occurred during the transition period. We then merge it with other socio-demographic variables obtained from the Quality of Government dataset (QoG) (Teorell et al., 2011), which provides country-level indicators on governance and development.

In the *dataset*, the variables `gwf_duration` and `gwf_fail` enable researchers to model regime duration and examine the determinants of authoritarian longevity. The variable `gwf_duration` indicates the *time that a regime remains at risk before either failing or being censored*—in other words, the number of days a regime survives until it either collapses or the observation period ends. If a regime failure is observed, `gwf_fail` is coded as 1; otherwise, it is coded as 0 (i.e., right-censored). We use `gwf_military` as the treatment variable, coded as 1 if a country is classified as a military regime—either group military, military–personal, or indirect military—and 0 otherwise. For socio-demographic control variables, we include GDP per capita (`wdi_gdpcapcur`) and population (`wdi_pop`), both obtained from the QoG dataset.

3.1 Fit Cox Proportional Hazard Model

We simply apply `coxph()` to fit the duration data. We should note that censoring is very common in duration models, so the `coxph()` function can easily accommodate censored observations. You simply supply a variable indicating censoring as the second argument to `Surv()`.

```
# install.packages("survival")
library(survival) # load the `survival` package
model <- coxph(
  Surv(gwf_duration, gwf_fail) ~ gwf_military + wdi_gdpcapcur + wdi_pop,
  data = new_df
)
```

After fitting the model, we can use `summary()` as usual to display the results.

```
summary(model)
```

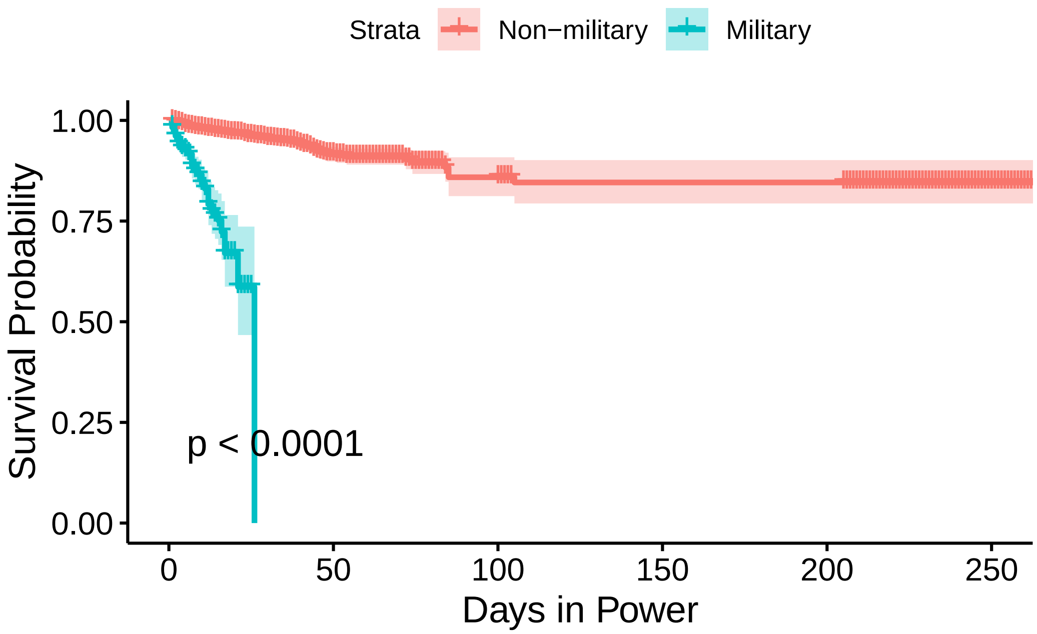
```
Call:
coxph(formula = Surv(gwf_duration, gwf_fail) ~ gwf_military +
    wdi_gdpcapcur + wdi_pop, data = new_df)

  n= 3141, number of events= 157
   (1450 observations deleted due to missingness)

                    coef  exp(coef)   se(coef)      z Pr(>|z|)
gwf_military   2.231e+00  9.307e+00  1.916e-01 11.642  < 2e-16 ***
wdi_gdpcapcur -5.076e-04  9.995e-01  9.224e-05 -5.503 3.73e-08 ***
wdi_pop       -7.013e-09  1.000e+00  3.263e-09 -2.149   0.0316 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
gwf_military     9.3072     0.1074    6.3932   13.5494
wdi_gdpcapcur    0.9995     1.0005    0.9993    0.9997
wdi_pop          1.0000     1.0000    1.0000    1.0000

Concordance= 0.811  (se = 0.018 )
Likelihood ratio test= 201.8  on 3 df,   p=<2e-16
Wald test            = 169  on 3 df,   p=<2e-16
Score (logrank) test = 236.5  on 3 df,   p=<2e-16
```

The coefficient is positive, indicating a higher hazard ratio. In a Cox proportional hazards model, the quantity of interest is the **hazard ratio**, which in this context represents the likelihood that a regime collapses at any given point in time. Therefore, if a regime is a military regime (`gwf_military` $= 1$), its risk of collapse at any given time is approximately 9.3 times higher than that of a non-military regime (`gwf_military` $= 0$), holding other variables constant.

Now, we want to explore the difference in survival probabilities between military and non-military regimes. To do this, we simply apply `ggsurvplot()`. Note that the survival probabilities shown here are not adjusted for control variables.

```
library(survminer) # load package
```

```
fit_km <-
  survfit(Surv(gwf_duration, gwf_fail) ~ gwf_military, data = new_df) # fit model

ggsurvplot(fit_km, conf.int = TRUE, pval = TRUE, # plot the survival probability
           legend.labs = c("Non-military", "Military"),
           xlab = "Days in Power", ylab = "Survival Probability")
```



The Kaplan–Meier curve shows that military regimes experience a rapid decline in survival probability during the early days of their rule, indicating regime instability—most collapse within just 50 days. In contrast, non-military regimes decline more gradually, suggesting greater stability, with many surviving beyond 250 days.

3.2 Counterfactual Predictions

3.2.1 Create simulated dataset for hypothesized scenario

Now we want to simulate what the average survival curve would look like if everyone in the sample had received the treatment versus if no one had received the treatment.

First, let's generate a new dataset (nd) to simulate average survival probabilities over time under two counterfactual scenarios:

```
library(marginaleffects)

min_time <- min(new_df$gwf_duration[new_df$gwf_fail == 1], na.rm = TRUE)
max_time <- max(new_df$gwf_duration[new_df$gwf_fail == 1], na.rm = TRUE)

nd <- datagrid(
  gwf_military = 0:1,                                                    ①
  gwf_duration = round(seq(min_time, max_time, length.out = 25)),        ②
  grid_type = "counterfactual",                                          ③
  model = model                                                          ④
)
```

① Specify groups to compare (military vs. non-military regimes)
② 25 equally spaced points in time between the first (`min_time`) and the last (`max_time`) occurring event time
③ Tells the function to hypothesize specific values of predictors (`gwf_military` == 1 or 0), regardless of the observed values in the original dataset.
④ Specify the model used to generate predictions.

To estimate the survival probability at different time points for both military and non-military regimes, we use the `avg_predictions()` function from the `marginaleffects` package along with our simulated data grid (`nd`). This function computes *adjusted survival probabilities* based on the fitted Cox model. These predictions are adjusted for all covariates in the model and represent the expected survival probabilities as if each observation were assigned to either the military or non-military regime type.

```
p <- avg_predictions(model,
                     type = "survival",                                  ①
                     by = c("gwf_duration", "gwf_military"),             ②
                     vcov = "rsample",                                   ③
                     newdata = nd)                                       ④
```

① Choose the survival probability scale.
② Calculate probability for each combination of `gwf_military` and `gwf_duration`. Every probability represents a point (survival probability).
③ Non-parametric bootstrapping.
④ Feed `avg_prediction()` with the new data that we just simulated.

```
tail(p)
```

```

 gwf_duration gwf_military Estimate 2.5 % 97.5 %
           96            0    0.737 0.583  0.851
           96            1    0.171 0.101  0.333
          101            0    0.737 0.583  0.851
          101            1    0.171 0.101  0.333
          105            0    0.737 0.583  0.851
          105            1    0.171 0.101  0.333

Type:  survival
```

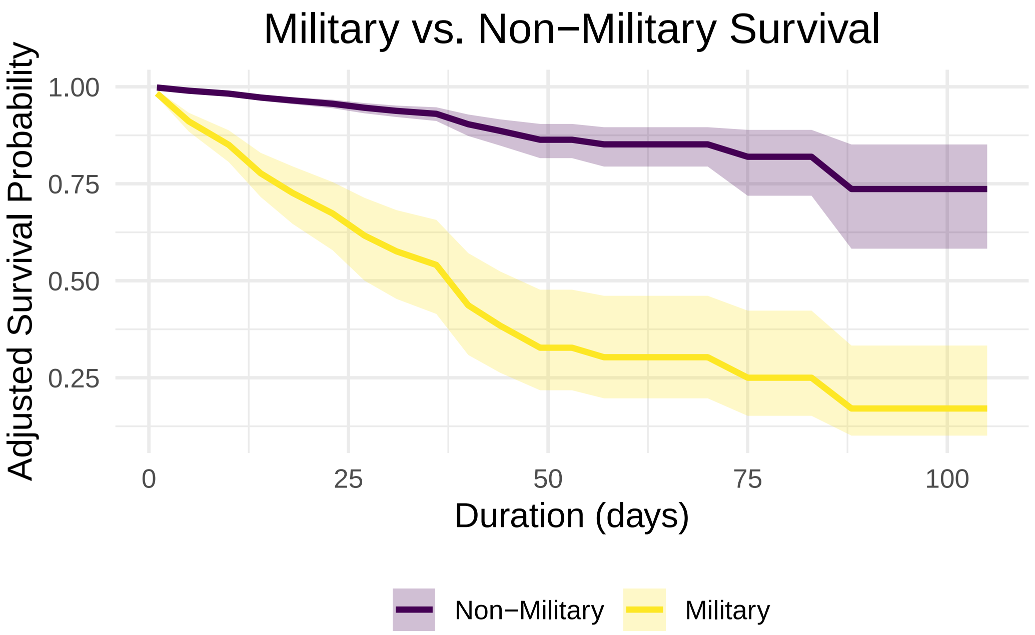
The estimate here represents the average survival probability. This analysis helps us answer the question: *Between military and non-military regimes, which type tends to survive longer on average?* For instance, at day 105, if a country is not a military regime (gwf_military = 0), about 73.7% of regimes are expected to survive. In contrast, if a country is a military regime (gwf_military = 1), only about 17% are expected to survive up to that point.

3.2.2 Plot the average adjusted survival probability

Since the output of avg_predictions() is a data frame, we can use it directly to plot the results, using ggplot2, base R, or any other plotting package. Here, we use the `ggplot2` package to display the adjusted survival curves:

```r
#install.packages("ggplot2")
library(ggplot2)
```

```r
# Convert the `gwf_military` to a factor variable
p$gwf_military <- factor(p$gwf_military, labels = c("Non-Military", "Military"))

# Plot the Survival Probability
ggplot(p, aes(x = gwf_duration,
              y = estimate,
              color = as.factor(gwf_military),
              fill = as.factor(gwf_military))) +
  geom_line(size = 1.1) +
  geom_ribbon(aes(ymin = conf.low, ymax = conf.high),
              alpha = 0.25, color = NA) +
  scale_color_viridis_d(
    option = "D",
    labels = c("Non-Military", "Military")
  ) +
  scale_fill_viridis_d(
    option = "D",
    labels = c("Non-Military", "Military")
  ) +
  labs(
    x = "Duration (days)",
    y = "Adjusted Survival Probability",
    color = "",
    fill = "",
    title = "Military vs. Non-Military Survival"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    legend.position = "bottom",
    legend.box = "horizontal",
    plot.title = element_text(hjust = 0.5)
  )
```



The resulting plot shows the survival curves for both treatment options of military regime, adjusted for GDP and population. Here we can observe that their CI does not overlap. If we only see the hazard ratio, we are unable to make such more nuanced distinctions regarding survival probability. Since the CI did not overlap much, difference between survival probabilities in the military regimes and non-military regimes are distinguishable.

4 Example 2: Veterans' Administration Lung Cancer Dataset

In this section, we use the Veterans' Administration Lung Cancer (VALC) dataset to demonstrate survival analysis. This dataset is embedded in `survival` package. Two variables in VALC allow us to model duration: (1) `time`, a numeric vector indicating survival time in days since treatment; and (2) `status`, a numeric vector indicating patient status, coded as 1 (dead) or 0 (alive). The main treatment variable, `trt`, specifies the type of lung cancer treatment, where 1 represents the standard treatment and 2 represents the test drug treatment.

4.1 Fit Cox Model with `coxph()`

```
# Load libraries
library(survival)
library(marginaleffects)
library(ggplot2)
library(survminer)
```

```
# Load and prepare the data
veteran <- survival::veteran
```

```r
# Fit Cox proportional hazards model
model_c <- coxph(Surv(time, status) ~ trt, data = veteran)
summary(model_c)
```

```
Call:
coxph(formula = Surv(time, status) ~ trt, data = veteran)

  n= 137, number of events= 128

       coef exp(coef) se(coef)     z Pr(>|z|)
trt 0.01774   1.01790  0.18066 0.098    0.922

    exp(coef) exp(-coef) lower .95 upper .95
trt     1.018     0.9824    0.7144      1.45

Concordance= 0.525  (se = 0.026 )
Likelihood ratio test= 0.01  on 1 df,   p=0.9
Wald test            = 0.01  on 1 df,   p=0.9
Score (logrank) test = 0.01  on 1 df,   p=0.9
```

The hazard ratio for the test treatment compared to standard treatment is 1.018, which suggests that patients receiving the test treatment face a 1.8% higher hazard of death relative to those receiving standard treatment. However, this effect is not statistically significant at p = 0.922, and the 95% confidence interval ranges from 0.714 to 1.450, which includes the null value of 1.0. Substantively, this means that we cannot reject the possibility that there is no difference in the risk of death between test treatment and standard treatment groups. In general, a hazard ratio equal to 1 indicates no difference in risk, values below 1 suggest a lower risk that is longer survival, and values above 1 suggest a higher risk that is shorter survival.

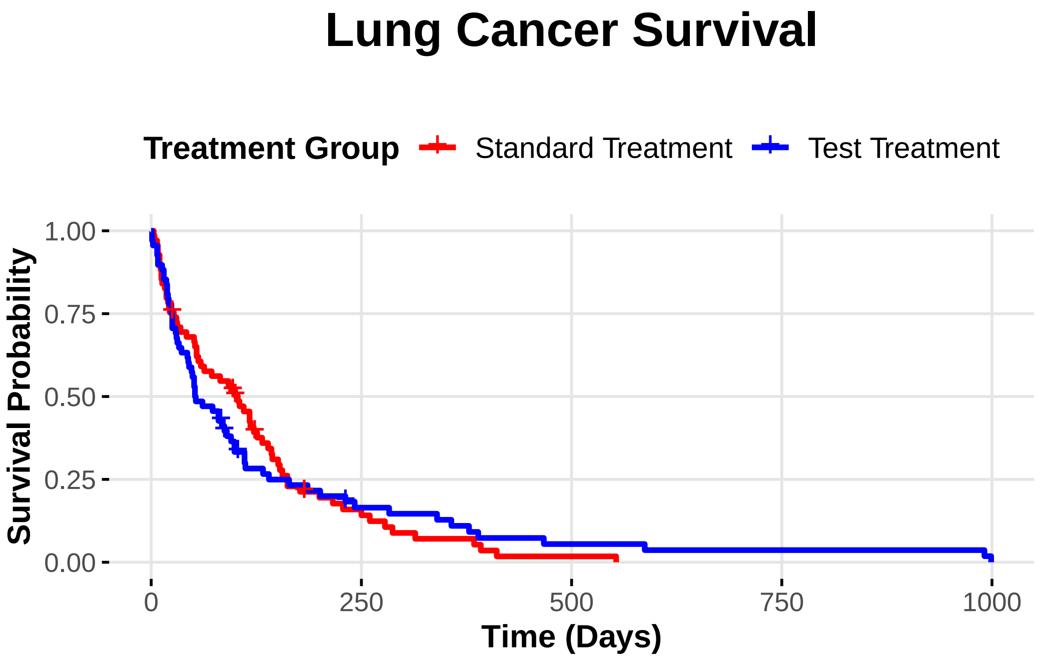
4.2 Survive Curves for Drug-Treatment and Standard-Treatment Groups

```r
# Estimate survival curves for each treatment group
surv_curves <- survfit(Surv(time, status) ~ trt, data = veteran) # Fit Cox model
surv_curves
```

```
Call: survfit(formula = Surv(time, status) ~ trt, data = veteran)

       n events median 0.95LCL 0.95UCL
trt=1 69     64  103.0      59     132
trt=2 68     64   52.5      44      95
```

After estimating the survival functions for each treatment group using the `survfit()` function, we can visualize the results to better illustrate the difference in survival probabilities between the standard and test treatments. The following code employs `ggsurvplot()` to produce Kaplan–Meier curves for both groups.

```r
# Visualize survival difference between standard and test treatments
ggsurvplot(surv_curves,
           data = veteran,
           palette = c("red", "blue"),
           legend.title = "Treatment Group",
           legend.labs = c("Standard Treatment", "Test Treatment"),
           xlab = "Time (Days)",
           ylab = "Survival Probability",
           title = "Lung Cancer Survival",
           risk.table = FALSE,
           conf.int = FALSE,
           ggtheme = theme_minimal() +
             theme(
               plot.title = element_text(size = 18, face = "bold", hjust = 0.5,
                                         margin = margin(b = 20)),
               legend.title = element_text(face = "bold", size = 12),
               legend.text = element_text(size = 11),
               legend.position = "top",
               axis.title = element_text(size = 12, face = "bold"),
               axis.text = element_text(size = 10),
               panel.grid.major = element_line(color = "grey90"),
               panel.grid.minor = element_blank()
             ))
```



The nearly overlapping curves show no meaningful difference in survival patterns between treatment group.

4.3 Conducting Counterfactual Comparisons with `marginaleffects()`

4.3.1 Average Comparison Across All Time Points

Here, we answer the following question with `avg_comparisons()`: *What's the average difference in survival probability between treatments?*

```r
avg_comp <- avg_comparisons(
  model_c,
  variables = "trt",
  type = "survival"
)
```

```r
avg_comp
```

```
 Estimate Std. Error       z Pr(>|z|)   S   2.5 % 97.5 %
 -0.00456     0.0464 -0.0981    0.922 0.1 -0.0956 0.0865

Term: trt
Type: survival
Comparison: +1
```

Patients receiving the test treatment have, on average, a 0.46% lower survival probability compared to patients receiving the standard treatment. However, this effect is not statistically significant, and the 95% confidence interval spans from -0.0956 to 0.0865, which includes zero. This means that the observed difference is extremely small and could easily be the result of random variation rather than a meaningful treatment effect. Therefore, the analysis does not provide evidence of a systematic survival advantage for the test treatment over the standard treatment.

Then, we can compare the average survival probability at different and specific time points. We use `comparison()` here and specify the time points we want in the `datagrid()` argument.

```r
# Compare at specific time points
time_comp <- comparisons(
  model_c,
  variables = "trt",
  newdata = datagrid(time = c(100, 200, 300, 400, 500)),
  type = "survival"
)
```

```r
time_comp
```

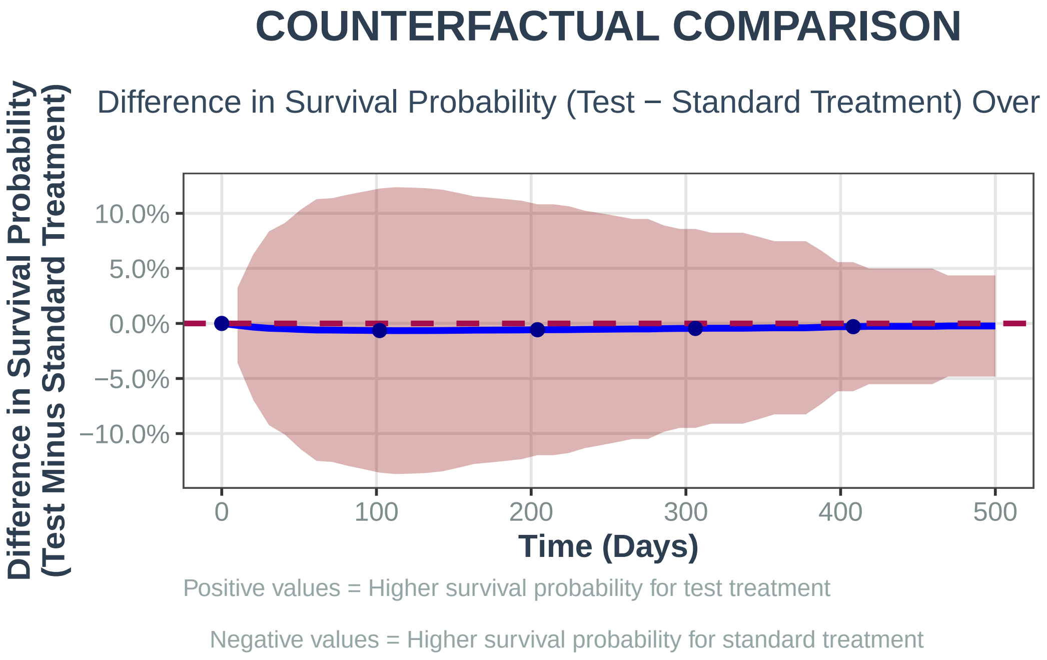
```
 time Estimate Std. Error       z Pr(>|z|)   S   2.5 % 97.5 %
  100 -0.00646     0.0658 -0.0982    0.922 0.1 -0.1355 0.1225
  200 -0.00580     0.0590 -0.0982    0.922 0.1 -0.1215 0.1099
  300 -0.00453     0.0461 -0.0982    0.922 0.1 -0.0949 0.0858
  400 -0.00294     0.0299 -0.0982    0.922 0.1 -0.0616 0.0557
  500 -0.00230     0.0234 -0.0982    0.922 0.1 -0.0482 0.0436

Term: trt
Type: survival
Comparison: +1
```

Here we answers the question: “What is the difference in survival probability between test treatment and standard treatment *at specific time points* (100, 200, 300, 400, and 500 days)?” We find that as we continue to go up in days (100, 200, 300, 400, or 500 days) patients receiving test treatment and standard treatment have virtually identical chances of survival. The tiny differences we see are so small and could be due to random chance rather than any real treatment effect. Therefore we can conclude that there is no meaningful relationship exists between treatment type and patient survival duration in this dataset.

Having known the survival probability at specific time points, the `marginaleffects()` package also allows to compare how the survival probability changes over time. We first computes counterfactual differences in survival probabilities between the two treatment groups (`trt`) across a sequence of time points (from 0 to 500 days). Then, we can use `ggplot2` to generate counterfactual comparison plot that visualizes the difference in survival probabilities between the test and standard treatment groups over time.

```r
# Compute counterfactual differences from 0 to 500 days
plot_data <- comparisons(
  model_c,
  variables = "trt",  # Specify treatment groups
  newdata = datagrid(time = seq(0, 500, length.out = 50)),
  type = "survival"
)
```

```r
# Generating counterfactual comparison plot
ggplot(plot_data, aes(x = time, y = estimate)) +
  geom_ribbon(aes(ymin = conf.low, ymax = conf.high),
              fill = "darkred", alpha = 0.3) +
  geom_line(color = "blue", size = 1.2) +
  geom_hline(yintercept = 0, linetype = "dashed",
             color = "#A40E4C", size = 1) +
  geom_point(data = plot_data[seq(1, nrow(plot_data), by = 10), ],
             color = "darkblue", size = 2) +
  labs(
    title = "COUNTERFACTUAL COMPARISON",
    subtitle = "Difference in Survival Probability (Test - Standard Treatment) Over Time",
    y = "Difference in Survival Probability\n(Test Minus Standard Treatment)",
    x = "Time (Days)",
    caption = "Positive values = Higher survival probability for test treatment\n
    Negative values = Higher survival probability for standard treatment"
  ) +
  theme_bw() +
  theme(
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5,
                              color = "#2c3e50", margin = margin(b = 15)),
    plot.subtitle = element_text(size = 12, hjust = 0.5, color = "#34495e",
                                 margin = margin(b = 20)),
    axis.title = element_text(size = 12, face = "bold", color = "#2c3e50"),
    axis.text = element_text(size = 10, color = "#7f8c8d"),
    plot.caption = element_text(size = 9, color = "#95a5a6", hjust = 0),
    panel.grid.major = element_line(color = "grey90"),
    panel.grid.minor = element_blank(),
    panel.border = element_rect(color = "grey30", fill = NA),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "white")
  ) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 0.1))
```



This figure illustrates the estimated counterfactual difference in survival probabilities between the test and standard treatment groups over time. The results show that the survival advantage of the test treatment remains negligible across the entire period, with the confidence intervals consistently overlapping zero. In other words, there is no statistically meaningful evidence that the test treatment improves or worsens patient survival relative to the standard therapy.

5 Reference

Janet M Box-Steffensmeier and Bradford S Jones. *Event history modeling: A guide for social scientists*. Cambridge University Press, 2004.

Barbara Geddes, Joseph Wright, and Erica Frantz. Autocratic breakdown and regime transitions: A new data set. *Perspectives on politics*, 12(2):313–331, 2014.

Jan Teorell, Marcus Samanni, Sören Holmberg, and Bo Rothstein. The quality of government dataset, version 6apr11. *University of Gothenburg: The Quality of Government Institute, http://www. qog. pol. gu. se*, 2011.