

Survival Analysis of Dialysis Patients

Shyamsundar

2025-01-25

Schema

What is survival analysis?

Survival analysis is used to analyze time-to-event data, where the event could be death, relapse, or another significant milestone. Other terms include event history analysis, duration analysis, reliability analysis, or failure time modeling.

For this dataset: - **Time origin**: Time since enrollment in the dialysis study. - **Event**: Death (1) or censored (0).

Why not use standard statistical methods?

1. Survival data are typically right-skewed.
2. **Censoring**: Some participants' endpoints are not observed (e.g., still alive or lost to follow-up).

Assumption: Censoring is non-informative, meaning it is unrelated to the likelihood of the event occurring.

Key Concepts

- Survival function: Probability of surviving beyond a given time.
- Probability density function (PDF): Rate at which events occur.
- Censoring types: Right, left, or interval censored.

Rationale for the Project

Visualizing survival graphs can be challenging in practice. This analysis incorporates survival ratio and survival difference plots to aid interpretation.

```
library(survival)
```

```
## Warning: package 'survival' was built under R version 4.4.2
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.2
```

```
##
```

```
## 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(survminer)
```

```
## Warning: package 'survminer' was built under R version 4.4.2
```

```
## Loading required package: ggpubr
```

```
## Warning: package 'ggpubr' was built under R version 4.4.2
```

```
##
```

```
## Attaching package: 'survminer'
```

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

```
##
```

```
## myeloma
```

```
# Load the dialysis dataset
```

```
dialysis_data <- read.csv("C:/Users/shyam/OneDrive/Desktop/dialysis survival dataset.csv")
```

```
# Rename problematic column
```

```
dialysis_data <- dialysis_data %>%  
  rename(time_months = `time_months`)
```

Kaplan-Meier Survival Estimation

```
# KM fit for diabetes and hypertension groups
```

```
diabetes_fit <- survfit(Surv(time = time_months, event = event) ~ disease_diabetes, data = dialysis_data)  
hypert_fit <- survfit(Surv(time = time_months, event = event) ~ disease_hypert, data = dialysis_data)
```

```
# Extract time points for survival analysis
```

```
manual_times <- seq(0, max(dialysis_data$time_months, na.rm = TRUE), by = 12)
```

```
# Survival probabilities for diabetes and non-diabetes groups
```

```
diabetes_surv <- summary(diabetes_fit, times = manual_times, extend = TRUE)$surv
```

```
nondiabetes_surv <- 1 - diabetes_surv

# Survival probabilities for hypertension and non-hypertension groups
hypert_surv <- summary(hypert_fit, times = manual_times, extend = TRUE)$surv
nonhypert_surv <- 1 - hypert_surv
```

Confidence Interval Calculation

```
# Confidence intervals for survival probabilities
ci_diabetes <- summary(diabetes_fit, times = manual_times, extend = TRUE)$std.err
ci_hypert <- summary(hypert_fit, times = manual_times, extend = TRUE)$std.err

# Log-transformed survival ratio and its variance
log_ratio <- log(hypert_surv / diabetes_surv)
log_ratio_var <- (ci_hypert / hypert_surv)^2 + (ci_diabetes / diabetes_surv)^2

# Confidence intervals for log-ratio
z_value <- qnorm(0.975) # 1.96 for 95% CI
lower_log <- log_ratio - z_value * sqrt(log_ratio_var)
upper_log <- log_ratio + z_value * sqrt(log_ratio_var)

# Back-transform to get confidence intervals for the ratio
lower_ci <- exp(lower_log)
upper_ci <- exp(upper_log)

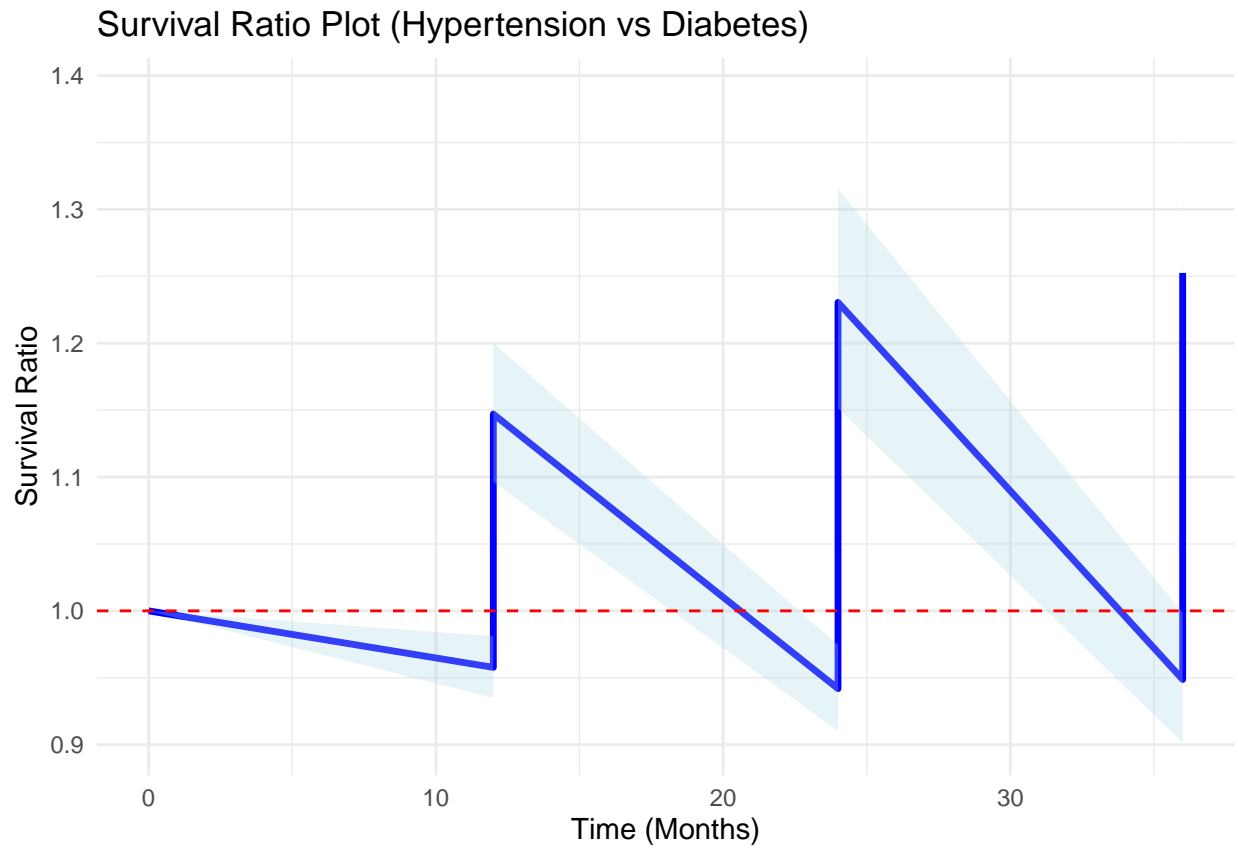
# Add CI to plot data
plot_data <- data.frame(
  time = manual_times,
  survival_ratio = hypert_surv / diabetes_surv,
  survival_diff = hypert_surv - diabetes_surv,
  lower_ci = lower_ci,
  upper_ci = upper_ci
)
```

Survival Ratio Plot

```
ggplot(plot_data, aes(x = time, y = survival_ratio)) +
  geom_line(color = "blue", size = 1.2) +
  geom_ribbon(aes(ymin = lower_ci, ymax = upper_ci), fill = "lightblue", alpha = 0.3) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  labs(
    title = "Survival Ratio Plot (Hypertension vs Diabetes)",
    x = "Time (Months)",
    y = "Survival Ratio"
  ) +
  theme_minimal()
```

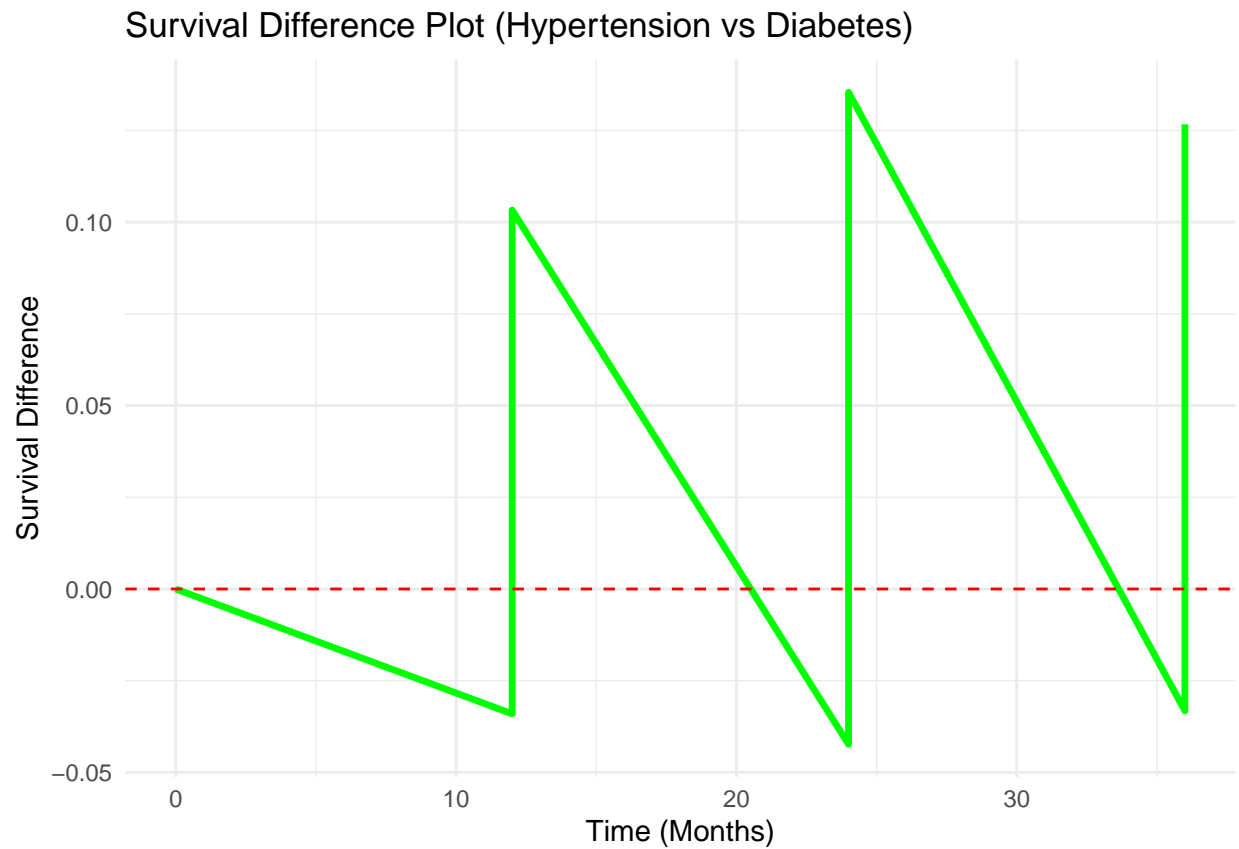
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
```

```
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



Survival Difference Plot

```
ggplot(plot_data, aes(x = time, y = survival_diff)) +
  geom_line(color = "green", size = 1.2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    title = "Survival Difference Plot (Hypertension vs Diabetes)",
    x = "Time (Months)",
    y = "Survival Difference"
  ) +
  theme_minimal()
```

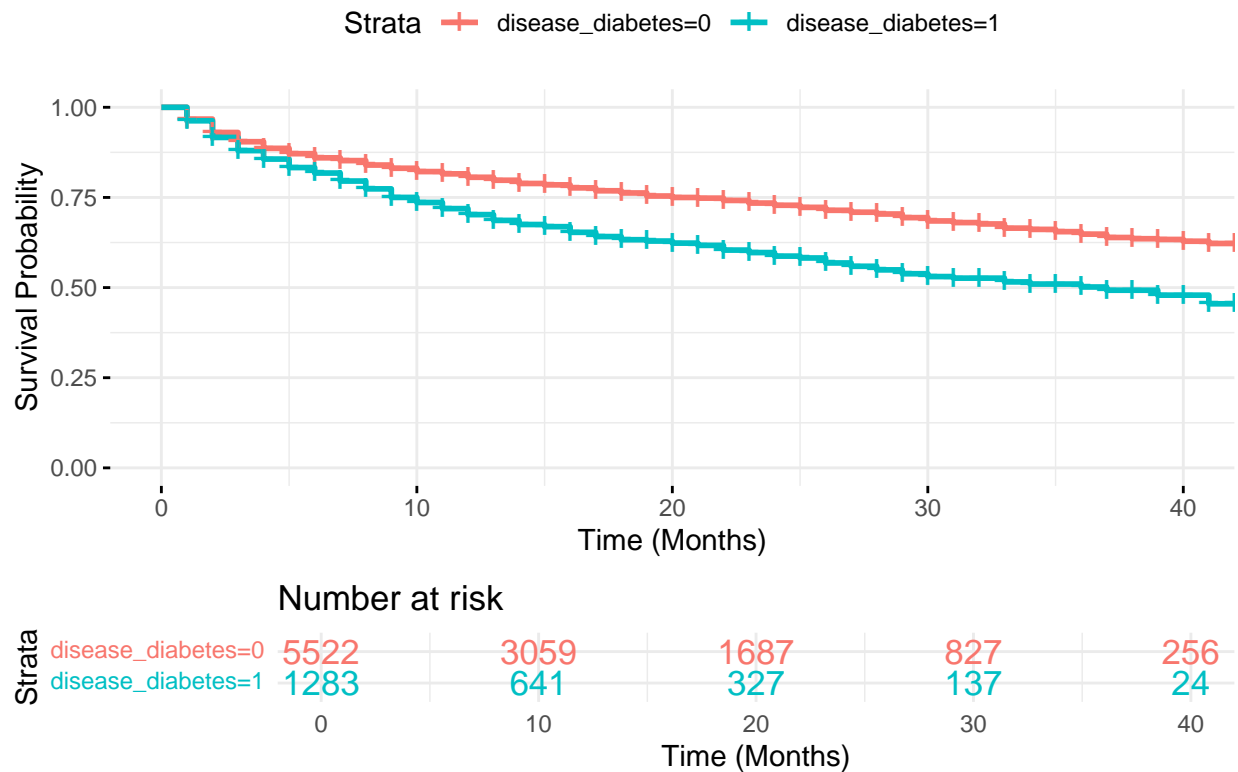


Risk Table

```
ggsurvplot(  
  list("Diabetes" = diabetes_fit, "Hypertension" = hypert_fit),  
  data = dialysis_data,  
  risk.table = TRUE,  
  risk.table.col = "strata",  
  ggtheme = theme_minimal(),  
  title = "Kaplan-Meier Curves with Risk Table",  
  xlab = "Time (Months)",  
  ylab = "Survival Probability"  
)
```

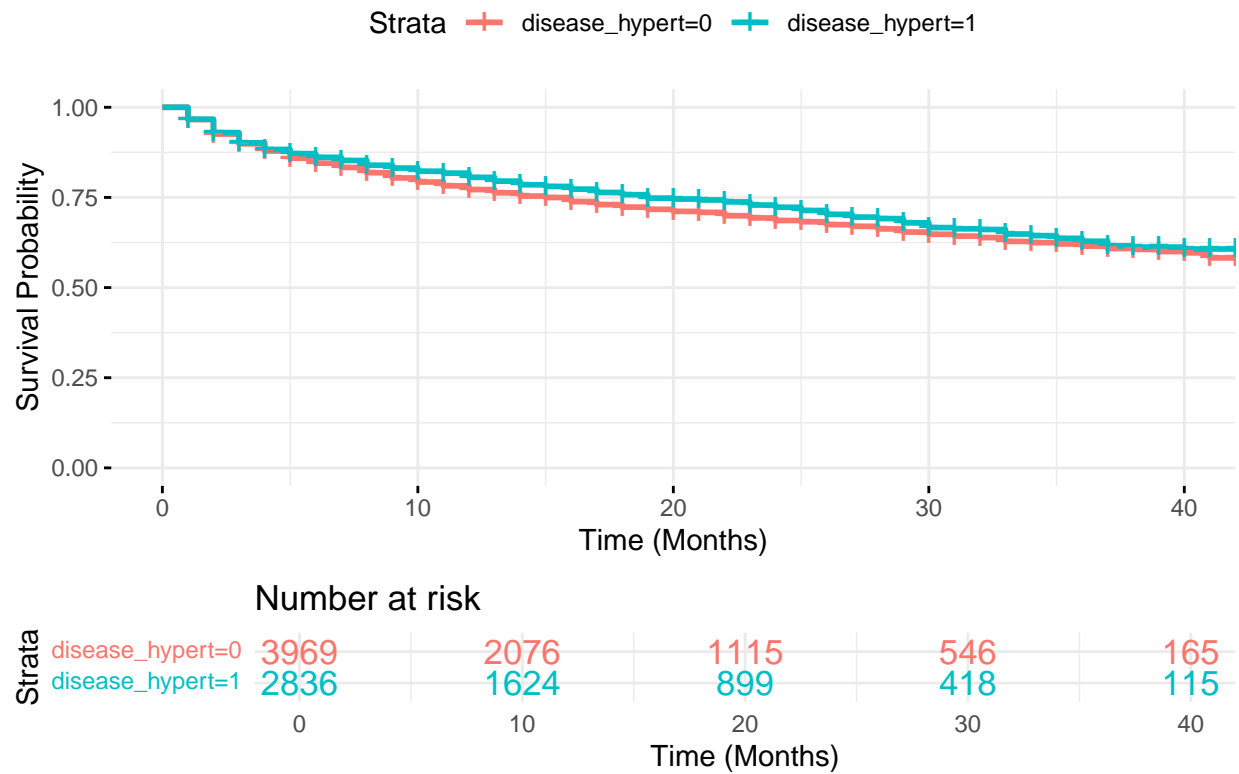
```
## $Diabetes
```

Kaplan–Meier Curves with Risk Table



\$Hypertension

Kaplan–Meier Curves with Risk Table



```
##
## attr("class")
## [1] "list"          "ggsurvplot_list"
```