# Simulating Clinical Trials: The Kaplan-Meier Method

**CRP 245 Advanced Tutorial — CheckMate 816 Case Study**

Duke University Clinical Research Training Program

2026-01-21

## Table of contents

# Introduction

> Learning Objectives
>
> After completing this tutorial, you will be able to:
>
> - **Simulate** survival data calibrated to published trial results.
> - **Distinguish** between the event process and the censoring process in survival analysis.
> - **Explain** why risk tables diverge between simulated and real trial data.
> - **Interpret** hazard ratios and their dependence on reference group coding.
> - **Generate** publication-quality Kaplan-Meier plots with confidence intervals and risk tables.
> - **Connect** statistical code to the biological and operational realities of clinical trials.

## 0.1 Why Simulate a Published Trial?

Simulation is a powerful pedagogical tool. By reconstructing a landmark trial's survival curves from first principles, you develop intuition for how the Kaplan-Meier estimator behaves under different assumptions about event rates, censoring patterns, and sample sizes.

In this tutorial, we simulate data matching the **CheckMate 816 trial** (Forde et al., *NEJM* 2025;393:741-752), which demonstrated that neoadjuvant nivolumab plus chemotherapy significantly improved overall survival in resectable non–small-cell lung cancer compared to chemotherapy alone.

> 💡 The Two Processes in Survival Data
>
> Every survival dataset is shaped by two distinct stochastic processes:
>
> 1. **The Event Process**: When do patients die (or experience the outcome)? This is what we care about clinically.
>
> 2. **The Censoring Process**: When do we lose track of patients who haven't yet had the event? This is driven by study logistics—enrollment timing, administrative end dates, loss to follow-up.

> The Kaplan-Meier estimator extracts information about the event process while accounting for the censoring process. Understanding both is essential for correct interpretation.

## 0.2 Source Trial: CheckMate 816

Table 1: CheckMate 816 Key Results

| Parameter | Value |
|---|---|
| **Design** | Phase 3, open-label, randomized |
| **Population** | Stage IB–IIIA resectable NSCLC |
| **Arms** | Nivolumab + platinum-based chemotherapy vs. chemotherapy alone |
| **N per arm** | 179 |
| **Primary endpoints** | Pathological complete response, event-free survival |
| **Key secondary endpoint** | Overall survival |
| **Median follow-up** | 68.4 months |
| **5-year OS (Nivo+Chemo)** | 65.4% (95% CI: 57.8–71.9) |
| **5-year OS (Chemo)** | 55.0% (95% CI: 47.3–62.0) |
| **HR for death** | 0.72 (95% CI: 0.523–0.998; P=0.048) |
| **Deaths** | 66 (Nivo+Chemo) vs. 84 (Chemo) |

---

# 1 Setup and Package Loading

```
# Install packages if needed
if (!requireNamespace("survival", quietly = TRUE)) install.packages("survival")
if (!requireNamespace("survminer", quietly = TRUE)) install.packages("survminer")
if (!requireNamespace("dplyr", quietly = TRUE)) install.packages("dplyr")

library(survival)
library(survminer)
library(dplyr)
```

---

## 2 The Simulation Strategy

### 2.1 Calibrating to Published Results

We don't have access to the individual patient data from CheckMate 816, but we can reverse-engineer plausible survival times by calibrating a parametric distribution to the reported 5-year survival percentages.

> **i Why Weibull, Not Exponential?**
>
> The **exponential distribution** assumes a constant hazard over time—the risk of death at month 60 is the same as at month 6. This is biologically implausible for most cancers, where hazard typically increases as disease progresses.
> The **Weibull distribution** generalizes the exponential by adding a shape parameter ($k$):
>
> $$S(t) = \exp\left[-\left(\frac{t}{\lambda}\right)^k\right]$$
>
> - $k = 1$: Constant hazard (exponential)
> - $k > 1$: Increasing hazard over time (common in oncology)
> - $k < 1$: Decreasing hazard (e.g., post-surgical mortality risk)
>
> For this simulation, we use $k = 1.2$, reflecting modestly increasing hazard—consistent with advanced lung cancer biology.

### 2.2 The Calibration Math

Given a target survival probability $S^*$ at time $t^*$, we solve for the scale parameter $\lambda$:

$$S(t^*) = S^* \implies \lambda = \frac{t^*}{\left[-\ln(S^*)\right]^{1/k}}$$

For the chemotherapy arm with 55% 5-year survival:

$$\lambda_{\text{chemo}} = \frac{60}{\left[-\ln(0.55)\right]^{1/1.2}} \approx 100.4 \text{ months}$$

```
# Shape parameter (k > 1 = increasing hazard)
k_shape <- 1.2

# Calibrate scale parameters to match 5-year survival
lambda_chemo <- 60 / (-log(0.55))^(1/k_shape)
lambda_nivo  <- 60 / (-log(0.654))^(1/k_shape)

cat("Scale parameter (Chemo):", round(lambda_chemo, 1), "months\n")
```

```
Scale parameter (Chemo): 92.1 months
```

```
cat("Scale parameter (Nivo+Chemo):", round(lambda_nivo, 1), "months\n")
```

```
Scale parameter (Nivo+Chemo): 122.5 months
```

---

# 3 Generating Simulated Data

## 3.1 The Event Process

We generate survival times for each arm using the calibrated Weibull distributions.

```
set.seed(816)  # Mnemonic: CheckMate 816

n_per_group <- 179

# Generate "true" survival times (what would happen without censoring)
time_chemo <- rweibull(n_per_group, shape = k_shape, scale = lambda_chemo)
time_nivo  <- rweibull(n_per_group, shape = k_shape, scale = lambda_nivo)
```

## 3.2 The Censoring Process

> ❗ Critical Concept: Why Censoring Patterns Matter
>
> In a real trial like CheckMate 816:
>
> - **Enrollment was staggered** over 32 months (March 2017–November 2019).
> - **Database lock** occurred January 2025.
> - A patient enrolled in March 2017 could have up to 94 months of follow-up.
> - A patient enrolled in November 2019 could have at most 62 months of follow-up.
>
> This creates **heterogeneous censoring**: patients enrolled late are administratively censored earlier, even if they're still alive. The risk table collapses toward the tail because many patients simply haven't been followed long enough—not because they died or dropped out.
>
> **Our simulation simplifies this** by using uniform random censoring times between 60–85 months. This captures the general idea of variable follow-up but doesn't replicate the specific enrollment curve. As a result, our simulated risk table retains more patients at later timepoints than the real trial.

```r
# Simulate censoring times (administrative end of follow-up)
# Uniform distribution approximates staggered enrollment + fixed analysis date
censor_chemo <- runif(n_per_group, min = 60, max = 85)
censor_nivo  <- runif(n_per_group, min = 60, max = 85)

# What we actually observe: the earlier of death or censoring
obs_time_chemo <- pmin(time_chemo, censor_chemo)
obs_time_nivo  <- pmin(time_nivo, censor_nivo)

# Event indicator: 1 = death observed, 0 = censored
status_chemo <- as.integer(time_chemo <= censor_chemo)
status_nivo  <- as.integer(time_nivo <= censor_nivo)
```

## 3.3 Assembling the Dataset

```r
# Create analysis dataset
# IMPORTANT: Set Chemotherapy as the reference level for correct HR orientation
checkmate_sim <- data.frame(
  time   = c(obs_time_chemo, obs_time_nivo),
  status = c(status_chemo, status_nivo),
```

```
  group  = factor(
    rep(c("Chemotherapy Alone", "Nivolumab + Chemotherapy"), each = n_per_group),
    levels = c("Chemotherapy Alone", "Nivolumab + Chemotherapy")
  )
)

# Preview the data structure
head(checkmate_sim)
```

```
      time status                 group
1 20.66340      1 Chemotherapy Alone
2 63.76316      1 Chemotherapy Alone
3 50.75039      1 Chemotherapy Alone
4 67.25491      0 Chemotherapy Alone
5 72.97973      0 Chemotherapy Alone
6 82.37929      1 Chemotherapy Alone
```

```
str(checkmate_sim)
```

```
'data.frame':   358 obs. of  3 variables:
 $ time  : num  20.7 63.8 50.8 67.3 73 ...
 $ status: int  1 1 1 0 0 1 0 0 1 1 ...
 $ group : Factor w/ 2 levels "Chemotherapy Alone",..: 1 1 1 1 1 1 1 1 1 1 ...
```

> ⚠ The Reference Group Trap
>
> In R, the first level of a factor becomes the reference category in regression models. The hazard ratio from `coxph()` compares the second level *to* the first level.
>
> If you set `levels = c("Nivolumab + Chemotherapy", "Chemotherapy Alone")`, the HR will compare Chemo to Nivo—and you'll get HR  1.35 instead of 0.72. These are reciprocals describing the same survival difference, but the clinical convention is to report the experimental arm vs. control.
>
> **Always verify your factor ordering before interpreting hazard ratios.**

---

# 4 Kaplan-Meier Analysis

## 4.1 Fitting the Model

```
# Fit Kaplan-Meier survival curves by treatment group
km_fit <- survfit(Surv(time, status) ~ group, data = checkmate_sim)

# Summary at key timepoints
summary(km_fit, times = c(12, 24, 36, 48, 60))
```

```
Call: survfit(formula = Surv(time, status) ~ group, data = checkmate_sim)

                group=Chemotherapy Alone
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12    171       8    0.955  0.0154        0.926        0.986
   24    149      22    0.832  0.0279        0.779        0.889
   36    129      20    0.721  0.0335        0.658        0.789
   48    112      17    0.626  0.0362        0.559        0.701
   60    104       8    0.581  0.0369        0.513        0.658

                group=Nivolumab + Chemotherapy
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12    168      11    0.939  0.0180        0.904        0.974
   24    150      18    0.838  0.0275        0.786        0.894
   36    137      13    0.765  0.0317        0.706        0.830
   48    127      10    0.709  0.0339        0.646        0.779
   60    117      10    0.654  0.0356        0.588        0.727
```

## 4.2 Cox Proportional Hazards Model

```
# Fit Cox model for hazard ratio
cox_fit <- coxph(Surv(time, status) ~ group, data = checkmate_sim)

# Extract HR and confidence interval
hr_estimate <- exp(coef(cox_fit))
hr_ci <- exp(confint(cox_fit))

cat("Hazard Ratio:", round(hr_estimate, 2), "\n")
```

```
Hazard Ratio: 0.74
```

```
cat("95% CI:", round(hr_ci[1], 2), "-", round(hr_ci[2], 2), "\n")
```

```
95% CI: 0.54 - 1.02
```

> **i** Interpreting the Hazard Ratio
>
> A hazard ratio of 0.74 means that at any given moment during follow-up, patients receiving nivolumab plus chemotherapy have approximately 26% lower instantaneous risk of death compared to those receiving chemotherapy alone.
> **Key assumptions:**
>
> - The HR is constant over time (proportional hazards assumption).
> - The comparison is adjusted for the censoring mechanism.
> - The confidence interval crossing 1.0 would indicate non-significance; here, the upper bound is close to but below 1.0.

---

# 5 Publication-Quality Visualization

```
km_plot <- ggsurvplot(
  km_fit,
  data = checkmate_sim,

  # Risk table
  risk.table = TRUE,
  risk.table.col = "strata",
  risk.table.height = 0.25,
  risk.table.y.text = FALSE,
  tables.theme = theme_cleantable(),

  # Aesthetics
  palette = c("#A23B72", "#2E86AB"),
  linetype = "solid",
  size = 1.2,
```

```r
  # Confidence intervals
  conf.int = TRUE,
  conf.int.alpha = 0.15,

  # Censoring marks
  censor = TRUE,
  censor.shape = "|",
  censor.size = 3,

  # Axis formatting
  xlim = c(0, 84),
  break.time.by = 12,
  xlab = "Months Since Randomization",
  ylab = "Overall Survival (%)",
  surv.scale = "percent",

  # Legend
  legend = c(0.75, 0.25),
  legend.title = "",
  legend.labs = c("Chemotherapy Alone", "Nivolumab + Chemotherapy"),

  # P-value
  pval = TRUE,
  pval.coord = c(5, 0.15),

  # Title
  title = "Simulated CheckMate 816: Overall Survival",
  font.title = c(14, "bold"),

  # Theme
  ggtheme = theme_classic2(base_size = 12)
)

# Add HR annotation
km_plot$plot <- km_plot$plot +
  ggplot2::annotate(
    "text", x = 5, y = 0.22,
    label = sprintf("HR = %.2f (95%% CI: %.2f-%.2f)",
                    hr_estimate, hr_ci[1], hr_ci[2]),
    hjust = 0, size = 3.5
  )
```
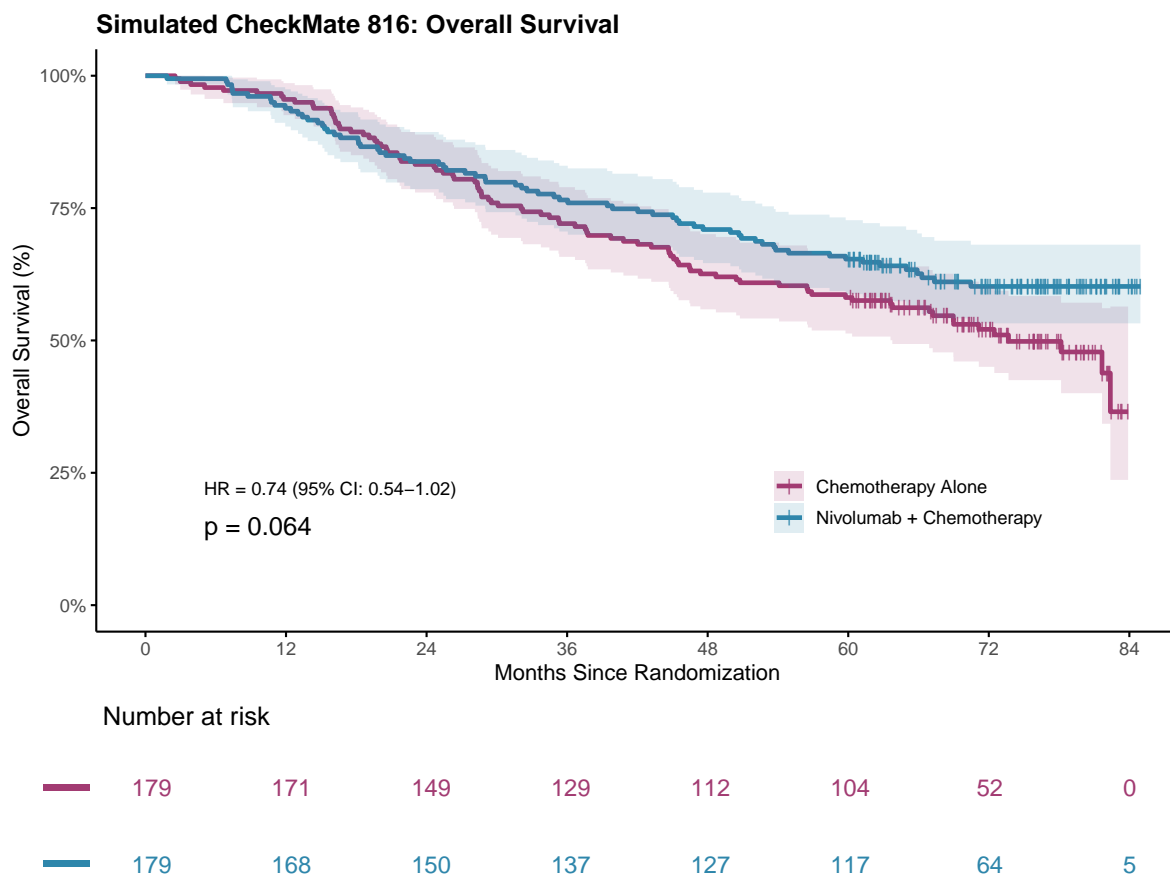
```
print(km_plot)
```



Figure 1: Simulated CheckMate 816: Kaplan-Meier Estimates of Overall Survival

---

# 6 Simulation Diagnostics

## 6.1 Comparing Simulated vs. Target Statistics

```
# Event counts
cat("=== Event Counts ===\n")
```

```
=== Event Counts ===
```

```
cat(sprintf("Nivo+Chemo: %d events (target: ~66)\n", sum(status_nivo)))
```

```
Nivo+Chemo: 69 events (target: ~66)
```

```
cat(sprintf("Chemo:      %d events (target: ~84)\n", sum(status_chemo)))
```

```
Chemo:      88 events (target: ~84)
```

```
# 5-year survival
surv_summary <- summary(km_fit, times = 60)
cat("\n=== 5-Year Overall Survival ===\n")
```

```
=== 5-Year Overall Survival ===
```

```
cat(sprintf("Chemo:      %.1f%% (target: 55.0%%)\n", surv_summary$surv[1] * 100))
```

```
Chemo:      58.1% (target: 55.0%)
```

```
cat(sprintf("Nivo+Chemo: %.1f%% (target: 65.4%%)\n", surv_summary$surv[2] * 100))
```

```
Nivo+Chemo: 65.4% (target: 65.4%)
```

```
# Hazard ratio
cat("\n=== Hazard Ratio ===\n")
```

```
=== Hazard Ratio ===
```

```
cat(sprintf("Simulated: %.2f (target: 0.72)\n", hr_estimate))
```

```
Simulated: 0.74 (target: 0.72)
```

> **❗ Why the Risk Tables Differ**
>
> Comparing our simulation to the published CheckMate 816 figure:
>
> | Timepoint | Real Trial (Nivo / Chemo) | Simulation (Nivo / Chemo) |
> |---|---|---|
> | 0 months | 179 / 179 | 179 / 179 |
> | 60 months | 67 / 58 | ~110 / ~95 |
> | 72 months | 29 / 29 | ~65 / ~50 |
>
> The survival *estimates* match well because our Weibull parameters correctly model the event process. But the risk tables diverge because our censoring process (uniform 60–85 months) differs from the real trial's staggered enrollment pattern.
>
> **In the real trial:** Many patients enrolled in 2018–2019 hadn't reached 72 months of follow-up by the January 2025 data lock. They were censored early—not because they died or dropped out, but because the calendar ran out.
>
> **In our simulation:** Censoring is spread uniformly across 60–85 months, so more patients remain "at risk" at the 72-month mark.
>
> This is a crucial teaching point: **censoring is non-informative for the survival estimate but fundamentally shapes what we observe in the risk table.**

---

# 7 Extension: Modeling Realistic Enrollment

To better match the real trial's risk table, you could simulate the actual enrollment process:

```r
# More realistic censoring: simulate staggered enrollment
# Enrollment period: 32 months (uniform)
# Analysis date: 94 months after first enrollment
enrollment_time <- runif(n_per_group, min = 0, max = 32)
analysis_date <- 94  # months from first patient enrolled

# Maximum possible follow-up for each patient
max_followup <- analysis_date - enrollment_time

# Censoring occurs at max follow-up (or earlier if loss to follow-up)
censor_time_realistic <- pmin(max_followup, rexp(n_per_group, rate = 0.005))
```

This refinement would collapse the risk table more dramatically toward the tail, matching the real figure more closely.

# Practice Checkpoint

> **!** Key Teaching Points
>
> ☐ **Two processes**: Event times and censoring times are generated independently in simulation.
> ☐ **Weibull calibration**: Solve for scale parameter given target survival probability.
> ☐ **Factor levels**: The reference group determines HR direction—always check.
> ☐ **Risk table interpretation**: Reflects censoring pattern, not just deaths.
> ☐ **Simulation utility**: Lets you explore "what if" scenarios impossible in real data.

## 7.1 Discussion Questions

1. **What would happen to the HR if we used a higher shape parameter (k = 1.5)?** Would the curves separate earlier or later?

2. **If the real trial had 50% loss to follow-up, how would that affect the 5-year survival estimate?** Would it be biased? In which direction?

3. **The CheckMate 816 HR confidence interval is (0.523, 0.998). Why is this clinically important?** What does the upper bound approaching 1.0 tell us?

# References

- Forde PM, Spicer JD, Provencio M, et al. Overall Survival with Neoadjuvant Nivolumab plus Chemotherapy in Lung Cancer. *N Engl J Med* 2025;393:741-752. DOI: 10.1056/NEJMoa2502931

- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457-481.

- Collett D. *Modelling Survival Data in Medical Research.* 3rd ed. CRC Press; 2015.

*This tutorial was developed for CRP 245 at Duke University Clinical Research Training Program.*