

# Chapter 4: Nonparametric Comparison of Survival Distributions

## Log-Rank Tests, Weighted Tests, and Stratification

Eric Delmelle | Lehigh University

### Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Log-Rank Test Basics</b>	<b>2</b>
2.1	Core Concept . . . . .	2
2.2	Simple Example . . . . .	2
<b>3</b>	<b>Weighted Log-Rank Tests</b>	<b>3</b>
3.1	Fleming-Harrington Family . . . . .	3
3.2	Pancreatic Cancer Example . . . . .	4
<b>4</b>	<b>Stratified Tests</b>	<b>5</b>
4.1	Smoking Cessation Example . . . . .	6
<b>5</b>	<b>When to Use Which Test</b>	<b>7</b>
<b>6</b>	<b>Exercises</b>	<b>7</b>
6.1	Exercise 4.1: Weighted Tests Comparison . . . . .	7
6.2	Exercise 4.2: Employment Stratification . . . . .	8
6.3	Exercise 4.3: Wilcoxon vs Log-Rank . . . . .	8
6.4	Exercise 4.4: Overall Survival . . . . .	9
<b>7</b>	<b>Key Points</b>	<b>10</b>

[Link to qmd \(quarto markdown\)](#)

# 1 Introduction

- **Log-rank test** compares survival curves between groups without assuming specific distributions.
- Nonparametric equivalent of the t-test for survival data.

**Key Applications:** - Treatment comparison in clinical trials - Prognostic factor analysis - Group comparisons with censored data

## 2 Log-Rank Test Basics

### 2.1 Core Concept

- At each event time, create a 2×2 table comparing observed vs expected events between groups.
- The test statistic sums these comparisons across all event times.

### 2.2 Simple Example

```
# Example: 16 patients, 8 in each group
tt <- c(4, 7, 8, 12, 15, 18, 20, 25,      # Control group
        6, 9, 14, 16, 19, 22, 28, 30)    # Treatment group

delta <- c(1, 0, 1, 1, 0, 1, 1, 0,        # Control: 5 events, 3 censored
           0, 1, 0, 0, 0, 0, 1, 1)       # Treatment: 3 events, 5 censored

trt <- c(rep(0, 8), rep(1, 8))           # 0=control, 1=treatment

# Log-rank test
survdif(Surv(tt, delta) ~ trt)
```

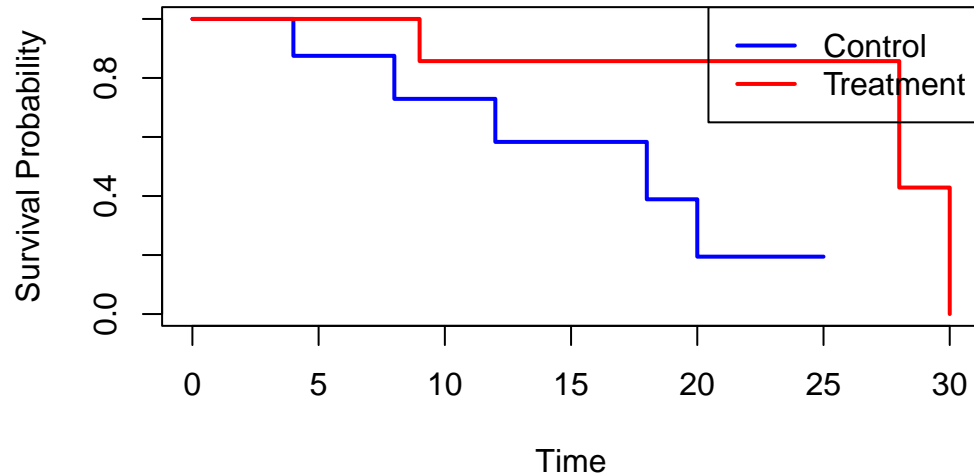
Call:

```
survdif(formula = Surv(tt, delta) ~ trt)
```

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
trt=0	8	5	2.66	2.06	3.71
trt=1	8	3	5.34	1.02	3.71

Chisq= 3.7 on 1 degrees of freedom, p= 0.05

```
# Visualize
km_fit <- survfit(Surv(tt, delta) ~ trt)
plot(km_fit, col = c("blue", "red"), lwd = 2,
     xlab = "Time", ylab = "Survival Probability")
legend("topright", legend = c("Control", "Treatment"),
     col = c("blue", "red"), lwd = 2)
```



**Result:**  $p = 0.05$  (significant difference)

### 3 Weighted Log-Rank Tests

Different weights emphasize early vs. late differences.

#### 3.1 Fleming-Harrington Family

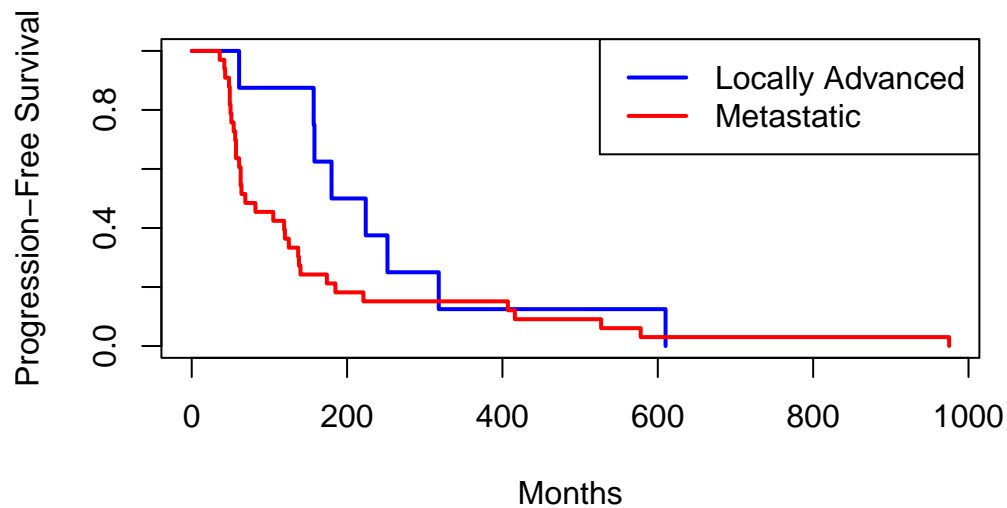
$$w_i = [\hat{S}(t_i)]^\rho$$

- $\rho = 0$ : Standard log-rank (equal weights)
- $\rho = 1$ : Prentice-Gehan (emphasizes early differences)

### 3.2 Pancreatic Cancer Example

```
# Use pancreatic2 dataset
data(pancreatic2)

# Survival curves
km_pancreatic <- survfit(Surv(pfs) ~ stage, data = pancreatic2)
plot(km_pancreatic, col = c("blue", "red"), lwd = 2,
     xlab = "Months", ylab = "Progression-Free Survival")
legend("topright", legend = c("Locally Advanced", "Metastatic"),
     col = c("blue", "red"), lwd = 2)
```



```
# Standard log-rank ( = 0)
cat("Standard Log-Rank:\n")
```

Standard Log-Rank:

```
survdif(Surv(pfs) ~ stage, rho = 0, data = pancreatic2)
```

Call:

```
survdif(formula = Surv(pfs) ~ stage, data = pancreatic2, rho = 0)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
stage=LA	8	8	12.3	1.49	2.25

```
stage=M 33      33      28.7      0.64      2.25
```

```
Chisq= 2.2 on 1 degrees of freedom, p= 0.1
```

```
# Prentice-Gehan ( = 1)
cat("\nPrentice-Gehan:\n")
```

Prentice-Gehan:

```
survdif(Surv(pfs) ~ stage, rho = 1, data = pancreatic2)
```

Call:

```
survdif(formula = Surv(pfs) ~ stage, data = pancreatic2, rho = 1)
```

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
stage=LA	8	2.34	5.88	2.128	4.71
stage=M	33	18.76	15.22	0.822	4.71

```
Chisq= 4.7 on 1 degrees of freedom, p= 0.03
```

**Key Finding:** - Standard test  $p = 0.134$  (not significant) - Prentice-Gehan  $p = 0.030$  (significant). - Why? The curves separate early but converge later - the weighted test detects the early difference.

## 4 Stratified Tests

Control for categorical confounders by stratifying.

$$\chi^2 = \frac{[\sum_{g=1}^G U_{0g}]^2}{\sum_{g=1}^G V_{0g}}$$

## 4.1 Smoking Cessation Example

```
data(pharmacoSmoking)

# Unstratified test
cat("Unstratified:\n")
```

Unstratified:

```
survdif(Surv(ttr, relapse) ~ grp, data = pharmacoSmoking)
```

Call:

```
survdif(formula = Surv(ttr, relapse) ~ grp, data = pharmacoSmoking)
```

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
grp=combination	61	37	49.9	3.36	8.03
grp=patchOnly	64	52	39.1	4.29	8.03

Chisq= 8 on 1 degrees of freedom, p= 0.005

```
# Create age groups
pharmacoSmoking$ageGroup <- ifelse(pharmacoSmoking$age <= 49, "49", "50+")

# Stratified by age
cat("\nStratified by Age:\n")
```

Stratified by Age:

```
survdif(Surv(ttr, relapse) ~ grp + strata(ageGroup), data = pharmacoSmoking)
```

Call:

```
survdif(formula = Surv(ttr, relapse) ~ grp + strata(ageGroup),
  data = pharmacoSmoking)
```

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
grp=combination	61	37	49.1	2.99	7.03
grp=patchOnly	64	52	39.9	3.68	7.03

Chisq= 7 on 1 degrees of freedom, p= 0.008

```
# Check age distribution
table(pharmacoSmoking$grp, pharmacoSmoking$ageGroup)
```

```

           49 50+
combination 31 30
patchOnly   35 29

```

**Result:** - Chi-square barely changes ( $8.0 \rightarrow 7.0$ ) - Indicates age is not a strong confounder here.

## 5 When to Use Which Test

Scenario	Test	Use When
Proportional hazards	Log-rank ( $\rho = 0$ )	Constant treatment effect
Early effect	Prentice-Gehan ( $\rho = 1$ )	Treatment works early
Confounding	Stratified	Categorical confounder
Multiple confounders	Cox regression	Many covariates

## 6 Exercises

### 6.1 Exercise 4.1: Weighted Tests Comparison

```
# Compare log-rank vs Prentice-Gehan
result_lr <- survdiff(Surv(ttr, relapse) ~ grp, rho = 0, data = pharmacoSmoking)
result_pg <- survdiff(Surv(ttr, relapse) ~ grp, rho = 1, data = pharmacoSmoking)

cat("Log-rank p-value:", round(1 - pchisq(result_lr$chisq, 1), 4), "\n")
```

Log-rank p-value: 0.0046

```
cat("Prentice-Gehan p-value:", round(1 - pchisq(result_pg$chisq, 1), 4), "\n")
```

Prentice-Gehan p-value: 0.0047

## 6.2 Exercise 4.2: Employment Stratification

```
# Check available variables
print(names(pharmacoSmoking)[1:10]) # First 10 variables
```

```
[1] "id"          "ttr"          "relapse"      "grp"          "age"
[6] "gender"      "race"         "employment"   "yearsSmoking" "levelSmoking"
```

```
# Stratify by employment (if variable exists)
if("employment" %in% names(pharmacoSmoking)) {
  survdiff(Surv(ttr, relapse) ~ grp + strata(employment), data = pharmacoSmoking)
} else {
  cat("Employment variable not found in standard form\n")
  # Show available factor variables
  factor_vars <- sapply(pharmacoSmoking, is.factor)
  print(names(pharmacoSmoking)[factor_vars])
}
```

Call:

```
survdiff(formula = Surv(ttr, relapse) ~ grp + strata(employment),
  data = pharmacoSmoking)
```

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
grp=combination	61	37	50.3	3.50	8.58
grp=patchOnly	64	52	38.7	4.54	8.58

Chisq= 8.6 on 1 degrees of freedom, p= 0.003

## 6.3 Exercise 4.3: Wilcoxon vs Log-Rank

```
# For pancreatic data (no censoring)
os_days <- as.numeric(pancreatic2$os)

# Wilcoxon test
wilcox_result <- wilcox.test(os_days ~ pancreatic2$stage)

# Log-rank test
logrank_result <- survdiff(Surv(os_days) ~ stage, data = pancreatic2)
```



```
cat("Wilcoxon p-value:", round(wilcox_result$p.value, 4), "\n")
```

Wilcoxon p-value: 0.2924

```
cat("Log-rank p-value:", round(1 - pchisq(logrank_result$chisq, 1), 4), "\n")
```

Log-rank p-value: 0.5145

## 6.4 Exercise 4.4: Overall Survival

```
# Overall survival analysis
survdif(Surv(os_days) ~ stage, rho = 0, data = pancreatic2) # Log-rank
```

Call:

```
survdif(formula = Surv(os_days) ~ stage, data = pancreatic2,
        rho = 0)
```

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
stage=LA	8	8	9.74	0.3093	0.425
stage=M	33	33	31.26	0.0963	0.425

Chisq= 0.4 on 1 degrees of freedom, p= 0.5

```
survdif(Surv(os_days) ~ stage, rho = 1, data = pancreatic2) # Prentice-Gehan
```

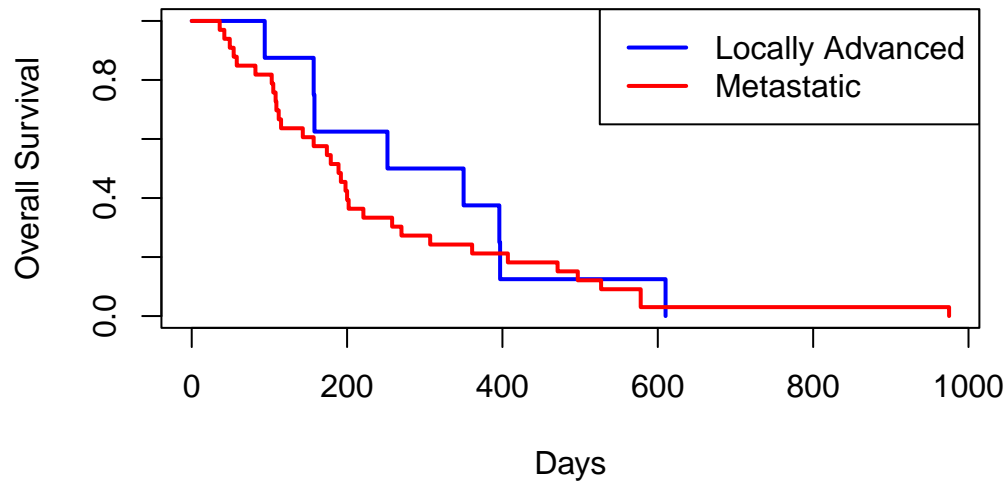
Call:

```
survdif(formula = Surv(os_days) ~ stage, data = pancreatic2,
        rho = 1)
```

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
stage=LA	8	3.32	4.9	0.513	1.02
stage=M	33	17.73	16.1	0.156	1.02

Chisq= 1 on 1 degrees of freedom, p= 0.3

```
# Visualize
km_os <- survfit(Surv(os_days) ~ stage, data = pancreatic2)
plot(km_os, col = c("blue", "red"), lwd = 2,
     xlab = "Days", ylab = "Overall Survival")
legend("topright", legend = c("Locally Advanced", "Metastatic"),
     col = c("blue", "red"), lwd = 2)
```



## 7 Key Points

- **Log-rank test:** Standard for comparing survival curves
- **Weighted tests:** Use when treatment effects vary over time
- **Stratification:** Controls categorical confounders
- **Test choice matters:** Different tests detect different patterns
- **Always visualize:** Plots reveal the nature of differences

The log-rank family provides robust nonparametric comparison methods, but choose the right variant for your specific research question.

[← Return to Course Materials Link to qmd \(quarto markdown\)](#)