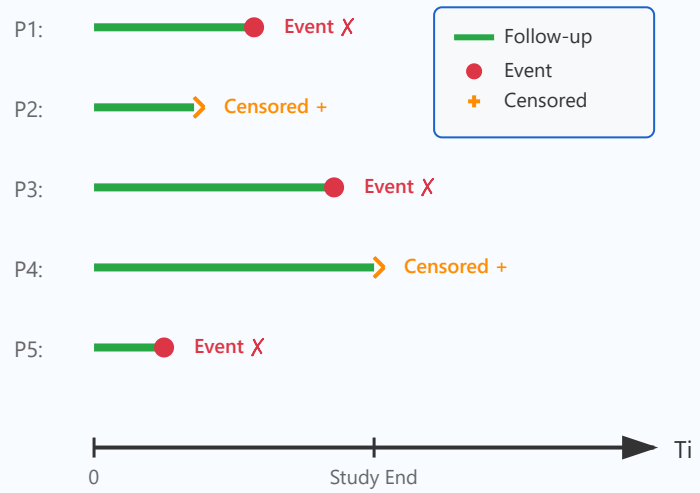
Survival Analysis

Time-to-event analysis: Death, disease recurrence, hospital readmission

Censoring Visualization



Censoring Types

- Right censoring (most common)
- Left censoring
- Interval censoring

Key Functions

- Survival function S(t)
- Hazard function h(t)
- Cumulative hazard H(t)

Advanced Topics

- Competing risks
- Recurrent events
- Time-varying covariates

Clinical Applications

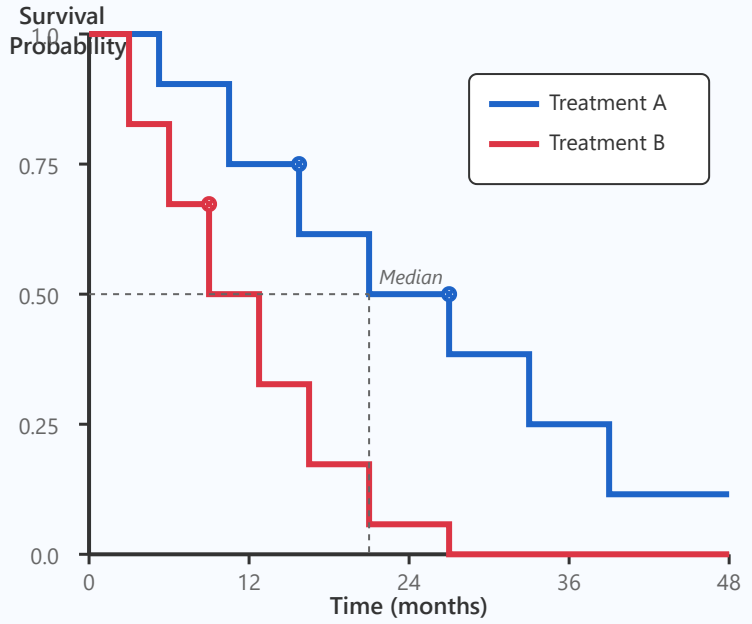
- Overall survival (OS)
- Progression-free survival (PFS)
- Time to treatment failure

Kaplan-Meier Survival Curve

Survival Curve Example

Kaplan-Meier Method

A **non-parametric survival estimation** method that accurately calculates survival probabilities even with censored data.



$$S(t) = \prod(1 - d_i/n_i)$$

- d_i: Number of events at time i
- n_i: Number at risk at time i
- Represented as a step function

Interpretation Points

- **Median survival**: Time when S(t) = 0.5
- **Curve comparison**: Use log-rank test
- **Confidence intervals**: Greenwood's formula
- Treatment A shows superior survival vs B

Cox Proportional Hazards Model

Hazard Ratio Visualization

Cox Proportional Hazards Model

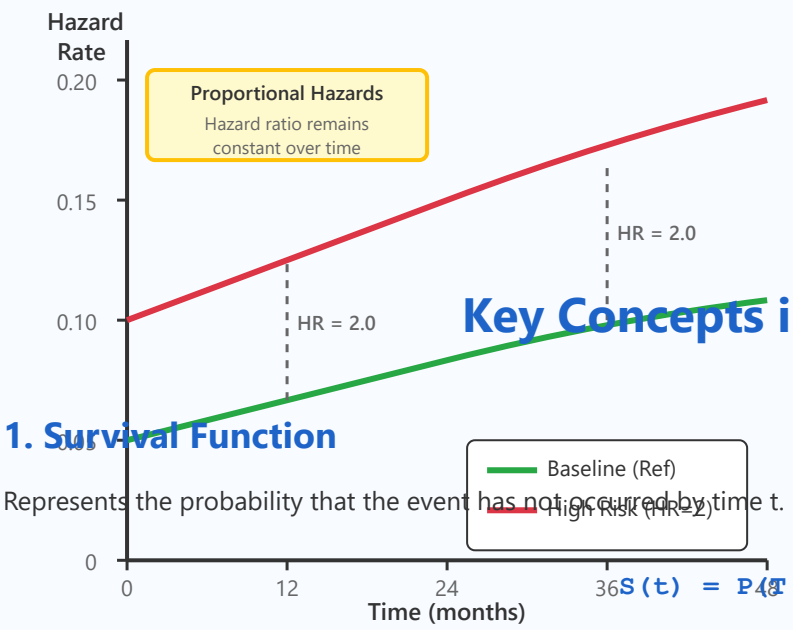
A **semi-parametric regression model** that simultaneously evaluates the effects of multiple covariates on survival.

$$h(t|X) = h_0(t) \times \exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p)$$

- $h_0(t)$: Baseline hazard function
- **exp(β)**: Hazard Ratio (HR)
- **Proportionality assumption**: HR is time-independent

Hazard Ratio (HR) Interpretation

Key Concepts in Survival Analysis



1. Survival Function

Represents the probability that the event has not occurred by time t.

$$S(t) = P(T > t) = 1 - F(t)$$

- S(0) = 1 (all subjects survive at start)
- S(∞) = 0 (eventually all experience event)
- Monotonically decreasing function

2. Hazard Function

The instantaneous risk of experiencing the event at time t, given survival until t.

$$h(t) = \lim_{\Delta t \to 0} P(t \leq T < t+\Delta t \mid T \geq t) / \Delta t$$

- **Constant**: Exponential distribution
- **Increasing**: Weibull distribution
- **Decreasing then increasing**: Log-normal

3. Cumulative Hazard

The accumulated hazard from time 0 to time t.

$$H(t) = \int_0^t h(u)du = -\ln[S(t)]$$

- Estimated by Nelson-Aalen estimator
- S(t) = exp[-H(t)]
- Interconvertible with survival function

4. Log-Rank Test

- Schoenfeld residuals test
- Log-log survival curve parallelism check

Clinical Research Application Example

Study Scenario: New Drug Clinical Trial

5. Censoring Handling

Objective: Evaluate whether a new cancer drug (Drug A) improves overall survival compared to standard treatment.

Population: 200 patients with advanced lung cancer (100 per arm)

Follow-up: Maximum 48 months

$$\chi^2 = [\Sigma(O_i - E_i)]^2 / \Sigma \text{Var}(O_i - E_i)$$

- Compares observed vs expected events

Step 1: Study Design

- **Primary endpoint**: Overall survival (OS)
- **Secondary endpoints**: Progression-free survival (PFS)
- **Covariates**: Age, sex, disease stage, ECOG score
- **Event definition**: Death from any cause

Step 2: Data Collection

Methods to utilize information from incomplete observations.

- **Non-informative censoring**: Independent of event
- Information until censoring included
- Removed from risk set afterwards
- Key strength of survival analysis

- From enrollment to death or last follow-up
- Censored: Lost to follow-up, withdrew, study end
- Events observed: 118 (59%)
- Censored: 82 (41%)

Step 3: Statistical Analysis

1 Kaplan-Meier Analysis

- Estimate survival curves for each arm
- Calculate median survival times
- Present 95% confidence intervals

Results:

Drug A: Median survival 26.5 months

Standard: Median survival 18.2 months

2 Log-Rank Test

- Compare survival curves between arms
- Consider entire follow-up period

Results:

$\chi^2 = 8.42$, $p = 0.004$

→ Statistically significant difference

Advanced Topics and Considerations

Violation of Proportional Hazards

Alternatives when the key Cox model assumption is not met:

1. Stratified Cox Model

- Stratify by variables violating proportionality
- Estimate separate baseline hazards per stratum

2. Time-Dependent Covariates

- Include time-varying variables
- Example: X(t) = X × g(t)

3. Accelerated Failure Time Model

Competing Risks Analysis

When multiple types of events can occur:

Example Scenarios

- Cancer death vs other-cause death
- Recurrence vs death
- Transplant: Transplanted vs died waiting

Analysis Methods

- **Cumulative Incidence Function (CIF)**: Probability of specific event
- **Fine-Gray Model**: Subdistribution hazard
- Kaplan-Meier may overestimate with competing risks

• Directly model survival time
• Use log-normal, Weibull distributions, etc.

ndar

Treatment effect remains significant after adjusting for other prognostic factors

Recurrent Events Analysis

Multiple events possible in same subject:

• Hospitalizations, recurrences, infections, etc.

Analysis Approaches

• **Anderson-Gill Model**: All events independent
• **PWP Model**: Sequential conditional events
• **Frailty Model**: Account for within-subject correlation

Sample Size Calculation

Considerations for survival study design:

$$\text{Events} = (Z_{1-\alpha/2} + Z_{1-\beta})^2 / [p_1(1-p_1)(\log HR)^2]$$

• Number of events determines power
• Consider follow-up duration and dropout
• Higher censoring requires more enrollment

Key Cautions and Checklist

⚠ Data Quality

✓ Is event definition clear?
✓ Are censoring reasons documented?
✓ Is the time origin clearly defined?

⚠ Result Reporting

✓ Present median follow-up time
✓ Display number at risk
✓ Report with confidence intervals

⚠ Model Assumptions

✓ Test proportional hazards assumption
✓ Verify non-informative censoring
✓ Check for outliers and influential observations

⚠ Interpretation

✓ Distinguish clinical vs statistical significance
✓ Consider adjustment for multiple testing
✓ Check for potential confounders