Chapter 17: Causal Survival Analysis

Contents

This chapter extends causal inference methods to **survival analysis** and **time-to-event outcomes**. We define causal effects for survival data, address challenges posed by censoring and competing events, and show how to estimate causal survival curves using IP weighting and the parametric g-formula.

This chapter is based on Hernán and Robins (2020, chap. 17, pp. 247-264).

Key challenge: Survival outcomes involve time, and individuals may be censored before experiencing the event. This requires careful definition of potential outcomes and attention to competing risks.

1 17.1 Hazards and Risks (pp. 247-250)

We begin by reviewing key concepts from survival analysis.

1.1 Basic Definitions

Definition 1.1 (Survival Analysis Terms)**. Event time** T: Time from baseline until an event occurs (e.g., death, disease onset)

Censoring time C: Time from baseline until censoring (e.g., loss to follow-up, study end)

Observed time $Y = \min(T, C)$: The time we actually observe

Event indicator $D = I(T \leq C)$: 1 if event occurred, 0 if censored

Risk at time t: $\Pr[T \leq t]$ (cumulative incidence)

Survival at time t: $S(t) = \Pr[T > t] = 1 - \Pr[T \leq t]$

Hazard at time t: $\lambda(t) = \lim_{dt \to 0} \frac{\Pr[t \leq T < t+dt | T \geq t]}{dt}$

1.2 Hazard vs Risk

Hazard: Instantaneous rate of event occurrence among those at risk

- Ranges from 0 to ∞
- Can increase, decrease, or remain constant over time
- Related to survival: $S(t) = \exp\left(-\int_0^t \lambda(s)ds\right)$

Risk (cumulative incidence): Probability of event by time t

- Ranges from 0 to 1
- Always non-decreasing
- More interpretable for most purposes

Relationship:

$$\text{Risk}(t) = 1 - \exp\left(-\int_0^t \lambda(s)ds\right)$$

Intuition: The hazard is like a velocity (rate of change), while risk is like a distance traveled (cumulative probability).

For causal inference, we typically focus on **causal risks** rather than causal hazards, though both can be defined.

2 17.2 From Hazards to Risks (pp. 250-252)

The **survival function** $S(t)$ and **cumulative hazard** $\Lambda(t)$ are key quantities.

2.1 Cumulative Hazard

$$\Lambda(t) = \int_0^t \lambda(s)ds$$

Interpretation: Total accumulated hazard up to time t.

2.2 Relationships

Survival from cumulative hazard:

$$S(t) = \exp(-\Lambda(t))$$

Risk from survival:

$$\Pr[T \leq t] = 1 - S(t) = 1 - \exp(-\Lambda(t))$$

2.3 Kaplan-Meier Estimator

Nonparametric estimator of survival function:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

where: - t_i are the ordered event times - d_i is the number of events at t_i - n_i is the number at risk just before t_i

Assumption: Censoring is **independent** of event time (given covariates).

Kaplan-Meier provides a nonparametric estimate of the survival curve without assuming a parametric model for the hazard.

Limitation: Doesn't adjust for confounders. For causal inference, we need IP weighting or parametric g-formula.

3 17.3 Why Censoring Matters (pp. 252-255)

Censoring creates missing data problems in survival analysis.

3.1 Types of Censoring

Administrative censoring: Study ends before all events occur

- Predictable, often at a fixed time
- Relatively benign if independent of event risk

Loss to follow-up: Participants drop out before event or study end

- Potentially problematic
- May be related to prognosis (informative censoring)

Competing risks: Individuals experience a different event that precludes the event of interest

- e.g., death from other causes when studying heart attack
- Requires special treatment

3.2 Independent Censoring Assumption

Definition 3.1 (Independent Censoring)**.** Censoring is **independent** if:

$$T \perp\!\!\!\perp C \mid L$$

where T is event time, C is censoring time, L are measured covariates.

Interpretation: Among individuals with the same covariate values, censoring is unrelated to their (unobserved) event time.

When violated: Standard survival methods (Kaplan-Meier, Cox regression) are biased.

Solution: Use IP weighting for censoring (Section 17.5).

Informative censoring: When censoring is related to prognosis even after adjusting for covariates.

Example: Patients who feel sicker may be more likely to drop out AND more likely to have events. If we can measure "feeling sick", we can adjust. If not, we have a problem.

Practical approach: Measure as many predictors of censoring as possible, then assume independence given those predictors.

4 17.4 The Hazard Ratio (pp. 255-257)

The **hazard ratio** is commonly used to quantify treatment effects in survival analysis.

4.1 Definition

Definition 4.1 (Hazard Ratio)**.** The **hazard ratio** comparing treatment a to a' is:

$$HR^{a,a'}(t) = \frac{\lambda^a(t)}{\lambda^{a'}(t)}$$

where $\lambda^a(t)$ is the hazard function under treatment a.

Cox proportional hazards model: Assumes $HR^{a,a'}(t) = HR^{a,a'}$ (constant over time).

4.2 Interpretation Challenges

Collapsibility: Unlike risk differences and risk ratios, hazard ratios are **non-collapsible**.

- Marginal HR $\neq$ average of conditional HRs
- Adjusted HR from Cox model is **not** a causal marginal effect
- Conditional on covariates, HR has complex interpretation

Built-in selection bias: Hazards condition on survival to time t, creating selection bias when effects are heterogeneous.

Why hazard ratios are tricky:

1. **Non-collapsibility**: Even under no confounding, adjusted HR $\neq$ unadjusted HR
2. **Selection bias**: By conditioning on $T \geq t$, we select a subset of the population that changes over time
3. **No causal interpretation**: The HR from a Cox model is not generally a causal parameter

Better approach for causal inference: Focus on **causal risk differences** or **causal risk ratios**, not hazard ratios.

5 17.5 IP Weighting of Survival Curves (pp. 257-260)

We can use **IP weighting** to estimate causal survival curves.

5.1 Causal Survival Curve

Definition 5.1 (Causal Survival Function)**.** The **causal survival function** under treatment a is:

$$S^a(t) = \Pr[T^a > t]$$

where T^a is the potential event time under treatment a.

Causal risk at time t: $\Pr[T^a \leq t] = 1 - S^a(t)$

5.2 IP Weighted Kaplan-Meier

Goal: Estimate $S^a(t)$ adjusting for confounding.

Method: IP weighted Kaplan-Meier estimator

1. Compute IP weights for treatment: $W^A = \frac{1}{\Pr[A|L]}$
2. Apply weighted Kaplan-Meier:

$$\hat{S}^a(t) = \prod_{t_i \leq t} \left(1 - \frac{\sum_{j:A_j=a} W_j^A I(Y_j = t_i, D_j = 1)}{\sum_{j:A_j=a} W_j^A I(Y_j \geq t_i)}\right)$$

Result: Estimate of causal survival curve under treatment a.

5.3 Handling Censoring

Joint weights for treatment and censoring:

$$W_i^{A,C} = \frac{1}{\Pr[A_i \mid L_i]} \times \frac{1}{\Pr[C_i > t \mid A_i, L_i, \bar{Y}_i(t)]}$$

where $\bar{Y}_i(t)$ represents the history of being at risk up to time t.

Stabilized weights can improve stability:

$$SW^{A,C} = \frac{\Pr[A]}{\Pr[A \mid L]} \times \frac{\Pr[C > t \mid A]}{\Pr[C > t \mid A, L, \bar{Y}(t)]}$$

Time-dependent weights: Censoring weights can change over time if censoring depends on time-varying covariates or outcome history.

Estimation:

1. Model $\Pr[C > t \mid A, L, \bar{Y}(t)]$ using pooled logistic regression
2. For each individual at each time, compute censoring probability
3. Create weights as product of treatment weights and censoring weights
4. Apply weighted Kaplan-Meier

This is implemented in various R packages (e.g., `ipw`, `WeightIt`).

6 17.6 The Parametric G-Formula for Survival Data (pp. 260-262)

The **parametric g-formula** can also be used for survival outcomes.

6.1 Discrete-Time Approach

Model: Hazard at each time point t

$$\Pr[T = t \mid T \geq t, A, L] = \text{expit}(\alpha_0(t) + \alpha_1 A + \alpha_2^\top L)$$

This can be fit using **pooled logistic regression**:

- Create one record per person per time period
- Include time as a predictor (e.g., dummy variables for each t)
- Fit logistic regression

G-formula algorithm:

1. Fit pooled logistic model for $\Pr[T = t \mid T \geq t, A, L]$
2. For each individual i and each time t:
   - Predict $\hat{p}_{it}^a = \Pr[T = t \mid T \geq t, A = a, L_i]$
3. Compute survival probabilities:
   - $\hat{S}_i^a(t) = \prod_{s=1}^{t}(1 - \hat{p}_{is}^a)$
4. Average over individuals:
   - $\hat{S}^a(t) = \frac{1}{n}\sum_{i=1}^{n} \hat{S}_i^a(t)$

6.2 Continuous-Time Approach

For continuous time, use **parametric survival models**:

- Weibull, exponential, log-normal, etc.
- Specify hazard function with parameters depending on A, L
- Predict survival curves for each individual under each treatment
- Average to get causal survival curves

Pooled logistic regression:

Approximates the continuous-time hazard when time intervals are small. As intervals $\to 0$, converges to Cox model.

Advantages over IP weighting:

- May be more efficient when outcome model is correct
- Easier to incorporate time-varying covariates
- Natural for discrete-time data

Disadvantages:

- Requires correct specification of outcome model
- More complex with time-varying treatments

7 17.7 Competing Risks (pp. 262-264)

Competing risks occur when multiple types of events can prevent observation of the event of interest.

7.1 Definition

Definition 7.1 (Competing Risks)**.** A **competing risk** is an event that precludes the occurrence of the event of interest.

Example: When studying heart attack incidence, death from cancer is a competing risk.

If someone dies from cancer, they can never have a heart attack (or at least we can't observe it).

7.2 Challenges

Censoring is not independent: Individuals who experience competing events may have different risk of the event of interest.

Standard survival methods fail: Treating competing events as censoring leads to biased estimates.

7.3 Approaches

1. Cause-specific hazards:

- Model hazard for each type of event separately
- Use cause-specific Cox models
- Provides hazard ratios, not easily interpretable as causal effects

2. **Subdistribution hazards** (Fine-Gray model):

- Models cumulative incidence function directly
- Individuals with competing events remain in risk set
- More interpretable but still has issues

3. **Parametric g-formula**:

- Model all event types jointly
- Simulate outcomes under each treatment
- Compute cumulative incidence of each event type
- **Recommended for causal inference**

Why g-formula for competing risks:

The g-formula naturally handles competing risks by modeling the joint distribution of all outcomes. Under an intervention to set $A = a$:

1. Predict which event type each person would experience first
2. Compute the proportion experiencing each event type by time t
3. These are the causal cumulative incidence functions

Practical implementation:

- Use multinomial logistic regression for multiple event types
- Or separate models for each event type with appropriate risk sets
- Simulate individual trajectories under each treatment
- Summarize to get causal curves

8 Summary

Key concepts:

1. **Survival analysis**: Study of time-to-event outcomes with censoring
2. **Hazard vs risk**: Instantaneous rate vs cumulative probability
3. **Censoring**: Missing data problem requiring careful assumptions
4. **Causal survival curves**: $S^a(t) = \Pr[T^a > t]$ under treatment a
5. **IP weighted Kaplan-Meier**: Adjusts for confounding via weighting
6. **Parametric g-formula**: Models hazards, predicts survival curves, averages
7. **Competing risks**: Multiple event types that preclude each other

Methods for causal survival analysis:

Method	Approach	Advantages	Disadvantages
IP weighted KM	Weight observations	Nonparametric, robust	Needs correct treatment/censoring models
Parametric g-formula	Model hazards	Efficient, handles competing risks	Needs correct outcome model
Cox model	Model conditional hazard	Familiar, flexible	HR not causally interpretable

Causal estimands:

- **Causal risk difference**: $\Pr[T^{a=1} \leq t] - \Pr[T^{a=0} \leq t]$
- **Causal risk ratio**: $\frac{\Pr[T^{a=1} \leq t]}{\Pr[T^{a=0} \leq t]}$
- **Causal survival ratio**: $\frac{S^{a=1}(t)}{S^{a=0}(t)}$

NOT generally causal:

- Hazard ratio from Cox model with covariates (non-collapsible, built-in selection)

Assumptions:

1. **Conditional exchangeability**: $T^a \perp\!\!\!\perp A \mid L$
2. **Independent censoring**: $T \perp\!\!\!\perp C \mid A, L$
3. **Positivity**: $\Pr[A = a \mid L] > 0$ and $\Pr[C > t \mid A, L] > 0$
4. **Correct models**: Treatment, censoring, and/or outcome models

Practical recommendations:

1. **Focus on risks**, not hazards, for causal inference
2. **Adjust for confounding** using IP weighting or g-formula
3. **Handle censoring** via IP weighting or modeling
4. **Use g-formula for competing risks**
5. **Check assumptions**, especially independent censoring
6. **Report survival curves**, not just summary measures

Common mistakes:

1. Interpreting adjusted hazard ratios as causal effects
2. Treating competing events as independent censoring
3. Ignoring time-varying confounding (see Part III)
4. Using methods that assume independent censoring when it's violated

Looking ahead: Part III extends these ideas to time-varying treatments, where survival analysis becomes even more complex but the g-formula and IP weighting remain powerful tools.

Hernán, Miguel A, and James M Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. https://miguelhernan.org/whatifbook.