CAUSAL SURVIVAL ANALYSIS

Barbra Dickerman, Joy Shi, Miguel Hernán
DEPARTMENTS OF EPIDEMIOLOGY AND BIOSTATISTICS

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Learning objectives
At the end of this lecture you will be able to

- Describe the peculiarities of survival analysis
- Construct survival curves via nonparametric hazard estimates
- Construct survival curves adjusted for confounders via parametric hazard estimates

☐ Key concepts
- Hazard
- Risk, survival

A new type of causal question

- So far we have been concerned with the causal effect of treatment on an outcome at a particular time point
  - effect of smoking cessation on weight gain in 1982
  - effect of smoking cessation on death status in 1992
- Today we discuss the causal effect of treatment on the time until the occurrence of an event of interest
  - effect of smoking cessation on the time until death, whenever death occurs
- This is an example of a **causal survival analysis**

Survival analysis
also known as failure time analysis

- The use of the word "survival" does not imply that the event of interest must be death
  - the event may be death, marriage, incarceration, cancer, flu infection, etc.
  - Here we will use the terms "survival" and "death" for simplicity
- Survival analyses require some special techniques
  - EPI289 describes techniques for time-fixed treatments
  - EPI207 for time-varying treatments

Study population

- 1629 cigarette smokers
- Aged 25-74 years when interviewed in 1971-75 and **alive on Jan 1, 1983** (baseline)
  - answered the general medical history questionnaire in 1971-75
  - known sex, age, race, weight, height, education, alcohol use, and smoking intensity between baseline and 1982
- 318 individuals died by the end of 1992
  - Follow-up: 10 years

Key variables

Treatment A	Quit smoking between baseline and 1982 1: yes, 0: no
Time of death T	Month of death Ranges from 1 (start of follow-up, in January 1983) to 120 (administrative end of follow-up)
Baseline (pre-treatment) covariates	Age, sex, race, alcohol use, intensity of smoking, weight…

So far we have defined the average causal effect as a difference of counterfactual means

if everybody had quit smoking

- $\mathrm{E}[Y^{a=1}]$
- $Y^{a=1}$ is a subject's outcome under $a=1$

if nobody had quit smoking

- $\mathrm{E}[Y^{a=0}]$
- $Y^{a=0}$ is a subject's outcome under $a=0$

☐ The average causal effect is $\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$?

- Ignore censoring for now

In survival analyses without losses to follow-up, we define the causal effect as $E[T^{a=1}] - E[T^{a=0}]$

True 0%

False 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Administrative end of follow-up

- The time when a follow-up study stops
  - 120 months in our study
- Time of death *T* is unknown for individuals who reach the end of follow-up alive
  - these individuals are administratively censored
  - cannot compute mean survival time $\mathrm{E}[T]$
- In our data with 1629 individuals:
  - 318 died before the end of 1992
  - 1,311 were administratively censored

See survival.R, lines 9-13

Administrative censoring is typically present

- Follow-up studies will rarely follow all individuals until all of them die
  - or develop the event of interest
- Survival analyses need measures that accommodate administrative censoring
  - For example, the survival probability
  - Not the mean time of death
- Of course, survival analyses also need to handle non-administrative censoring by loss to follow-up
  - Same as any other analyses

Survival probability (or just Survival)

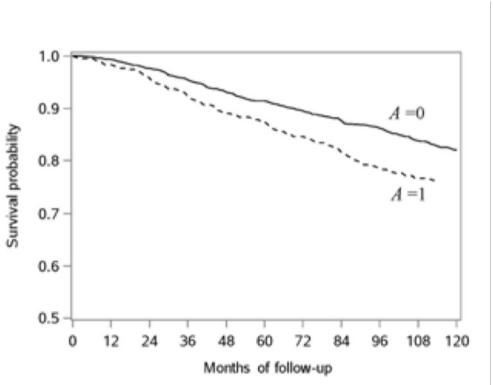
- Pr $[T>k]$
  - where $k = 0, 1, 2\ldots k_{end}$ is time of follow-up
  - Probability of not having the event by time k
- Survival curve
  - Plot of survival probabilities for $k = 0, 1, 2\ldots k_{end}$
  - Starts at 1 at time $k=0$ and decreases monotonically
- The survival curve can be plotted separately for the treated $A=1$ and the untreated $A=0$

In randomized experiments

- The survival probability at time k in the treated is the survival probability if everybody had been treated
  - Pr $[T>k|A=1]$ = Pr $[T^{a=1}>k]$
- And same in the untreated
  - Pr $[T>k|A=0]$ = Pr $[T^{a=0}>k]$
- Comparing survivals between $A=1$ and $A=0$ is therefore a natural way to quantify causal effects in survival analyses
  - Pr $[T>k|A=1]$ − Pr $[T>k|A=0]$ for all k

In the smoking cessation data

- Survival at end of follow-up k_{end} was
  - 0.76 in A=1
  - 0.82 in A=0
- Survival difference
  - −0.06
- Sometimes curves compared via the log-rank test
  - p-value = 0.005 suggests that difference not due to chance
    - But not randomized, so differences possibly due to confounding



See survival.R, lines 15-27

An alternative measure: Cumulative incidence (or Risk)

- $\Pr[T \leq k]$
  - where $k = 0, 1, 2 \ldots k_{end}$ is time of follow-up
  - Probability of having the event by time k
  - 1 − Survival
- Cumulative incidence curve
  - Plot of risks for $k = 0, 1, 2 \ldots k_{end}$
  - Starts at 0 at time k=0 and increases monotonically
- In randomized experiments, comparing risks between A=1 and A=0 is a natural way to quantify causal effects in survival analyses
  - Because $\Pr[T \leq k|A=a] = \Pr[T^a \leq k]$

Another widely used measure: Hazard

- $\Pr[T=k \mid T>k-1] = \lambda_T(k)$
  - where $k = 0, 1, 2 \ldots k_{end}$ is time of follow-up
  - Probability of having the event between $k-1$ and k
- Technically, this is a discrete-time hazard
  - The hazard in a study in which time is measured in discrete intervals rather than continuously
  - In real-world studies, time is indeed measured in discrete intervals (years, months, days…)
- In randomized experiments
  - $\lambda_T(k|A=a) = \lambda_{T^a}(k)$

Hazards vs. risks Over time…

- … the risk stays flat or increases
  - monotonically increasing
  - because its denominator (number of individuals at baseline) is constant across times k and its numerator (all events between baseline and k) is cumulative
- … the hazard may increase or decrease
  - because its denominator (number of individuals alive at $k-1$) varies across times k and its numerator includes only recent events during interval $k-1$ to k

In our smoking cessation example

- The hazard at the end of follow-up k_{end} was
  - 0% among quitters
    - because the last death happened at 113 months in this group
  - 1/986 = 0.10% among non-quitters
- The hazard curves between 0 and 120 months had roughly the shape of a letter *M*
  - In contrast, the cumulative incidence curves are monotonically increasing

Hazard ratio

$\lambda_T(k|A=1) \ / \lambda_T(k|A=0)$

- *The* hazard ratio is likely the most commonly used effect measure in randomized trials
  - $\lambda_T(k|A=1) \ / \lambda_T(k|A=0) = \lambda_{T^{a=1}}(k) \ / \lambda_{T^{a=0}}(k)$
- A consequence of the popularity of the Cox proportional hazards model
  - $\lambda_T(k|A) = \lambda_T(k|A=0)\exp(\beta_1 A)$
  - where β_1 is the log hazard ratio
- A Cox model is a semiparametric approach
  - No restrictions about the baseline hazard $\lambda_T(k|A=0)$
  - Hazard ratio $\exp(\beta_1)$ is assumed to be constant

EXAMPLE: Women's Health Initiative
Randomized trial of ~16,000 postmenopausal women

- Estrogen plus progestin hormone therapy vs. placebo
- Outcome: coronary heart disease
- Intention-to-treat analysis

- Hazard ratio by period of follow-up
  - All years 1.23
  - 0-2 years 1.51
  - >2-5 years 1.31
  - >5 years 0.67

The hazard ratio of ~0.7 after 5 years indicates that treatment becomes beneficial after 5 years

True 0%

False 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

The hazard ratio is a flawed measure of causal effect

- Built-in selection bias
  - See WHI example
- Not such a thing as **the** hazard ratio
  - Because the hazards vary over time, their ratio also does
- Often survival analyses report a single hazard ratio
  - a weighted average of the time-varying hazard ratios
  - not easily comparable across studies: the magnitude of the hazard ratio depends on the length of follow-up
  - testing for proportional hazards is often underpowered

Hazards are helpful to estimate survival if there is non-administrative censoring

- Using the product-limit or Kaplan-Meier method
  - used for the nonparametric survival curves shown before
- To describe the method, we define
  - D_k as an indicator of death by time k (1 if $T \leq k$, 0 if $T > k$)
  - C_k as an indicator of non-administrative censoring by k
- Hazard at k among those who are uncensored by k
  - $\Pr[D_k=1 | D_{k-1}=0, C_k=0]$
- One minus the hazard at k is the probability of surviving the interval between $k-1$ and k among those at risk
  - uncensored and alive at the start of the interval

Kaplan-Meier method (review)

K	N_k	D_k	C_k	λ_k	$1-\lambda_k$	S_k
0	100	0	0	0/100=0	1	1
1	100	10	0	10/100=0.1	0.9	1*0.9=0.9
2	90	45	0	45/90=0.5	0.5	1*0.9*0.5=0.45

Kaplan-Meier method (review)

- Survival at k is the product of (one minus the hazard) between 0 and k
  - For k=2, $\Pr[D_2=0] = \Pr[D_1=0]\Pr[D_2=0|D_1=0, C_2=0]$
  - For any k, $\Pr[D_k=0] = \prod^{k}_{m=1} (1 - \Pr[D_m=1|D_{m-1}=0, C_m=0])$
- This survival at k is the counterfactual survival that would have been observed in the absence of censoring
  - Under the assumption that there is no selection bias due to censoring, e.g., censoring at random

Typically, hazards cannot be estimated nonparametrically

- We cannot estimate the hazard at each interval as
  - # deaths in the interval / # individuals at risk
- Because we need to adjust for confounding and selection bias
- Parametric estimation of hazards is often a useful intermediate step for estimating survivals and risks
- Let's see how
  - for simplicity, first in a setting without confounding and censoring
  - See Chapter 17 for more discussion

Parametric estimation of survival Steps

0. Arrange data

1. Parametric estimation of hazards

2. Computation of survivals

3. Estimation of variance

Step #0: Arrange data with a person-time structure

- One row per person-time
  - In our smoking cessation example, each row is a person-month
- From time k=0 to the minimum of T and k_{end}

ID	Month	A	D_{k+1}
1	0	1	0
1	1	1	0
…	…	…	…
1	119	1	0
2	0	0	0
2	1	0	0
…	…	…	…
2	16	0	0
2	17	0	1

Step #1: Parametric estimation of hazards

- Fit a pooled logistic regression model to the data

$$\text{logit}(\Pr[D_{k+1}=1|D_k=0,A]) = \theta_{0,k} + \theta_1 A$$

  - the time-varying intercept $\theta_{0,k}$ is a function of time, e.g.,

$$\theta_{0,k} = \theta_0 + \theta_4 k + \theta_5 k^2$$

  - which allows the hazard to vary over time
- $\exp(\theta_1)$ is approximately the hazard ratio of death for A=1 vs. A=0
  - when the hazard is small at all times k

Step #1: Parametric estimation of hazards

- Allow for time-varying hazard ratio by adding product terms between A and time k

$$\text{logit}(\Pr[D_{k+1}=1|D_k=0,A]) = \theta_{0,k} + \theta_1 A + \theta_2 Ak + \theta_3 Ak^2$$

- That is, we do not impose a single hazard ratio
  - not a proportional hazards model

Step #2: Computation of survivals (or risks)

- Remember: one minus the predicted value from the hazards model

$$\text{logit}(\Pr[D_{k+1}=1|D_k=0,A]) = \theta_{0,k} + \theta_1 A + \theta_2 Ak + \theta_3 Ak^2$$

  estimates the conditional survival in each interval

- Predict the conditional survival for each k under each treatment level A
  - Plot the survival curves for $A=1$ and $A=0$

Step #3: Estimation of variance

- Simplest way: bootstrapping
  - Remember: The lazy statistician's method
  - We can use bootstrapping to calculate an approximate 95% confidence interval of the difference of standardized survivals
- Smoking cessation example
  - See R/SAS programs for implementation of these steps

But our smoking cessation study is not a randomized experiment

- Need to construct survival curves that are adjusted for confounding
  - So that, under conditional exchangeability, the survival curve is the counterfactual survival curve in $A=a$ that would have been observed if everybody had been assigned to treatment level a
- Let's see how that can be done using
  - IP weighting
  - Standardization / G-formula

IP weighting for survival analysis
Steps

0. Arrange data

1. Estimation of IP weights

2. Parametric estimation of hazards

3. Computation of survivals

4. Estimation of variance

Step #1:
Estimation of IP weights

Fit a logistic regression model

$$\text{logit Pr}[A{=}1|\, L] = \theta_0 + \theta_1 L$$

- θ_1 is a vector parameter, we assume exchangeability conditional on the same 9 variables L as in previous sections
- and a separate logistic model without L

Compute the estimated stabilized IP weights SW^A

- Predicted values from the first model are used for the denominator
- Predicted values from the second model are used for the numerator

Step #2: Parametric estimation of hazards

- Fit a weighted pooled logistic regression model

$$\text{logit Pr}[D_{k+1}=1|D_k=0, A] = \theta_{0,k} + \theta_1 A + \theta_2 Ak + \theta_3 Ak^2$$

  - $\theta_{0,k} = \theta_0 + \theta_4 k + \theta_5 k^2$ is a time-varying intercept to allow the hazard to vary over time
  - the product terms $\theta_2 Ak + \theta_3 Ak^2$ allow the hazard ratio to vary over time
- Each person-time receives the IP weight SW^A

Step #3: Computation of survivals

- Survival at k is the product of conditional probabilities of having survived each interval between 0 and k (one minus the hazards)
- For example, for k=2

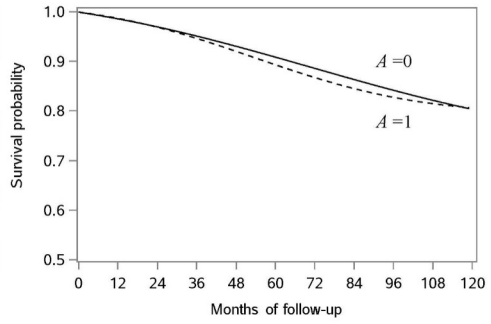
$$\text{Pr}[D_2=0] = \text{Pr}[D_1=0]\text{Pr}[D_2=0|D_1=0]$$

- Risk is one minus survival probability

$$\text{Pr}[D_2=1] = 1 - \text{Pr}[D_2=0] = 1 - \text{Pr}[D_1=0]\text{Pr}[D_2=0|D_1=0]$$

IP weighted survival curves in smoking cessation example

- 10-year survival estimates
  - 80.8% under smoking cessation (A=1)
  - 80.5% under no smoking cessation (A=0)
- Difference in 10-year survival
  - 0.3%



See survival.R, lines 82-145

Step #4: Estimation of variance

- Bootstrapping
- Difference in 10-year survival (95% CI)
  - 0.3% (-4.2 to 4.1)

If there is selection bias due to non-administrative censoring

- Then estimate IP weights for censoring SW^C
- and conduct the analysis with weights $SW^A \times SW^C$

Standardization/G-formula for survival analysis Steps

0. Arrange data

1. Parametric estimation of conditional hazards

2. Computation of weighted average of survivals

3. Estimation of variance

Step #1: Parametric estimation of conditional hazards

Fit a pooled logistic regression model

$$\text{logit}\ \Pr[D_{k+1}=1|D_k=0, A, L] = \theta_{0,k} + \theta_1 A + \theta_2 Ak + \theta_3 Ak^2 + \theta_4 L$$

- $\theta_{0,k} = \theta_0 + \theta_5 k + \theta_6 k^2$ is a time-varying intercept to allow the hazard to vary over time
- the product terms $\theta_2 Ak + \theta_3 Ak^2$ allow the hazard ratio to vary over time
- θ_4 is a vector parameter

We assume exchangeability conditional on the same 9 variables L as in previous sections

Aside

- Parametric estimation of the hazards can be conditional on propensity score

$$\text{logit}(\Pr[D_{k+1}=1|D_k=0,A,PS]) = \theta_{0,k} + \theta_1 A + \theta_2 Ak + \theta_3 Ak^2 + \theta_4 PS + \theta_5 PS^2 + \ldots$$

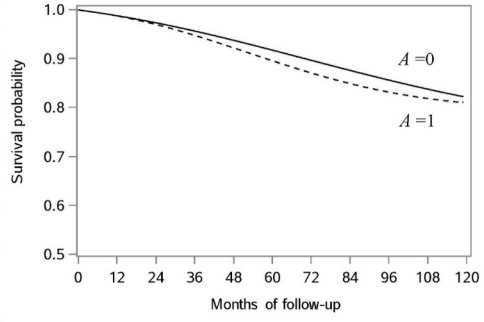
Step #2: Computation of weighted average of survivals

Sum over all combinations of values of L

- Again this integral can be approximated by using the empirical distribution of the confounders
- That is, compute the average of estimated survivals for each individual in the population under treatment (A=1) and under no treatment (A=0)

Standardized survival curves in smoking cessation example

- 10-year survival estimates
  - 80.4% under smoking cessation (A=1)
  - 80.6% under no smoking cessation (A=0)
- Difference in 10-year survival
  - 0.2%



Step #4: Estimation of variance

- Bootstrapping
- Difference in 10-year survival
  - 0.2% (-4.6 to 4.1)

See survival.R, lines 148-196

Readings

- *Causal Inference, What If*. Chapter 17

Progress report

1. Introduction to modeling
2. Stratified analysis:
   - Outcome regression
   - Propensity scores
3. Standardization
4. Inverse probability weighting
   - Marginal structural models
5. Instrumental variable estimation
6. G-estimation
7. Causal Survival Analysis
8. The bias of traditional methods