Survival Analysis: An Under Used But Very Useful Methodology

Margaret K. Keiley, Ed.D.
Auburn University
keilemk@auburn.edu

SRCD
4/7/2017
Survival Analysis: Types of Research Questions

“Whether” and, if so, “when” critical events in a life-course occur:

- Sexual initiation
- Marriage
- Birth of first child
- First arrest
- Dropping out of high school

And what predicts those event occurrences
We used the Widom data from the National Data Archive on Child Abuse and Neglect at Cornell [http://www.ndacan.cornell.edu/]

We utilized the *Urban Poverty and Family Life Survey of Chicago* (UPFLS; Wilson, 1987) that was conducted in 1987 via personal and telephone interviews by the National Opinion Research Center. A stratified probability sample of census tracts was used that consisted of individuals from impoverished areas in the city of Chicago.
FIGURE 22.2. Fitted survival function for the age of first sexual intercourse from Model 5 that includes the effects of gender (male, female), race (AA, EA, Hispanic), and the interaction of gender with time. The best age to have children is included in this model, but we present only the prototypical profiles for those who feel the best age to have children is a younger age, 19. For those who feel the best age would be older, for example, age 26, all of the prototypical profiles would merely be shifted upward.
Survival Analysis

Allows us to answer questions about the occurrence and timing of events

- Time to relapse after smoking cessation
- Time to separation or divorce
- Time to attainment of developmental milestones

Outcome can be continuous or dichotomous

We are focusing on dichotomous outcomes in this workshop

Most survival analyses focus on a dichotomous outcome (event occurrence vs. not)
Other Features of Survival Analysis Studies

Data can be collected:
- Prospectively or retrospectively
- Over a short or long period of time
- In an experimental or observational study

The target event can:
- Occur once
- Occur repeatedly over time
- Be within or beyond the individuals control
- Positive or negative
The “Whether” and “When” Test

Researchers should use survival analysis if they are interested in:

*Whether* an event occurs
*When* that event occurs

In other words, if your research question includes either “whether” or “when” you will likely need to use survival analysis
Examples of Questions Addressed by Survival Analysis

Relapse to alcohol consumption
- When are individuals most at risk to relapse?
- Who is at greatest risk of relapse?

Length of stay in teaching
- When are teachers most likely to leave teaching for the first time?
- Who is most likely to leave?

Age at first suicide ideation
- When do undergraduate students first think about suicide?
- Who is most at risk to do so?
Three Required Methodological Features for Survival Analysis

You must have a target event

Dropout, relapse, separation/divorce, birth of first child, entry into non-maternal care, relocation...

You must identify the “beginning of time”

Initial starting point when no one under study has experienced the target event

You must identify a metric for clocking time

A meaningful scale in which event occurrence is recorded (e.g., days, weeks, months, semesters, years, grades...)

Defining Event Occurrence

Represents the transition from one “state” to another “state”

From teaching (state 1) to not teaching (state 2)
From not drinking (state 1) to relapse to drinking (state 2)
From being partnered (state 1) to being separated or divorced (state 2)

Each “state” must be mutually exclusive and exhaustive
Defining Event Occurrence

Some “states” can be occupied only once
- First word
- High school graduation/college graduation
- First marriage
- First sexual intercourse

Other “states” can be occupied multiple times
- Relapse to drinking
- Clinical depressive symptoms
- Pregnancy
- Marriage
Identifying the “Beginning of Time”

The moment in time when everyone in the population occupies **one and only one** of the possible “states”

On the day teachers are hired, they are teaching
On the day students enter high school, they are enrolled
On the day individuals are married, they are a married couple
On the first day of drinking after being sober, individuals have relapsed
Identifying the “Beginning of Time”

The point in time when everyone in the population is *at risk* of experiencing the event:

- Day one for a newly hired teacher (to be fired)
- First day out of rehab for alcoholism (to experiencing relapse)
- First Wedding day (to experiencing first divorce)
Measuring Time: Continuous or Discrete?

**Continuous:** If the exact time of an event is known (e.g., age of teen mother in days from her own birth to the birth of her first child)

**Discrete:** Events occur (or can be remembered to have occurred, in the case of retrospective data) across discrete periods, or separable and often large units of time (e.g., age of first sexual intercourse)

I focus on discrete-time event history analysis for this workshop, as this framework allows for a more straightforward and simple explanation of the method of event history analysis itself.

For more information on continuous-time event history analysis see Singer and Willett (2003) or Allison (2010)
Why Common Statistical Methods Fail to Answer these Questions Well

Whether and when an event occurs is unknown for some people under investigation

These cases are CENSORED
The outcome is unknown
Problem: Some Individuals Will Not Experience the Target Event

You could set these individuals aside
   But they provide important information about the individuals least likely to experience the event

You could include them in your analysis
   But they need some sort of value for the outcome, right?
Censoring

Censored cases:

They are people in the sample who do not experience the target event while they are being observed or, they may do so after data collection ends.

They are the ones who are least likely to experience the event, the participants who “survive” the longest.

So they are very important!!!
Why Common Statistical Methods Fail to Answer these Questions Well

Whether and when an event occurs is unknown for some people under investigation

These cases are CENSORED

The outcome is unknown
Censored Cases

Age of First Alcoholic Drink

Beginning of Data Collection (Age 10)

End of Data Collection (Age 15)
Common Strategies for Dealing with Censoring and Their Problems

Ignore the censored cases completely, treating them as if they were missing

- Reduces statistical power
- Negatively biases estimates of time to event

Imputing unknown event times

- Event time equal to length of data collection
- Underestimates the true length of time-to-event

Dichotomization

- Collapses perfectly good continuous duration data to create a single new dichotomous outcome that may be meaningless!

Individuals who first experienced depression before or after age 25
Those with initial depression in early childhood would be pooled with those who did not become depressed until their 20s then compared with those who did not become depressed until after 25!
Such individuals may differ enormously in the causes of depression
Survival Analysis, Event History Analysis, Hazard Modeling:

A data-analytic method that deals evenhandedly with both the non-censored and the censored observations
Questions

What is the “whether” and “when” question that you will try to answer with your data?

What is the “beginning of time?”

What is the metric in which the passage of time is recorded?

How is the target event defined?

Who drops out of the study and how do you denote these respondents?

UCLA Website for all the programs and data needed for examples in their book:

http://www.ats.ucla.edu/stat/examples/alda.htm

UCLA Website for their Seminar Notes in Power Point:

http://www.ats.ucla.edu/stat/seminars/alda/default.htm

keilemk@auburn.edu:
For help or more information about this methodology.
I am glad to help with your own analyses
Secondary Data Archives

National Center for Educational Statistics
http://nces.ed.gov/

Center for Disease Control and Prevention
http://www.cdc.gov/nchs/data_access/ftp_data.htm

Murray Center at Harvard
http://dvn.iq.harvard.edu/dvn/dv/mra

Social Science Data Archive -- Amazing amount of public data available here
http://www.sscnet.ucla.edu/issr/da/

Odum Institute for Research in Social Science -- more data !!! Click the link and use ‘advanced search’
http://arc.irss.unc.edu/dvn/

Inter-University Consortium for Political and Social Research
http://www.icpsr.umich.edu/icpsrweb/ICPSR/

National Data Archive on Child Abuse and Neglect
http://www.ndacan.cornell.edu/
Some of My Publications About Survival Analysis


