

Survival Analysis — A Framework for Event Occurrence

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

GCM, 2010

Survival Analysis — A Framework for Event Occurrence

- 1 Introduction
- 2 Should You Conduct a Survival Analysis?
 - Time to Relapse in Alcoholism
 - Length of Stay in Teaching
 - Age at First Suicide Ideation
 - Time to End of Remission in Leukemia
 - Time to Infection in Kidney Dialysis Patients
- 3 Framing a Research Question about Event Occurrence
 - Defining Event Occurrence
 - Identifying the “Beginning of Time”
 - Specifying a Metric for Time
- 4 Censoring — How Complete are the Data?
 - How and When Does Censoring Arise?
 - Different Types of Censoring
 - How Does Censoring Affect Statistical Analysis?

Introduction

In this module, we examine the basic framework for Survival Analysis, which analyses basic information about if and when events occur over a span of time.

Should You Conduct a Survival Analysis?

A common characteristic of survival analysis studies is an interest in succinctly describing *whether* and/or *when* events occur.

Early applications concentrated on the occurrence of negative events, like death, recidivism, divorce.

However, the techniques are appropriate for positive and neutral events as well.

Let's look at some examples.

Time to Relapse in Alcoholism

Example (Time to Relapse in Alcoholism)

- Cooney, Kadden, Litt, and Getter (1991) evaluated the comparative efficacy of two regimens supporting alcoholics in their attempts to avoid relapse after treatment in a 21-day inpatient program.
- One regimen emphasized *coping skills*, the other emphasized *interaction skills*.
- Each person was followed for up to two years.
- Research interest centered on whether, and when, the newly released patients relapsed to alcohol use.

Time to Relapse in Alcoholism

Example (Time to Relapse in Alcoholism)

- The outcome — time to first day of “heavy drinking,” defined as consuming 3 or more ounces of alcohol in a 24-hour period, was measured as the number of days between the day of release from the program.
- The object of the study was to predict time to relapse using two variables:
 - ① Level of psychopathology
 - ② Type of aftercare regimen

Length of Stay in Teaching

Example (Length of Stay in Teaching)

- In many fields where teachers are in high demand, there are high turnover rates.
- Singer (1993) examined career histories of 3941 special education teachers newly hired in Michigan between 1072 and 1978.
- The goals: To determine
 - ① Determine how long special education teachers stay in their jobs
 - ② Identify factors associated with their stay or leave decisions

Age at First Suicide Ideation

Example (Age at First Suicide Ideation)

- Suicide is a major cause of death among adolescents
- The rates are increasing
- Bolger, Downey, Walker, and Steininger (1989) focused not on suicide but on suicidal ideation.
- 391 undergrads aged 16–22 were asked
 - ① “Have you ever thought of committing suicide?”
 - ② “At what age did the thought first occur to you?”

Age at First Suicide Ideation

Example (Age at First Suicide Ideation)

- Trends were summarized.
- Demographic factors such as gender, age, and race were measured and their impact on occurrence and time of suicidal ideation assessed.

Time to End of Remission in Leukemia

Example (Time to End of Remission in Leukemia)

- Freireich et al. (1963) report the results of a clinical trial of a drug 6-mercaptopurine (6-MP) versus a placebo in 42 children with acute leukemia.
- The trial was conducted at 11 American hospitals. Patients were selected who had a complete or partial remission of their leukemia induced by treatment with the drug prednisone. (A complete or partial remission means that either most or all signs of disease had disappeared from the bone marrow.)
- The trial was conducted by matching pairs of patients at a given hospital by remission status (complete or partial) and randomizing within the pair to either a 6-MP or placebo maintenance therapy.
- Patients were followed until their leukemia returned (relapse) or until the end of the study (in months).

Time to Infection in Kidney Dialysis Patients

Example (Time to Infection in Kidney Dialysis Patients)

- 1 Nahman et al. (1992) designed a study to assess the time to first exit site infection (in months) in patients with renal insufficiency.
- 2 43 patients utilized a surgically placed catheter (Group 1), and 76 patients utilized a percutaneous placement of their catheter (Group 2).
- 3 Cutaneous exit site infection was defined as a painful cutaneous exit site and positive cultures, or peritonitis, defined as a presence of clinical symptoms, elevated peritoneal dialytic fluid, elevated white blood cell count (100 white blood cells / μ l with >50% neutrophils), and positive peritoneal dialytic fluid cultures.

Framing the Methodology of a Survival Analysis

Studies that are amenable to a survival analysis share several common methodological features:

- 1 A well-defined *target event* under study
- 2 A well-defined *starting point* for the study at which time *no one in the study has experienced the target event*
- 3 A meaningful metric for recording time

Defining Event Occurrence

Event occurrence represents an individual's transition from one well-defined state to another. These states must be defined precisely. In some cases, definitions are clear and uncontroversial, while in others (e.g., recidivism of sex offenders), there are numerous possibilities that need to be sorted out.

Example (Event Occurrence Examples)

- 1 A recently treated ex-alcoholic is (a) abstinent (b) starts drinking
- 2 A leukemia patient is (a) in remission or (b) no longer in remission
- 3 A heart transplant patient is (a) alive or (b) dead

Identifying the “Beginning of Time”

The “Beginning of time” is a moment when *everyone* in the population occupies one, and only one, of the possible states. This moment can be arbitrary, at least if it has no relation to event occurrence.

Example (Beginning of Time)

- 1 The day after surgery, a surviving heart transplant patient is alive.
- 2 At the time of release from prison, no sex offender has yet re-offended.

Specifying a Metric for Time

Time should be measured in well-defined units: Days, Months, Years, Minutes. Singer and Willett speak of *continuous-* versus *discrete-time* models for survival data.

Continuous vs. Discrete-Time Models

In continuous survival analysis models, time is treated as inherently continuous. In the practical sense, of course, all time measurement is discrete. The issue is one of degree. In continuous survival models, we try to arrange it so that the number of time points at which occurrence of the event of interest is assessed is large enough that ties are virtually impossible.

In discrete survival models,

- 1 *Case 1.* Events are “collected” at various well defined intervals and counted, or
- 2 *Case 2.* The measurement times are small in number relative to the number of subjects, and so ties are inevitable.

The Problem of Censored Data

We speak of data being *censored* in survival analysis when the event of interest is not observed during the period that data are gathered.

This can occur for a number of reasons.

How and When Does Censoring Arise?

The two major reasons for censoring are:

- 1 Some individuals will never experience the event at all
- 2 Some individuals will experience the event, but not during data collection

Censoring can occur at one or multiple time points, depending on how a study is conducted.

Noninformative vs. Informative

Noninformative censoring occurs for reasons independent of the event occurrence or, indeed, the risk of event occurrence. It is similar to MAR missing data.

Informative censoring occurs when non-observation of the individual is due to occurrence or imminent occurrence of the event of interest.

Example (Informative Censoring)

If someone is about to relapse into alcoholism and stops answering the experimenter's calls during the period of data gathering as a result, the censoring is *informative*.

Right- vs. Left-Censoring

- 1 *Right-Censoring* occurs because the event of interest is not observed during the study. This is by far the most common form of censoring.
- 2 *Left-Censoring* occurs because the beginning of time is not known for the individual.

Left-Censoring

Example (Left-Censoring)

A study of coal miners examined “attendance spells,” over a calendar year. Spells began when a miner returned to work after an absence, and ended at the next absent day. Each miner had a left-censored spell represented by the last period of attendance in the previous year that was terminated by an absence in the current year.

How Does Censoring Affect Statistical Analysis?

In the teacher study discussed previously, many of the teachers were still teaching when the data gathering period ended. Those who entered the study latest (1978) had been teaching for 7 years when the study terminated prior to 1985, while those who entered the study earliest had been teaching for 12 years. If you restrict the study to those for whom the event was observed, you obtain a picture shown in the foreground of Figure 9.1 in Singer and Willett (page 321). The bars in the background show the minimum teaching time for those who were right-censored. Clearly, these cases completely alter the statistical picture.

How Does Censoring Affect Statistical Analysis?

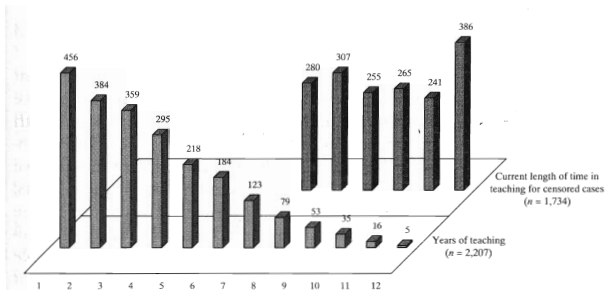


Figure 9.1. Distribution of the number of years in teaching by censoring status, for the 3941 special educators.