Biostatistics – Spring 2026
Lecture 07: Introduction to Survival Analysis

Dr. Zaid T. Al-Khaledi
Department of Statistics and Informatics
University of Mosul

March 25, 2026

Introduction

In previous lectures, we studied rates, proportions, and population-based measures. Now we move to a different type of biomedical outcome:

Not only whether an event happens, but **when** it happens.

This type of analysis is called **survival analysis**. Although the name comes from medical survival studies, the same methods are used for many outcomes such as:

- time until death,
- time until relapse,
- time until recovery,
- time until infection,
- time until machine failure.

By the end of this lecture, you should be able to:

- explain what time-to-event data are,
- define survival time and event,
- understand censoring,
- define the survival function,
- define the hazard function conceptually,
- distinguish survival analysis from ordinary methods.

1. What Is Survival Analysis?

Survival analysis is a branch of statistics used to analyze **time-to-event data**.

The variable of interest is not only the event itself, but the time from a starting point until the event occurs.

General form:

$$T = \text{time from origin until the event occurs.}$$

Examples of time origin:

- date of diagnosis,
- date of surgery,
- start of treatment,
- start of follow-up.

Examples of events:

- death,
- recurrence of disease,
- discharge from hospital,
- device failure.

2. Why Can We Not Use Ordinary Statistical Methods?

Survival data have a special feature:

For some individuals, the event is not observed during the study period.

This creates incomplete observations. Because of that, ordinary methods such as the mean of observed times may be misleading.

For example, if a patient is still alive when the study ends, we know only that the survival time is greater than the follow-up time. We do not know the exact event time.

This is why survival analysis methods are needed.

3. Survival Time and Event Indicator

For each subject, we usually record:

$$t_i = \text{observed follow-up time}$$

and

$$\delta_i = \begin{cases} 1, & \text{if the event occurred,} \\ 0, & \text{if the observation is censored.} \end{cases}$$

Thus, every subject contributes two pieces of information:

- the observed time,
- whether that time ended with an event or with censoring.

4. Censoring

Censoring means that the exact event time is not fully observed.

4.1 Right censoring

This is the most common type in medical research.

A subject is right censored if:

- the study ends before the event occurs,
- the subject is lost to follow-up,
- the subject withdraws from the study,
- the subject dies from another unrelated cause (depending on study definition).

Interpretation: We only know that the true event time is greater than the observed time.

4.2 Example

Suppose a patient enters the study and is followed for 12 months. If no event occurs by month 12 and the study ends, then the patient is censored at 12 months.

This does **not** mean the event never occurs. It means only that it was not observed during follow-up.

5. Basic Functions in Survival Analysis

5.1 Survival function

The survival function is:

$$S(t) = P(T > t).$$

It represents the probability that a subject survives beyond time t.

Important properties:

- $0 \leq S(t) \leq 1$
- $S(0) = 1$ in most studies
- $S(t)$ decreases as time increases

Interpretation

If

$$S(12) = 0.80,$$

this means that the probability of surviving beyond 12 months is 0.80, or 80%.

5.2 Distribution function

The cumulative probability of experiencing the event by time t is:

$$F(t) = P(T \leq t) = 1 - S(t).$$

So:

- $S(t)$ describes surviving beyond time t,
- $F(t)$ describes having the event by time t.

5.3 Hazard function

The hazard function describes the **instantaneous risk** of the event at time t, among those who have survived up to time t.

Conceptually, it answers:

> Among individuals still under observation and event-free at time t, how risky is the next instant?

We usually denote it by

$$h(t).$$

At this stage, the main idea is conceptual:

- survival function describes remaining alive,
- hazard function describes risk of failure at a given time.

6. Example of Time-to-Event Data

Suppose five patients are followed after surgery.

Patient	Time (months)	Status
1	4	1
2	7	0
3	10	1
4	10	0
5	14	1

Here:

- Patient 1 had the event at month 4.
- Patient 2 was censored at month 7.
- Patient 3 had the event at month 10.
- Patient 4 was censored at month 10.
- Patient 5 had the event at month 14.

This table shows why survival analysis needs both time and status.

7. Main Biomedical Questions in Survival Analysis

Survival analysis is used to answer questions such as:

- What proportion of patients survive beyond 1 year?
- Does one treatment have better survival than another?
- At what time is the risk of death highest?
- How does censoring affect estimation?

8. Common Student Mistakes

- Thinking survival analysis applies only to death.
- Ignoring censored observations.
- Confusing censoring with the event itself.
- Believing that censored subjects provide no information.
- Mixing survival probability with hazard.

9. Summary

- Survival analysis studies time until an event occurs.
- Survival data often contain censoring.
- Right censoring is very common in medicine.
- The survival function is $S(t) = P(T > t)$.
- The hazard function describes instantaneous risk.
- Specialized methods are needed because ordinary methods are not enough.

Homework (HW)

HW1

State whether each of the following is a survival-analysis outcome:

1. Time from diagnosis to death.
2. Blood pressure measured once at baseline.
3. Time from surgery to recurrence.
4. Sex of the patient.

HW2

For each situation below, say whether the observation is an event or censoring:

1. A patient dies 8 months after entering the study.
2. A patient is still alive when the study ends after 12 months.
3. A patient moves to another city and is no longer followed.
4. A patient relapses 5 months after treatment.

HW3

Suppose $S(24) = 0.65$. Interpret this in words.