ARISTOTLE UNIVERSITY
OF THESSALONIKI

FACULTY OF HEALTH SCIENCES - SCHOOL OF MEDICINE
MSc Health Statistics and Data Analytics

Survival Analysis

Eirini Pagkalidou, PhDc
pagalidou@auth.gr

THESSALONIKI 2021-22

Learning objectives

- Understand what survival data is and how to handle censoring
- Prepare survival curves using the life table and Kaplan-Meier methods
- Estimate median survival times and survival rates at specified times
- Compare survival curves between two or more groups using the log-rank test

Survival Analysis

Also called “time to event analysis”

Survival Time is defined as the **time** starting from an already defined point to the occurrence of the event of interest **(survival time)**.

Examples:

- time to death
- time to cancer metastasis
- time to failure of a light bulb
- time to complete a PhD

Properties

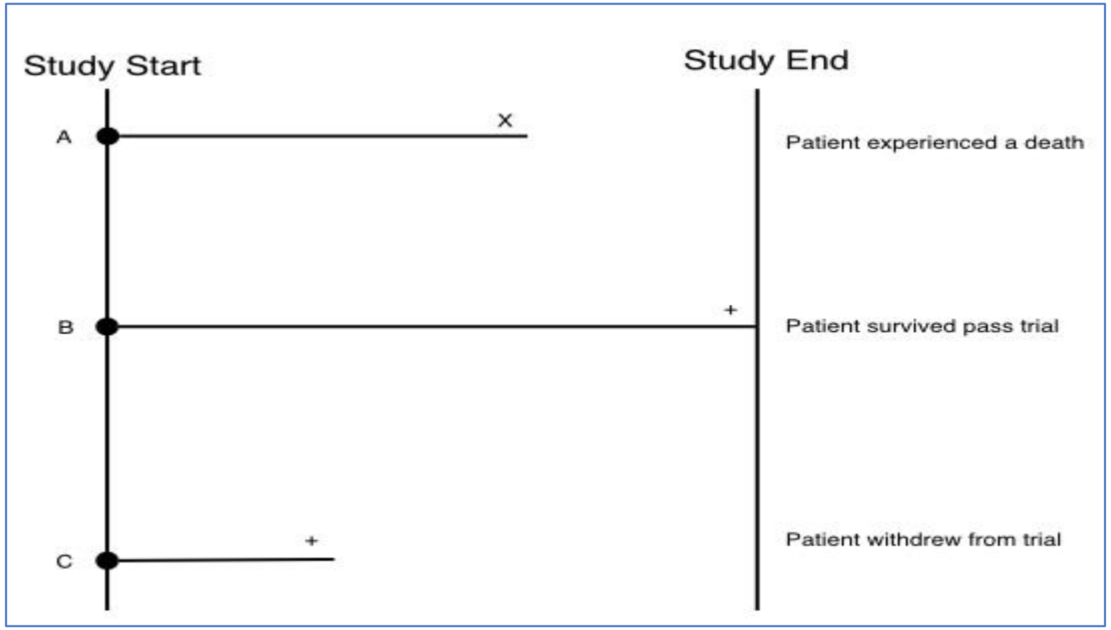
- Survival time is almost never normally distributed.

- In many studies, some of the times to an event are not recorded (for various reasons) – These observations are called **censored,** and the corresponding phenomenon is called **censoring**.

- Thus, such data cannot be handled by already known parametric or non-parametric methods.

Censoring

- **Censoring** is present when there is incomplete information about a subject’s event time, but we don’t know the exact event time.
- The most common case is having **right censoring**



✓ A person does not experience the event before the study ends

✓ A person is lost to follow-up during the study period or withdraws from the study

Survival function *S(t)*

Survival Function $S(t)$=pr(T > t)

- Is **the probability that a subject survives longer than time t**.

Example: If t=100 years, S(t=100) = probability of surviving beyond 100 years.

- The graphical representation of S(t) is known as survival curve **(Kaplan-Meier Curve).**

Kaplan-Meier Curve

Assumptions

- Observations should be independent
- Groups (if there are any) should be independent
- Accurate time calculation
- In the case of two or more groups, curves should not intersect with each other.

Example for the construction of a Kaplan-Meier Curve

Example:

Survival analysis for patients with different types of cancer

Group A: **Astrocytoma (n=20)**
6,13,21,30, 31+,37,38, 47+,49,50,63,79,80+,82+,82+,86,98,149+,202,219 weeks

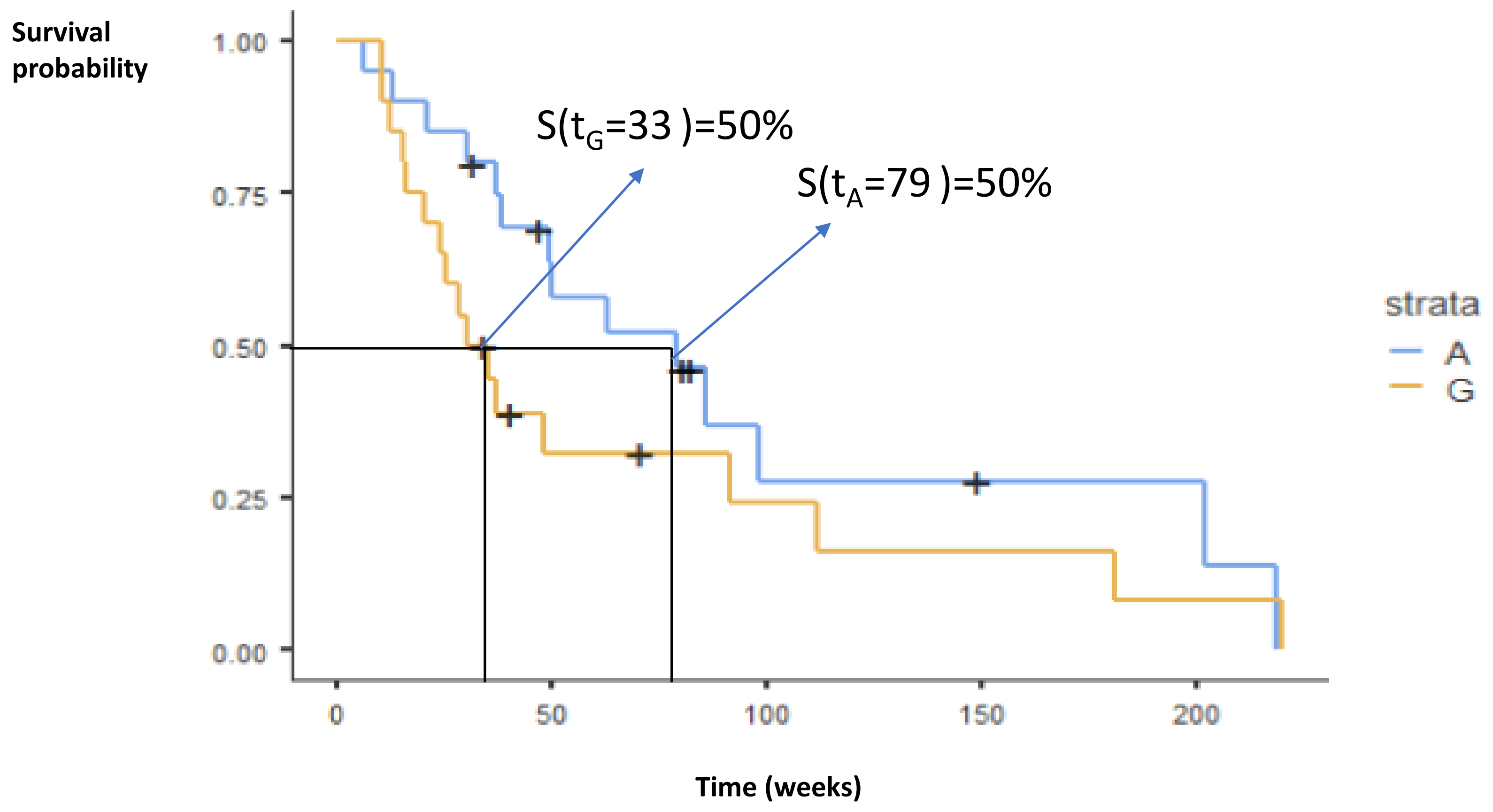
Group B: **Glioblastoma (n=20)**
10,10,12,15,16,20,24,25,28,30,34+,35,37,40+,48,70+,91,112,181,220 weeks

+ censored data

Life table-Group A

Time (weeks) i	Live at start n_i	Died d_i	Censored	At risk	Probability of surviving $(n_i - d_i)/n_i$	Cumulative Probability $S(t_i)=((n_i - d_i)/n_i)*S(t_{i-1})$
1	20	0	0	20	1=(20-0)/20	1
6	20	1	0	19	0.950=(20-1)/20	0.950
13	19	1	0	18	0.947=(19-1)/19	0.947*0.950=0.90
21	18	1	0	17	0.944=(18-1)/18	0.944*0.90=0.85
30	17	1	0	16	0.941=(17-1)/17	0.941*0.85=0.80
31	16	0	1	16	1=(16-0)/16	1*0.80=0.80
37	15	1	0	14	0.933=(15-1)/15	0.933*0.80=0.75
....						

Comparison of two survival curves



Log-Rank Test

- We use log-rank test to compare survival curves

- This is a non parametric test

  H_0: There is no difference between the curves

  H_1: There is a difference between the curves

Log-rank test gives p=0.203>0.05 thus there is no significant difference between the two survival curves.