July 21 - 24, 2021 | Garki Hospital Abuja | Resource Person

Research Methodology Boot Camp

with Epi Info Training

Dr. Adamu Onu
MBBS, FWACP (FM)
MS Epidemiology & Biostatistics
PhD Public Health (Epidemiology)

Target Audience Clinical Researchers, Post-Part 1 Residents, and Others

Important Information

- Limited slots are available on a first come, first served basis
- Laptop running Windows 10 required
- Organized as morning lecture sessions and afternoon hands on coaching sessions

Highlights

- Research Methodology
- Research Design
- Data Management
- Sample Size Calculations
- Test Statistics
- Interpretation of Results
- Report Writing
- Hands-on training sessions
- Statistical consulting sessions

For further details contact
Email: epimetrix@gmail.com
Phone: +234 803 474 9930

epimetrix
health systems
data for health decisions

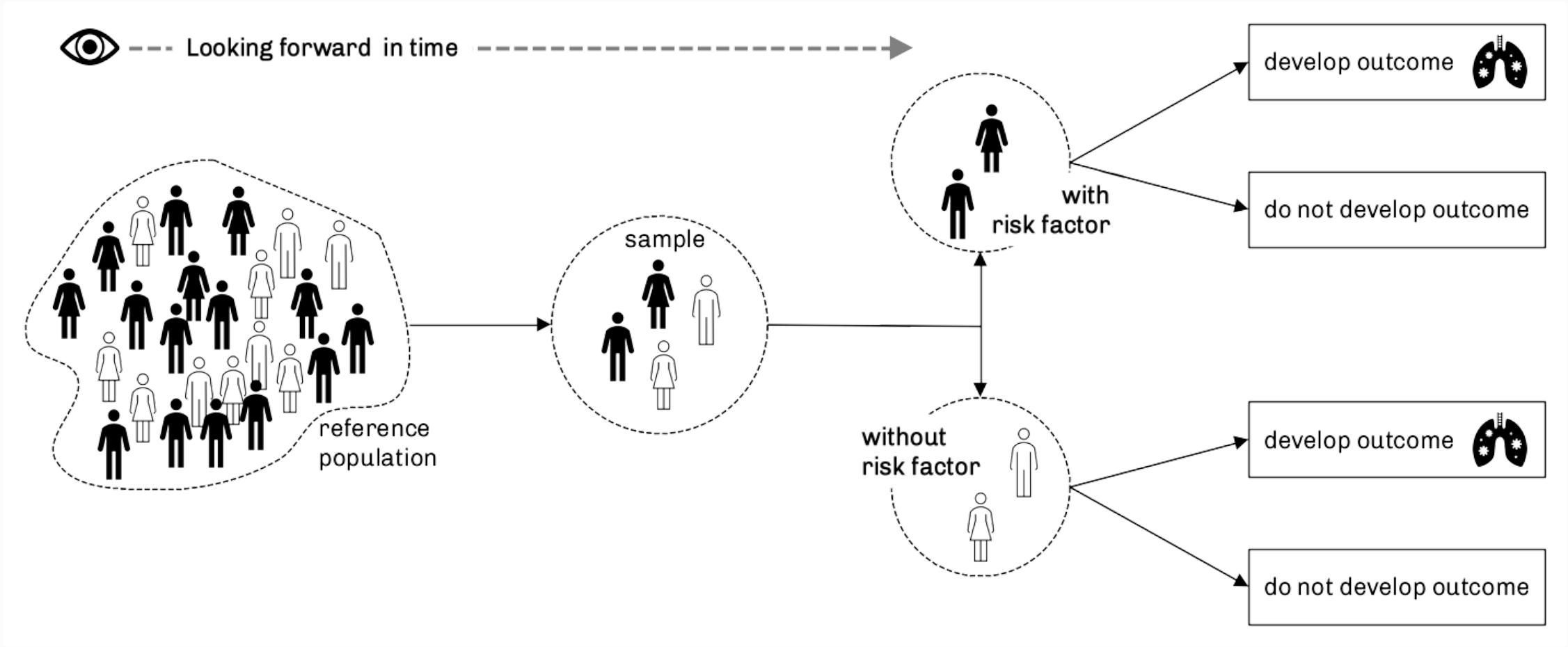
Survival Analysis

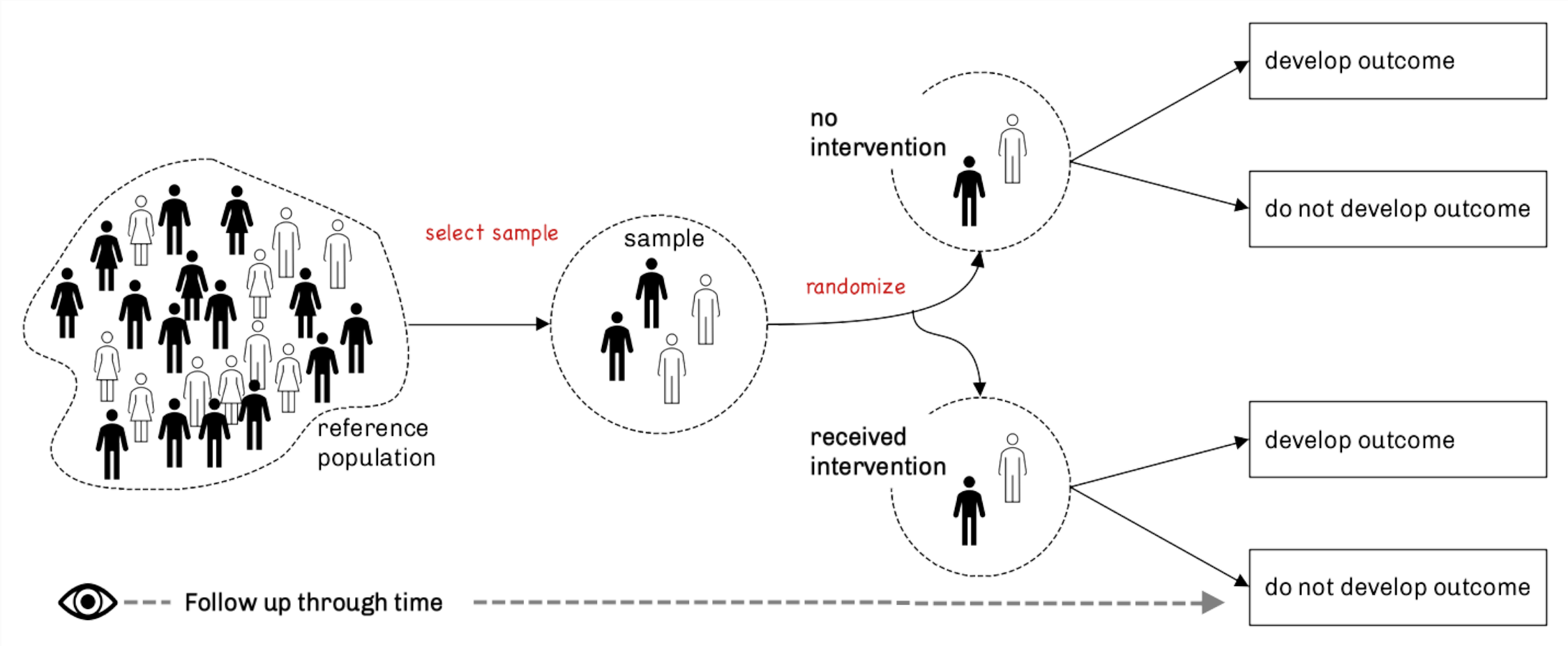
Overview

- What is survival analysis?
- Terminology and data structure
- Survival/hazard functions
- Parametric versus semi-parametric regression techniques
- Introduction to Kaplan-Meier (non-parametric) methods

What is survival analysis?

- Statistical methods for analyzing longitudinal data on the occurrence of events.
- Events may include death, injury, onset of illness, recovery from illness (binary variables) or transition above or below the clinical threshold for a meaningful continuous variable (e.g. CD4)
- Accomodates data from randomized clinical trial or cohort study design

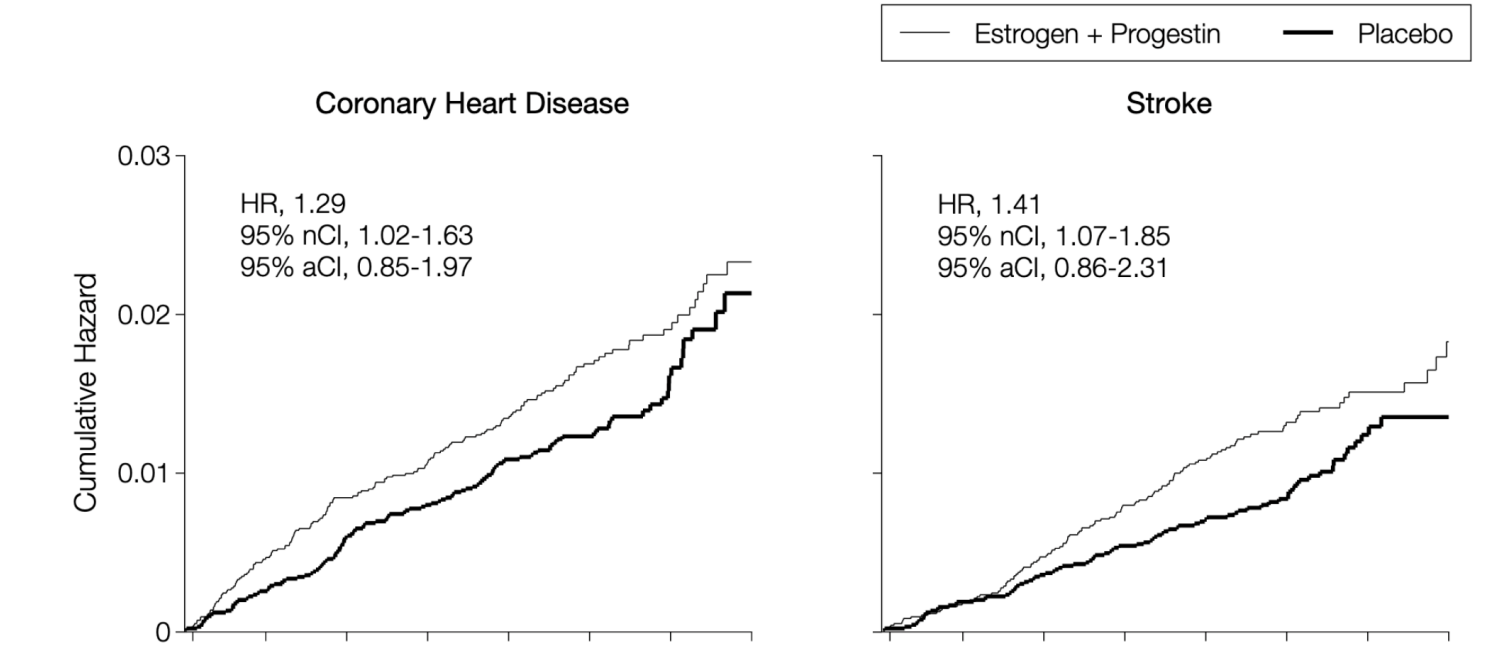

Looking forward in time
reference population
sample
with risk factor
without risk factor
develop outcome
do not develop outcome
develop outcome
do not develop outcome


reference population
select sample
sample
randomize
no intervention
received intervention
develop outcome
do not develop outcome
develop outcome
do not develop outcome
Follow up through time

Examples of survival analysis in medicine

JAMA. 2002 Jul 17;288(3):321-33. doi: 10.1001/jama.288.3.321.

Figure 3. Kaplan-Meier Estimates of Cumulative Hazards for Selected Clinical Outcomes



Survival analysis

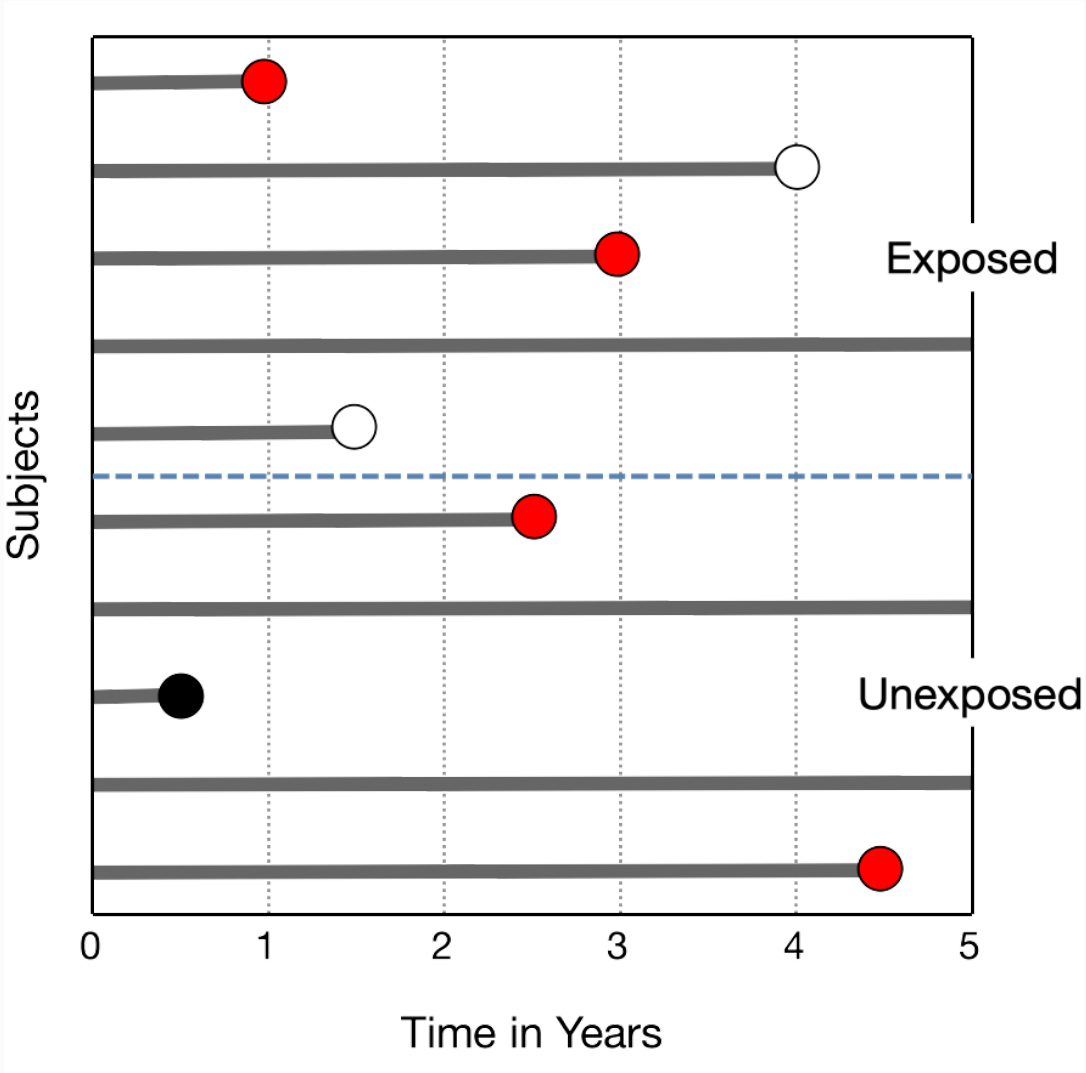
- Encompasses a wide variety of analytical methods for assessing **time to event**.
- Prototypical event is death “survival analysis.”
- Estimates the effects of one or more factors on survival of a cohort of disease-specific patients.
- Special methods are needed because time to event is rarely normally distributed.
- In different fields known as event history analysis, failure-time analysis, transition analysis, duration analysis, reliability analysis.

Objectives of survival analysis

- Estimate time-to-event for a group of individuals
  - Time until second heart attack for a group of MI patients
- To compare time-to-event between two or more groups:
  - treated vs. placebo MI patients in an RCT
- To assess the relationship of covariables to time-to-event
  - Does weight, insulin resistance, or cholesterol influence survival time of MI patients?

Why use survival analysis?

- Why not compare mean time-to-event between your groups using a *t* test or linear regression?
  - *Ignores censoring*
- Why not compare proportion of events using risk/odds ratios or logistic regression
  - *Ignores time*



Survival analysis

- Similar to open cohort study with person-time assessment, but with two major differences:
  - time to event is measured precisely for each subject as opposed events per time interval
  - the outcome event is usually much more frequent for example, assumes the outcome event will occur (e.g., death) *sooner or later* in a large proportion of the subjects.

Survival analysis

- Takes into consideration that survival times are rarely normally distributed and that survival should be estimated non-parametrically.
- Also considers time-dependent covariates.
- Involves concepts of **censoring**.
- Most often use non-parametric methods:
  - Kaplan-Meier estimates and plots
  - log-rank tests
  - Cox regression.

Survival analysis: terms

Time-to-event

- The time from entry into a study until a subject has a particular outcome

Censoring

- Subjects are said to be censored if they are lost to follow up or drop out of the study, or if the study ends before they die, or have an outcome of interest (they are counted as alive or disease-free for the time they are enrolled in the study).

Data structure: survival analysis

Two-variable outcome:

- Time variable (t_i)
  - time at last ‘disease’-free observation *or* time of event
- Censoring variable (c_i)
  - $c_i = 1$ if had the event
  - $c_i = 0$ if no event by time t_i

Censoring

- Censoring is when the study subject is not under observation.
- Because time is followed very closely, censoring is a critically important feature is survival analysis.

Right censoring

- Occurs when a subject is removed from the study (observation) before the outcome event occurs.
  - A patient has not yet experienced the outcome by the end of the study period.
  - A patient is lost to follow up before the end of the study period.
  - A patient experiences a different event (typically death) that makes further follow up impossible or removes the patient from risk of the main outcome event (competing risks).

Left censoring

- Initial time of risk is unknown.

Interval censoring

- Subjects come in and out of observation.

Survival distributions

- T_i the event time for an individual, is a *random* variable having a probability distribution
- Different models for survival data are distinguished by different choice of distribution for T_i

Survival and hazard

- Survival function or probabily – *S(t)* is the probability that an individual survives from the time of origin (e.g., diagnosis) to a specified future time *t*.
- This probability incorporates a certain period of time.
- Hazard function or rate *h(t)* is the instantaneous event rate for an individual who has already survived to time *t*.

$$h(t) = -\frac{d \ln S(t)}{dt}$$

Kaplan-Meier (Product-Limit) method

- Cumulative incidence after 2 case occurrences.

$$\begin{aligned}\text{Risk} &= 1 - \left(\frac{10-1}{10}\right) \cdot \left(\frac{9-1}{9}\right) \\ &= 1 - 0.8 \\ \text{Risk} &= 1 - \text{Survival}\end{aligned}$$

- Risk is estimated for the follow-up time corresponding to each new case occurrence.
- Also known as Kaplan-Meier method of risk estimation and is basis for survival analyses.

Kaplan-Meier (Product-Limit) method

- The survival function *S(t)* is defined as the probability of surviving at least to time *t*.
- The graph of *S(t)* against time is called the **survival curve** and is estimated by the Kaplan-Meier method from observed survival times without the assumption of an underlying probability distribution (nonparametric).
  - Survival = 1 - Risk
  - Risk = 1 - Survival

Kaplan-Meier (Product-Limit) method

Patient	Survival Time (days)	Outcome	Treatment
1	1	Died	2
2	1	Died	2
3	4	Died	2
4	5	Died	2
5	6	Unknown	2
6	8	Died	1
7	9	Survived	2
8	9	Died	2
9	12	Died	1
10	15	Unknown	1
11	22	Died	2
12	25	Survived	1
13	37	Died	1
14	55	Died	1
15	72	Survived	1

- Considers “censoring”, lost to follow up or competing risks.
- Risk is estimated as 1 – survival for the follow-up time corresponding to each new case occurrence (i.e., time to event).

Kaplan-Meier (Product-Limit) method

Patient	Survival Time (days)	Outcome	Treatment
1	1	Died	2
2	1	Died	2
3	4	Died	2
4	5	Died	2
5	6	Unknown	2
6	8	Died	1
7	9	Survived	2
8	9	Died	2
9	12	Died	1
10	15	Unknown	1
11	22	Died	2
12	25	Survived	1
13	37	Died	1
14	55	Died	1
15	72	Survived	1

- Survival at one day:
  - $(15 - 2) \div 15 = 0.867$
- Survival at four days:
  - $0.867 \times (13 - 1)/13 = 0.8000$
- Survival at 37 days
  - $0.4023 \times (3 - 1)/3 = 0.2682$

Patient	Survival Time (days)	Outcome	Treatment	Beginning	Fail	Net Lost	Survivor Function
1	1	Died	2	15	2	0	0.8667
2	1	Died	2	15	2	0	0.8667
3	4	Died	2	13	1	0	0.8000
4	5	Died	2	12	1	0	0.7333
5	6	Unknown	2	11	0	1	0.7333
6	8	Died	1	10	1	0	0.6600
7	9	Survived	2	9	1	1	0.5867
8	9	Died	2	9	1	1	0.5867
9	12	Died	1	7	1	0	0.5029
10	15	Unknown	1	6	0	1	0.5029
11	22	Died	2	5	1	0	0.4023
12	25	Survived	1	4	0	1	0.4023
13	37	Died	1	3	1	0	0.2682
14	55	Died	1	2	1	0	0.1341
15	72	Survived	1	1	0	1	0.1341

Logrank test

- Most widely used method for comparing two or more survival curves. Tests for the equality of survival distributions.

Cox regression models

$$\ln h(t) = h_0 + bx$$

- Uses regression methods for investigating the effect of several variables upon the time a specified event takes to happen.
- Useful in studies with unequal periods of close observation and binary outcome (open cohort studies or survival analyses).
- Provides stable coefficient estimates with decreasing variance estimates for increasing periods.

Cox model assumptions

- Outcome events occur independently.
- Hazard rates may vary over time, but the multiplicative effects of exposure variables (i.e., hazard rate ratios) are constant over the entire period of observation.
- Censoring of subjects should not be related to the probability of an event occurring.