

# Introduction to Survival Analysis

Professor Dr. Syed Hatim Noor  
Unit Biostatistics and Research Methodology  
School of Medical Sciences  
Universiti Sains Malaysia

## Topics to be discussed

- What is survival analysis?
- When to use survival analysis?
- Univariable method: Kaplan-Meier survival curve
- Multivariable methods:
  - Cox-proportional hazards regression model
- Assessment of adequacy of analysis
- Examples and hands-on exercises

Syed Hatim Noor

2

## What is survival analysis?

- Concerns time to event data
- Common examples: death, onset of a disease, relapse of a disease, pregnancy, extraction of tooth after filling, expulsion of intra-uterine device
- Kinds of survival studies : clinical trials, prospective cohort studies, retrospective cohort studies

Syed Hatim Noor

3

## Importance of survival analysis

- Model time to failure or event
- Able to account for censoring
- Can compare survival between groups
- Assess relationship between covariates and survival time
- Can handle data with subjects entering the study at different times, lost to follow-up or withdrawl

Syed Hatim Noor

4

## Censoring

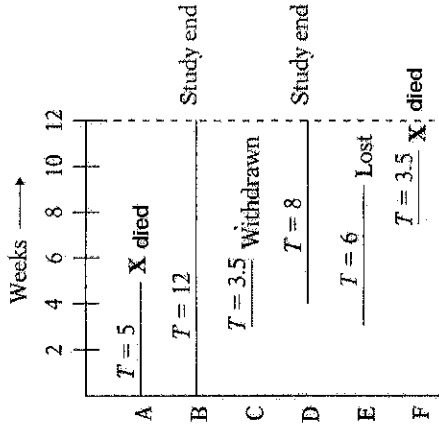
- Distinguishes survival analysis from other classical types of statistical analysis
- The random variable of interest – “time to an event”

Syed Hatim Noor

5

## Censored observation

- A subject who does not experience event of interest during study period
- Incomplete follow-up
  - Lost to follow-up
  - Withdrawal
  - still alive at study closure



Syed Hatim Noor

6

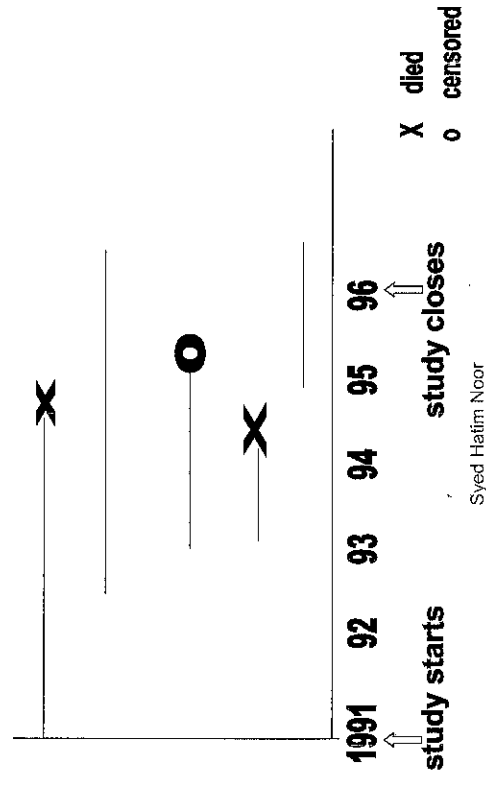
## Illustration of random censoring

- Calendar time
- Study time or internal time

Syed Hatim Noor

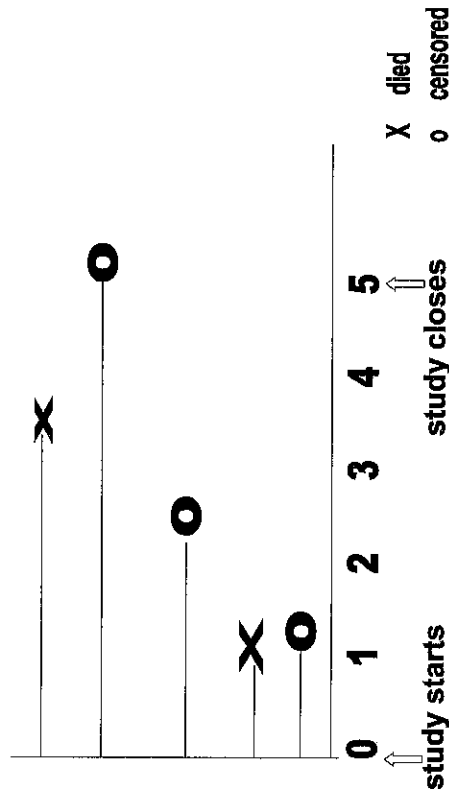
7

## Illustration by calendar time



Syed Hatim Noor

## Illustration by study time



Syed Hatim Noor

10

## Importance of censored data

- If the status of death / censored is not considered, then the researchers will underestimate survival time

Syed Hatim Noor

## Justification to use survival analysis

- Examples
  - Time to death after diagnosing as cancer colon
  - Time in remission after treatment of cancer
  - Time to develop AIDS after receiving HIV contaminated blood transfusion, Etc.
- When one believes that explanatory variable(s) explains the differences in time to an event
- Especially when follow-up is incomplete

Syed Hatim Noor

11

## Methodological characteristics

- A well-defined starting point of each patient (e.g. date of diagnosis or treatment)
- A well-defined endpoint (death or closing date of study of patient withdraws from study) for each patient
- Definition of when the endpoint occurs

Syed Hatim Noor

12

## Reasons for special techniques

- The censored observations do not provide the exact time of death
- The follow-up time varies for case to case

Syed Hatim Noor

13

## Definitions and notations

- Failure time variable is described in terms of either survival function  $S(t)$  or hazard function  $h(t)$

Syed Hatim Noor

14

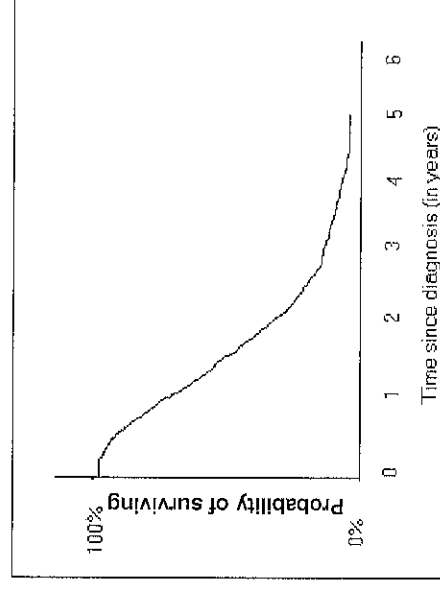
## Survival function $\{S(t)\}$

- Survival function,  $S(t)$  defines an estimate of the probability of cases surviving longer than specified time  $t$
- $S(t) = Pr(T \geq t)$ 
  - this is what we normally want to know

Syed Hatim Noor

15

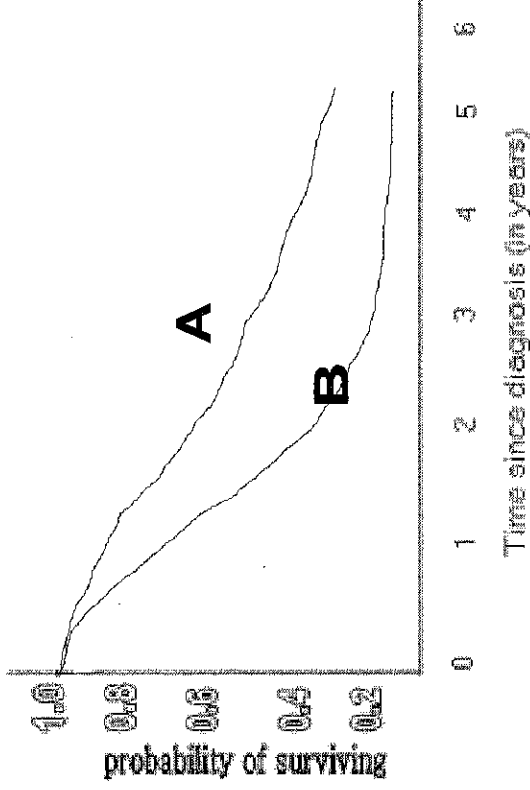
## Survival distribution



Syed Hatim Noor

16

## Survival function



17

Syed Hatim Noor

18

## Hazard function $\{h(t)\}$

- A measure of failure per unit time
- Example : failure (death) just after time  $t$ , among those who are still alive at  $t$
- Indicates the expectation that a case will terminate at any particular time period

## Approach to survival analysis

- Analogous to other statistics
  - Univariable statistics
    - Log Rank test, Simple Cox Regression
  - Multivariable statistics
    - Multiple Cox Regression

Syed Hatim Noor

19