Introduction to Survival Analysis

Tuan V. Nguyen
Professor and NHMRC Senior Research Fellow
Garvan Institute of Medical Research
University of New South Wales
Sydney, Australia

Introduction to survival analysis

- What is survival analysis
- Terminology
- Concepts of survivor function, hazard function

What is survival analysis?

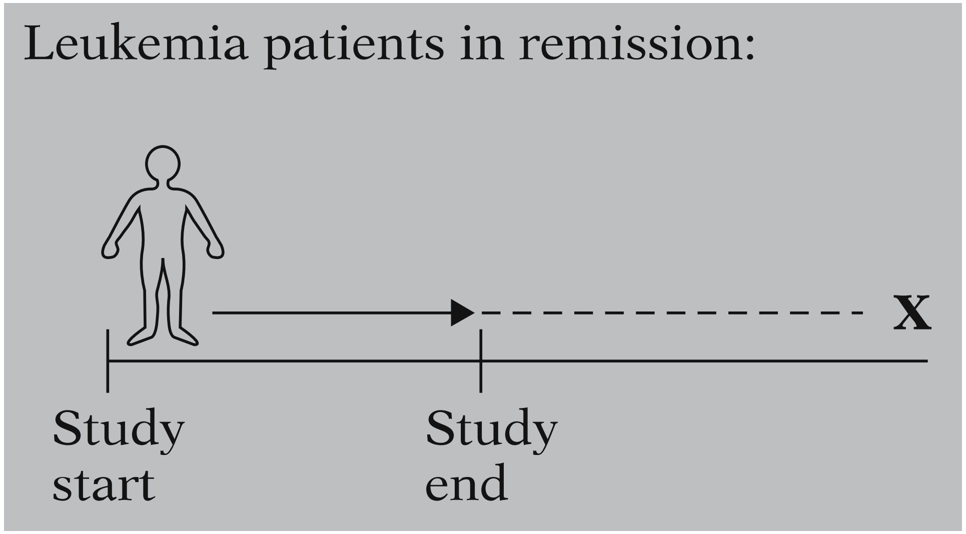
- Statistical technique for analyzing ***prospective study***
- Outcome: **time to an event**
- Time: years, months, weeks, days, etc
- Event: death, disaese, relapse, recovery, etc
- One event, multiple events

Examples of outcome

- Leukemia patients / time to remission (weeks)
- Disease-free cohort until fracture (years)
- Elderly population/time until death (years)
- Heart transplants/time until death (months)
- Time until divorce (marriage)
- Etc

Censored and Failure

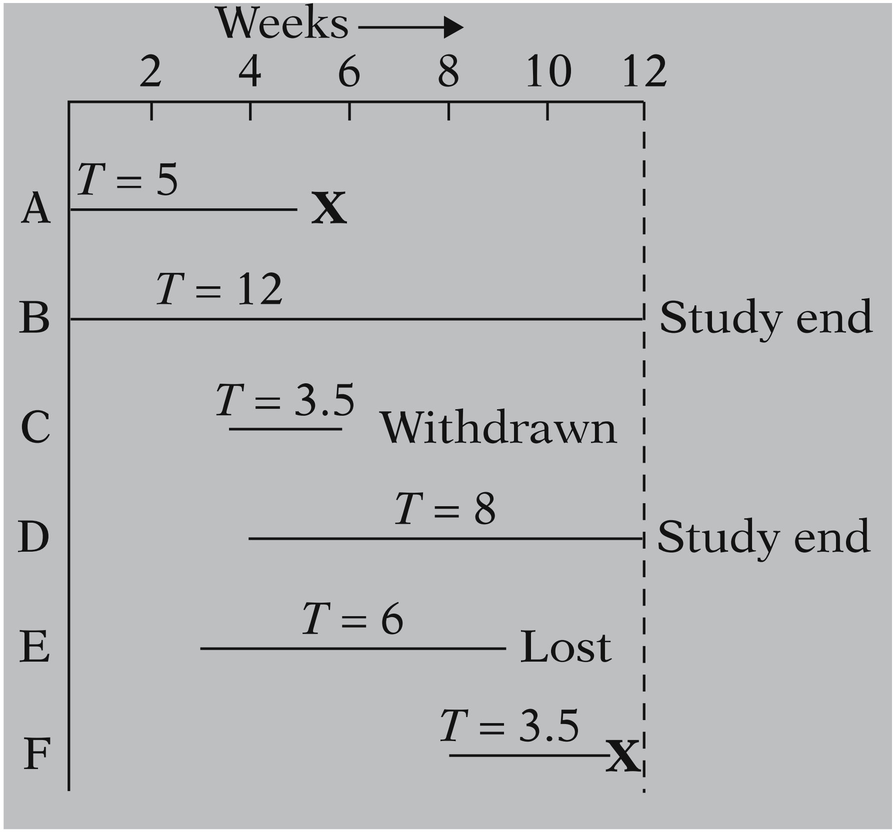
- Censoring: key analytic problem
- Censoring occurs when
  - We have some information about an individual survival time
  - Don't know exactly



Why censoring occurs?

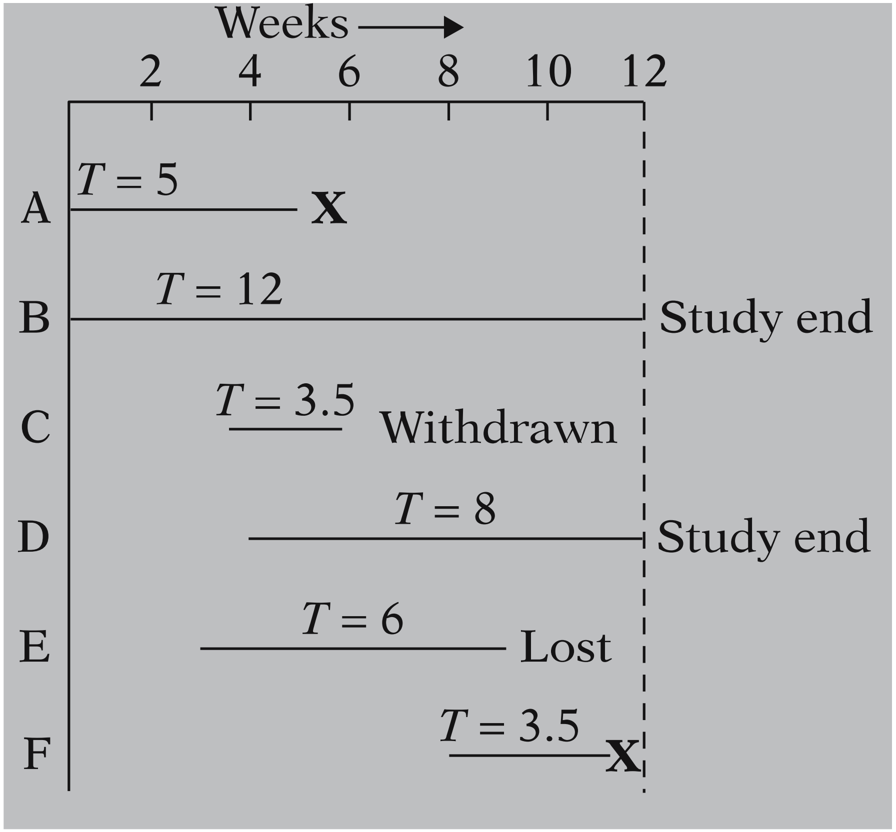
- No event (until the end of study)
- Loss to follow-up
- Withdraws

Censoring



- X denotes an event
- Person A: got the event at week 5 (not censored)
- Person B: followed for 12 weeks, no event until the end of the study
- Person C: entered the study between 2nd and 3rd week, and withdrew at week 6 (survival time 3.5 weeks)
- etc

Survival time

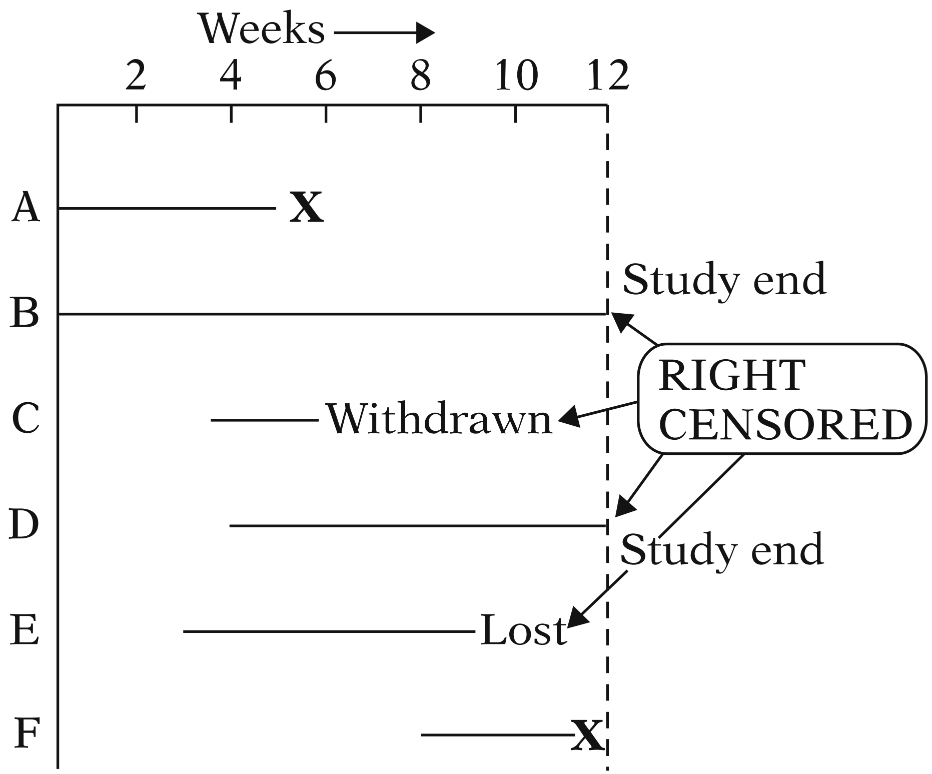


Person	Survival time	Failed (1); censored (0)
A	5.0	1
B	12	0
C	3.5	0
D	8.0	0
E	6.0	0
F	3.5	1

Two types of censoring data

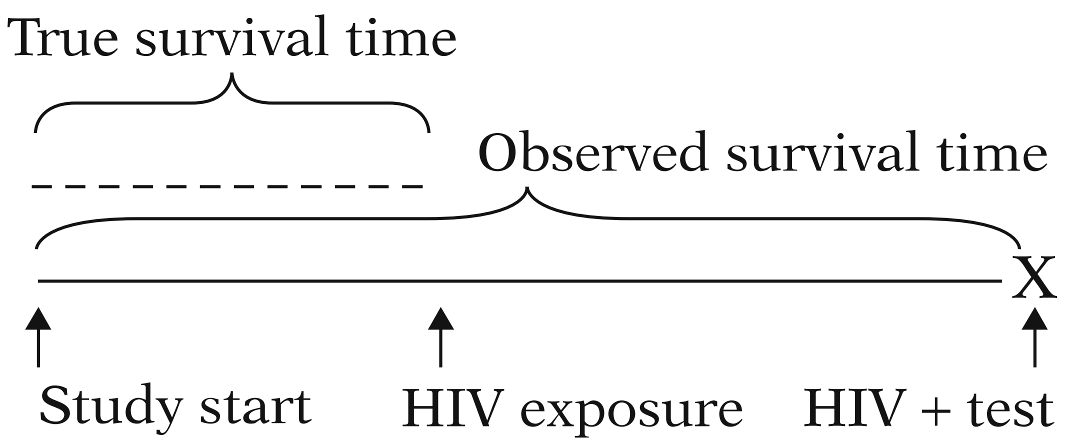
- **Right censored**
- **Left censored**

Right censored data



Exact survival time is incomplete at the right side of the follow-up

Left censored data



True time <= Observed time

When a person's true survival time is LESS than or equal to that person's observed survival time

Terminologies and notations

- Survivor function
- Status variable
- Hazard function

Survivor function

- T = survival time
- $T \geq 0$
- T is a random variable
- t = specific value for T
- "Survived > 5 years" → $T > t=5$
- S(t) : survivor function

Status variable

$$\delta = (0,\, 1) \text{ random variable}$$

$$= \begin{cases} 1 & \text{if failure} \\ 0 & \text{if censored} \end{cases}$$

- **Study ends**
- **Lost to follow-up**
- **Withdraws**

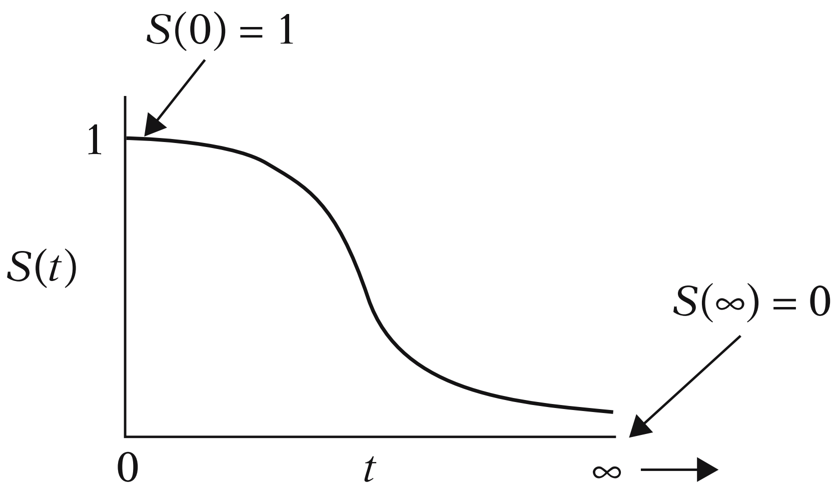
Survival function, S(t)

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

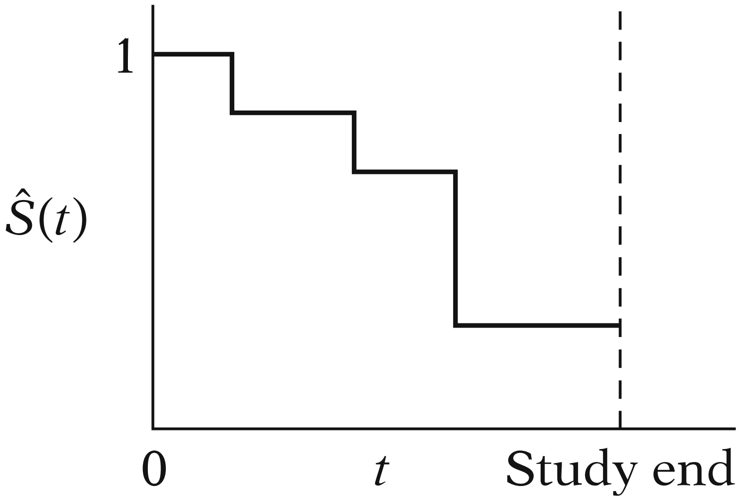
t	**S(t)**
1	$S(1) = P(T > 1)$
2	$S(2) = P(T > 2)$
3	$S(3) = P(T > 3)$
.	.
.	.
.	.

Step function

Theoretical $S(t)$:



$\hat{S}(t)$ in practice:



Hazard function, h(t)

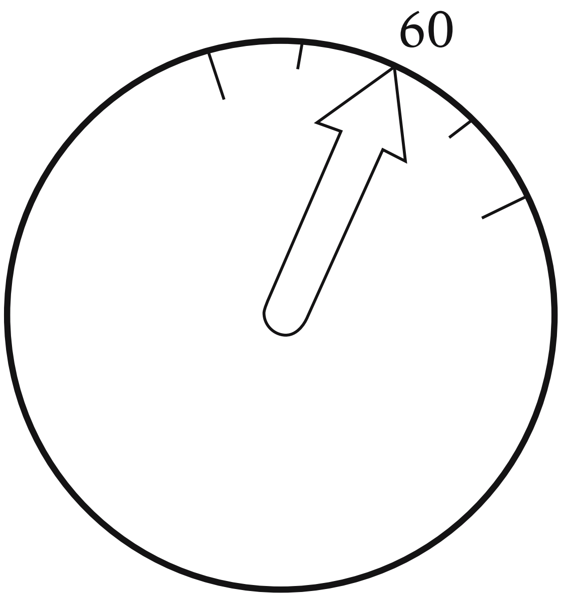
- **S(t) : NOT failing**
- **h(t) : Failing**

- **h(t): the probability of getting an event at time t, GIVEN surviving up to time t**

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t}$$

Hazard function = Velocity

- **Driving a car, 60 km/h**
- **The distance travelled exactly 60 km in 1 hour**
- **How fast the car is going at THAT moment**
- **Hazard h(t): instantaneous potential at time t.**



Hazard function: conditional probability

Given

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

- **Hazard function – conditional probability**
- **P(t ≤ T < t+dt | T ≥ t)**
- **P(person fails in the interval t to t+δt, IF suvived up to time t)**
- **P(A will survive until 80 years if A has survived to 60)**
- **P(A will survive until 80 years if I has survived to 50)**

Hazard function: conditional failure rate

$$\lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t}$$

Probability per unit time

Rate: 0 to ∞

Because of GIVEN (if), hazard function is a rate rather than a probability

Hazard function: conditional failure rate

$P = P(t \leq T \leq t+dt \mid T \geq t)$

P = 1/3

P	**dt**	**P / dt = rate**
1/3	½ day	(1/3) / (1/2) = 0.67/day
1/3	1/14 week	(1/3) / (1/14) = 4.67/week
.	.	.
.	.	.

Instantaneous potential

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t}$$

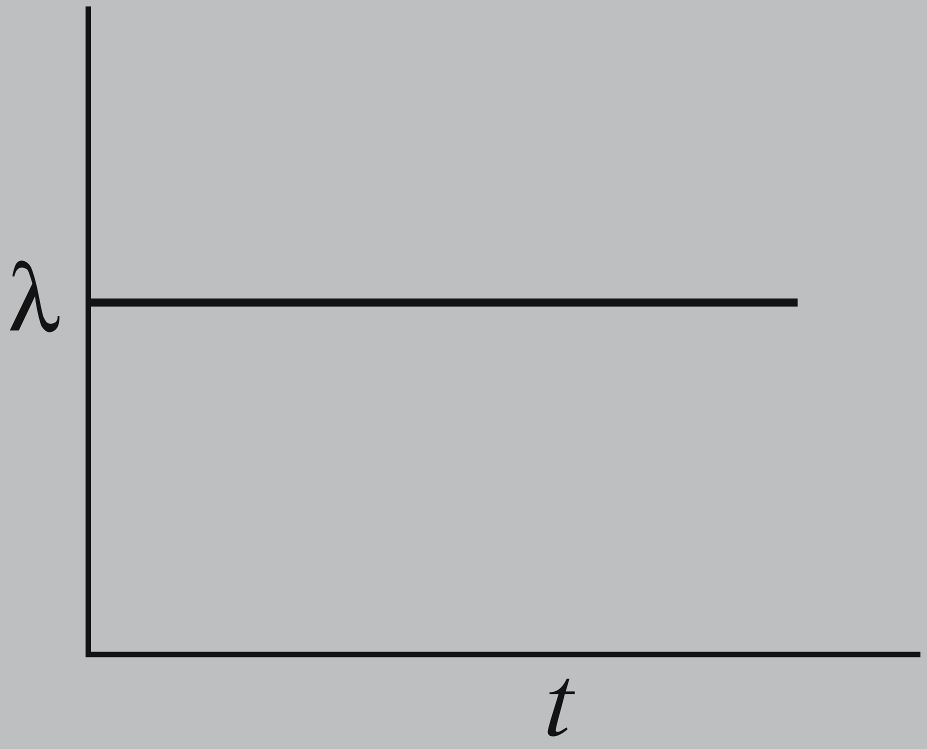
Gives instantaneous potential

- When limit near 0 (dt → 0), it is an instantaneous rate at time t per unit time
- h(t) gives the instantaneous potential for failing at time t per unit time (given survived up to time t)

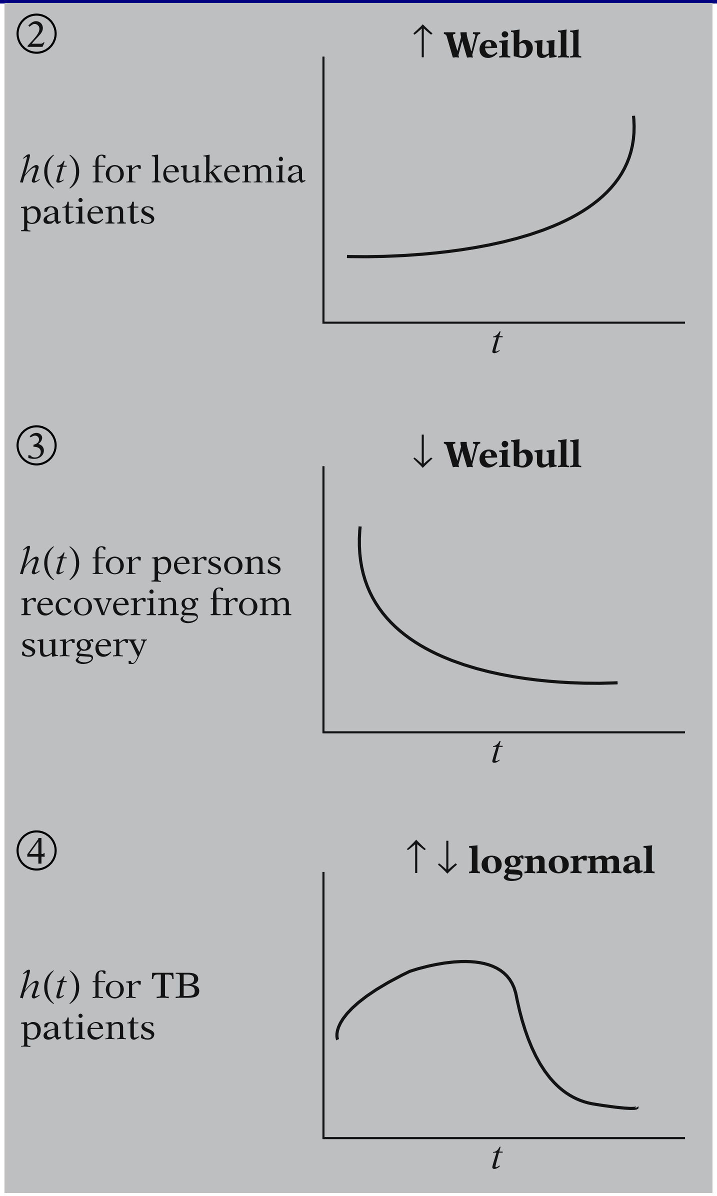
Constant hazard

Constant hazard
(**exponential model**)

$h(t)$ for healthy persons



More on hazard functions



- Increasing Weibull model
- Decreasing Weibull model
- Lognormal survival

Uses of hazard function

- A measure of instantaneous potential
- Useful for identifying a specific model form (eg Weibull, lognormal)
- "Vehicle" for mathematical modeling of survival data (eg survival model is written in terms of hazard function)

Relationship between S(t) and h(t)

- **H(t) is sometimes denoted by λ (lambdat)**

$$h(t) = \lambda \text{ if and only if } S(t) = e^{-\lambda t}$$

General formulae:

$$S(t) = \exp\left[-\int_0^t h(u)du\right]$$

$$h(t) = -\left[\frac{dS(t)/dt}{S(t)}\right]$$

Goals of survival analysis

- To estimate and interpret survivor/hazard functions from survival data
- To compare survivor/hazard functions
- To assess the relationship of predictors to survival time
- Develop prognostic models