Survival Analysis

Dr Saikat Das

Christian Medical College Vellore

Specific learning objective (SLO)

1. What is the statistical relevance of survival analysis
2. What are the three basic concepts in survival analysis (might be little mathematical)
3. Understand the two ways of describing survival data
4. Learn how to do intergroup comparison
5. *If possible, learn* regression analysis

Introduction to survival analysis

- Statistical concept of rate

- In most of the statistical analysis we assume rate is constant over the period of study or with in the time period specified

Data on the occurrence of MI

Table 24.2 First ten lines of the computer dataset from the Caerphilly study. Analyses of the Caerphilly study are by kind permission of the MRC Steering Committee for the Management of MRC Epidemiological Resources from the MRC Epidemiology Unit (South Wales).

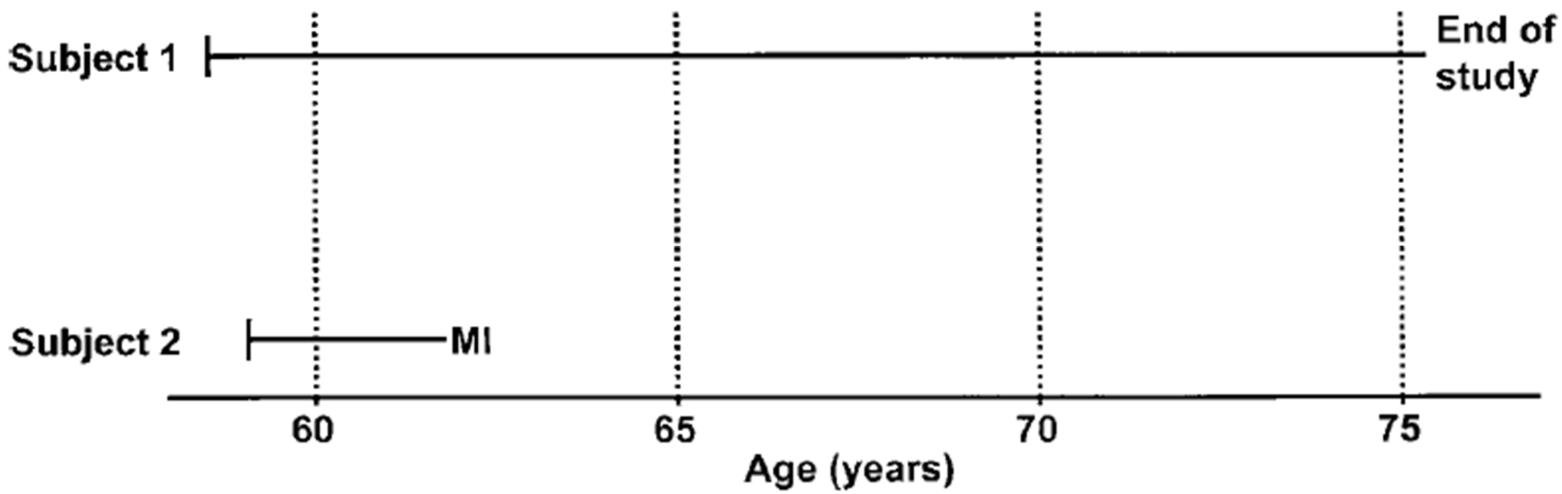
id	dob	examdate	exitdate	years	MI	cursmoke
1	20/May/1929	17/Jun/1982	31/Dec/1998	16.54	0	1
2	9/Jul/1930	10/Jan/1983	24/Dec/1998	15.95	0	0
3	6/Feb/1929	23/Dec/1982	26/Nov/1998	15.93	0	1
4	24/May/1931	7/Jul/1983	22/Nov/1984	1.38	1	0
5	9/Feb/1934	3/Sep/1980	19/Dec/1998	18.29	0	0
6	14/Mar/1930	17/Nov/1981	31/Dec/1998	17.12	0	0
7	13/May/1933	30/Oct/1980	27/Dec/1998	18.16	0	1
8	23/May/1924	24/Apr/1980	24/Jan/1986	5.75	1	1
9	20/Jun/1931	11/Jun/1980	12/Dec/1998	18.50	0	1
10	12/May/1929	17/Nov/1979	20/Jan/1995	15.18	1	0

Age grouping for convenience

Table 24.10 Follow-up time split into 5-year age bands for the first two subjects in the Caerphilly study.

Date at start of interval	Date at end of interval	Age group	Age at start of interval	Age at end of interval	Years in interval	MI
Subject 1, born 22 Aug 1923, recruited 1 Mar 1982, exit (at end of follow-up) 31 Dec 1998						
1 Mar 1982	21 Aug 1983	55–59	58.52	60	1.48	0
22 Aug 1983	21 Aug 1988	60–64	60	65	5	0
22 Aug 1988	21 Aug 1993	65–69	65	70	5	0
22 Aug 1993	21 Aug 1998	70–74	70	75	5	0
22 Aug 1998	31 Dec 1998	75–79	75	75.36	0.36	0
Subject 2, born 8 May 1923, recruited 30 May 1982, exit (on date of MI) 27 Feb 1985						
30 May 1982	7 May 1983	55–59	59.06	60	0.94	0
8 May 1983	27 Feb 1985	60–64	60	61.81	1.81	1

Age grouping visually



Can we always assume rate is constant?

- Longitudinal study in which there is a clear event: Diagnosis , intervention of treatment: can assume

- Consider the variables like the following

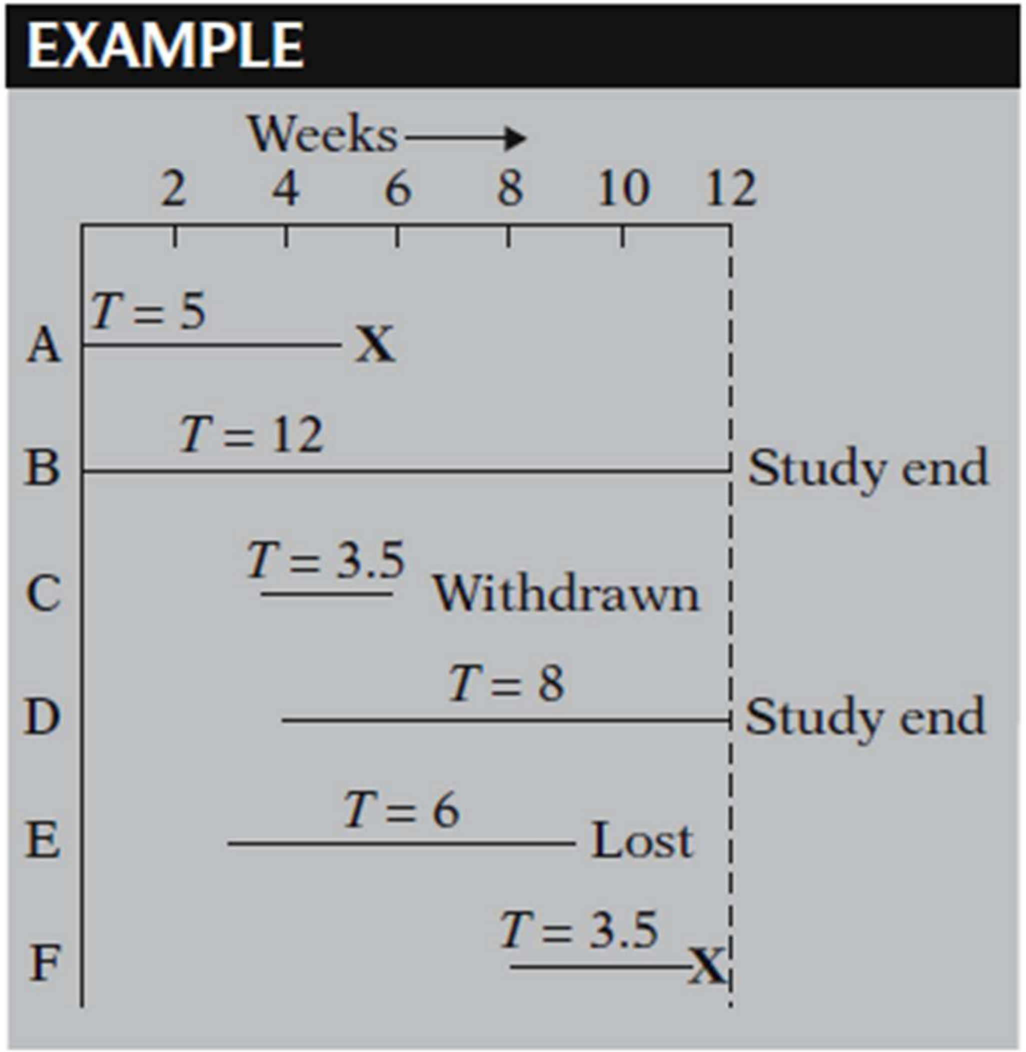
1. Risk of death following a surgery
2. Recurrence rate of a tumor

Survival data

Survival data has three distinct features:

- time of event (such as death, recurrence, new primary)
- time variable does not follow a Normal distribution
- events could not have happened yet (censored)

Example of key features



Key features in survival analysis

1. Hazard function h(t)

2. Survivor function S(t)

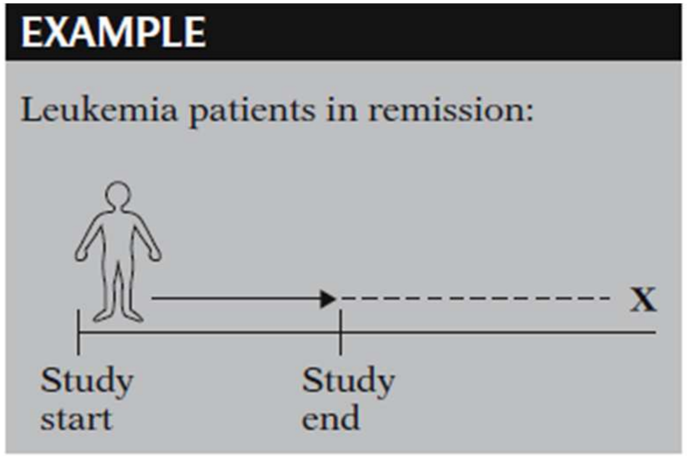
3. Censoring

Will discuss 1 and 2 later, let us see 3 first

Censoring

- Survival time of subjects who are known to have survived up to a certain point of time and beyond that survival status is not known
- Censoring probability is unrelated to survivor probability
- censoring occurs when we have some information about individual survival time, but **we don’t know the survival time exactly**

Censoring



1. Study ends
2. Lost to follow up
3. Withdrawing

Mathematical parameters of survival analysis

T = survival time random variable

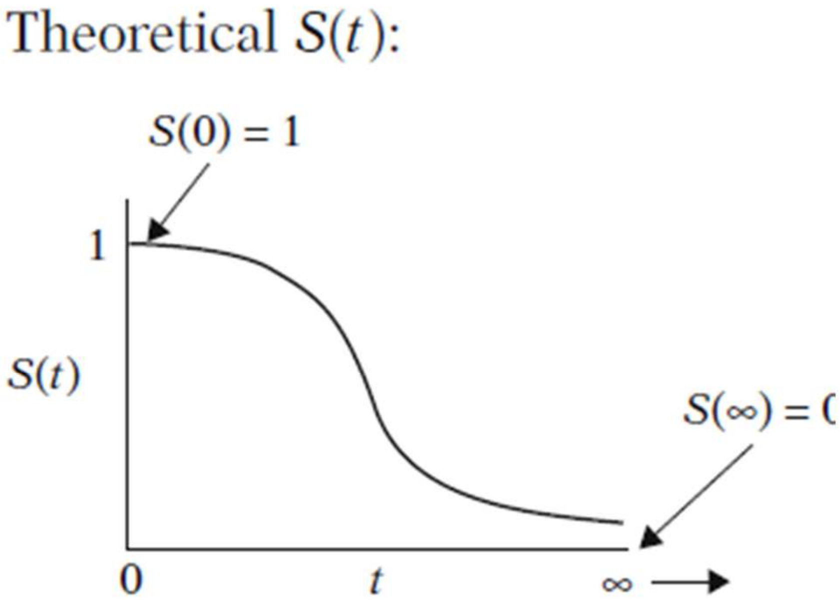
t = specific value of T

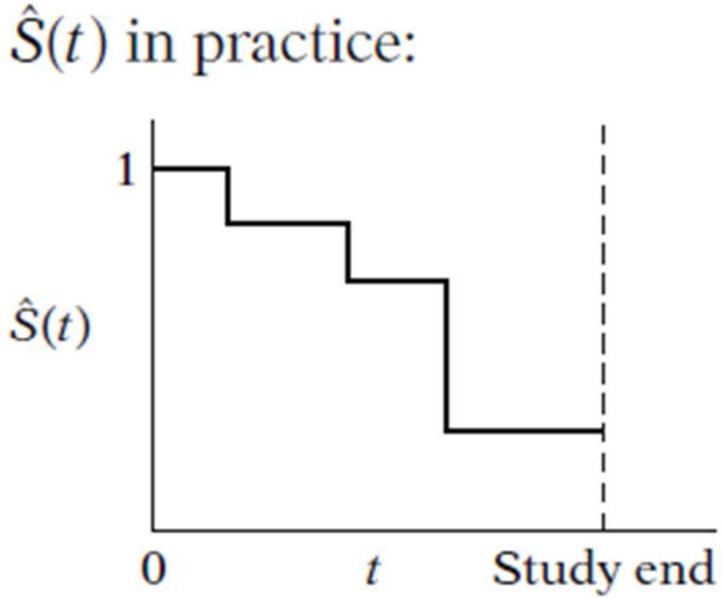
δ = (0, 1) variable for failure/censorship

$S(t)$ = survivor function

$h(t)$ = hazard function

Survival function





Hazard Function

EXAMPLE

①
Constant hazard
(exponential model)

$h(t)$ for healthy persons

Hazard function ≡ conditional failure **rate**

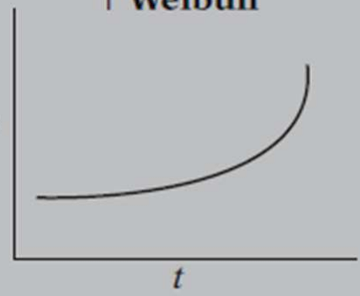
$$\lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t}$$

EXAMPLE (continued)

②

↑ Weibull

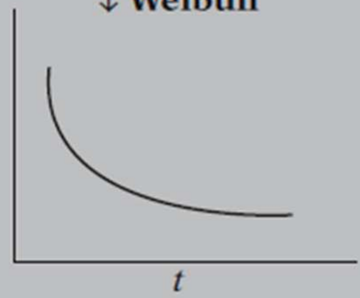
$h(t)$ for leukemia patients



③

↓ Weibull

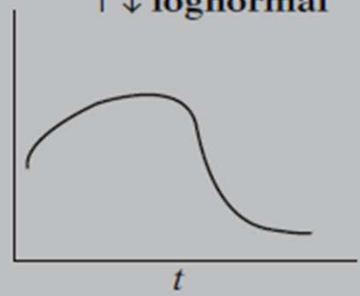
$h(t)$ for persons recovering from surgery



④

↑↓ lognormal

$h(t)$ for TB patients



Relationship between survival and hazard function

$$S(t) \longleftrightarrow h(t)$$

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

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

Representing survival analysis

- Life table

- Kaplan Meier method

Life table

1. Cohort life table (Refer the data set 1)

2. Current (period) life table (Refer the data set 2)

3. Complete life table

4. Abridged life table

Exercise on life table

- Based on data set 1

Drawbacks of life table

- Assumption due to time grouping
- No information of individual survival data

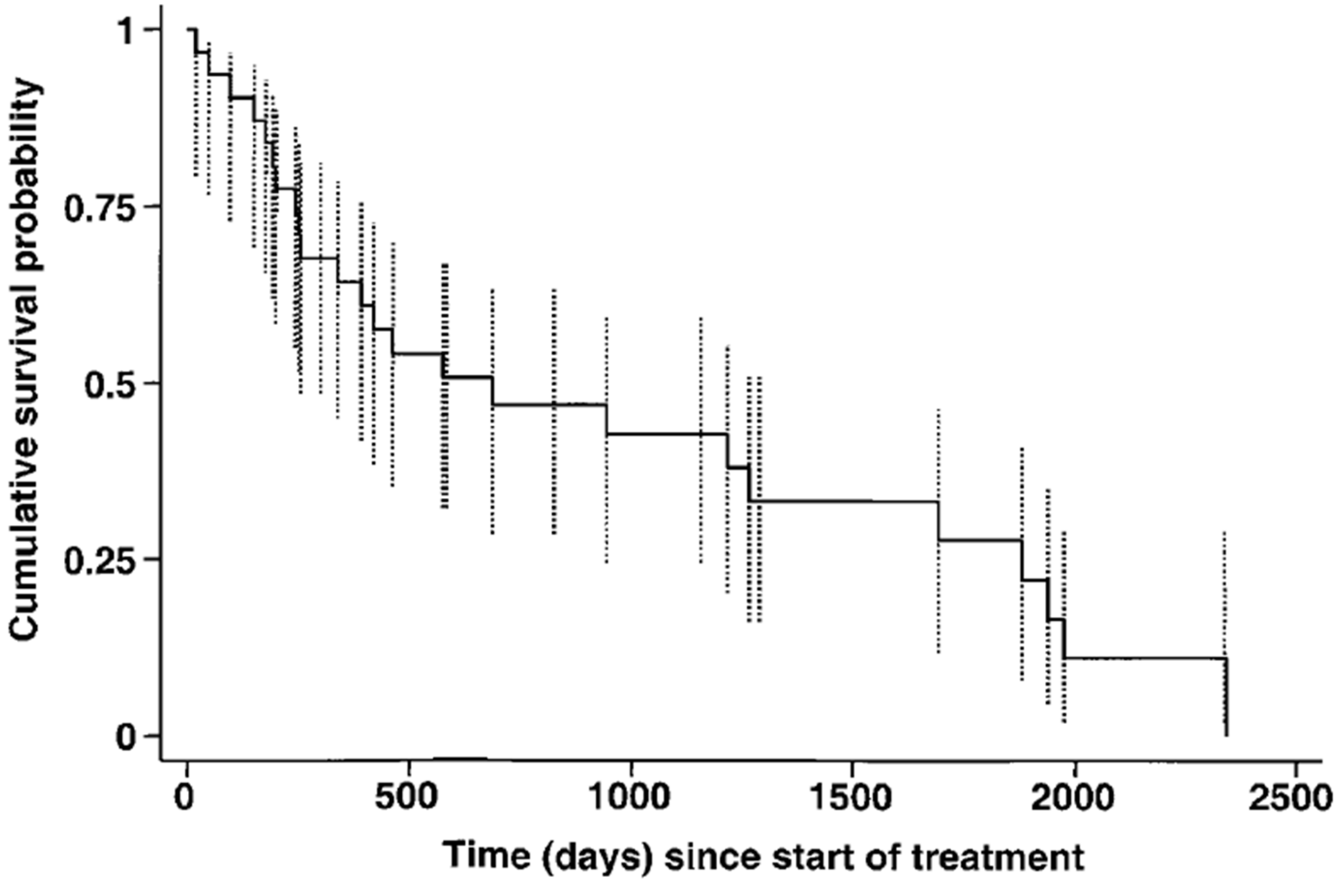
Look at dataset 3

- Biliary cirrhosis data

Kaplan Meier estimate of survival function

$$s_t = 1 - r_t = \frac{n_t - d_t}{n_t}$$

Plot as per Kaplan Meier estimate



Tutorial 1

- Follow through how a Kaplan Meier survival curve is drawn using SPSS

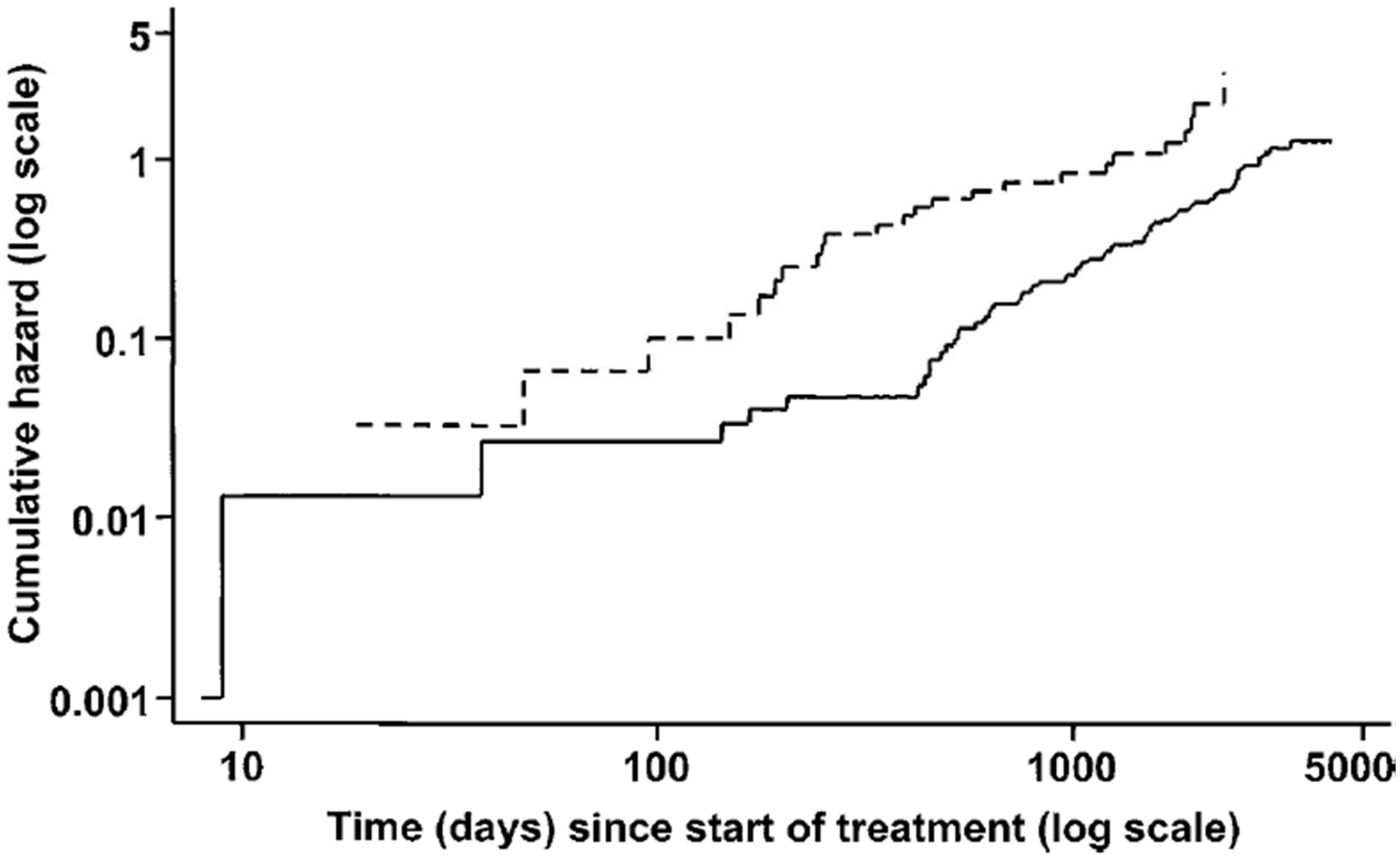
Proportional hazard assumption: comparing survival curve

$$\frac{h_1(t)}{h_0(t)} = \text{constant}$$

$$H(t) = \Sigma \frac{d_i}{n_i}, \text{ summed over all times up to and including } t$$

$$\frac{H_1(t)}{H_0(t)} = \frac{h_1(t)}{h_0(t)} = \text{constant}$$

Plot of hazard ratio (x-axis in log scale)



Comparisons of Hazards

1. Mantel-Cox method (Log rank test)
2. Modification of Mantel-Haenszel procedure
3. Constructs separate 2X2 table at each time event occurs
4. Combines the contribution of each table over the period of time assuming the hazard ratio is constant

2 X 2 Table

Table 18.3 Notation for the 2 × 2 table in stratum i.

	Outcome		
	Experienced event: D (Disease)	Did not experience event: H (Healthy)	Total
Group 1 (exposed)	d_{1i}	h_{1i}	n_{1i}
Group 0 (unexposed)	d_{0i}	h_{0i}	n_{0i}
Total	d_i	h_i	n_i

Mantel-Haenszel estimate of Odds Ratio

$$\mathrm{OR}_i = \frac{d_{1i} \times h_{0i}}{d_{0i} \times h_{1i}} \qquad w_i = \frac{d_{0i} \times h_{1i}}{n_i}$$

$$\mathrm{OR}_{MH} = \frac{\Sigma(w_i \times \mathrm{OR}_i)}{\Sigma w_i} = \frac{\Sigma \dfrac{d_{1i} \times h_{0i}}{n_i}}{\Sigma \dfrac{d_{0i} \times h_{1i}}{n_i}}$$

Mantel-Cox estimate of Hazard ratio

Here usually d0*i* or d1*i* is 0 or 1 and vice versa

$$\mathrm{HR}_{\mathrm{MC}} = Q/R, \text{ where}$$

$$Q = \Sigma \frac{d_{1i} \times h_{0i}}{n_i} \text{ and } R = \Sigma \frac{d_{0i} \times h_{1i}}{n_i}$$

Mantel Cox X^2 (log rank) test

$$\chi^2_{\text{MC}} = \frac{U^2}{V};\ \text{d.f.} = 1,\ \text{where}$$

$$U = \Sigma(d_{1i} - E_{1i}),\ \text{and}\ E_{1i} = \frac{d_i \times n_{1i}}{n_i}$$

Understand the steps

- Go through the dataset 4

Tutorial 2

- Find out the Mantel-Cox estimate and do the log rank test

What did we learn

1. Properties of survival analysis and parameters (Hazard function and survival function)

2. Life table and Kaplan Meier Curve

3. How to do intergroup comparison by log rank test using SPSS

Practical interpretation

- Look at the IPASS data (Survival curve in Figure 2)

- Now read the methods section (highlighted)

Cox Regression

- Semi-parametric
- Cox models the effect of predictors and covariates on the hazard rate but leaves the baseline hazard rate unspecified.
- Also called proportional hazards regression
- Does NOT assume knowledge of absolute risk.
- Estimates *relative* rather than *absolute* risk.

Cox regression vs.logistic regression

Distinction between rate and proportion:

- Incidence (hazard) rate: number of new cases of disease per population at-risk per unit time (or mortality rate, if outcome is death)
- Cumulative incidence: proportion of new cases that develop in a given time period