Accelerating Impact: Immersive Summer Bootcamp in Implementation Science and Biostatistics

Georgian Implementation Science Fogarty Training (GIFT) Program

Ilia State University & Yale University

Yale Center for Analytical Sciences

Survival Analysis

- Pediatric Melanoma: Risk Factor and Survival Analysis of the Surveillance, Epidemiology and End Results Database

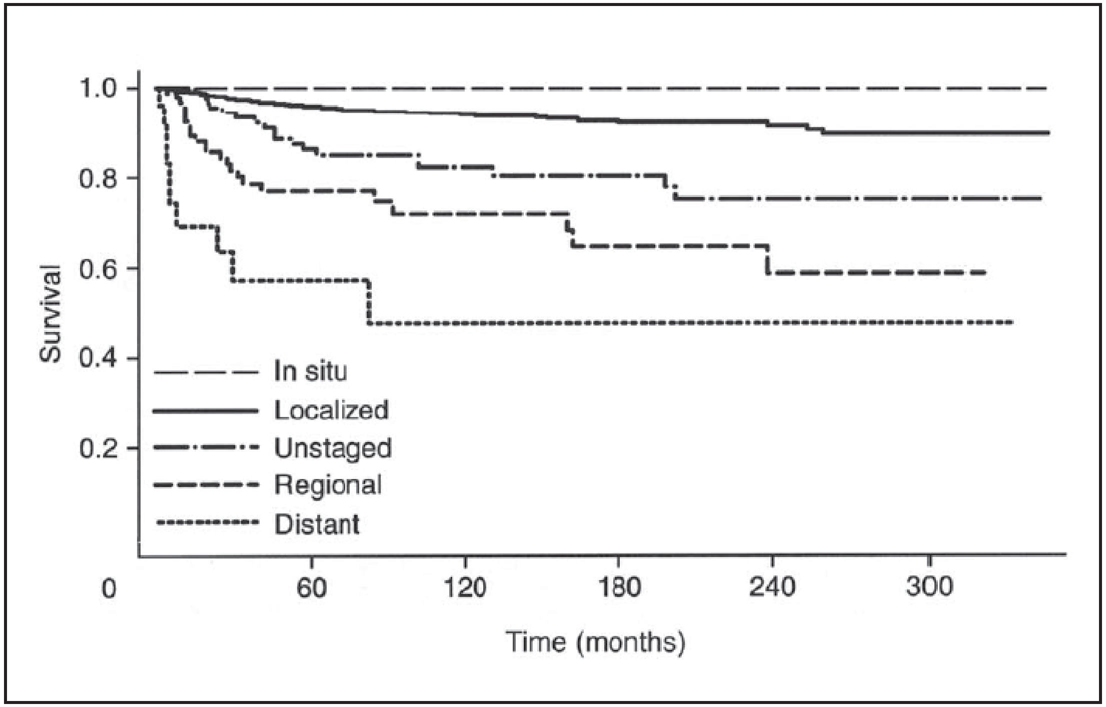


Fig 5. Mortality from malignant melanoma in children (age < 20 years) of all races by extent of disease from the Surveillance, Epidemiology and End Results 12 database (1973 to 2001).

- Strouse et al (2005); *Journal of Clinical Oncology*: 23(21)

Intensity of Renal Support in Critically Ill Patients with Acute Kidney Injury

The VA/NIH Acute Renal Failure Trial Network*

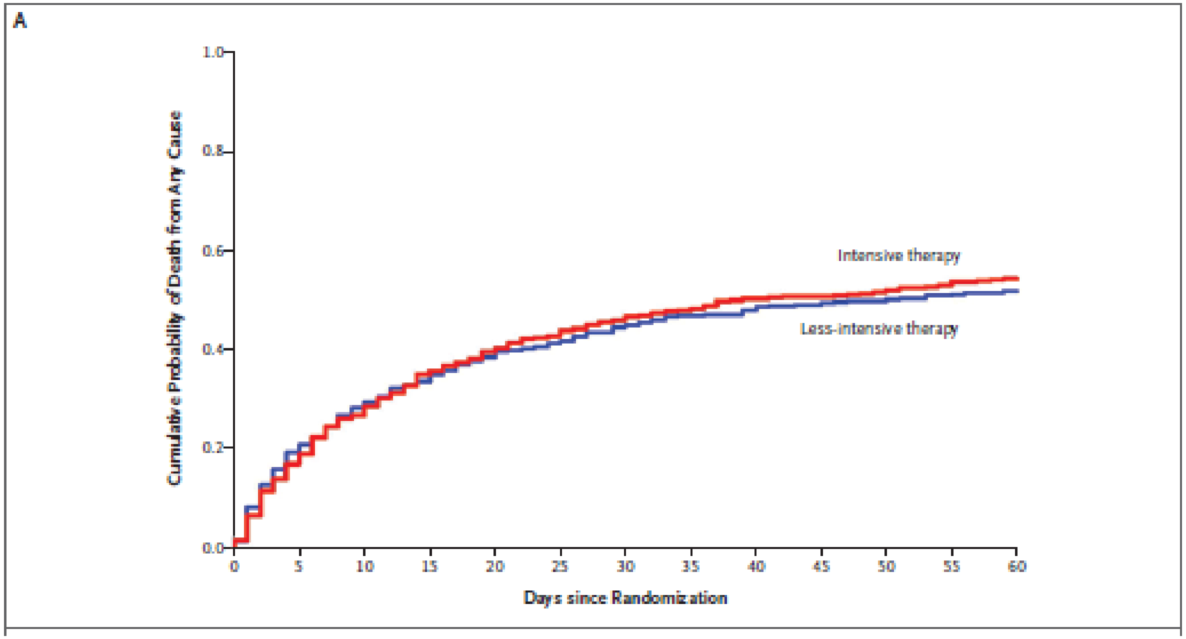


Figure 2. Kaplan–Meier Plot of Cumulative Probabilities of Death (Panel A) and Odds Ratios for Death at 60 Days, According to Baseline Characteristics (Panel B).

Panel A shows the cumulative probability of death from any cause in the entire study cohort. Panel B shows odds ratios (and 95% confidence intervals [CI]) for death from any cause by 60 days in the group receiving the intensive treatment strategy as compared with the group receiving the less-intensive treatment strategy, as well as P values for the interaction between the treatment group and baseline characteristics. P values were calculated with the use of the Wald statistic. Higher Sequential Organ Failure Assessment (SOFA) scores indicate more severe organ dysfunction. There was no significant interaction between treatment and subgroup variables, as defined according to the prespecified threshold level of significance for interaction (P=0.10). Sex was not recorded for one patient receiving less-intensive therapy.

Survival Analysis

- Collection of methods that are used to describe how or why certain events occur
  - Also called event history analysis, failure-time analysis, reliability analysis, duration analysis, transition analysis, hazard analysis
- Event – a qualitative change that can be localized in time (e.g. death, AIDS, marriage, promotion)
- Event history – a longitudinal record of when events occurred for set of individuals

Objectives of Survival Analysis

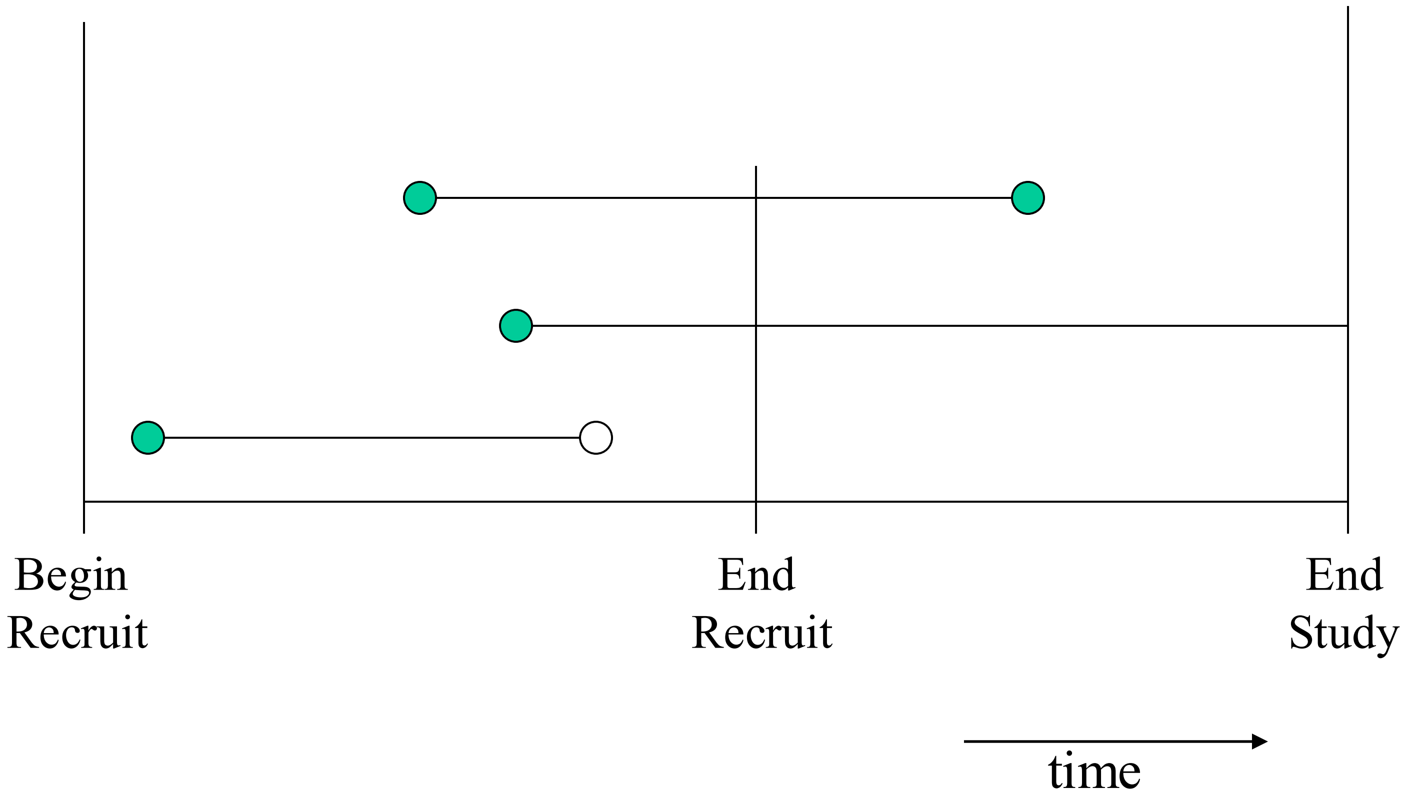
- Estimate survival time/time to event
  - Example – time to death due to lung cancer diagnosis
- Compare time to event between 2 or more groups
  - Example – time to death between patients receiving chemotherapy and surgery vs no treatment
- Examine the relation between independent variables and survival time
  - Example – do type of treatment, smoking, coffee drinking, gender influence survival in cancer patients

Lung Cancer and Mortality

	Death	No Death	
No Trt	50	0	50
Chemo + Surgery	50	0	50
	100	0	100

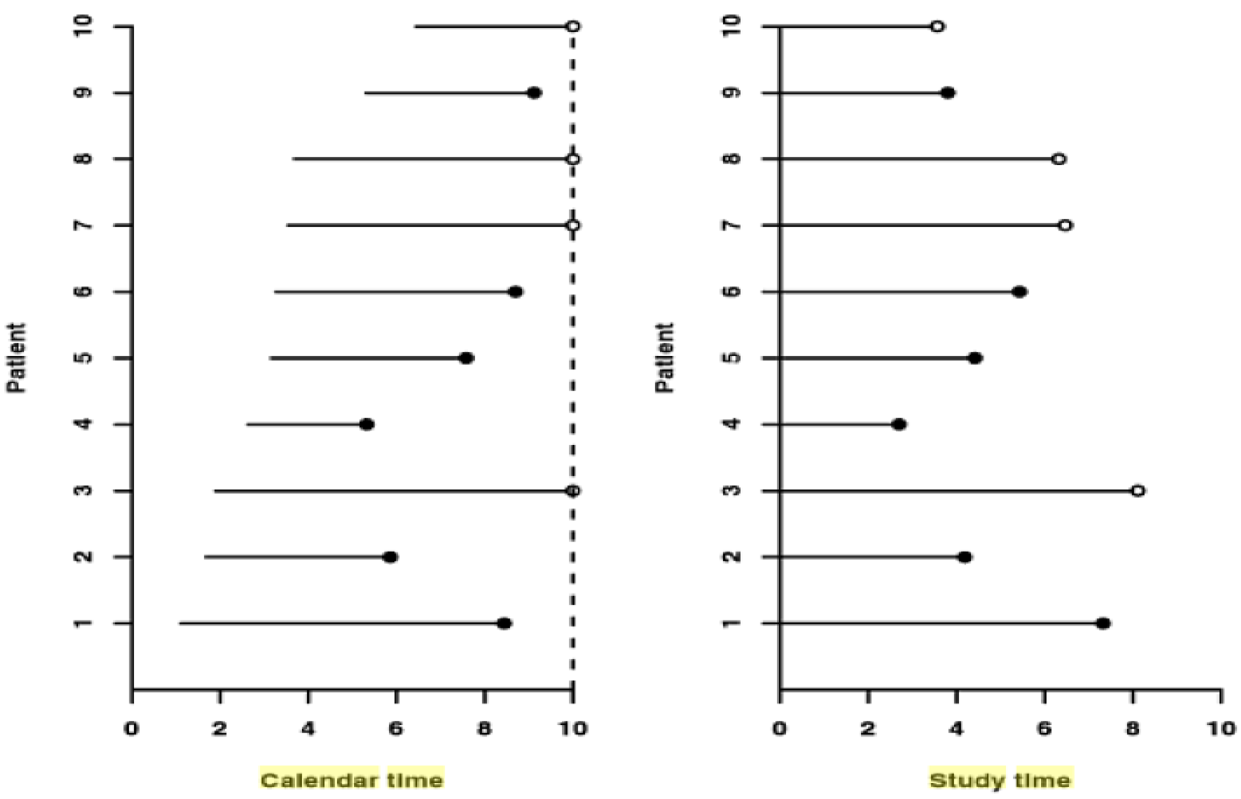
Survival Analysis

- Evaluate time to event while taking into account censored observations



Calendar Time vs Study Time

- Example: Following 10 patients over 10 months



Survival Analysis

- Combination of two outcome variables
  - Event/no event – δ_i
  - Time followed – t_i

- How do we define the outcome?
  - What is the event?
  - When do we start the clock?
    - Time of first diagnosis
    - Time of start of treatment
    - Time of end of primary episode
    - Time of entry into a cohort

Motivation for Survival Analysis

- Time to an event is a continuous variable, why not use a **t-test** or **linear regression**?
  - Everyone has to be followed to the outcome of interest
  - Survival times tend to be *positively skewed*
  - *Log* transformation of T to normalize it still ignores subjects without an event during a follow up!
- An event is a binary outcome, why not use a **logistic regression**?
  - Losing information from knowledge of timing of event
- **Survival Analysis** evaluates time to event while taking into account censored observations

Survival Analysis

- Failure rate (hazard) –when rate is constant

$$r = \frac{\#\text{ of failures (deaths)}}{\text{Total Follow-up Time}} = \frac{\sum \delta_i}{\sum t_i}$$

- CI for failure rate

$$e^{\ln r \pm z_{\alpha/2}\sqrt{1/\sum \delta_i}}$$

Survival Analysis

- Example – time to death for cancer patients
  - 6* 6 10 11* 13 35* - * censored observation

$$\sum \delta_i = 3$$

$$\sum t_i = 6 + 6 + 10 + 11 + 13 + 35 = 81$$

$$r = \frac{3}{81} = 0.037 \text{ per p - m}$$

- 90% CI

$$e^{\ln 0.037 \pm 1.645 \sqrt{1/3}} \qquad (0.014, 0.096)$$

Survival Curves

- Estimation of the probability of surviving past a given time.

Survival Curves
Actuarial Method

- Breakdown time into equal intervals and quantify
  - l_i - # under follow-up at beginning of interval
  - d_i - # of outcomes during the interval
  - w_i - # censored during the interval

Survival Curves

Actuarial Method

6* 6 10 11* 13 35*

	Interval	d_i	w_i	l_i	l_i^1	p_i	P_i
x_1	(0, 6]	1	1	6	5.5	$1-1/5.5=0.82$	
x_2	(6, 12]	1	1	4	3.5	$1-1/3.5=0.71$	
x_3	(12, 18]	1	0	2	2	$1-1/2=0.50$	
x_4	(18, 24]	0	0	1	1	$1-0/1=1.00$	
x_5	(24, 30]	0	0	1	1	$1-0/1=1.00$	
x_6	(30, 36]	0	1	1	0.5	$1-0/0.5=1.00$	

- The conditional probability of surviving an interval given that alive at beginning of interval

$$p_i = P(\text{alive at } x_i \mid \text{alive at } x_{i-1}) = 1 - \frac{di}{l_i'}$$

Survival Curves

Actuarial Method

	Interval	d_i	w_i	l_i	l_i^1	p_i	P_i
x_1	(0, 6]	1	1	6	5.5	$1-1/5.5=0.82$	0.82
x_2	(6, 12]	1	1	4	3.5	$1-1/3.5=0.71$	0.58
x_3	(12, 18]	1	0	2	2	$1-1/2=0.50$	0.29
x_4	(18, 24]	0	0	1	1	$1-0/1=1.00$	0.29
x_5	(24, 30]	0	0	1	1	$1-0/1=1.00$	0.29
x_6	(30, 36]	0	1	1	0.5	$1-0/0.5=1.00$	0.29

S(t) ← P_i

- Use the conditional probabilities to calculate the cumulative survival (i.e. the probability of surviving to a particular time)

$$P_i = P(\text{alive at } x_i) = p_i \times p_{i-1} \times p_{i-2} \times \ldots p_1$$

Survival Curves
Actuarial Method

- Assumptions
  - Withdrawals occur randomly during interval
  - The probability of surviving one interval is independent of probability of surviving other intervals
  - Survival probabilities are the same for those entering the study early as those entering late – average survival is not changing during the course of the study
    - Example – if estimating survival from cancer in a lengthy observational study, the casemix of cancer can change due to earlier detection where less aggressive tumors are included
  - Censoring is not informative (unrelated to event)
    - any subject that is censored has the same survival prospects as those that continue to be followed
  - Observations are independent

Informative Censoring in Presence of Covariates

- If you have covariate information, censoring should be independent at each level of the covariate (i.e. among those with a given value of the covariate, censored subjects must have similar prospects for survival as uncensored)
  - May be dependent on different levels of the covariate
  - For example, earlier in a trial older subjects (with increased risk of event) may not be enrolled but conditional on age censoring is uninformative – need to adjust for age in the analysis

Survival Curves

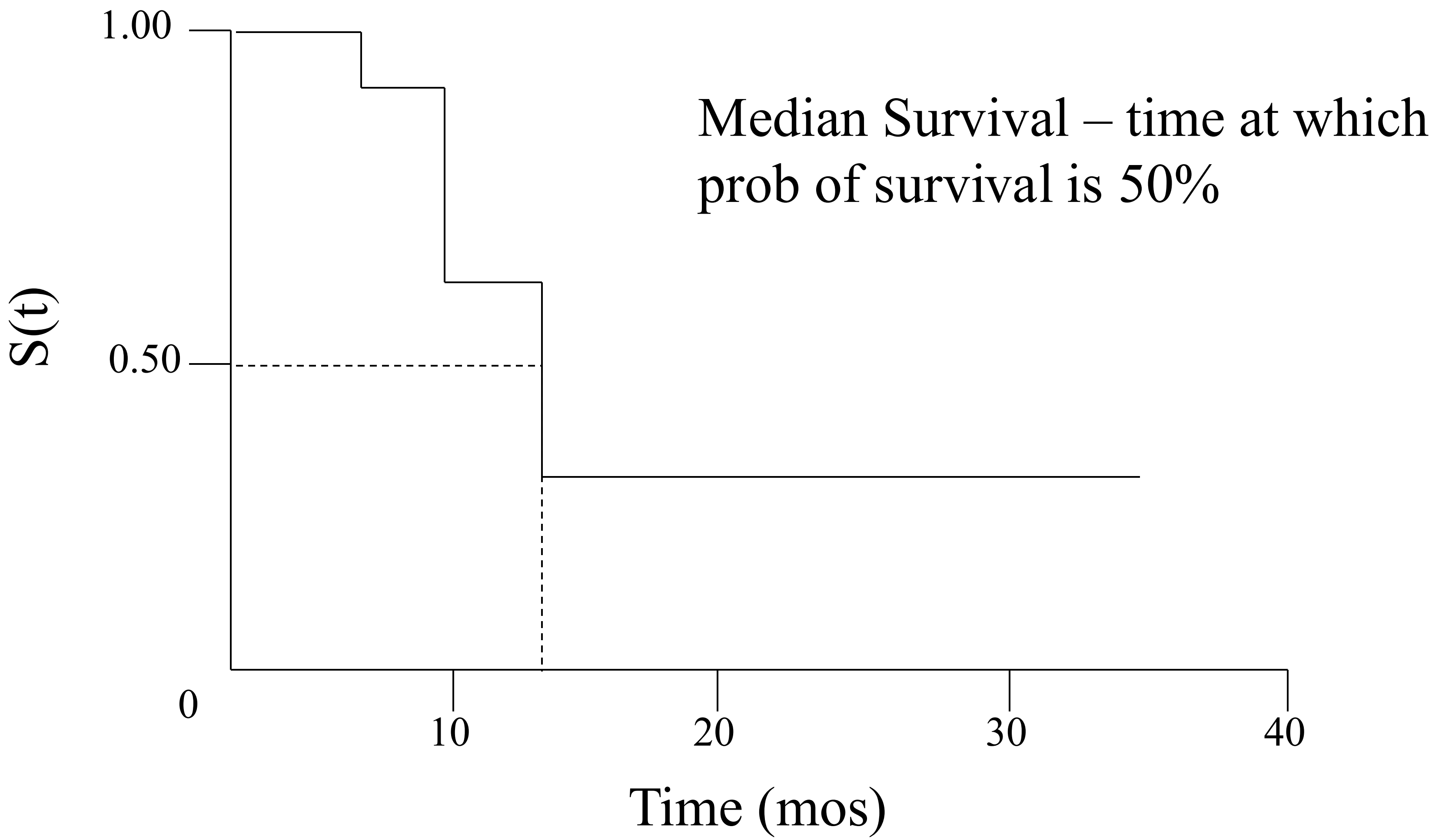
Kaplan-Meier (Product Limit) Method

- Estimate survival each time an event occurs

$p_i = 1 - \frac{d_i}{l_i}$

Time	d_j	w_i	l_i	p_i	P_i
0	0	0	6	1.0	1.0
6	1	1	6	$1-\frac{1}{6}=0.83$	0.83
10	1	0	4	$1-\frac{1}{4}=0.75$	0.62
13	1	0	2	$1-\frac{1}{2}=0.50$	0.31

Plot of Survival Curve



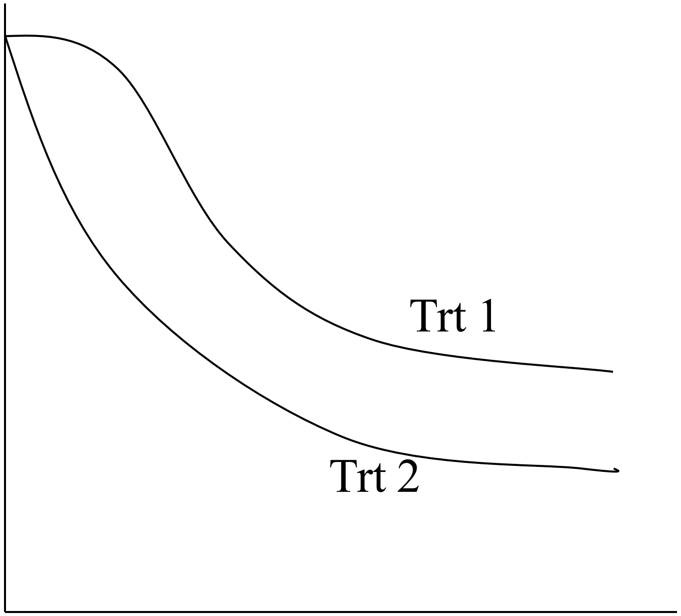
K-M Survival Curve

- Properties
  - If last observation is an event, K-M estimate of survival is 0
  - If last observation is censored, K-M estimate of survival is undefined after that time
- Assumptions
  - A censored value is not informative (unrelated to event)
  - The probability of surviving one interval is independent of probability of surviving other intervals
  - Survival probabilities are the same for those entering the study early as those entering late
  - Observations are independent

Comparing Survival Curves

- Log-rank (Mantel-Haenszel Test)
  - H_0: Survival curves are the same
  - H_a: Survival curves differ
- At each event time, calculate

$$\frac{(O_1 - E_1)^2}{E_1} \text{ and } \frac{(O_2 - E_2)^2}{E_2}$$



Comparing Survival Curves

Example

- Patients with hypernephroma
  - Nephrectomy 77* 18 68 35 8 (days)
  - No Nephrectomy 8 17* 6 12 21
- Make a contingency table for each event time

Time 6

	Death	No Death	
Nephr	0	5	5
No Nephr	1	4	5
	1	9	10

$$E_N = \frac{5(1)}{10} = 0.5$$

$$E_{NN} = \frac{5(1)}{10} = 0.5$$

Comparing Survival Curves

Example

Nephr	8	18	35	68	77*
No Nephr	6	8	12	17*	21

Time 8

	Death	No Death	
Nephr	1	4	5
No Nephr	1	3	4
	2	7	9

$$E_N = \frac{5(2)}{9} = \frac{10}{9}$$

$$E_{NN} = \frac{4(2)}{9} = \frac{8}{9}$$

Time 12

	Death	No Death	
Nephr	0	4	4
No Nephr	1	2	3
	1	6	7

$$E_N = \frac{4(1)}{7} = \frac{4}{7}$$

$$E_{NN} = \frac{3(1)}{7} = \frac{3}{7}$$

Comparing Survival Curves

Example

Nephr	8	18	35	68	77*
No Nephr	6	8	12	17*	21

Time 18

	Death	No Death	
Nephr	1	3	4
No Nephr	0	1	1
	1	4	5

$$E_N = \frac{4(1)}{5} = \frac{4}{5}$$

$$E_{NN} = \frac{1(1)}{5} = \frac{1}{5}$$

Time 21

	Death	No Death	
Nephr	0	3	3
No Nephr	1	0	1
	1	3	4

$$E_N = \frac{3(1)}{4} = \frac{3}{4}$$

$$E_{NN} = \frac{1(1)}{4} = \frac{1}{4}$$

Comparing Survival Curves

Example

Nephr	8	18	35	68	77*
No Nephr	6	8	12	17*	21

Time 35

	Death	No Death	
Nephr	1	2	3
No Nephr	0	0	0
	1	2	3

$$E_N = \frac{3(1)}{3} = 1$$

$$E_{NN} = 0$$

Time 68

	Death	No Death	
Nephr	1	1	2
No Nephr	0	0	0
	1	1	2

$$E_N = \frac{2(1)}{2} = 1$$

$$E_{NN} = 0$$

Comparing Survival Curves

Example

$$\mathrm{Var}(a_i) = \frac{(a+c)(b+d)(a+b)(c+d)}{N^2(N-1)}$$

	Nephr		No Nephr	
Time	a_i	$E(a_i)$	c_i	$E(c_i)$
6	0	0.5	1	0.5
8	1	1.11	1	0.89
12	0	0.57	1	0.43
18	1	0.80	0	0.20
21	0	0.75	1	0.25
35	1	1.00	0	0.00
68	1	1.00	0	0.00
	4	5.73	4	2.27

Nephr $\mathrm{Var}(a_i)$
0.25
0.43
0.24
0.16
0.19
0
0
1.27

Comparing Survival Curves
Example

- Strength of Association

$$HR = \frac{O_1/E_1}{O_2/E_2} = \frac{4/5.73}{4/2.27} = 0.48$$

  - Subjects in the nephrectomy group have a 52% less chance of dying at the next time point compared to non-nephrectomy group

  - Another interpretation: calculate HR/(1+HR)=0.48/1.48=0.32
    - For a randomly selected neph patient and a randomly selected non-neph patient, the probability of the neph patient dying first is 0.32

Comparing Survival Curves
Example

- Approximation

$$T = \frac{(4 - 5.73)^2}{5.73} + \frac{(4 - 2.27)^2}{2.27} = 1.84$$

- Mantel-Haenszel Log Rank

$$\chi^2 = \frac{[\sum a_i - \sum E(a_i)]^2}{\sum Var(a_i)} = \frac{(4 - 5.73)^2}{1.27} = 2.36$$

- Compare to $\chi^2_{\alpha,\ p\text{-}1\ d.f.}$ ($\chi^2_{0.05,1\ d.f.}$=3.84)

Comparing Survival Curves
Log Rank

- Properties
  - Optimal if failure rate for one group is constant multiple of failure rate for another group
  - Can also compare >2 groups
  - Can also adjust for a confounder (stratified log rank)

AN R EXAMPLE OF SURVIVAL ANALYSIS

Cox Regression

- Model the hazard rate (conditional rate)
  - P(event at t | no event to time t)
    - Instantaneous risk –event rate ***at t*** for those at risk
    - Example. Event is seizure, event time is time since withdrawal of seizure meds measured in months
      - A hazard of 0.02 at t=24 mos means that at 2-years, patients are having seizures at a rate of 2% per month; or at 2-years the chance of having a seizure in 1 month is 2%

Cox Regression

- Hazard for an individual at time t is the product of:
  - Baseline hazard function
  - Linear function of the independent predictors

$$h(t) = h_0(t)e^{(\beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k)}$$

$$\ln h(t) = \ln h_0(t) + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k$$

↑

Hazard rate at time t for individual with covariates=0

Cox Regression

- Interpretation
  - 0/1 indicator for gender
    - x_1=0=M / x_1=1=F

$$\ln h(t)_F = \ln h_0(t) + \beta_1(1) = \ln h_0(t) + \beta_1$$
$$\ln h(t)_M = \ln h_0(t) + \beta_1(0) = \ln h_0(t)$$

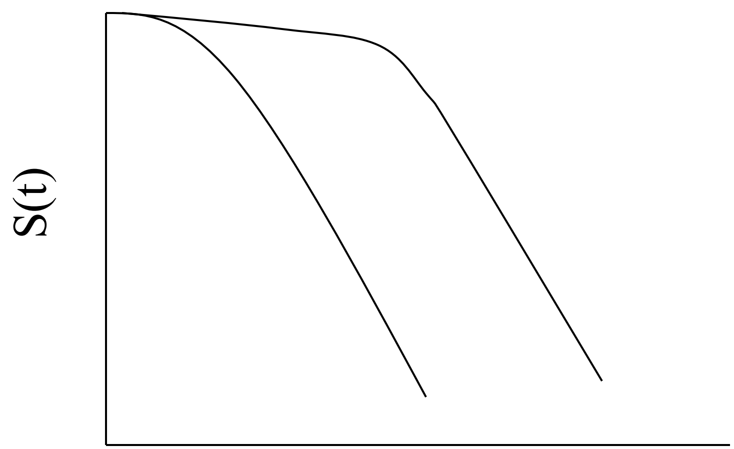
$$\frac{h(t)_F}{h(t)_M} = \frac{h_0(t)e^{\beta_1}}{h_0(t)} = e^{\beta_1}$$

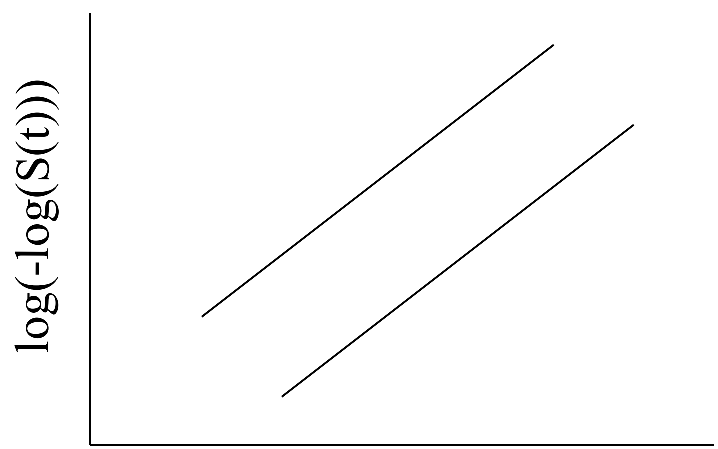
Cox Regression

- Assumption - Hazards are proportional

$$H(t;\ x=1) = C \times H(t;\ x=0)$$

$$\ln H(x=1) = \ln C + \ln H(x=0)$$





Cox Regression

- Confidence interval for HR

$$e^{\beta_{\mathrm{k}} \pm z_{\alpha/2}(s.e.\beta_k)}$$

Cox Regression

- Parameters (β) are estimated using *Partial Likelihood*
- Hypothesis testing:
  - H_0: $\beta_1 = 0$ ($HR_1 = 1$) vs. H_a: $\beta_1 \neq 0$ ($HR_1 \neq 1$)
- test statistic is *Wald test*:

  - Under H_0, the ratio $\sim N(0,1)$

Wald Test for β

$$W = \frac{\hat{\beta}_j}{\hat{SE}(\hat{\beta}_j)} \stackrel{\cdot}{\sim} N(0,1) \text{ under } H_0$$

  - $W^2 \sim \chi^2_1$: The Wald chi-square is often reported by software packages (including SAS)

Cox Regression

Leukemia Example: Data

Placebo			6-MP Chemotherapy		
Survival Time	δ	logWBC	Survival Time	δ	logWBC
1	1	2.8	6	1	2.31
1	1	5	6	1	4.06
2	1	4.91	6	1	3.28
2	1	4.48	6	0	3.2
3	1	4.01	7	1	4.43
4	1	4.36	9	0	2.8
4	1	2.42	10	1	2.96
5	1	3.49	10	0	2.7
5	1	3.97	11	0	2.6
8	1	3.52	13	1	2.88
8	1	3.05	16	1	3.6
8	1	2.32	17	0	2.16
8	1	3.26	19	0	2.05
11	1	3.49	20	0	2.01
11	1	2.12	22	1	2.32
12	1	1.5	23	1	2.57
12	1	3.06	25	0	1.78
15	1	2.3	32	0	2.2
17	1	2.95	32	0	2.53
22	1	2.73	34	0	1.47
23	1	1.97	35	0	1.45

- Survival time = Weeks until go out of remission
- Covariates:
  - X_1: Group Status (treatment or placebo)
  - X_2: logWBC

Cox Regression

SAS: Leukemia Example

Table: Model 1

Effect	$\hat{\beta}$	$\hat{SE}(\hat{\beta})$	W^2	Wald P-value	$\hat{HR}$	95% CI
Group	1.509	0.410	13.58	0.0002	4.523	(2.027, 10.094)

$-2l = 172.759$

Table: Model 2

Effect	$\hat{\beta}$	$\hat{SE}(\hat{\beta})$	W^2	Wald P-value	$\hat{HR}$	95% CI
Group	1.294	0.422	9.399	0.0022	3.648	(1.595, 8.342)
logWBC	1.604	0.329	23.732	$<.0001$	4.974	(2.609, 9.486)

$-2l = 144.559$

Table: Model 3

Effect	$\hat{\beta}$	$\hat{SE}(\hat{\beta})$	W^2	Wald P-value	$\hat{HR}$	95% CI
Group	2.355	1.681	1.963	0.1612	.	.
logWBC	1.803	0.447	16.286	$<.0001$	.	.
Group× logWBC	-0.342	0.520	0.434	0.5103	.	.

$-2l = 144.131$

AN R EXAMPLE OF COX REGRESSION