Lecture 14

Survival Data Analysis

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

In this lecture…

- Survival data and censoring
- Functions for describing survival
- Kaplan-Meier (KM) curves (review)
- Log-rank test (review)
- Hazard function
- Proportional hazards model
- Cox model
- The Cox model in STATA (STCOX)

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Survival data examples

- Years to death post surgery
- Weeks to relapse post rehab
- Minutes to infection post exposure
- Days to full recovery post surgery

Different types of outcome with different effective time-scales

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Survival data have special features

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Example

Survival (in days) of patients with lymphoma

6, 19, 32, 42, 42, 43*, 94, 126, 169*, 207, 211*, 227, 253, 255*, 270*, 310*, 316*, 335*, 346*

… so what makes survival data special?

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Example

Survival (in days) of patients with lymphoma

6, 19, 32, 42, 42, 43*, 94, 126, 169*, 207, 211*, 227, 253, 255*, 270*, 310*, 316*, 335*, 346*

… so what makes survival data special?

Answer: * = still alive at end of follow up

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Right Censoring

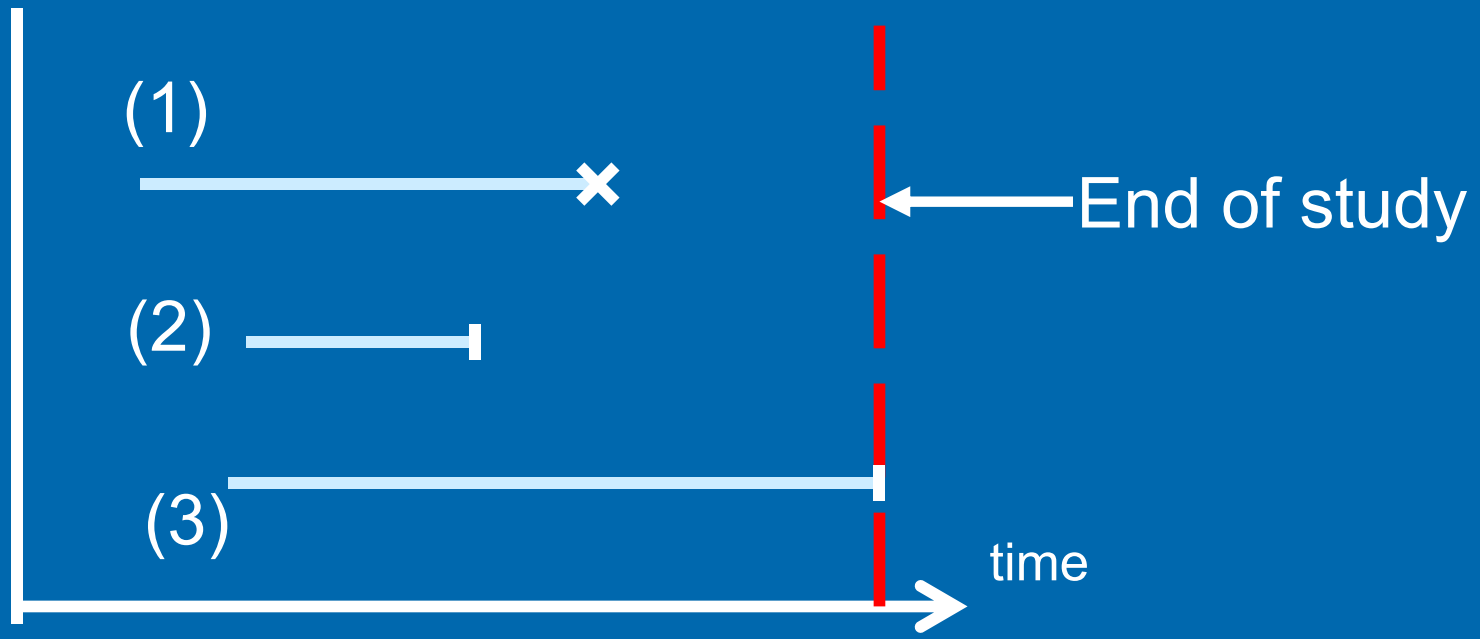
Definition:

A survival time is "right censored at time t" if we only know know that it is greater than t

Example: A subject is followed for 18 months. Follow up ends and the subject is alive. The subject is then *right censored* at 18 months

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Right censoring – not everyone is followed to death



(1) Subject followed to death (not censored)
(2) Subject right censored by loss to follow up
(3) Administratively censored: end of study

Assume censoring is independent of treatment (non-informative, independent censoring)

Pediatric Kidney Transplant

Example of Survival Data

- United Network for Organ Sharing (UNOS) database
- 9750 children under 18 yrs with kidney tx,1990-2002
- Outcome: time to death post transplant
- *38,000 patient years, 429 deaths*
- What are predictors of post tx mortality? *e.g., donor source: cadaveric v. living*

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Features of UNOS Data

- Risk of death depends on length of follow-up (an individual has more chance of dying during the study if followed for 10 vs. 5 yrs)
- Follow-up ranges from 1 day to 12.5 yrs
- Most children are alive at end of study *438 of 9750 subjects have events (4.5%)*
- Thus, 95.5% of subjects are *right censored*

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

UNOS Data

$Y_i = \texttt{time}$ years since transplant)

$$\delta_i = \texttt{indic} \begin{cases} 1: & \text{Death at } Y_i \\ 0: & \text{Alive as of } Y_i \end{cases}$$

$$X_i = \texttt{txtype} \begin{cases} 1: & \text{Cadaveric} \\ 0: & \text{Living} \end{cases}$$

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Survival Data in STATA

- **Declare the data to be survival data**
- **Use** stset **command**
- stset **Y**, failure(δ)
  UNOS data:
  stset time, failure(indic)
- **In** STATA, **code δ carefully**
  1 should be event, 0 a censored observation read stset **output summary to check**

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

How can we analyze survival data?

What happens if we try regression methods from earlier lecture?

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Treat as Continuous?

Question: What is the mean survival time?

- **Outcome: Time to Death**
- **Problem:**

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Treat as Continuous?:

Question: What is the mean survival time?

- **Outcome:**
  - **Time to Death**
- **Problem:**
  - **Most subjects still alive at study end**
- **Average time of death can be highly misleading**
  - **Example:**
    - 100 subjects censored at 500 months and only 2 deaths at 1 and 3 months
    - Average death time: 2 months....misleading!

Mean time to death is not generally a useful summary for survival data!

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Treat as Binary

Question: what proportion of subjects survive?

Proportion of subjects "alive"

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Treat as Binary

Question: what proportion of subjects survive?

Proportion of subjects "alive"

Requires an arbitrary cut-off time (e.g., 1 yr)

What if a subject is censored at 360 days?
their data is wasted: not followed for a year

Deaths after 1 year are ignored, *but most deaths may occur after 1 year!*

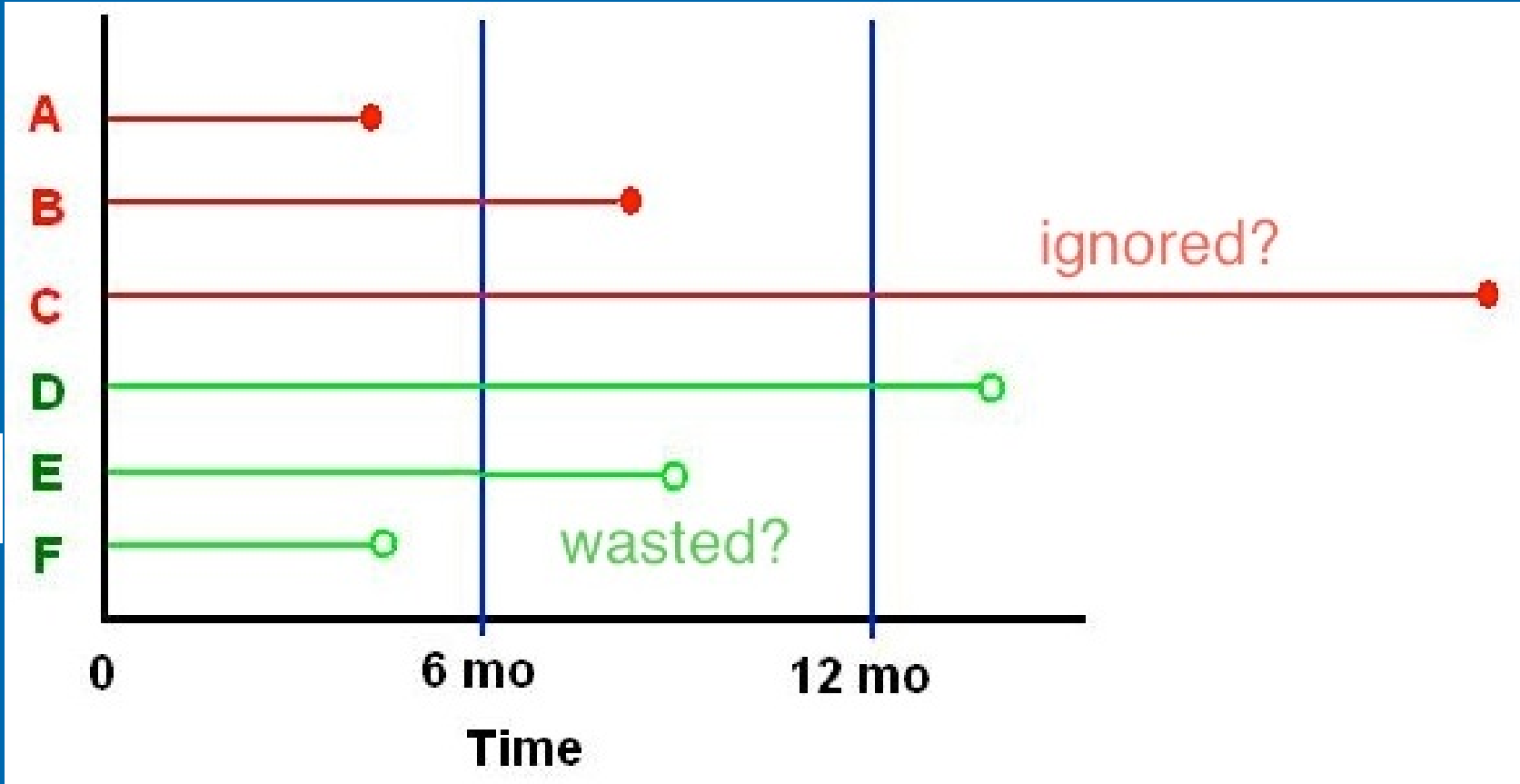
sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Treat as Binary

e.g., cut-off at 6 months or 1 year?

observed

censored



Right *and* left censoring with no mechanism to deal with it!

Aims of a Survival Analysis

- Summarize the distribution of survival times
  - Tool: Kaplan-Meier estimates (of survival distribution)
- Compare survival distributions between groups

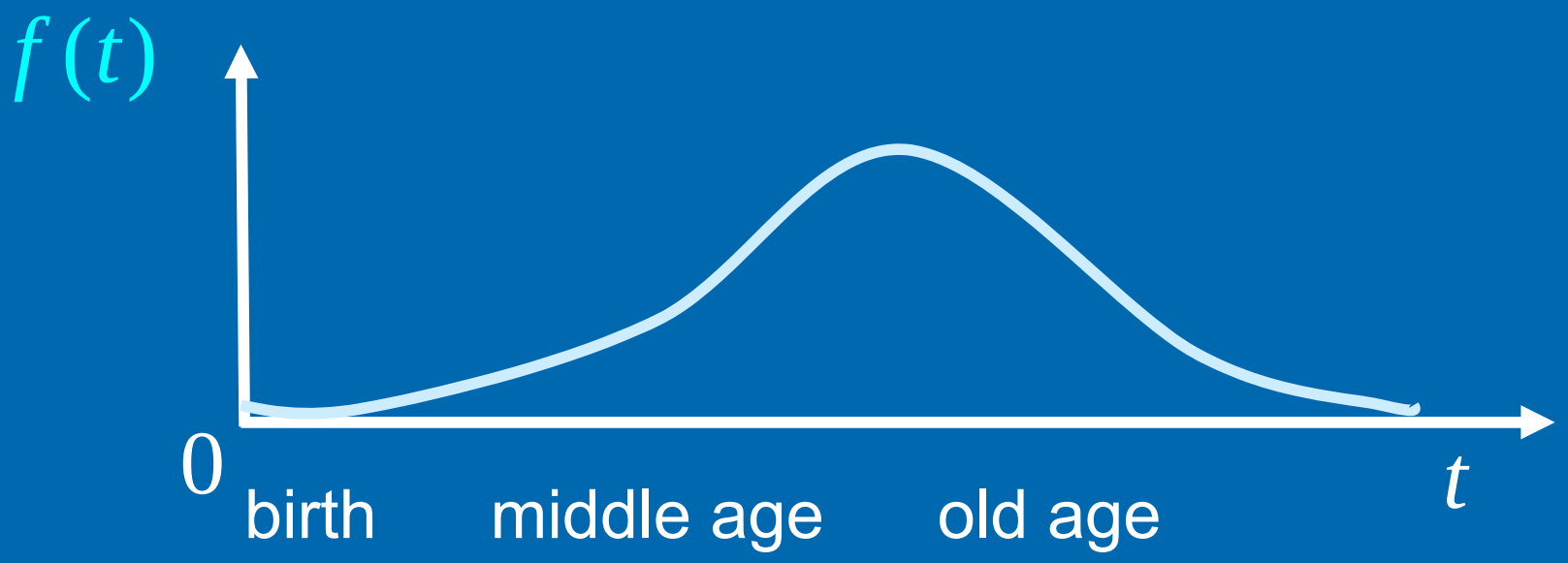
  Tool: logrank test
- Investigate predictors of survival
  - Tool: Cox regression model

Kaplan-Meier and logrank covered previously but will review

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Functions for Describing Survival

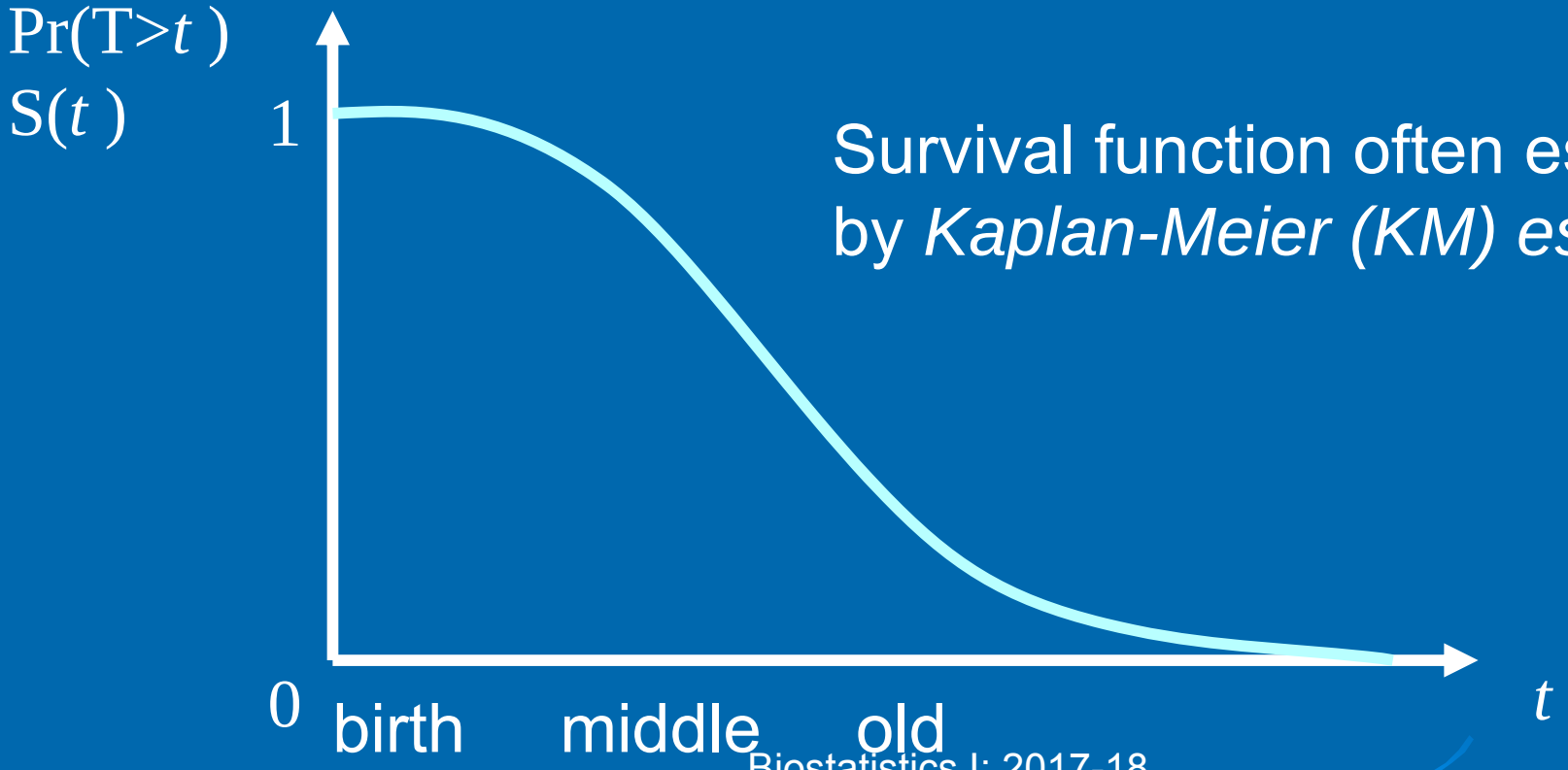
Probability density (distribution) - of death



sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Functions for Describing Survival

Survival function: the probability an individual will survive longer than a particular time t



Survival function often estimated by *Kaplan-Meier (KM) estimator*

sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Introduction to Kaplan-Meier

Non-parametric estimate of the survival function:

Simply, the empirical probability of surviving past certain times in the sample (taking into account censoring).

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Introduction to Kaplan-Meier

- Non-parametric estimate of the survival function.
- Commonly used to describe survivorship of study population/s.
- Commonly used to compare two study populations.
- Intuitive graphical presentation.

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Estimating the Survival function with the Kaplan-Meier (KM) estimator

The probability of death in any particular time interval can be estimated by:

$$\frac{\#\text{ observed deaths}}{\#\text{ at risk}}$$

E.g. in a study of 2000 genetically bred mice, 230 died of heart failure between 3 and 6 weeks; estimate probability of death between 3 and 6 weeks to be 230/2000

Kaplan-Meier

The survival curve is high at the start (because everyone is alive early on) and then the KM approach assumes that at the times we observe deaths in our dataset the survival curve should drop.

To generate a KM estimated survival curve we consider intervals of "zero" length in time:

$$\Pr(\text{death at time } t) = \frac{\#\text{ observed deaths at time } t}{\#\text{ at risk at time } t}$$

This leads to an instant drop in the survival curve of size "Pr(death at time *t*)" every time there is a death.

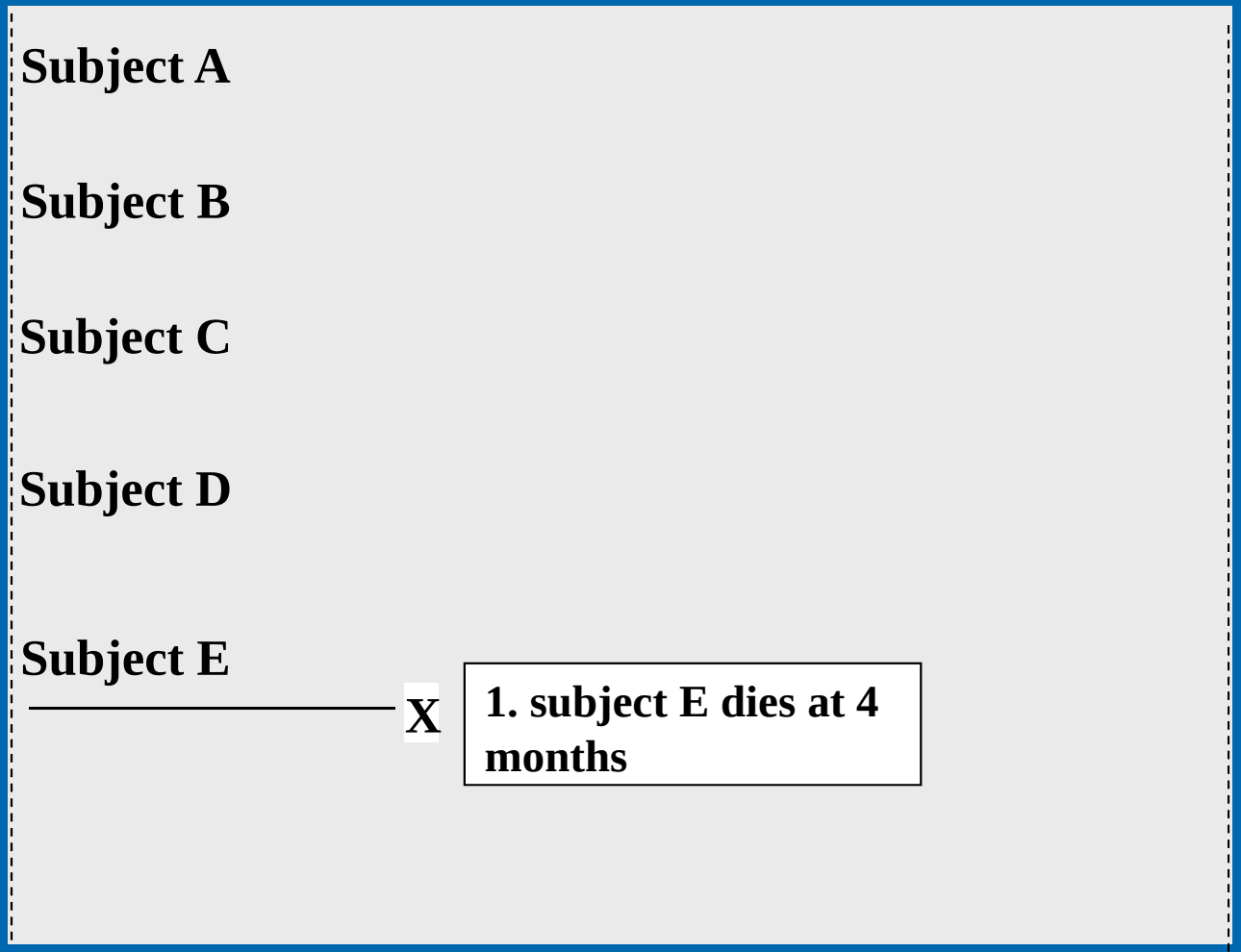
sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Kaplan-Meier

Note that the “number at risk” in the denominator accounts for the censored data! Once a subject is censored they are no longer in the “at risk” group.

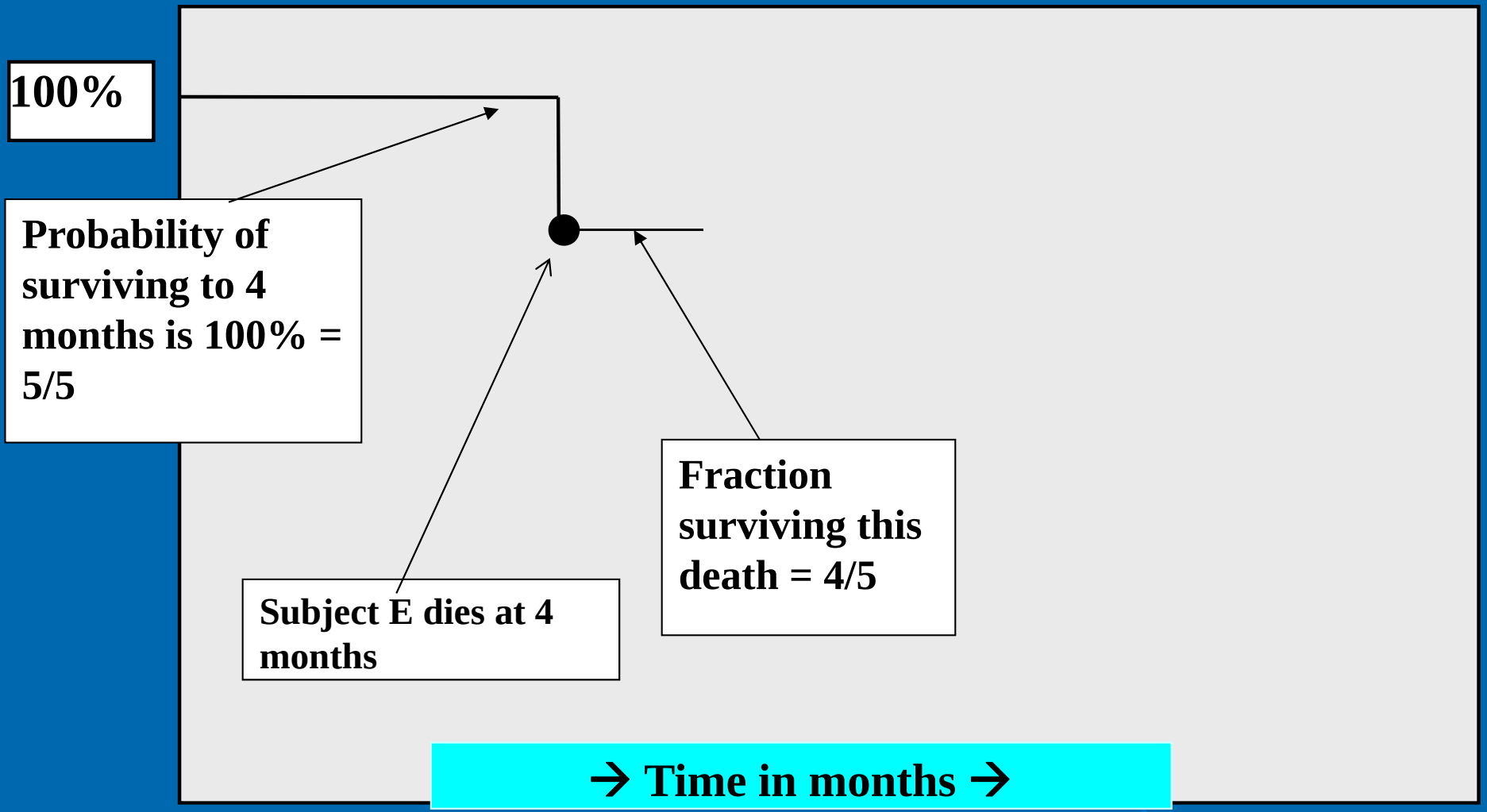
The subsequent heights of drops for each death in the KM survival curve increases with each individual lost to follow up (and each death).

Survival Data (right-censored)



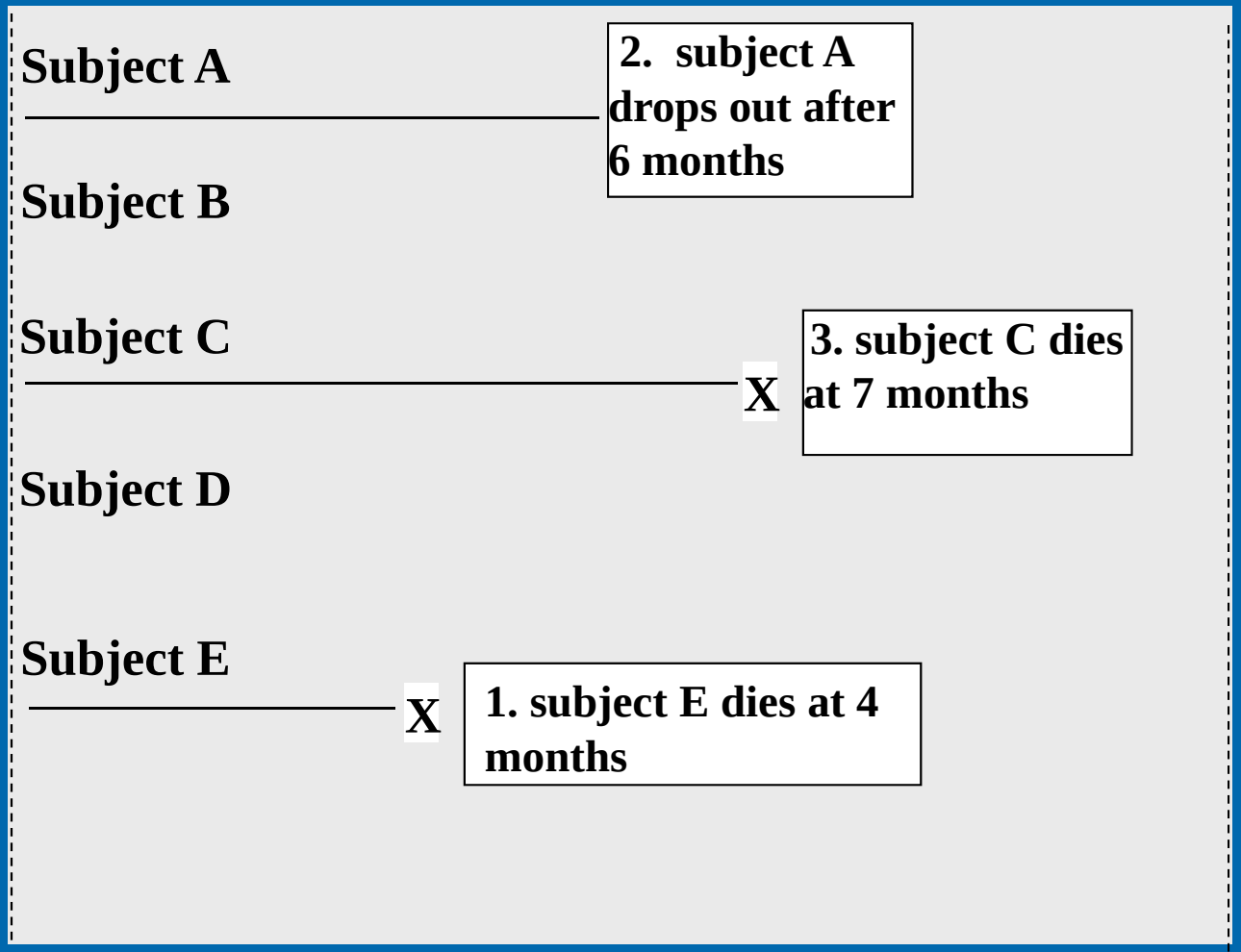
sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Corresponding Kaplan-Meier Curve



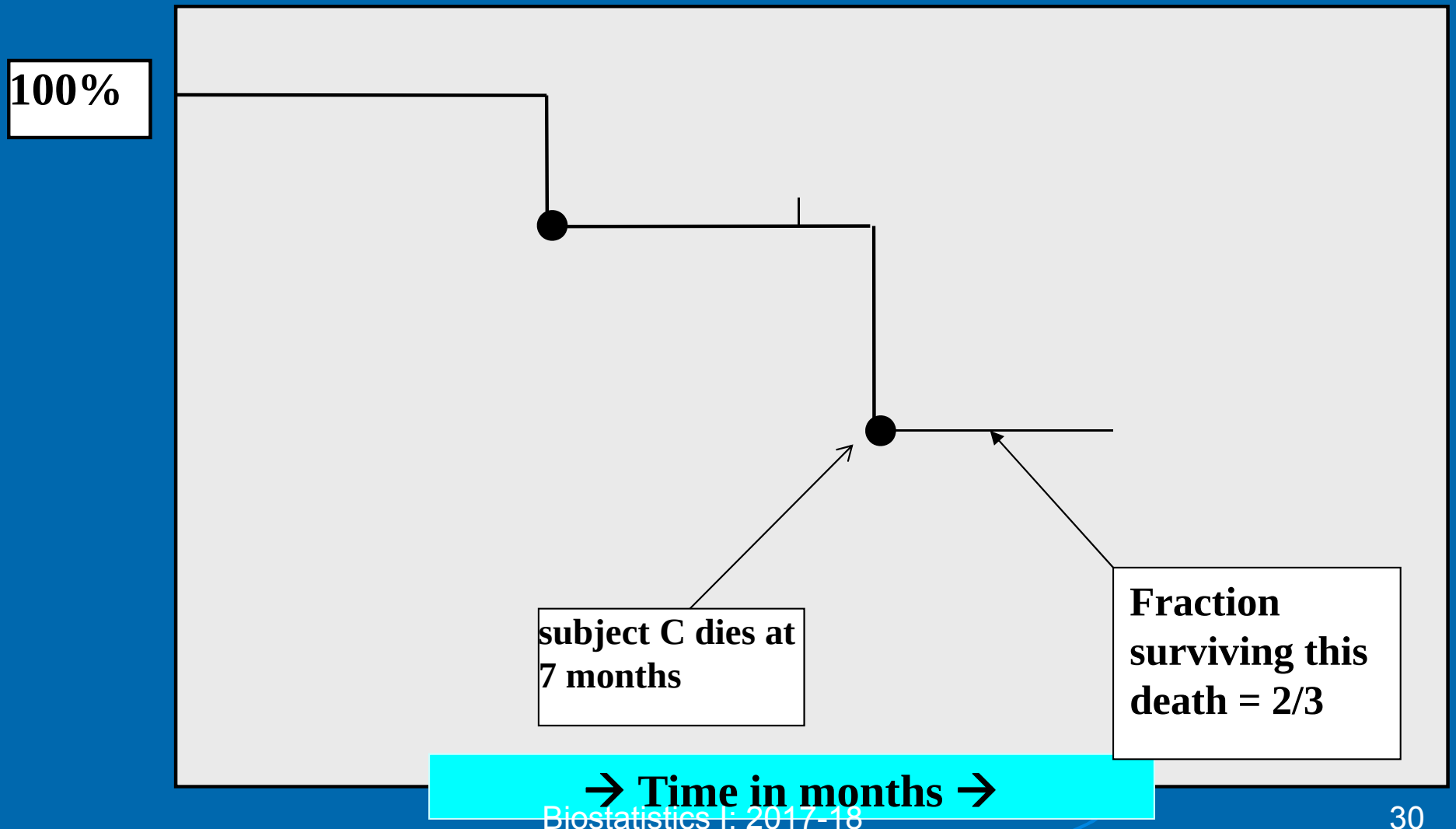
sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Survival Data

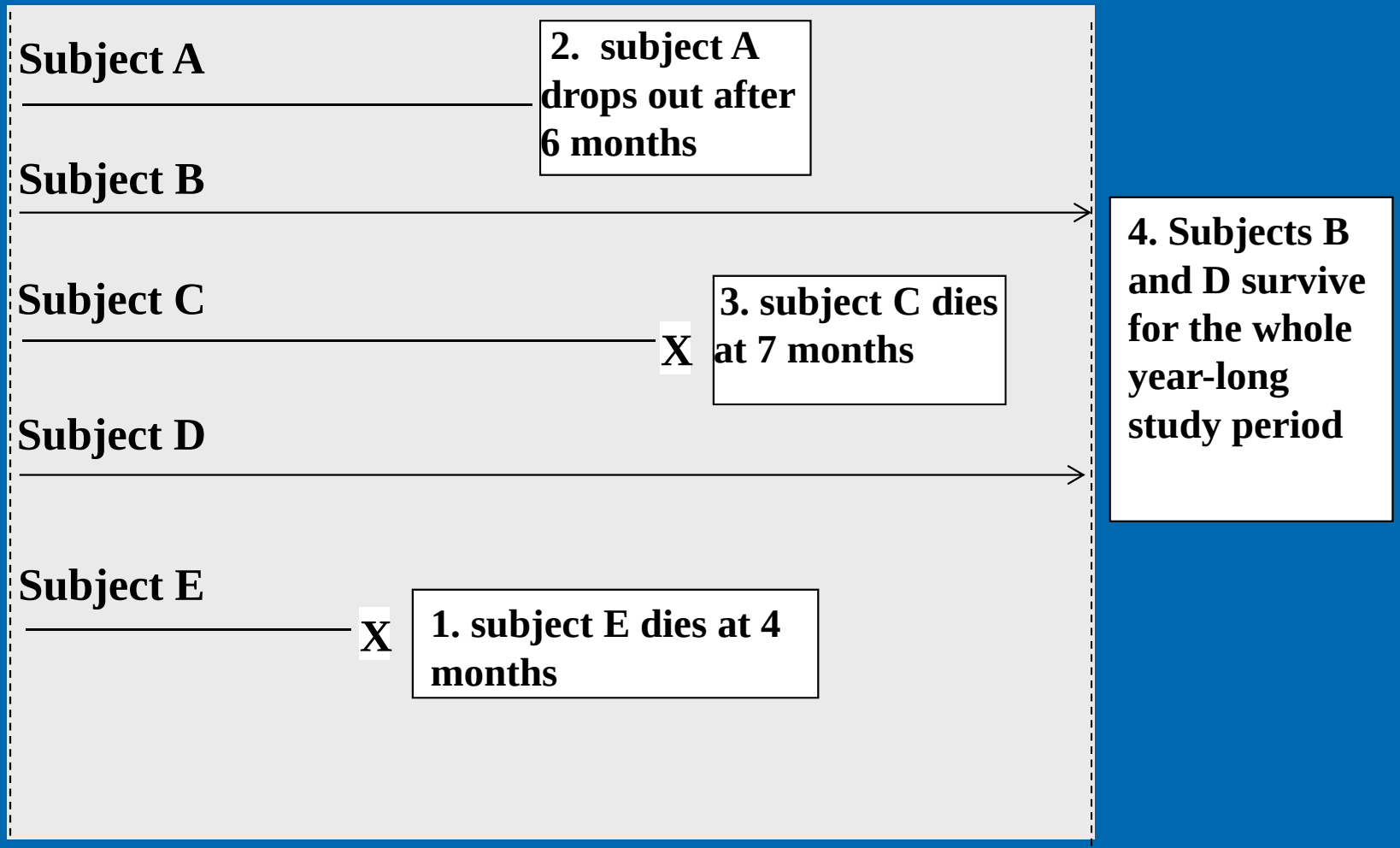


sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Corresponding Kaplan-Meier Curve



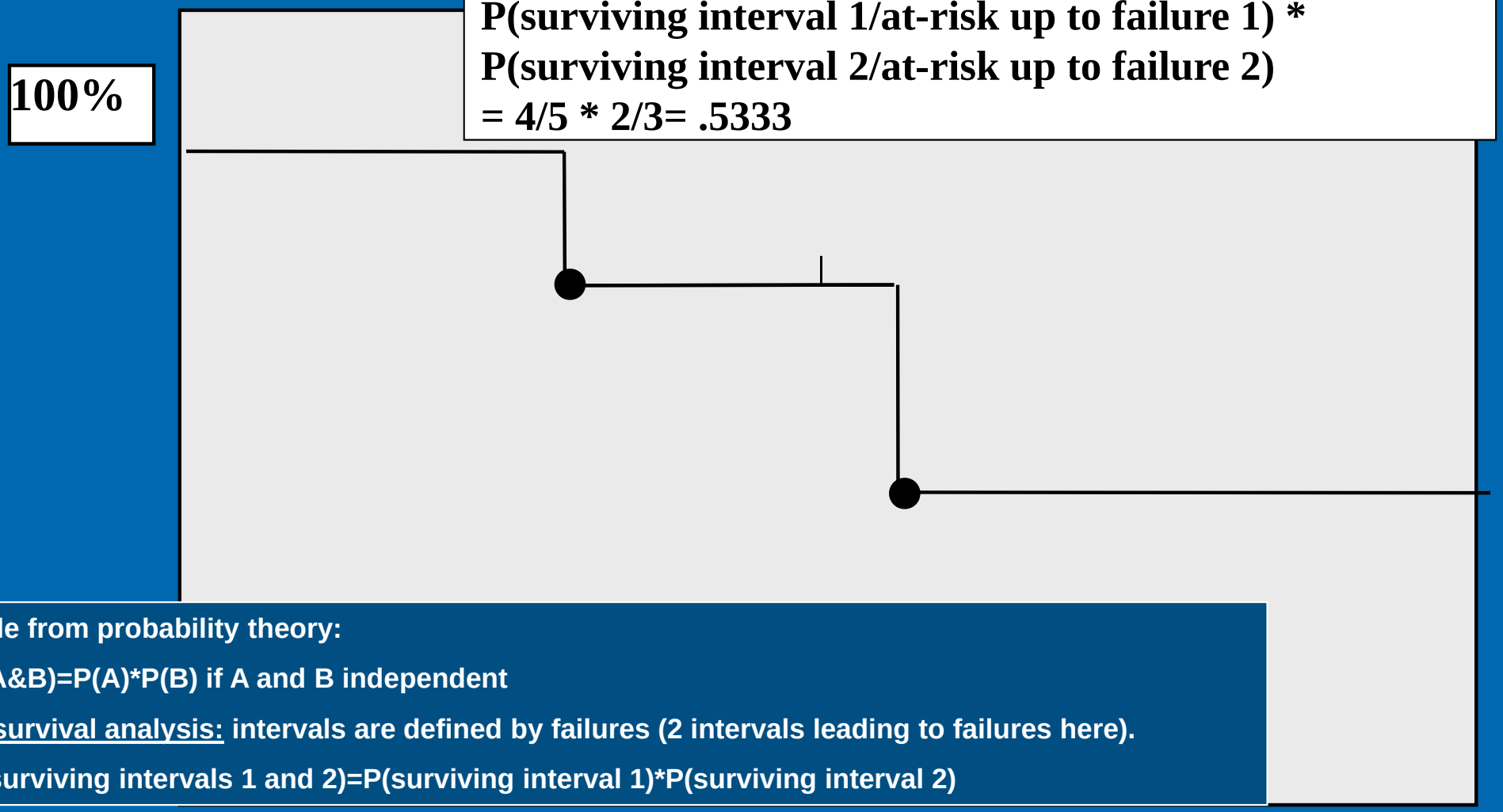
Survival Data



sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Corresponding Kaplan-Meier Curve

∴ Product limit estimate of survival =
P(surviving interval 1/at-risk up to failure 1) *
P(surviving interval 2/at-risk up to failure 2)
= 4/5 * 2/3= .5333



Rule from probability theory:

P(A&B)=P(A)*P(B) if A and B independent

<u>In survival analysis:</u> intervals are defined by failures (2 intervals leading to failures here).

P(surviving intervals 1 and 2)=P(surviving interval 1)*P(surviving interval 2)

The product limit estimate

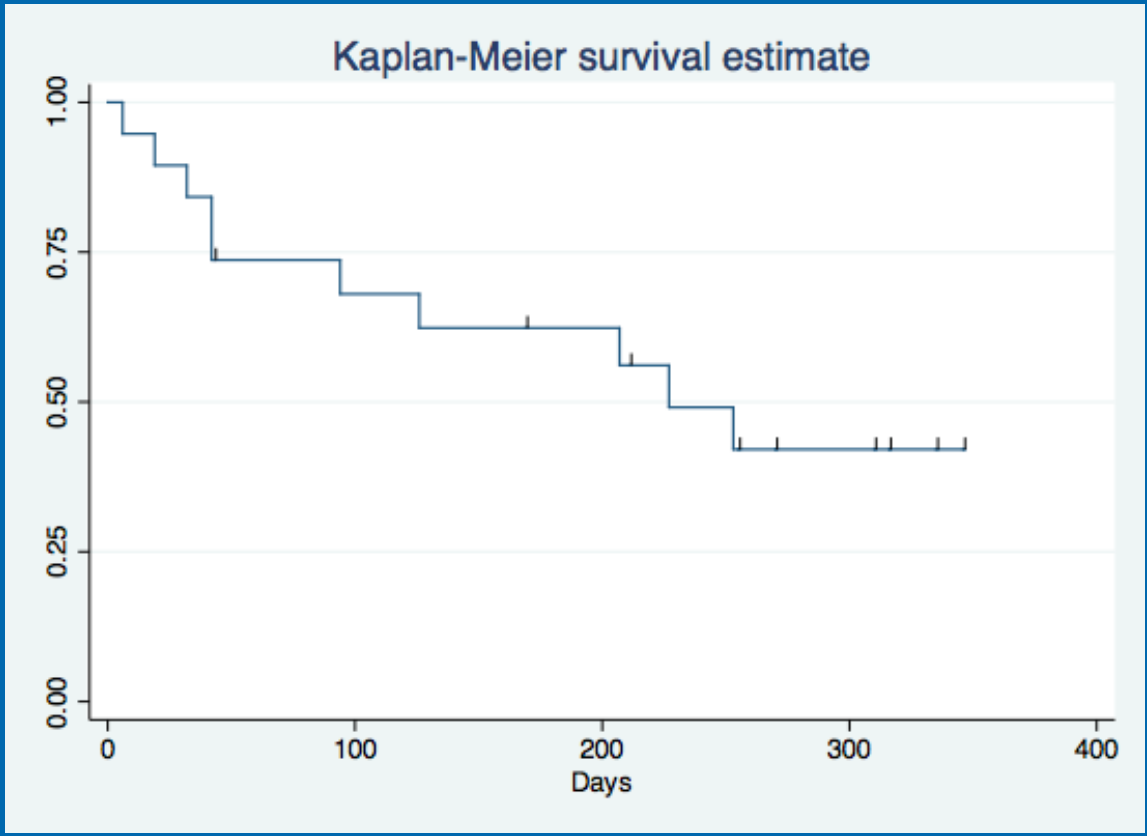
- The probability of surviving in the entire year, taking into account censoring = (4/5) (2/3) = 53%
- NOTE: > 40% (2/5) because the one drop-out survived at least a portion of the year.
- AND <60% (3/5) because we don't know if the one drop-out would have survived until the end of the year.

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Kaplan-Meier

Lymphoma data

Days	Beg. Total	Fail	Net Lost	Survivor Function
6	19	1	0	0.9474
19	18	1	0	0.8947
32	17	1	0	0.8421
42	16	2	0	0.7368
43	14	0	1	0.7368
94	13	1	0	0.6802
126	12	1	0	0.6235
169	11	0	1	0.6235
207	10	1	0	0.5611
211	9	0	1	0.5611
227	8	1	0	0.4910
253	7	1	0	0.4209
255	6	0	1	0.4209
270	5	0	1	0.4209
310	4	0	1	0.4209
316	3	0	1	0.4209
335	2	0	1	0.4209
346	1	0	1	0.4209



sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Logrank test for comparing Two Survival Curves

Logrank test compares different KM estimated survival functions (e.g. between groups)

The idea of the test is that the proportion of deaths within each group at any time should not be too different than for the combined data from all groups under the null hypothesis: “All groups have the same survival distribution”.

KM and logrank in STATA

stset

- **command declares outcome as survival** ***doesn't need to be specified again***

sts list, by(txtype)

- ***print Kaplan-Meier by variable txtype***

sts graph, by(txtype)

graph Kaplan-Meier by variable txtype

sts test txtype

calculates logrank test for variable txtype

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Aims of a Survival Analysis

- Summarize the distribution of survival times
  - Tool: Kaplan-Meier estimates (of survival distribution)
- Compare survival distributions between groups
  - Tool: logrank test
- Investigate predictors of survival
  - Tool: Cox regression model

Kaplan-Meier and logrank covered previously

Regression by Outcome

Outcome Data	Regression	Data Summary
Continuous	Linear	Mean
Binary	Logistic	Odds
Survival	Cox	Hazard

sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Functions for Describing Survival:

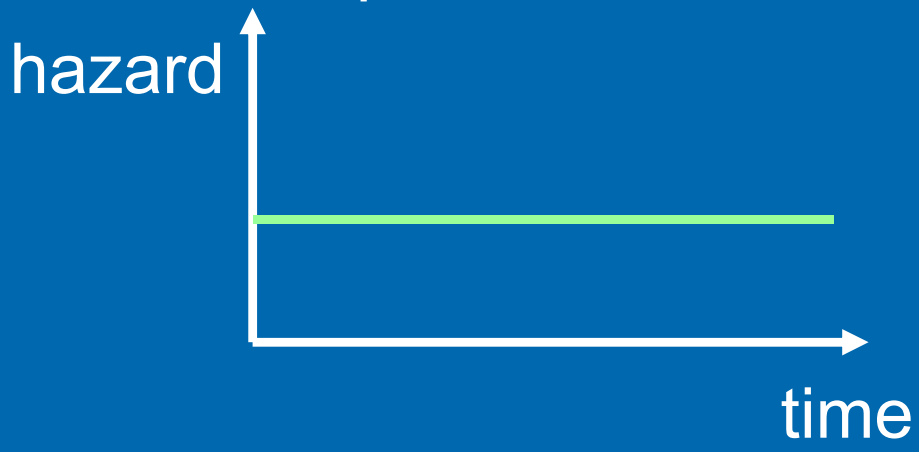
Hazard Function

- Rate of failure per (small) unit time
- Hazard is like an instantaneous (daily) death rate: h(t) = # die at day t / # followed to t
- Rate of death among those alive (at risk)
- Easily estimated for censored data
- A measure of "risk"
  higher hazard => greater risk of death
- Outcome doesn't have to be death

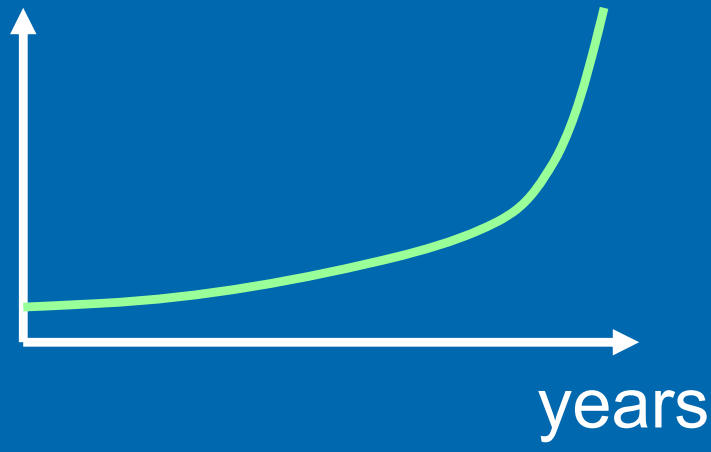
sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Hazard function examples

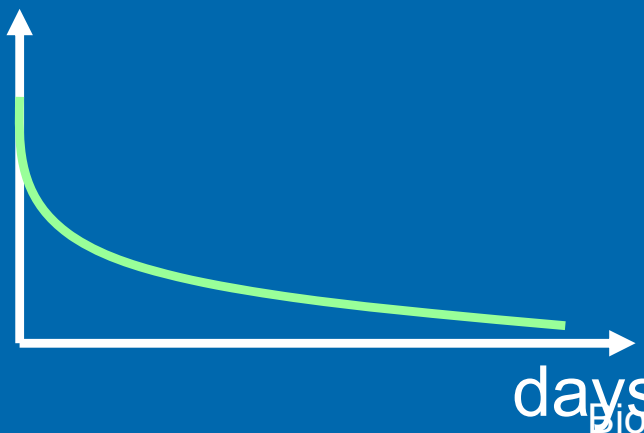
Exponential survival



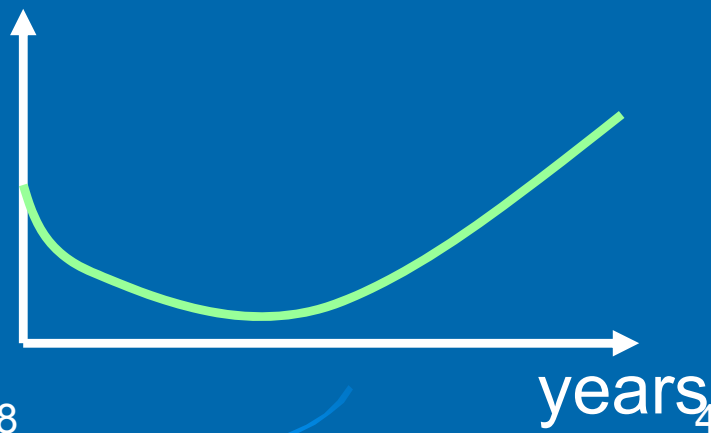
Normal aging



Post heart attack risk



Post heart bypass surgery



UNOS Data: Daily Death Rate

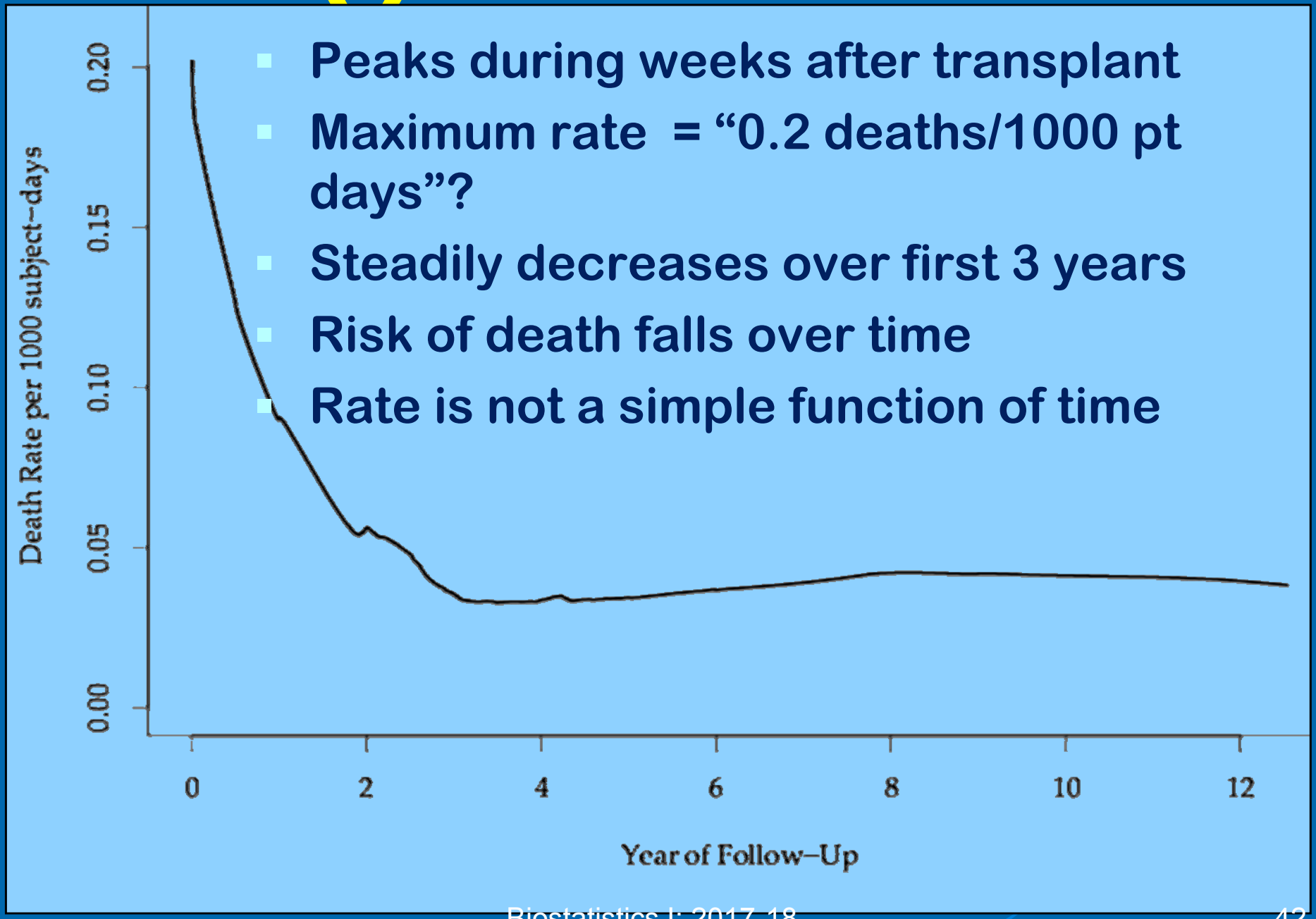
Day of FU	No. Fol.	No. Died	No. Cens.	Death Rate	Death Rate per 1000 Subj-Days
1	9752	7	14	7/9752	0.72
2	9731	5	8	5/9731	0.51
3	9718	5	12	5/9718	0.51
4	9701	7	41	7/9701	0.72
5	9653	3	54	3/9653	0.31
6	9596	2	57	2/9596	0.21
7	9537	0	50	0/9537	0.00
8	9487	4	49	4/9487	0.42
9	9434	1	49	1/9434	0.11
10	9384	3	28	3/9384	0.32

Let's smooth these

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

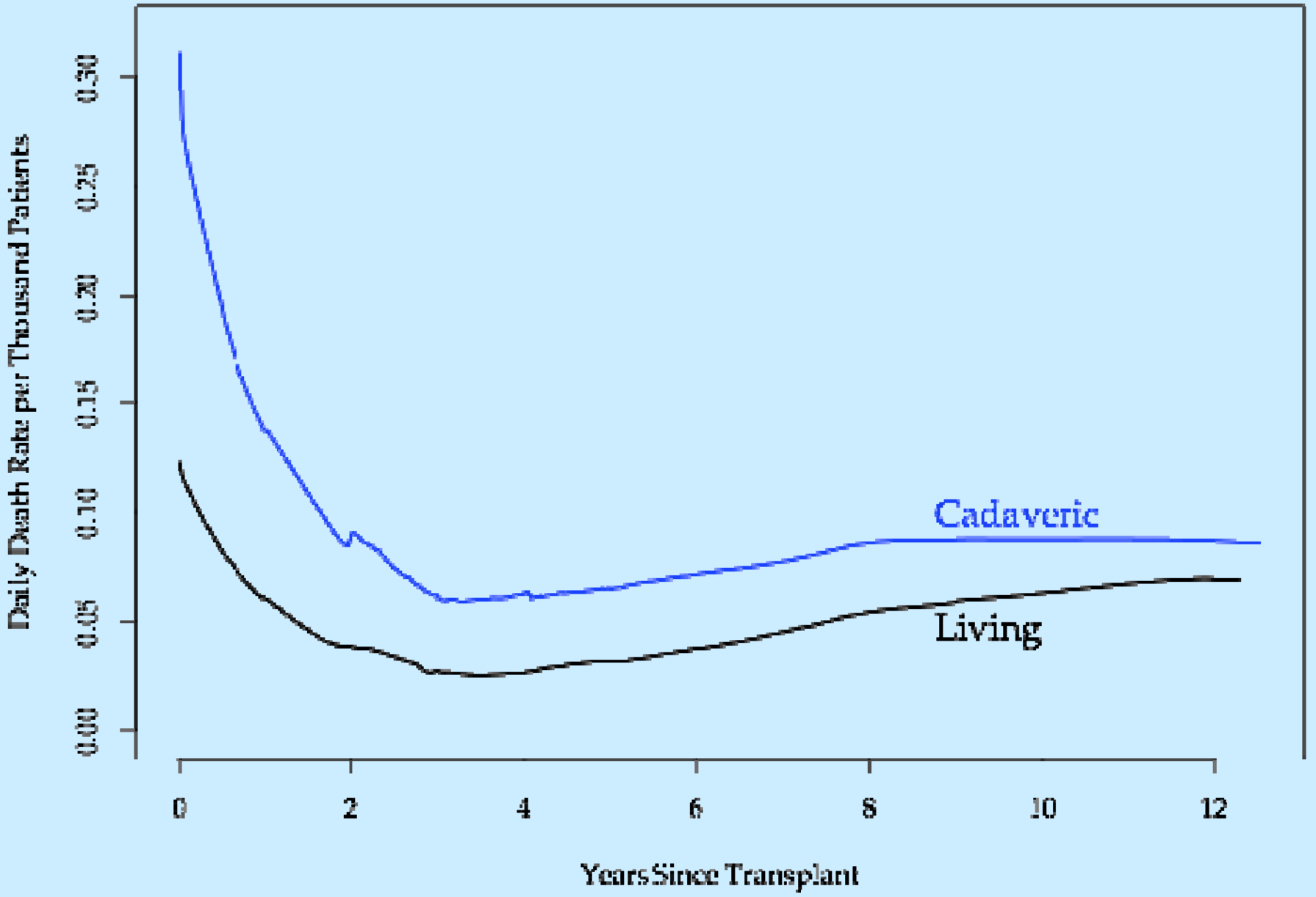
h(t) for UNOS data

- Peaks during weeks after transplant
- Maximum rate = “0.2 deaths/1000 pt days”?
- Steadily decreases over first 3 years
- Risk of death falls over time
- Rate is not a simple function of time



sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Hazard by Type of Tx



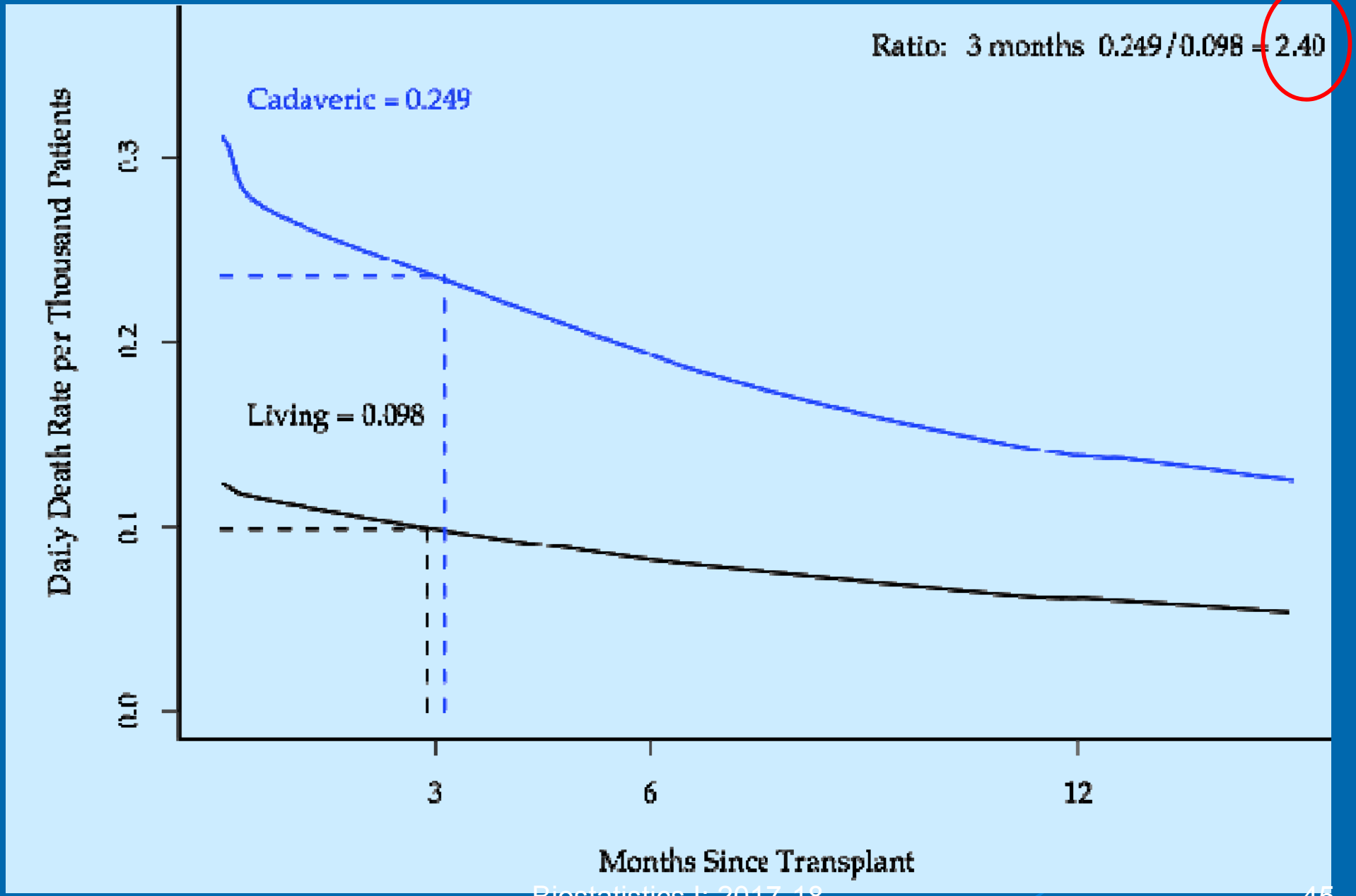
sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Comparing Hazards

- Hazards are between 0 and infinity (c.f. odds)
- Reasonable to divide when comparing (c.f. odds)
- Leads to "hazard ratio"
- Consider the hazard ratio at different times…

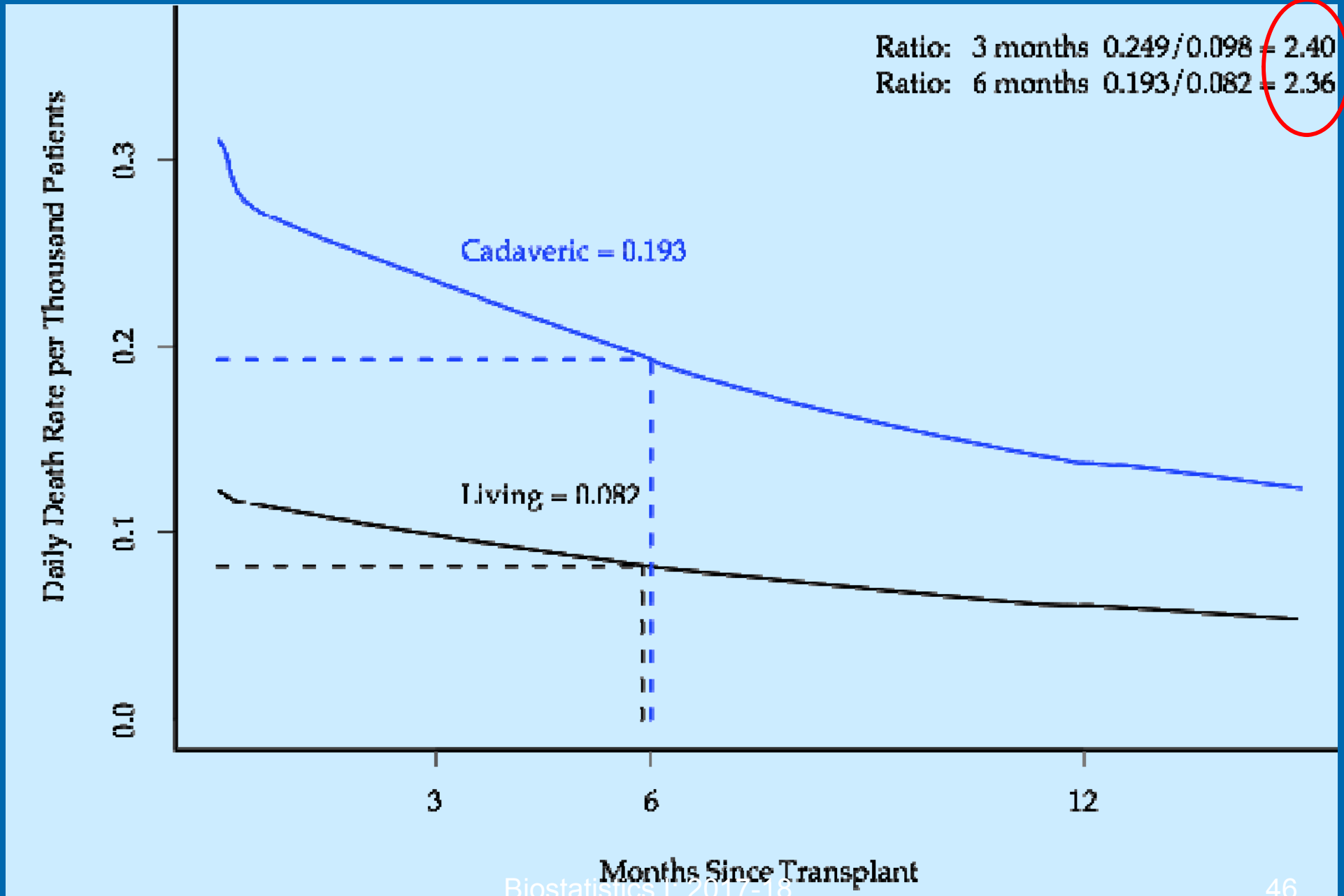
sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Hazard at 3 months



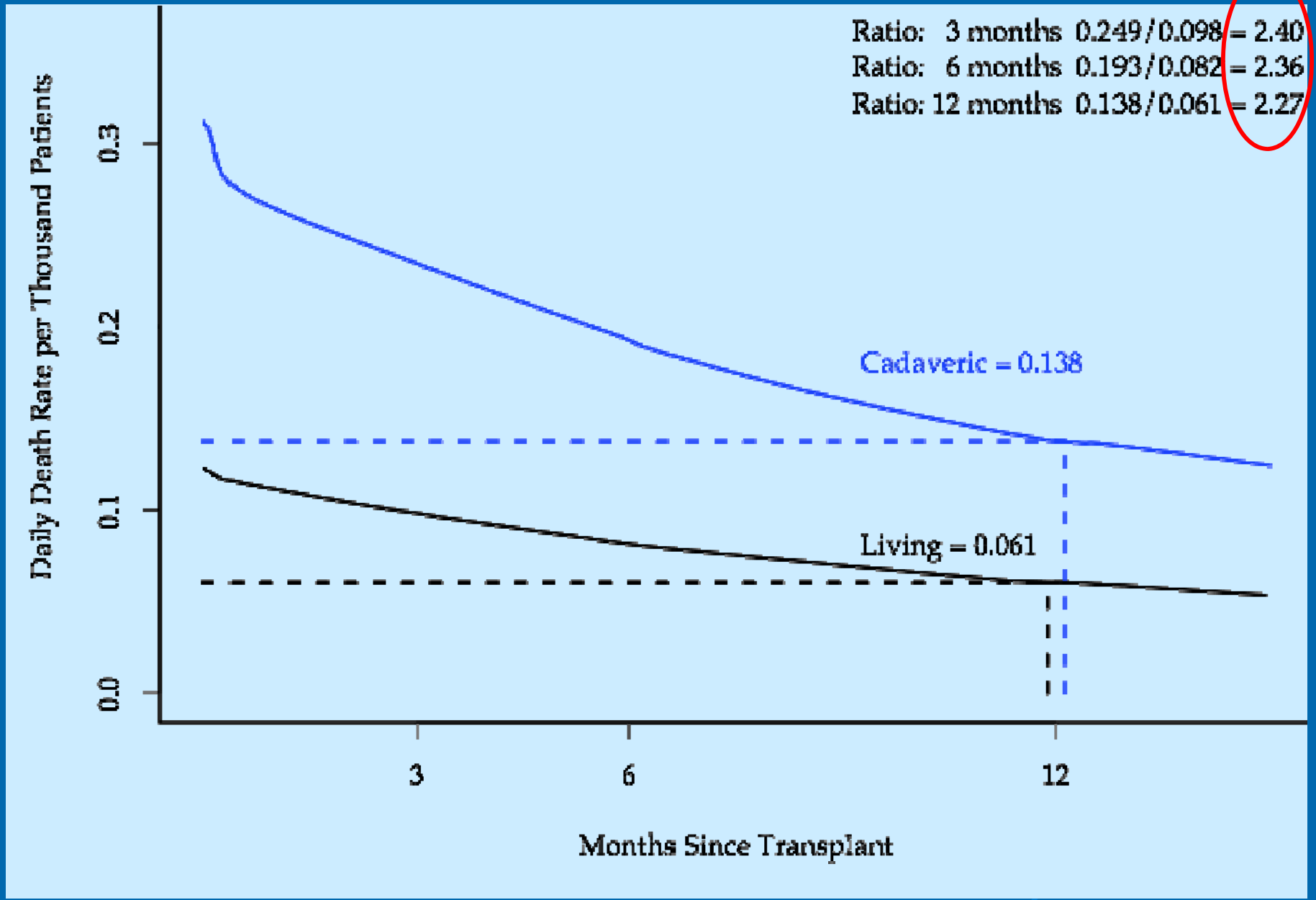
sawilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Hazard at 6 months



sawitopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Hazard at 12 months



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Smoothed Rates

	$1000 \times$ Haz Rate		Relative
Time	Cadaveric	Living	Rate
3mo	0.235	0.098	2.40
6mo	0.193	0.082	2.36
1yr	0.138	0.061	2.27
2yr	0.088	0.038	2.30
3yr	0.061	0.027	2.25
4yr	0.063	0.026	2.37
5yr	0.065	0.032	2.03

i.e., hazard ratio

Estimated hazard ratio differs greatly over time?

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Hazard Ratio

- $h_0(t)$: hazard for living recipients at t (col. 3)
- $h_1(t)$: hazard for cadaveric recipients at t (col. 2)
- $h_1(t)/h_0(t)$: relative short-term risk (col. 4) "**hazard ratio" at time t**
- **If $h_1(t) = r h_0(t)$** for all t, hazards are proportional, i.e., *hazards have the same shape* (proportional curves)
- ***r* is the relative hazard**
- ***r* is a useful quantity** for discussing predictor effects

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Hazard Ratio in UNOS Data

- Hazards changed with time:
  - Cadaveric :0.24 at 3 mos, 0.07 at 5 years
  - Living :0.10 at 3 mos, 0.03 at 5 years
- But, their ratio (estimated hazard ratio) didn't vary much: between 2.03 and 2.40
- "Cadaveric recipients have a little over twice the death rate of living recipients"
- Proportional hazards model: gives predictor effects based on hazards, i.e., regression estimates in terms of hazard ratios.

The Proportional Hazards Model for Survival Data

- Let $x = (x_1, \ldots, x_p)$:
  - the set of predictors
- $h(t|x)$:
  - hazard of someone with predictors x
- $h(t|x) = h_0(t) \exp(\beta_1 x_1 + \ldots + \beta_p x_p)$

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

The Proportional Hazards Model for Survival Data

- $h(t|x)/h_0(t) = \exp(\beta_1 x_1 + \ldots + \beta_p x_p)$
- $\log(h(t|x)) = \log(h_0(t)) + \beta_1 x_1 + \ldots + \beta_p x_p$ because $\log(a/b) = \log(a) - \log(b)$

Formulation is much like logistic regression but change *odds* to *hazards*

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Cox Model

- The "baseline" hazard $h_0(t)$ is unspecified; *plays the role of intercept*
- Predictor effects in terms of hazard ratios; *relative rates of failure*
- **Key point:** don't need to know $h_0(t)$ to understand the effects of predictors
- The effect of one unit increase in predictor x_p is to multiply hazard by $\exp(\beta_p)$ (holding all other predictors constant)….

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

+1 unit change in x_p

$$h(t|x) = h_0(t)\exp(\beta_1 x_1 + \ldots + \beta_p V) \qquad x_p = V$$

versus

$$h(t|x) = h_0(t)\exp(\beta_1 x_1 + \ldots + \beta_p (V+1)) \qquad x_p = V+1$$

$$\text{Ratio} = \frac{h_0(t)\exp(\beta_1 x_1 + \ldots + \beta_p (V+1))}{h_0(t)\exp(\beta_1 x_1 + \ldots + \beta_p V)}$$

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

+1 unit change in x_p

$$\text{Ratio} = \frac{\cancel{h_0(t)}\exp(\beta_1 x_1 + \ldots + \beta_p(V+1))}{\cancel{h_0(t)}\exp(\beta_1 x_1 + \ldots + \beta_p V)}$$

$$= \frac{h_0(t)\exp(\beta_1 x_1 + \ldots + \beta_p (V+1))}{h_0(t)\exp(\beta_1 x_1 + \ldots + \beta_p V)}$$

Ratio does not depend on t !

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

+1 unit change in x_p

$$\text{Ratio} = \frac{h_0(t)\exp(\beta_1 x_1 + \ldots + \beta_p(V+1))}{h_0(t)\exp(\beta_1 x_1 + \ldots + \beta_p V)}$$

$$= \exp(\beta_1 x_1 + \ldots + \beta_p(V+1) - (\beta_1 x_1 + \ldots + \beta_p V))$$

because exp(a)/exp(b) = exp(a-b)

$$= \exp(\beta_p(V+1) - \beta_p V)$$ b/c same other predictors

$$= \exp(\beta_p)$$ b/c $\beta_p V$ terms cancel

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Hazard Ratio

- β is the regression coefficient
  If no effect of a predictor variable then β=0
- HR for a unit increase in a predictor is exp(β)
  If no effect of variable then exp(β)=1
- Can readily switch from summarizing models in terms of coefficients, β, or in terms of hazard ratios, exp(β) -- Hazard ratios work better for interpretation, but math simpler based on coefficients
- A useful way to discuss predictor effects (increases or decreases Hazard by a factor)

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Stata Command

- First use the command stset to declare the data as survival
- Then fit Cox model with

  stcox *predictorlist*

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Stata Output

```
. stcox txtype

No. of subjects =         9750                 Number of obs    =      9750
No. of failures =          438
Time at risk    =  38236.09865
                                               LR chi2(1)       =     49.23
Log likelihood  =   -3762.6976                 Prob > chi2      =    0.0000


------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      txtype |   1.974683    .195328     6.88   0.000     1.626672    2.397149
------------------------------------------------------------------------------
```

(living *x=0; cadaveric x=1)*

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

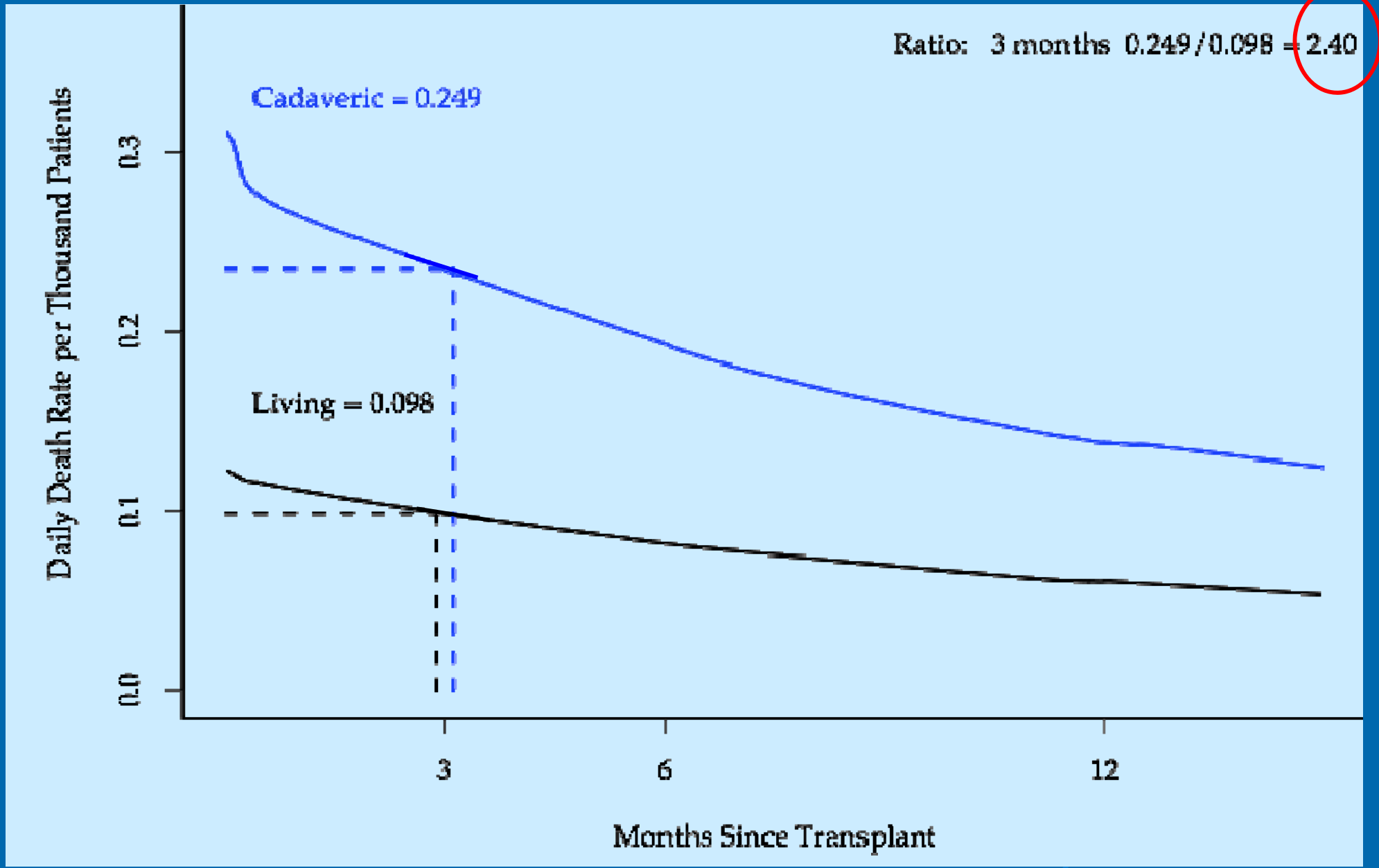
Stata Output

```
. stcox txtype

No. of subjects =         9750                 Number of obs    =      9750
No. of failures =          438
Time at risk    =  38236.09865
                                               LR chi2(1)       =     49.23
Log likelihood  =   -3762.6976                 Prob > chi2      =    0.0000


------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      txtype |   1.974683   .195328     6.88   0.000     1.626672    2.397149
------------------------------------------------------------------------------
```

Estimated hazard ratio: *hazard of death about double for cadaveric recipients*
(living *x=0; cadaveric x=1)*

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Hazard over 12 months



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Stata Output

```
. stcox txtype

No. of subjects =         9750                 Number of obs    =      9750
No. of failures =          438
Time at risk    =  38236.09865
                                               LR chi2(1)       =     49.23
Log likelihood  =   -3762.6976                 Prob > chi2      =    0.0000


------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      txtype |   1.974683   .195328     6.88   0.000     1.626672    2.397149
------------------------------------------------------------------------------
```

SE of hazard ratio

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Stata Output

```
. stcox txtype

No. of subjects =         9750                 Number of obs    =      9750
No. of failures =          438
Time at risk    =  38236.09865
                                               LR chi2(1)       =     49.23
Log likelihood  =   -3762.6976                 Prob > chi2      =    0.0000


------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      txtype |   1.974683   .195328     6.88   0.000     1.626672    2.397149
------------------------------------------------------------------------------
```

Wald test p-value < 0.05: *group hazards are statistically significantly different*

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Stata Output

```
. stcox txtype

No. of subjects =         9750                 Number of obs    =      9750
No. of failures =          438
Time at risk    =  38236.09865
                                               LR chi2(1)       =     49.23
Log likelihood  =   -3762.6976                 Prob > chi2      =    0.0000


------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      txtype |   1.974683    .195328     6.88   0.000     1.626672    2.397149
------------------------------------------------------------------------------
```

Likelihood ratio test p-value < 0.05: *group hazards are statistically significantly different*

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Stata Output

```
. stcox txtype

No. of subjects =         9750                 Number of obs    =      9750
No. of failures =          438
Time at risk    =  38236.09865
                                               LR chi2(1)       =     49.23
Log likelihood  =   -3762.6976                 Prob > chi2      =    0.0000


------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
      txtype |   1.974683    .195328     6.88   0.000     1.626672    2.397149
------------------------------------------------------------------------------
```

95% CI for Hazard ratio: *A plausible range for the hazard ratio is between 1.63 and 2.40.*

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Interpretation

"The hazard ratio of mortality for the recipient of a cadaveric kidney is about 2.0 relative to a living organ ($p < 0.001$). The 95% CI for the hazard ratio is 1.6 to 2.4"

Survival by HLA loci

Number of matching HLA loci range from 0 to 6

```
. stcox i.hla
No. of subjects =         9517                 Number of obs    =      9517
No. of failures =          424
Time at risk    =  37439.62467                 LR chi2(6)       =     47.20
Log likelihood  =   -3629.6294                 Prob > chi2      =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
      hlamat |
           1 |   .9543573   .1697015    -0.26   0.793      .673523    1.352289
           2 |    .697097   .1288643    -1.95   0.051      .485223    1.001486
           3 |   .4644883   .0788633    -4.52   0.000      .333007    .6478825
           4 |   .4249741    .090394    -4.02   0.000     .2800966    .6447883
           5 |   .5870101   .1700124    -1.84   0.066     .3327492    1.035558
           6 |   .3834296   .1396405    -2.63   0.008     .1877968    .7828581
------------------------------------------------------------------------------
```


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Interpretation

HLA is a significant predictor of mortality Chi2=47, $p < 0.0001$ by the likelihood ratio test.

# Loci	HR compared to 0 matching loci	% Change in Hazard of Death
1	0.95	-5%
2	0.70	-30%
3	0.46	-54%
4	0.42	-58%
5	0.58	-42%
6	0.38	-62%

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Effect of Age

```
. stcox age
No. of subjects =         9742                  Number of obs    =      9742
No. of failures =          437
Time at risk    =  38217.71783
                                                LR chi2(1)       =     23.62
Log likelihood  =   -3766.4598                  Prob > chi2      =    0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .9581318   .0083224    -4.92   0.000     .9419582    .9745831
------------------------------------------------------------------------------
```


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Interpretation

Age is a significant predictor of mortality $p < 0.001$, by the Wald test (Z= -4.92) or likelihood ratio test.

Each one-year increase in age (at transplant) reduces the hazard of mortality by 4%, 95% CI (2% to 6% reduction)

Why the Cox Model?

- Can be fitted without an explicit model for the hazard
- Can model the effect of a continuous predictor
- Can model multiple predictors: *continuous, binary, categorical*
- Can adjust for confounders: *adjust by adding confounders to the model*
- Can incorporate interaction, mediation: *create and add product terms (**simpler in Stata**11)*
- Can detect and estimate predictors for patient-level prognosis

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Comparison with other forms of regression

- *Same issues as in linear and logistic regression:* predictor selection

- *differences:* interpretation, assumptions, model checking

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostatistics, Epidemiology and Population Health

Summary of the lecture

- **Survival Data**
  - characterized by censoring
  - requires special methods
- **Cox Model**
  - based on hazard functions
  - assumes proportional hazards
  - uses hazard ratio for predictor effects
  - does not require model for baseline hazard
  - important similarities with other regressions