JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

When Time Is of Interest: The Case for Survival Analysis

John McGready
Johns Hopkins University

Lecture Topics

- Why another set of methods?
- Event times versus censoring times
- Estimating the survival curve—the Kaplan Meier method
- Statistically comparing survival curves

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section A

Motivating the Need

Survival Analysis

- Statistical methods for the study of time to an event

- Accounts for . . .
  - Time that events occur
  - Different follow-up times
  - Loss to follow-up

HIV Progression among IVDUs

- From article* abstract:
  - "**Objectives**: We sought to examine whether there were differential rates of HIV incidence among aboriginal and non-aboriginal injection drug users in a Canadian setting."
  - "**Methods**: Data were derived from two prospective cohort studies of injection drug users in Vancouver, British Columbia . . . we compared HIV incidence among aboriginal and non-aboriginal participants."
  - "**Results**: Aboriginal ethnicity was independently associated with elevated HIV incidence."

Notes: * Wood, E., et al. (2003). Burden of HIV infection among aboriginal injection drug users in Vancouver, British Columbia, *American Journal of Public Health*. 98: 3.

HIV Progression among IVDUs

- From article text:
  - "Participants were eligible for our study if they were recruited between May 1996 and December 2005."
  - "As previously described, the date of HIV seroconversion was estimated by using the midpoint between the last negative and the first positive antibody test results. Participants who remained persistently HIV seronegative were censored at the time of their most recent available HIV antibody test result prior to December 2005." (end of study)

Source: Wood, E., et al. (2003). Burden of HIV infection among aboriginal injection drug users in Vancouver, British Columbia, *American Journal of Public Health*. 98: 3.

HIV Progression among IVDUs

- Event of interest: HIV seroconversion

- Time frame for tracking HIV seroconversion among participants who were HIV negative at time of enrollment
  - "Clock" starts at time of enrollment"
  - "Clock" stops at either:
    - Seroconversion (event observed)
    - End of study (no event observed)
    - Loss to follow-up prior to seroconversion (no event observed)

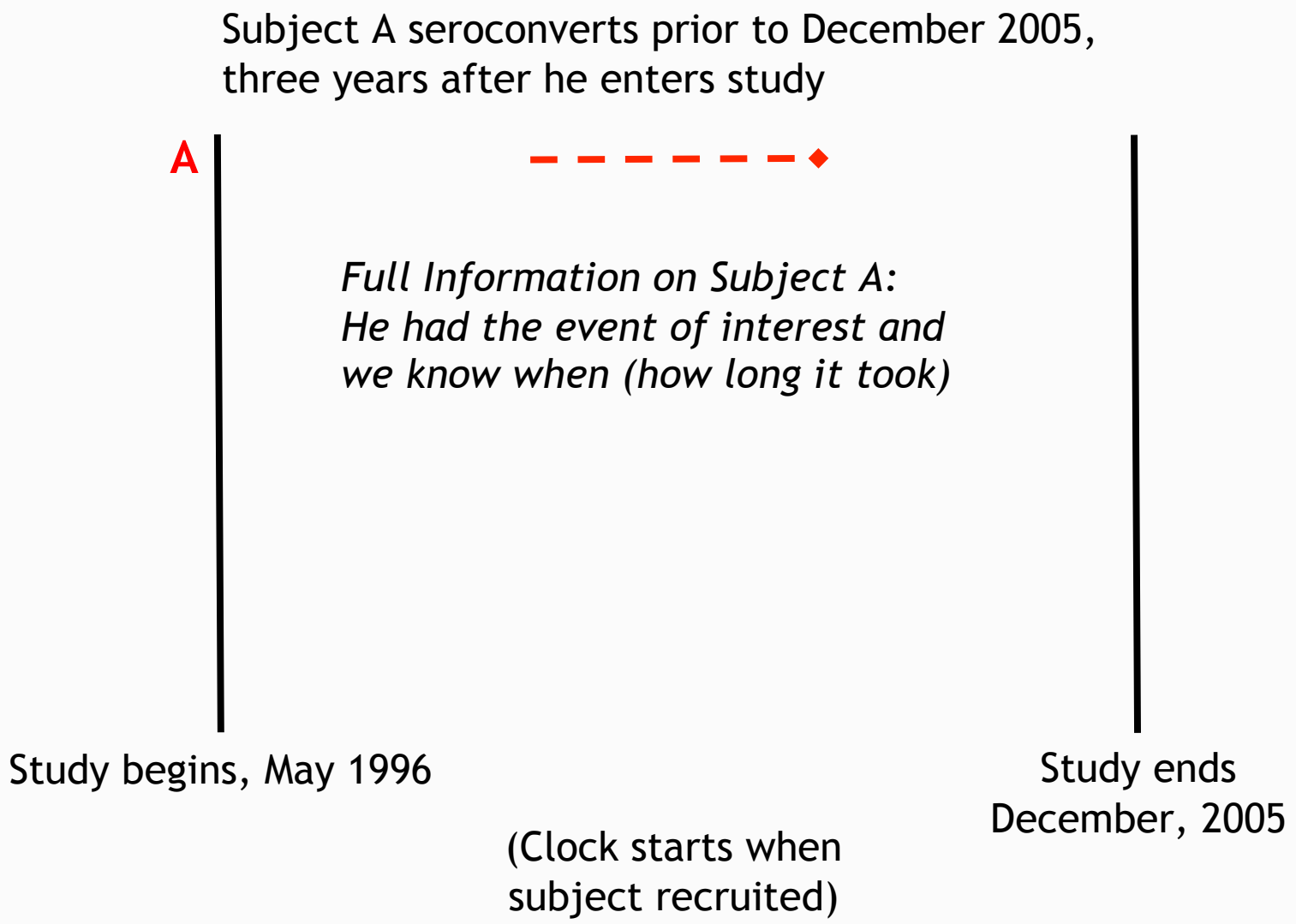
- Researchers interested in both frequency of event AND time to event

HIV Progression among IVDUs

- Graphic of possible scenarios

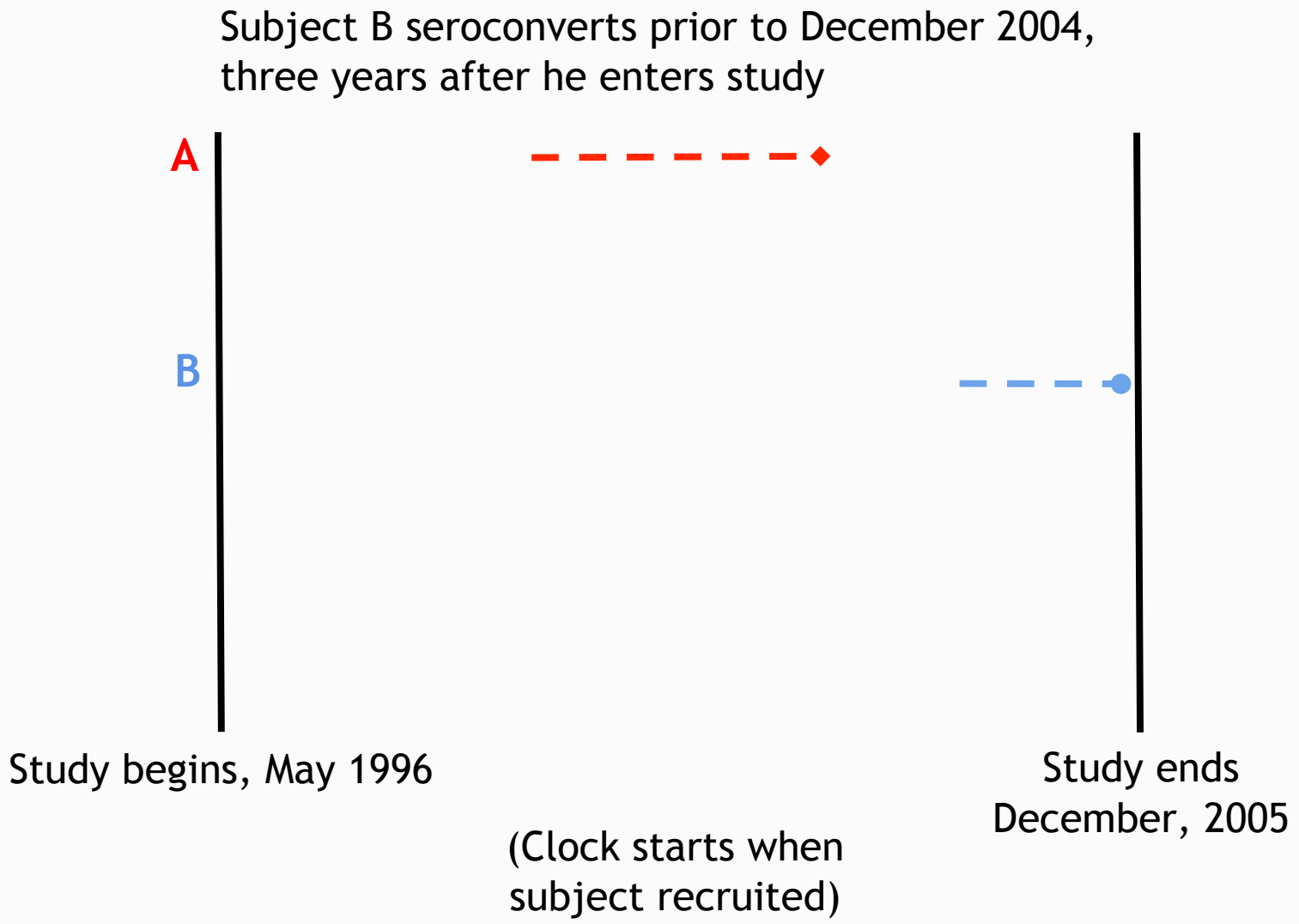
HIV Progression among IVDUs

- Graphic of possible scenarios



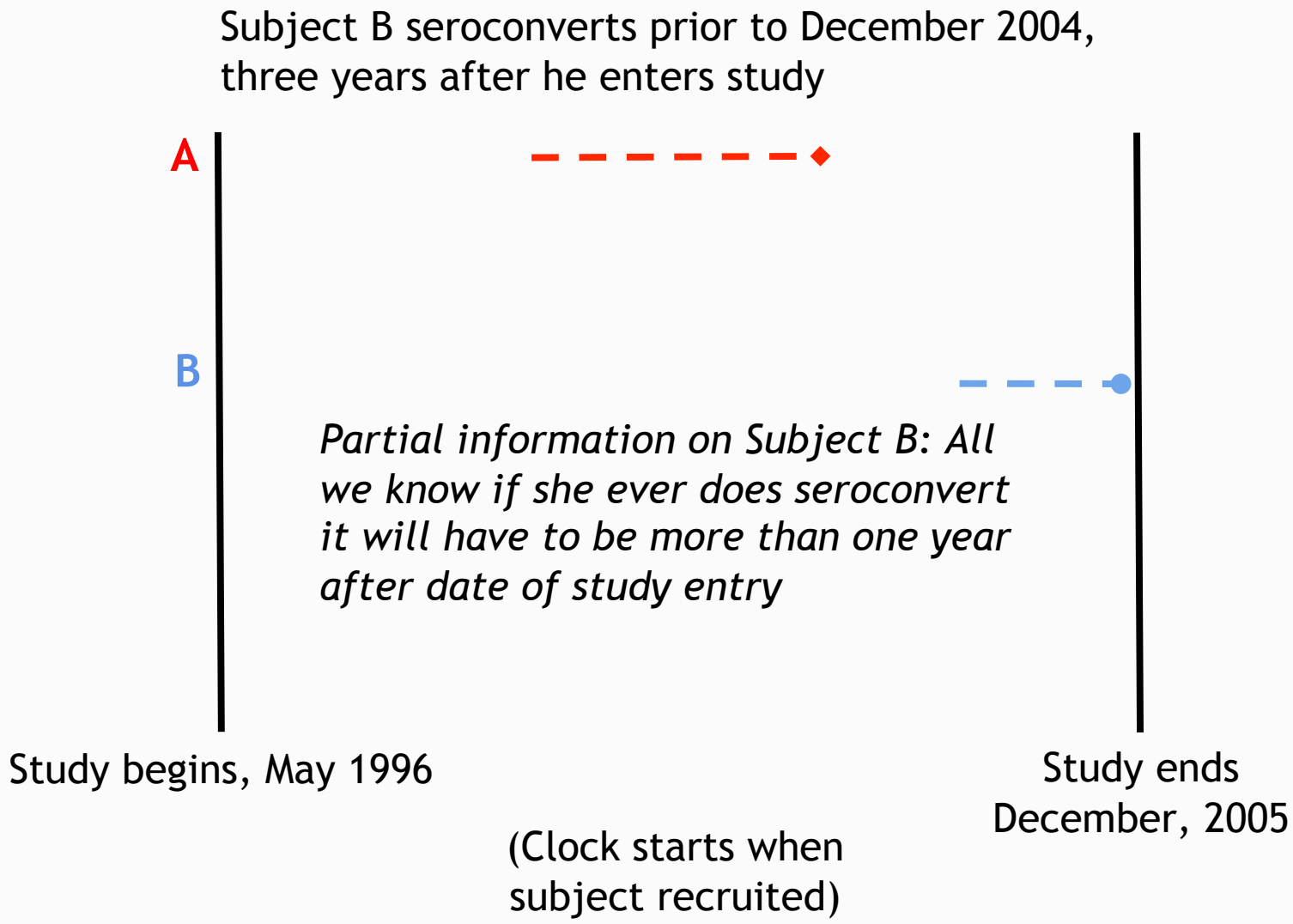
HIV Progression among IVDUs

- Graphic of possible scenarios



HIV Progression among IVDUs

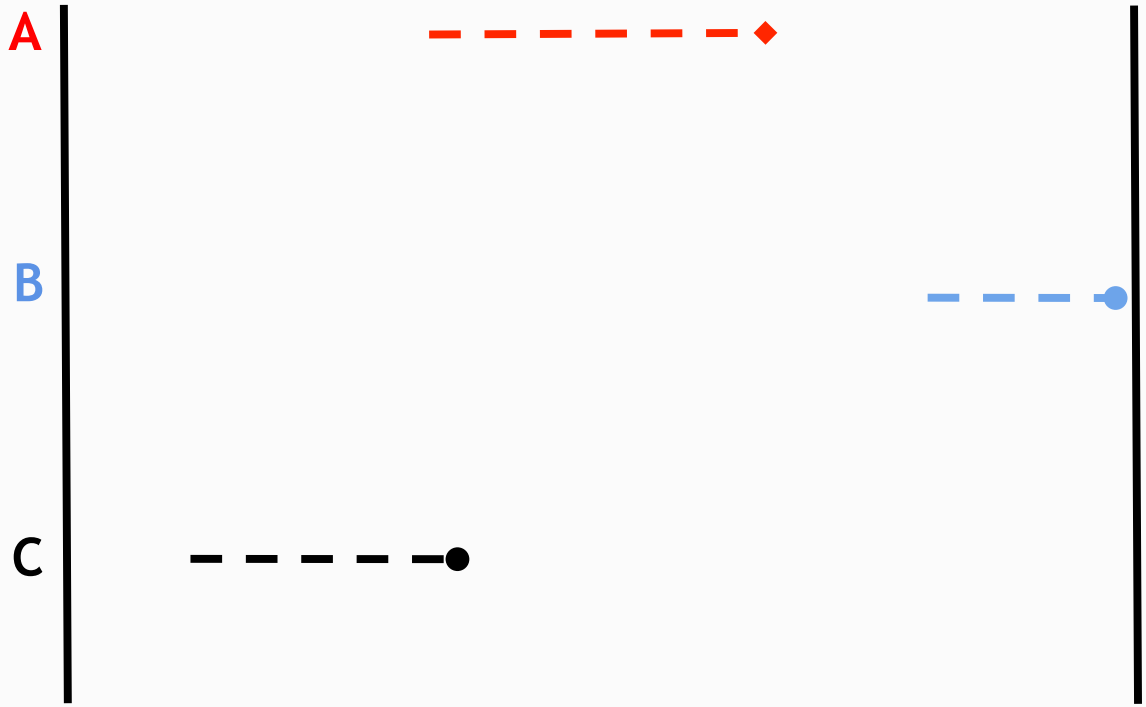
- Graphic of possible scenarios



HIV Progression among IVDUs

- Graphic of possible scenarios

Subject C enters in November 1997; He has all negative HIV tests until last follow-up visit in November 1999



Study begins, May 1996

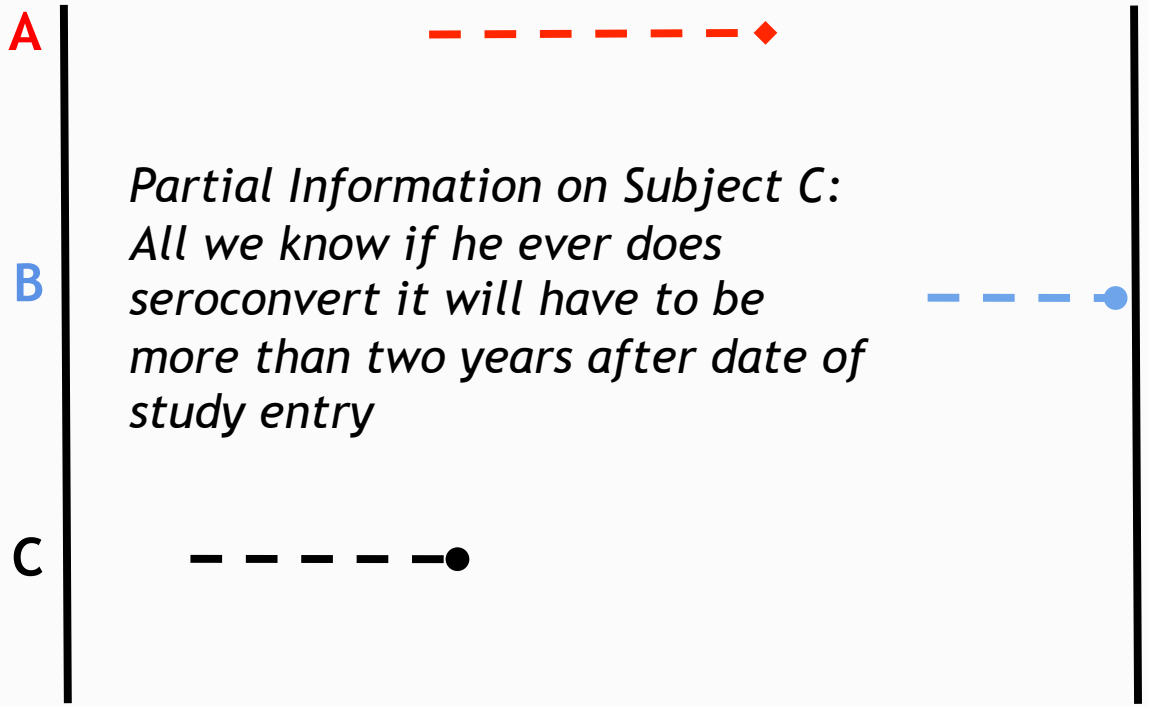
(Clock starts when subject recruited)

Study ends December, 2005

HIV Progression among IVDUs

- Graphic of possible scenarios

Subject C enters in November 1997; He has all negative HIV tests until last follow-up visit in November 1999



Study begins, May 1996

(Clock starts when subject recruited)

Study ends December, 2005

Chemotherapy Example

- Suppose we have designed a study to estimate survival after chemotherapy treatment for patients with a certain cancer
- Patients received chemotherapy between 1990 and 1994 and were followed until death or the year 2000, whichever occurred first
- In this study the event of interest is death
- The time clock starts as soon as the subject finishes his/her chemotherapy treatments

Chemotherapy Example

- Three results from study:
  - Patient one enters in 1990, dies in 1995: patient one survives five years
  - Patient two enters in 1991, drops out in 1997: patient two is lost to follow-up after six years
  - Patient three enters in 1993 and is still alive at end of study: patient three is still alive after seven years

Why Is Survival Analysis Tricky?

- Patient:
  - 1: 1990 → 1995 5 years
  - 2: 1991 → 1997 6+ years
  - 3: 1993 → 2000 7+ years

- Patients two and three are called censored observations

- We need a method which can incorporate information about censored data into an analysis

Interested in Time: Why Not Treat as Continuous

- Patient:
  - 1: 1990 → 1995 5 years
  - 2: 1991 → 1997 6+ years
  - 3: 1993 → 2000 7+ years

- Suppose we wanted to estimate the mean time to death for the three patients listed above: suppose we average the three death/ censoring times

- This average would systematically underestimate the average of the three persons, because two of the three numbers are underestimates of time to death after finishing chemotherapy

Interested in Occurrence of Event: Binary?

- Event of interest is binary
  - Why not just summarize total proportion who had the event before the end of study, treating those censored as "non-events"
  - Suppose we have designed a study to compare survival after two different chemotherapy treatments for patients with a certain cancer
  - Patients randomized to one of two chemotherapy groups: after assignments, received chemotherapy between 1990 and 1994 and were followed until death or the year 2000, whichever occurred first

Interested in Occurrence of Event: Binary?

- At end of study, 40% of patients in each of the two chemotherapy groups had died
  - Exactly the same proportion (do we even need a p-value?)
  - Does this show that neither treatment is “superior” in terms of prolonging survival?

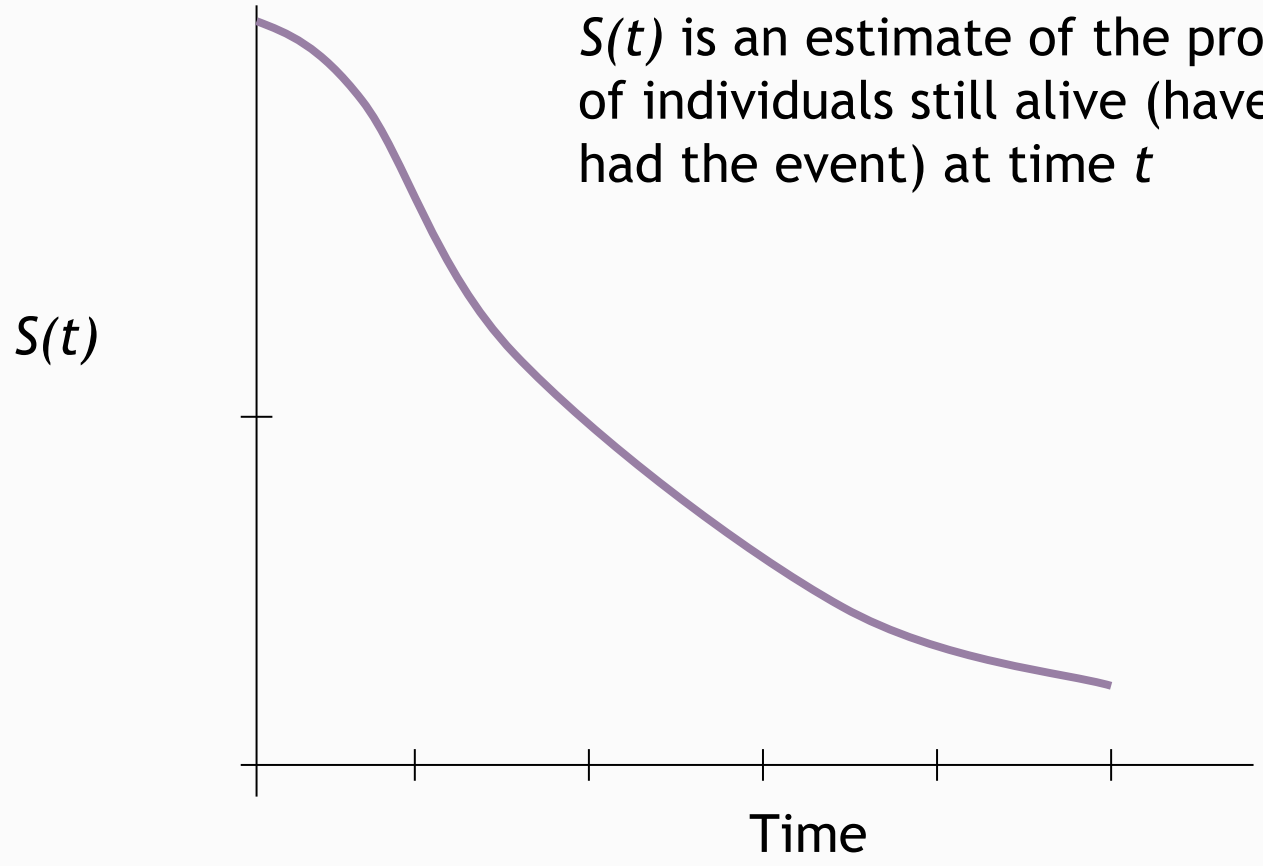
- Suppose in the first chemotherapy group, most of the 40% died within a year of stopping the treatment; in the second group, most of the 40% died between five to six years after stopping treatment:
  - Timing of the event is very different between the two groups even though the end percentages are similar

Another Method Needed

- Another method is needed to analyze time to event data in the presence of censoring
- This method needs to utilize time in its analysis, but also differentiate between event times (full time information) and censoring times (partial time information)
- This method will produce a summary statistic that captures both the binary portion (event y/n) and the time portion of the “story”

Summary Statistics

- The method we will discuss in the next section produces the following "summary statistic" for a sample of time-to-event data
  - The survival curve



S(t) is an estimate of the proportion of individuals still alive (have not had the event) at time *t*

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

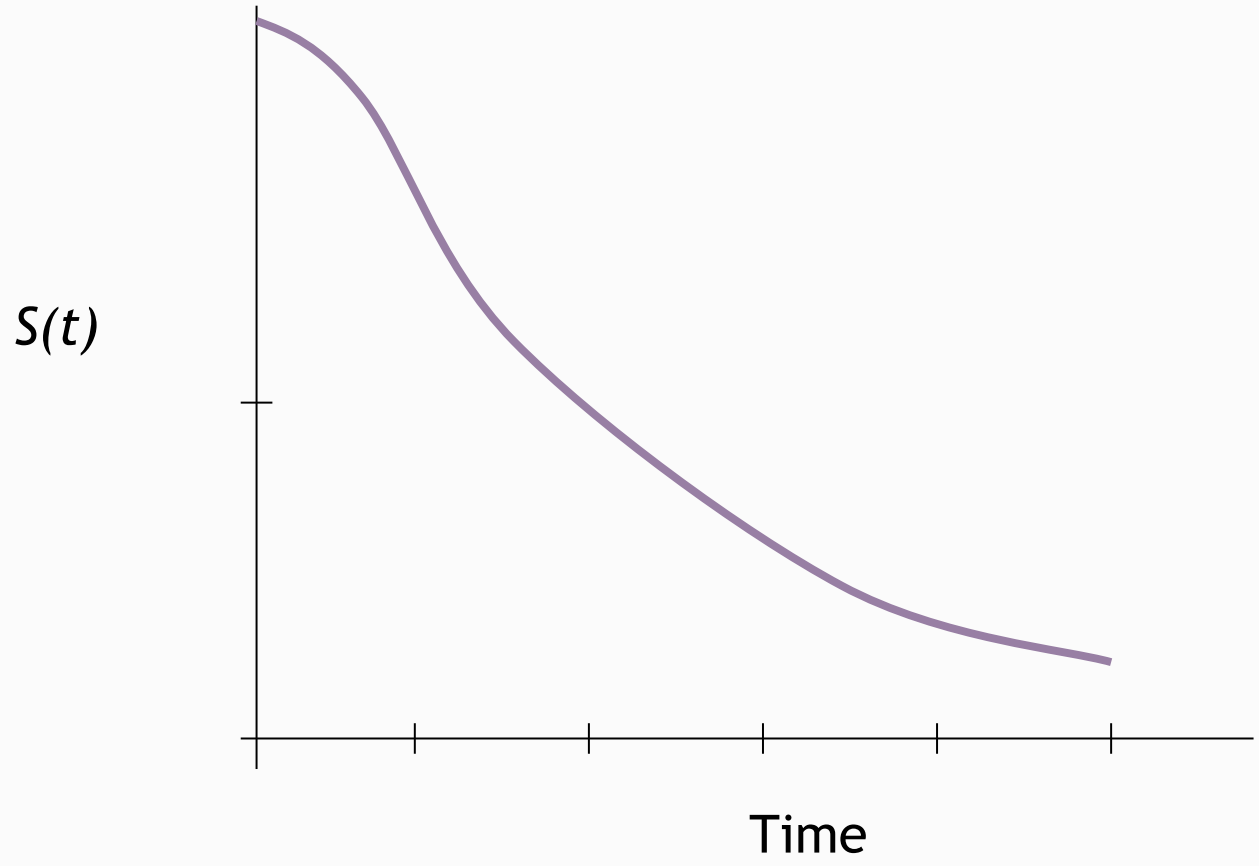
Section B

Estimating the Survival Curve: The Kaplan Meier Approach

Central Problem

- Estimation of the "survival curve"

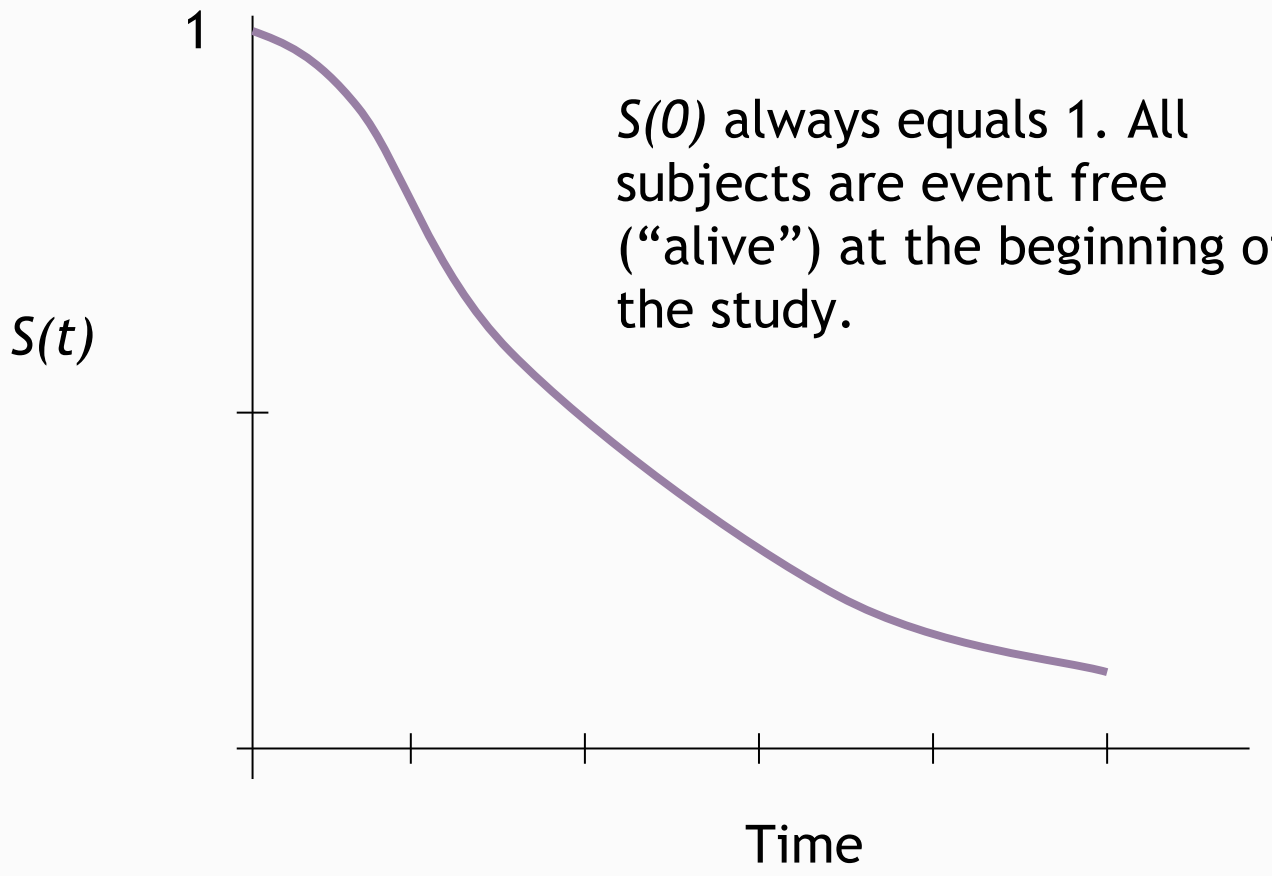
- *S(t)* = proportion remaining event free (surviving) at least to time t or beyond



Central Problem

- Estimation of the "survival curve"

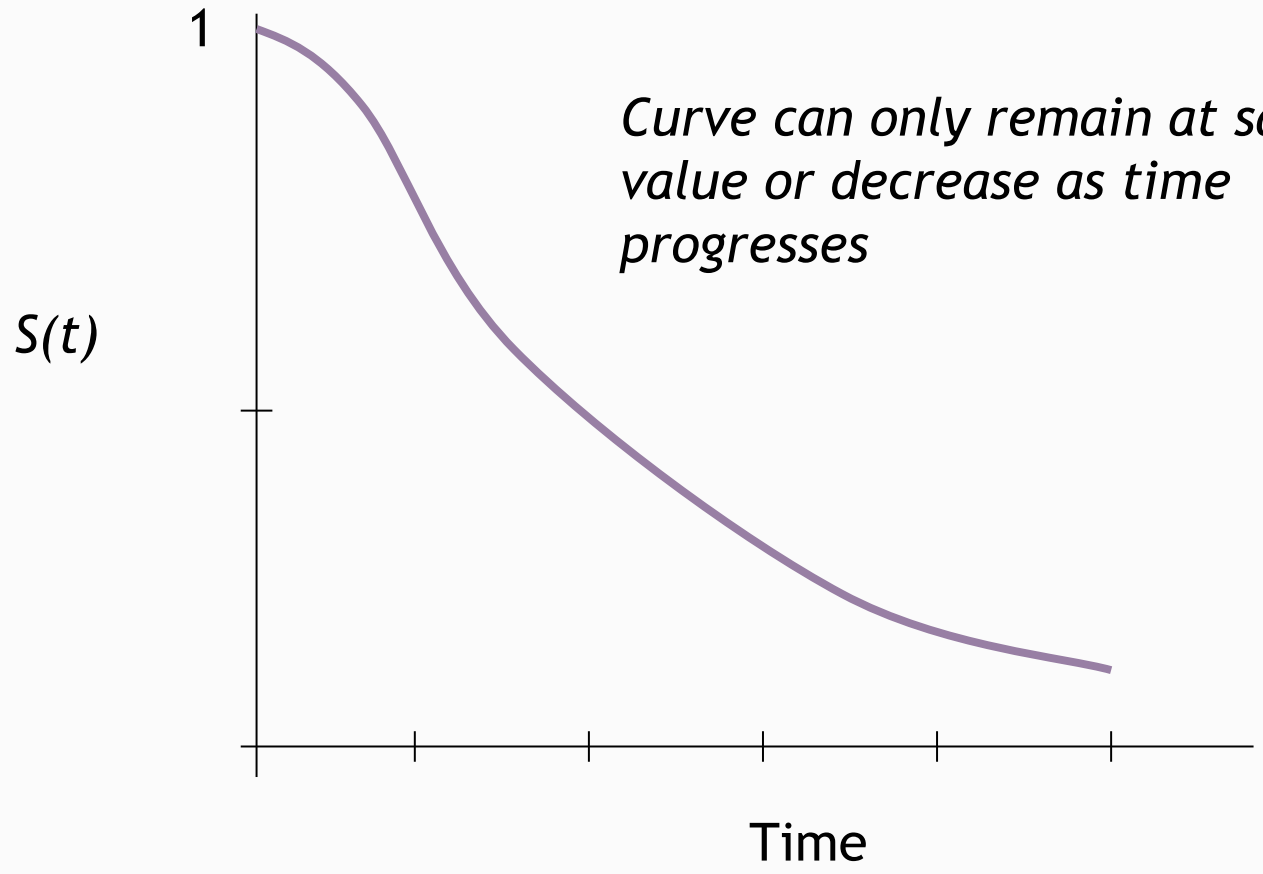
- *S(t)* = proportion remaining event free (surviving) at least to time *t* or beyond



Central Problem

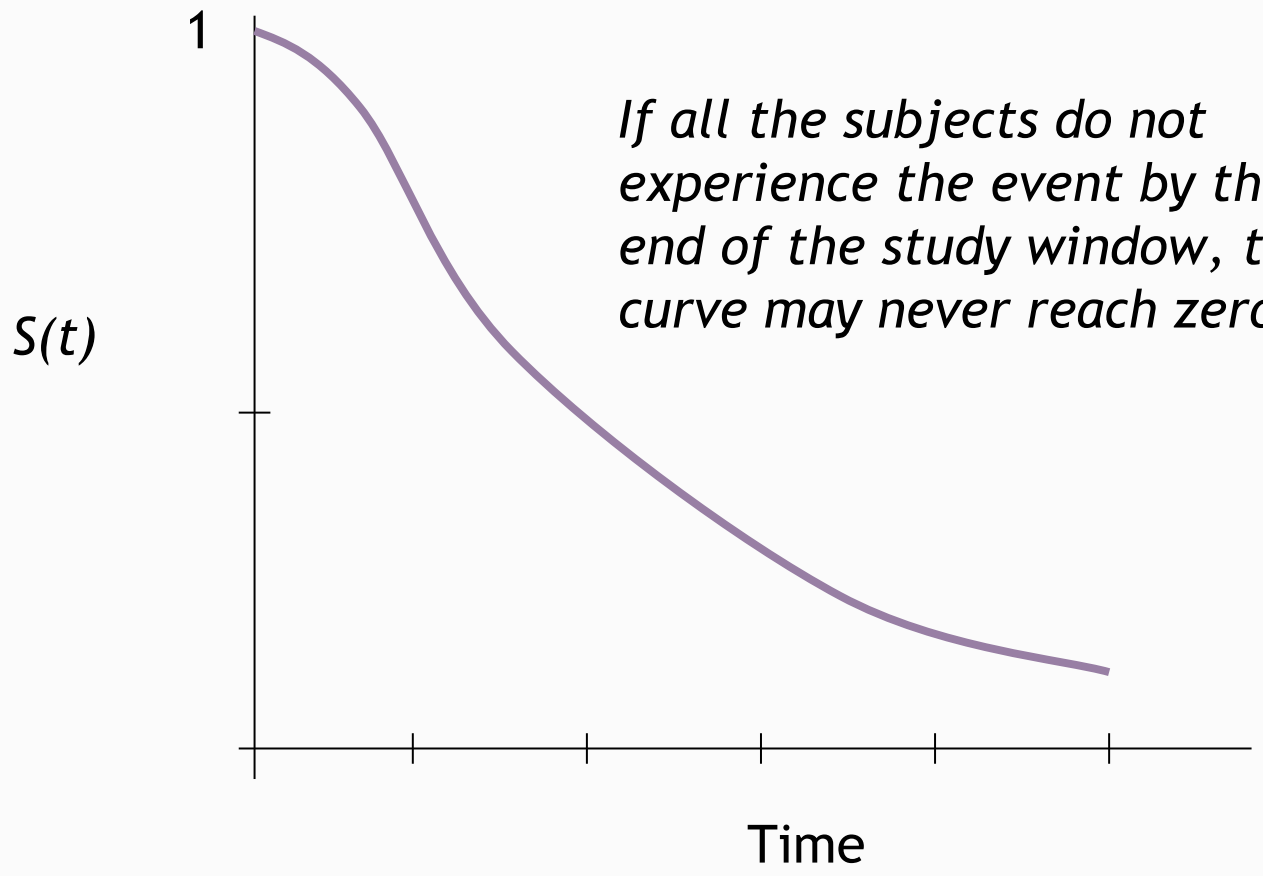
- Estimation of the "survival curve"

- *S(t)* = proportion remaining event free (surviving) at least to time t or beyond



Central Problem

- Estimation of the "survival curve"
  - *S(t)* = proportion remaining event free (surviving) at least to time *t* or beyond
  - We can estimate *S(t)* from a sample of data: out statistic is $\hat{S}(t)$



If all the subjects do not experience the event by the end of the study window, the curve may never reach zero

Approaches

- Life table method
  - Grouped in intervals

- Kaplan-Meier (1958)
  - Ungrouped data
  - Small samples

Kaplan-Meier Estimate

- Example: time (months) from primary AIDS diagnosis for a sample of 12 hemophiliac patients under 40 years old at time of HIV seroconversion*
  - Event times (n = 12):
  - 2 3+ 5 6 7+ 10 15+ 16 16 27 30 32

Notes: * Example based on data taken from Rosner, B. (1990). Fundamentals of biostatistics, 6th ed. (2005). Duxbury Press. (based on research by Ragni, et al. (1990). Cumulative risk for AIDS in *Journal of Acquired Immune Deficiency Syndromes*, Vol. 3.

Kaplan-Meier Estimate

- $\hat{S}(t)$ = 1, to start

- After starting at time 0, curve can be estimated at each event time t, but not at censoring times

$$\hat{S}(t) = \left(\frac{N(t) - E(t)}{N(t)} \right) \times \hat{S}(Previous\ Event\ Time)$$

- *E(t)* = # events at time *t*

- *N(t)* = # subjects at risk for event at time *t*

Kaplan-Meier Estimate

- Curve can be estimated at each event, but not at censoring times

$$\hat{S}(t) = \left(\frac{N(t) - E(t)}{N(t)} \right) \times \hat{S}(Previous\ Event\ Time)$$

Proportion of original sample making it to time *t*

Kaplan-Meier Estimate

- Curve can be estimated at each event, but not at censoring times

$$\hat{S}(t) = \left(\frac{N(t) - E(t)}{N(t)} \right) \times \hat{S}(Previous\ Event\ Time)$$

Proportion surviving to time *t* who survive beyond time *t*

Kaplan-Meier Estimate

- Start estimate at first (ordered) event time
  - 2 3+ 6 6 7+ 10 15+ 15 16 27 30 32

$$\hat{S}(2) = \left(\frac{N(2) - E(2)}{N(2)} \right) = \frac{12 - 1}{12} = \frac{11}{12} = .92$$

Kaplan-Meier Estimate

- Can estimate *S(t)* at each subsequent event time
  - (Censoring times inform estimate about number at risk of having the event at a time *t* until censoring occurs)
  - 2 3+ 6 6 7+ 10 15+ 15 16 27 30 32

$$\hat{S}(6) = \left(\frac{N(6) - E(6)}{N(6)} \right) \times \hat{S}(2) = \left(\frac{10 - 2}{10} \right) \times .92 = .80 \times .92 = .74$$

Kaplan-Meier Estimate

- Can estimate *S(t)* at each subsequent event time
  - (Censoring times inform estimate about the number at risk of having the event at a time *t*)
  - 2 3+ 6 6 7+ 10 15+ 15 16 27 30 32

$$\hat{S}(10) = \left(\frac{N(10) - E(10)}{N(10)} \right) \times \hat{S}(6) = \left(\frac{7-1}{6} \right) \times .74 = .86 \times .74 = .64$$

Kaplan-Meier Estimate

- Continue through final event time

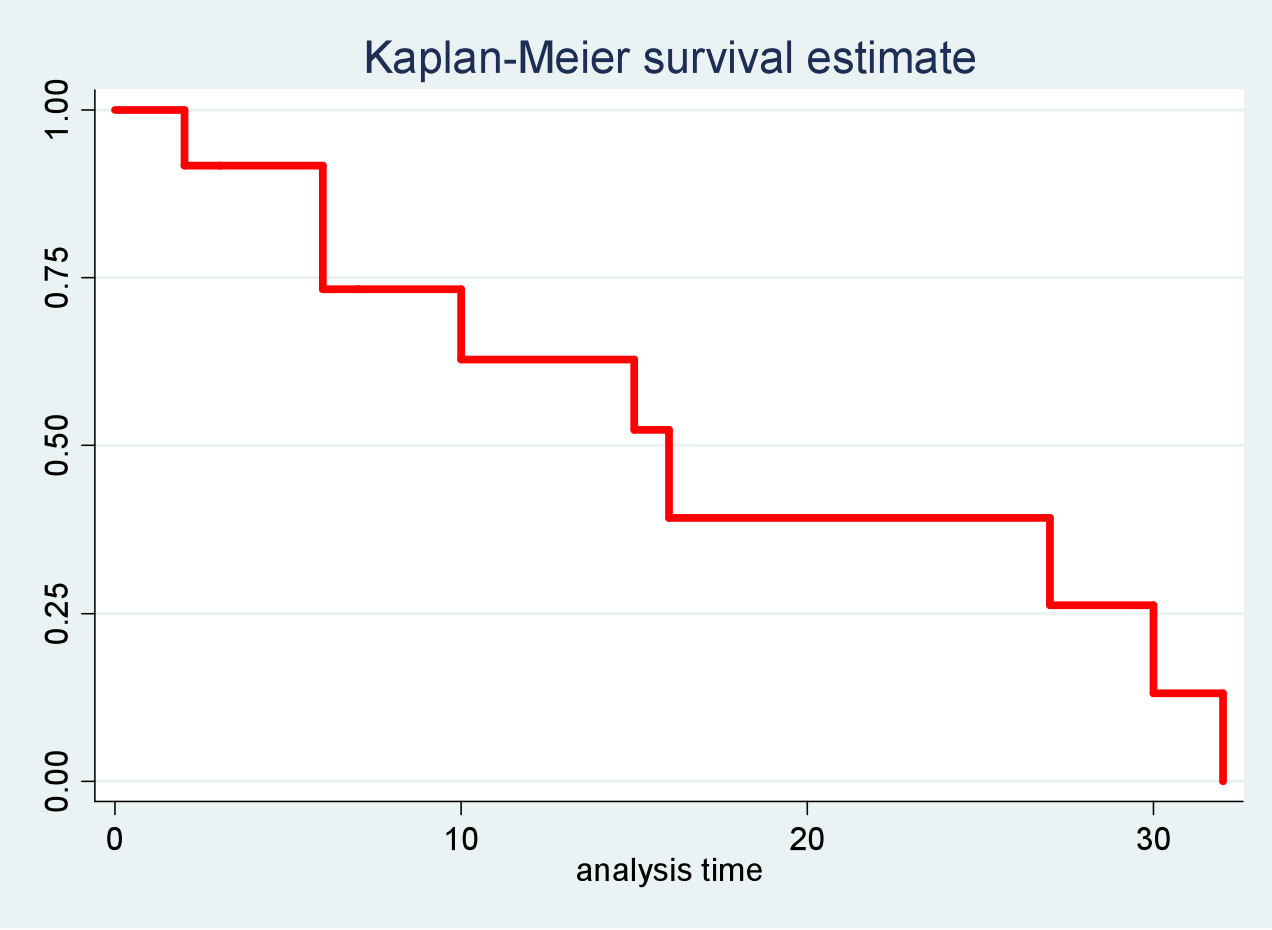
t	$\hat{S}(t)$
2	.92
6	.74
10	.64
15	.52
16	.39
27	.26
30	.13
32	0

Kaplan-Meier Estimate

- Graph is a step function
- “Jumps” at each observed event time
- Nothing is assumed about curved shape between each observed event time

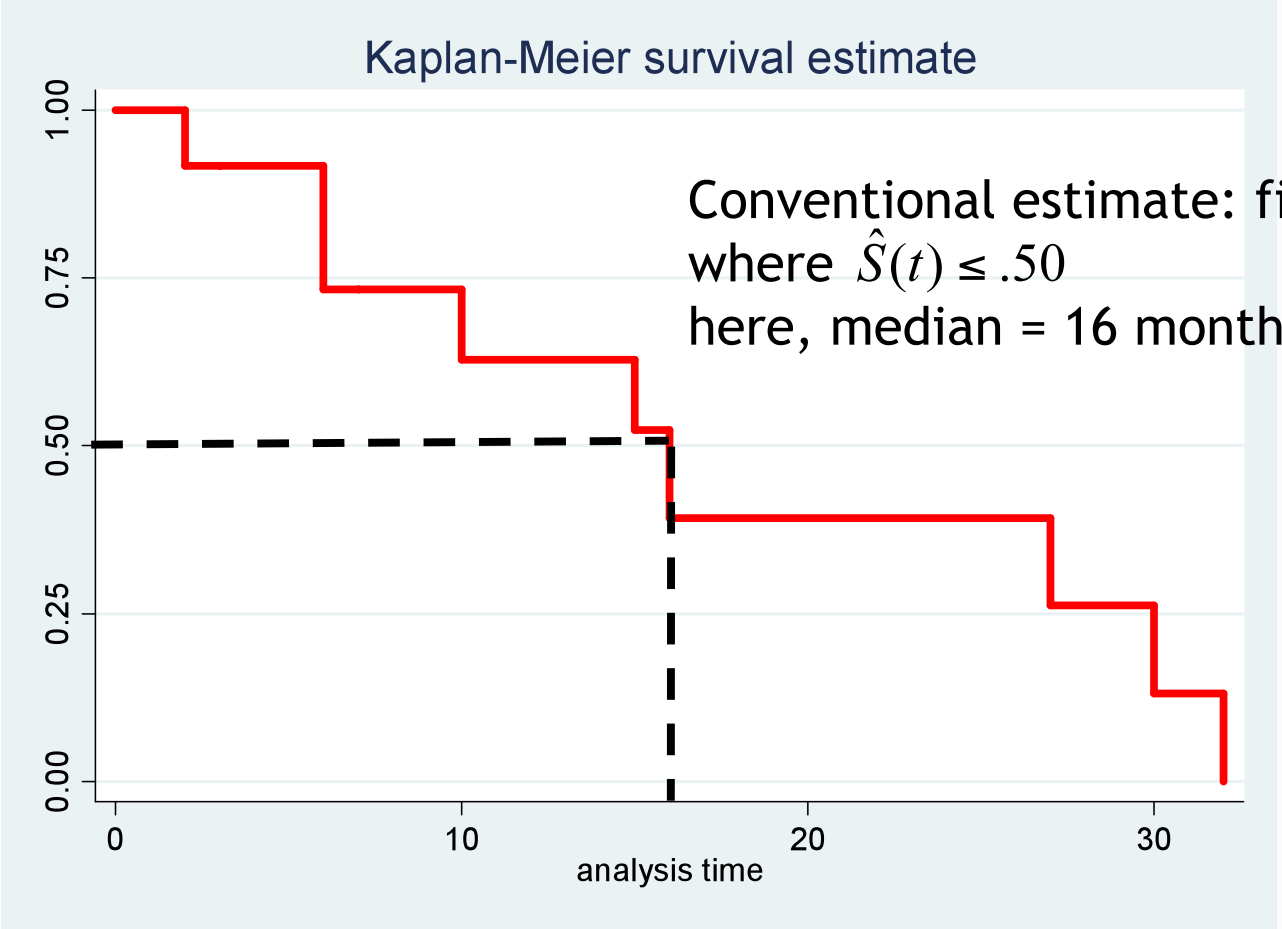
Kaplan-Meier Estimate

- Kaplan-Meier estimate graphically presented



Kaplan-Meier Estimate

- You can use these to estimate single number summary statistics, like the median survival time (median time remaining event free)



Kaplan-Meier Estimate

- Example
  - Time days to resuming smoking in first month following completion of five one-hour group coaching sessions on smoking cessation (10 subjects)
  - 15 3+ 30+ 5 10+ 30+ 7 1 24+ 27

Kaplan-Meier Estimate

- Example
  - Time days to resuming smoking in first thirty day period following completion of five one-hour group coaching sessions on smoking cessation (10 subjects): ordered times
  - 1 3+ 5 7 10+ 15 24+ 27 30+ 30+

Kaplan-Meier Estimate

- Example
  - Time days to resuming smoking in first thirty day period following completion of five one-hour group coaching sessions on smoking cessation (10 subjects): ordered times
  - 1 3+ 5 7 10+ 15 24+ 27 30+ 30+
  - $$\hat{S}(1) = \left(\frac{N(1) - E(1)}{N(1)} \right) = \frac{10 - 1}{10} = \frac{9}{10} = .90$$

Kaplan-Meier Estimate

- Example
  - Time days to resuming smoking in first thirty day period following completion of five one-hour group coaching sessions on smoking cessation (10 subjects): ordered times
  - 1 3+ 5 7 10+ 15 24+ 27 30+ 30+
  - $\hat{S}(5) = \left(\frac{N(5) - E(5)}{N(5)} \right) \times \hat{S}(1) = \left(\frac{8-1}{8} \right) \times .90 = .88 \times .90 = .79$

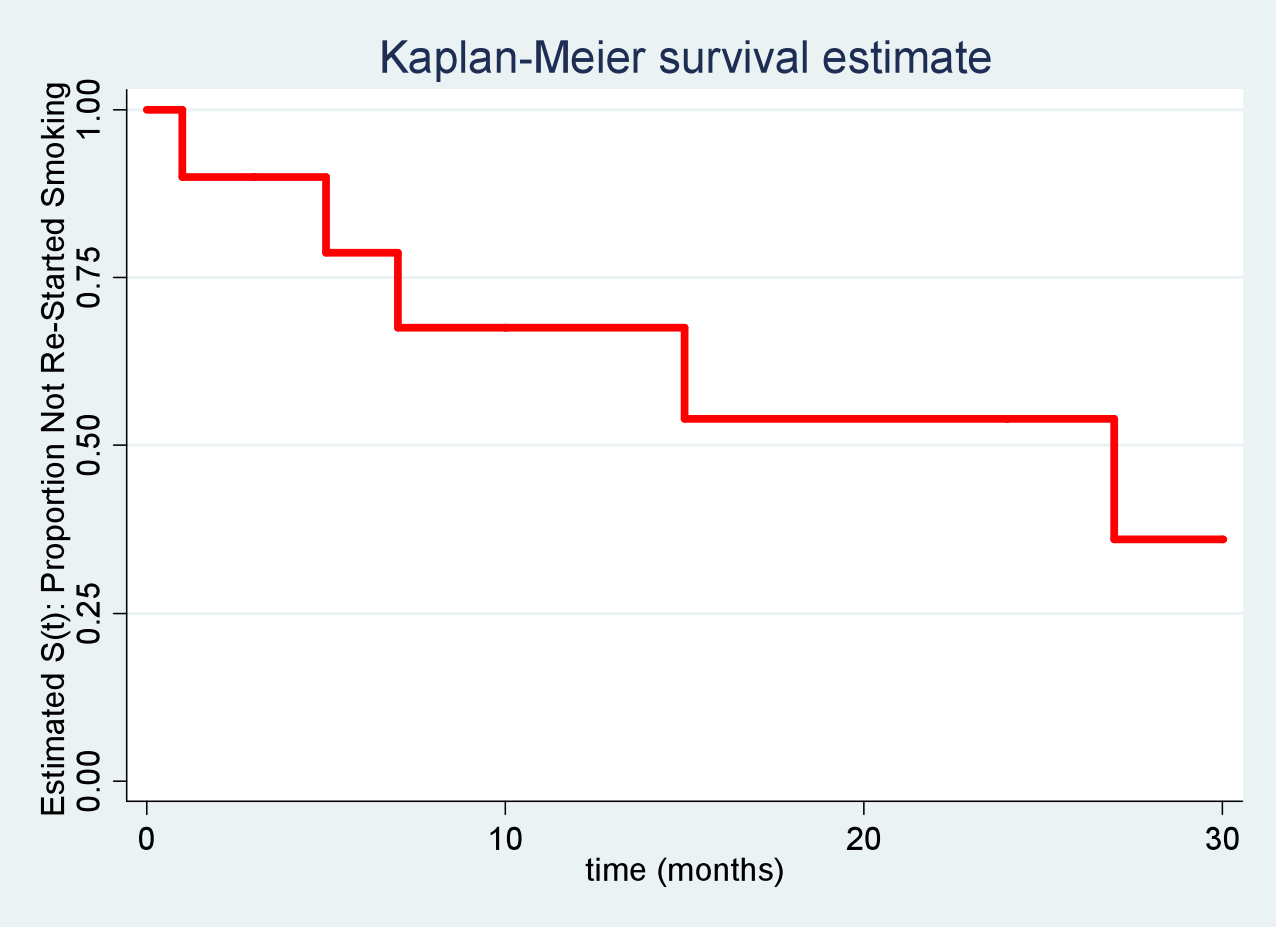
Kaplan-Meier Estimate

- Continue through final event time: notice this estimated curve never reaches 0 because largest time values are censoring times

t	$\hat{S}(t)$
1	.90
5	.79
7	.68
15	.54
27	.36

Kaplan-Meier Estimate

- Graphical presentation



Big Assumption

- Independence of censoring and survival
- Those censored at time *t* have the same prognosis as those not censored at *t*
- Examples of possible violations
  - Time to tumor—animal
  - Occupational health—loss to follow up

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section C

Statistical Inference on Survival Curves

Comparing Survival Curves

- The estimate survival curve $\hat{S}(t)$ is just an estimate based on a sample from a larger population: how to quantify uncertainty on a curve?

- One approach: can put confidence intervals around each change estimated at each event time
  - This can be cumbersome to read/interpret when there are many event times
  - Not very efficient approach for comparing the survival curves between multiple populations based on multiple random samples (ex: drug versus placebo)

Comparing Survival Curves

- Common statistical tests
  - Generalized Wilcoxon (Breslow, Gehan)
  - Log-rank

- Both compare two survival curves across multiple time points to answer the question—"is overall survival different between the groups?"
  - $H_o : S_1(t) = S_2(t)$
  - $H_A : S_1(t) \neq S_2(t)$

Comparing Survival Curves

- Wilcoxon (Breslow, Gehan) more sensitive to early survival differences
- Log-rank more sensitive to later survival differences
- Both: compute difference between what is observed at each event time and what would be expected under the null hypothesis
  - These differences are aggregated across all event times into one overall "distance" measure (i.e., how far sample curves differ from null after accounting for sampling variability)
  - The Wilcoxon and log-rank tests aggregate these event-time specific differences slightly differently
  - Both tests give a p-value and generally these p-values are similar
- Neither
  - Give overall measure of association (like a relative risk, etc.) or confidence interval

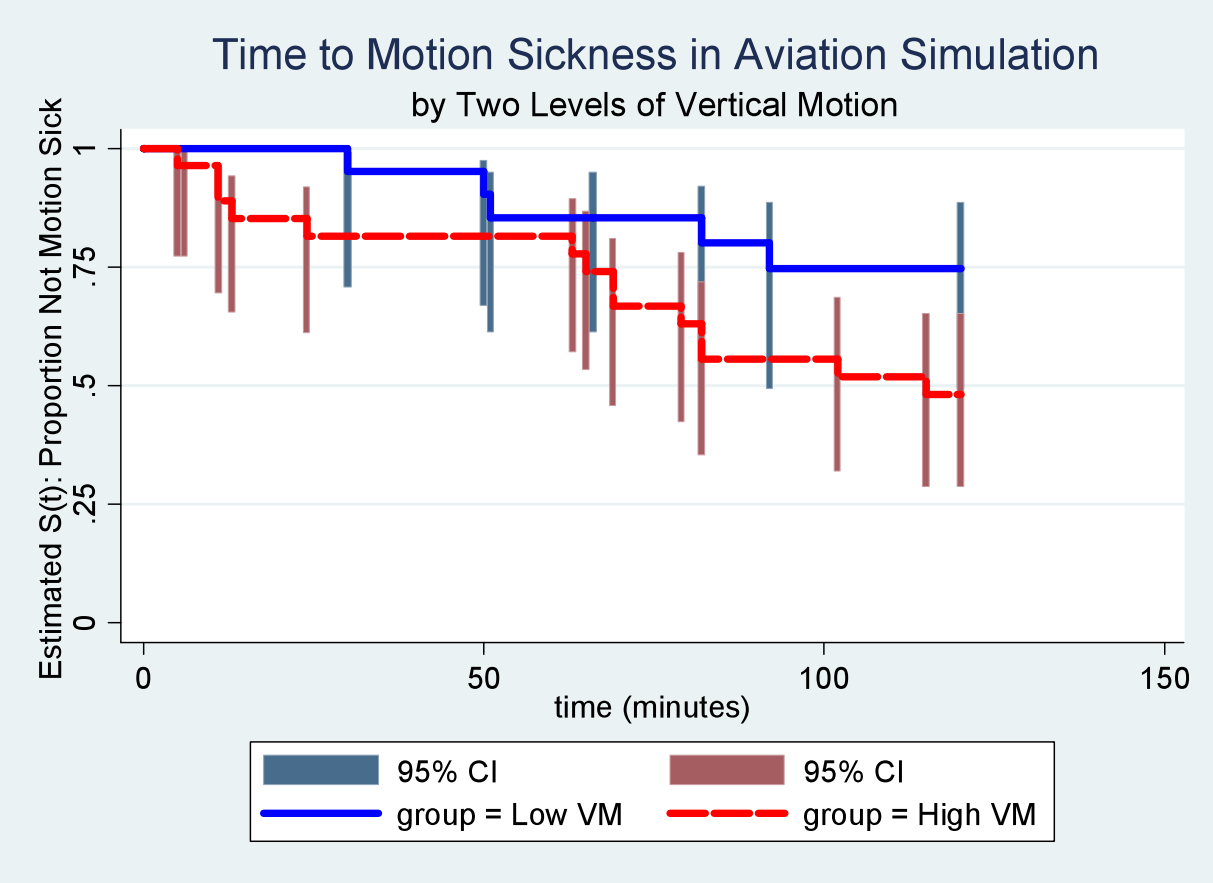
Examples of Logrank and Breslow-Gehan Test

- Time to motion sickness*: simulation designed to measure impact of intensity of prolonged vertical motion exposure on motion sickness
  - Group 1 subjected (21 persons) to low vertical motion for up to two hours
  - Group 2 (28 persons) subject to high vertical motion for up to two hours
  - Event of interest motion sickness (first vomiting episode)
  - Some subjects “dropped out” prior to the end of two hours without vomiting

Note: * Example based on data taken from Altman, D. (1991). Practical statistical for medical research, 1st ed. Chapman and Hall (based on research by Burns, K.C. (1990). *Motion sickness . . . aviation space environmental medicine*, 56, 21-7.

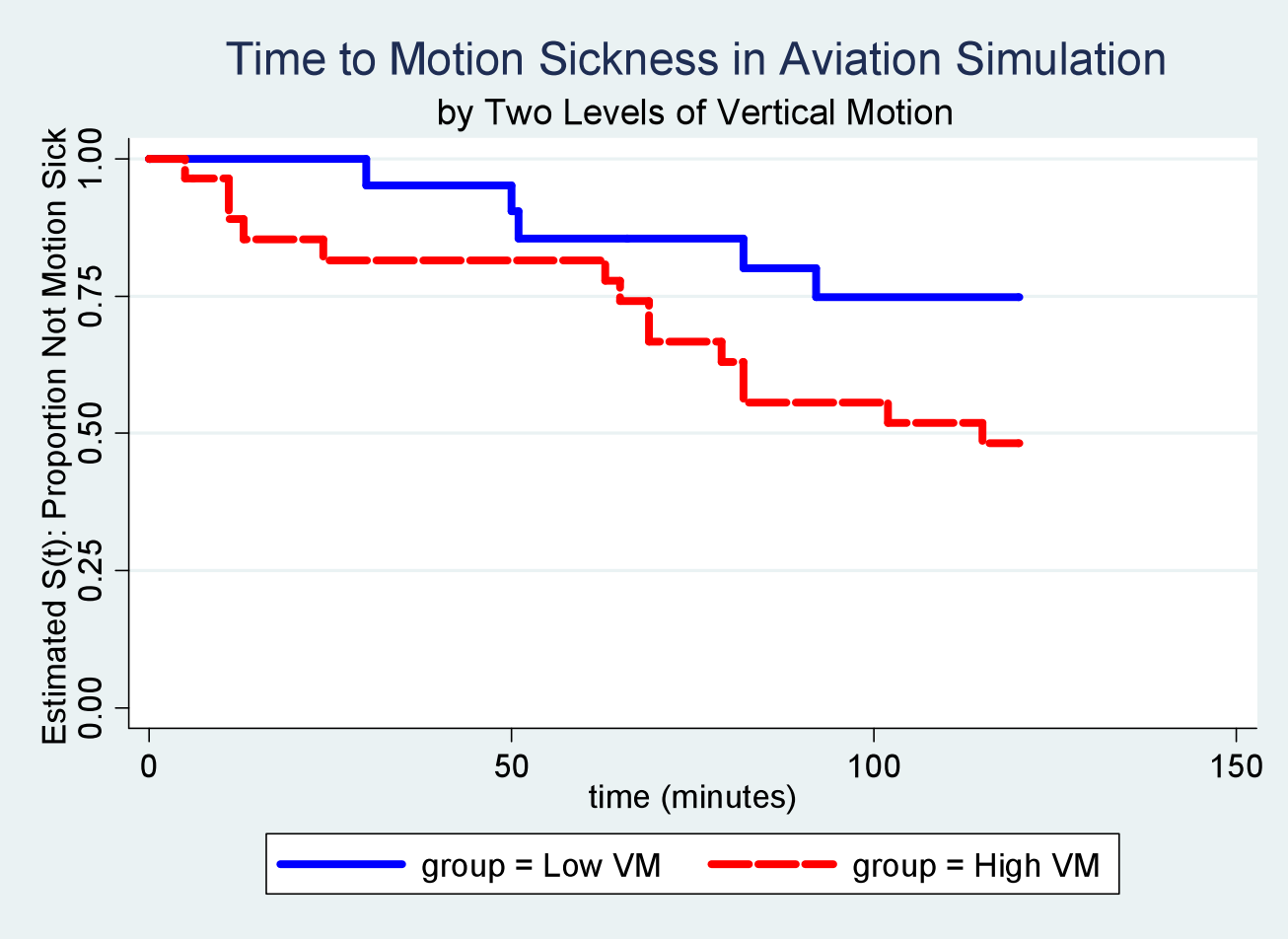
Examples of Logrank and Breslow-Gehan Test

- Time to motion sickness
  - Kaplan-Meier curves for time to motion sickness for each group, with 95% CIs (hard to see, but these get wider with increased time)



Examples of Logrank and Breslow-Gehan Test

- Time to motion sickness
  - Kaplan-Meier curves for time to motion sickness for each group, without 95% CIs



Testing VM Intensity/Motion Sickness Relationship

- Hypothesis test setup
  - H_o: $S_{LVM}(t) = S_{HVM}(t)$
  - H_A: $S_{LVM}(t) \neq S_{HVM}(t)$

- Log-rank results:
  - p = .073

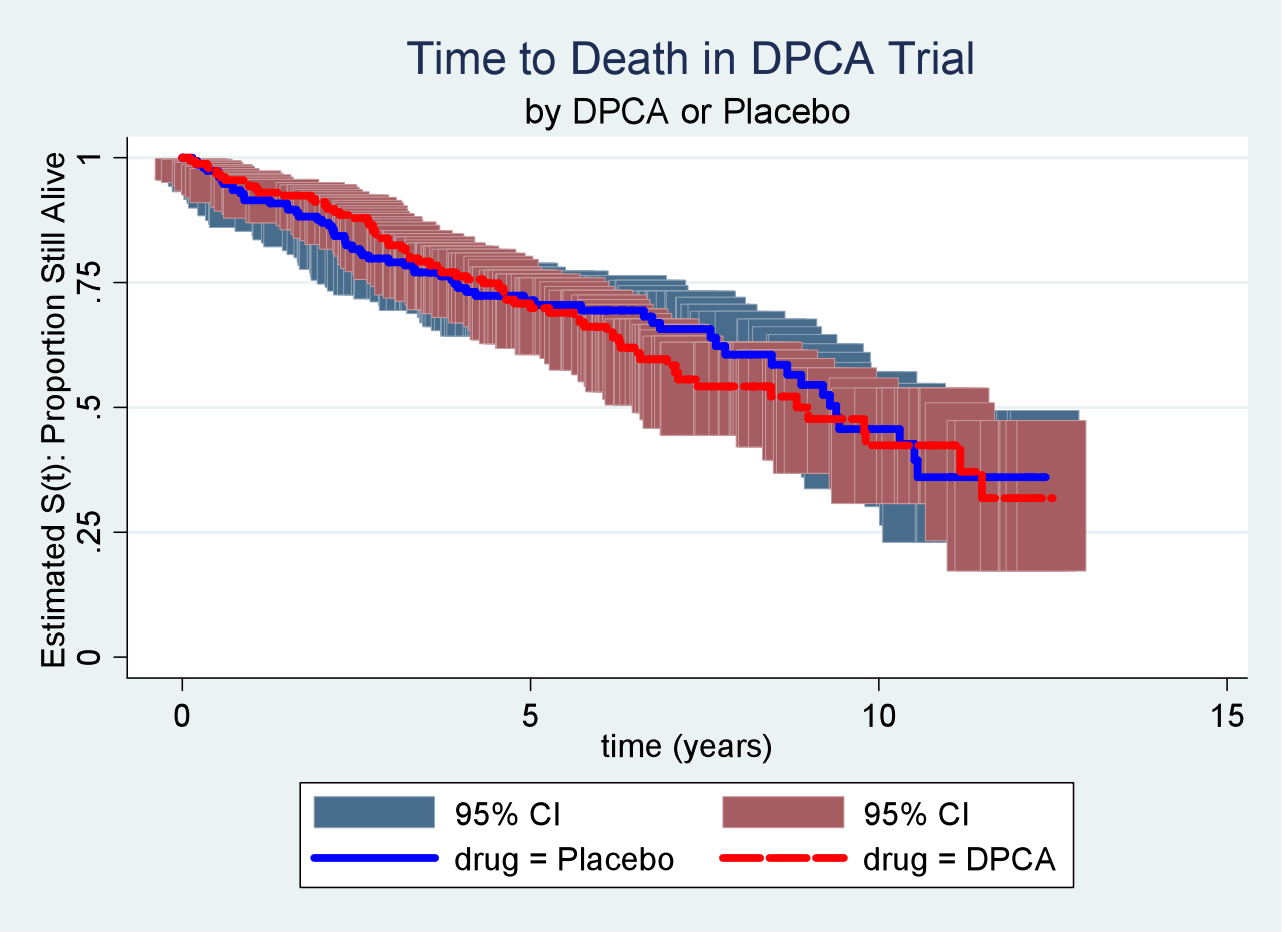
- Breslow/Wilcoxon/Gehan results:
  - p = .075

Examples of Logrank and Breslow-Gehan Test

- Clinical trial: between January 1974 and May 1984 a double-blinded randomized trial on patients with primary biliary cirrhosis (PBC) of the liver was conducted at the Mayo clinic (Rochester, MN)
  - A total of 312 patients were randomized to either DPCA (n = 154) or placebo (n = 158)
  - Patients were followed until they died from PBC or until censoring—either administrative censoring (withdrawn alive at the end of the study), death not attributable to PBC, liver transplantation, or lost to follow-up

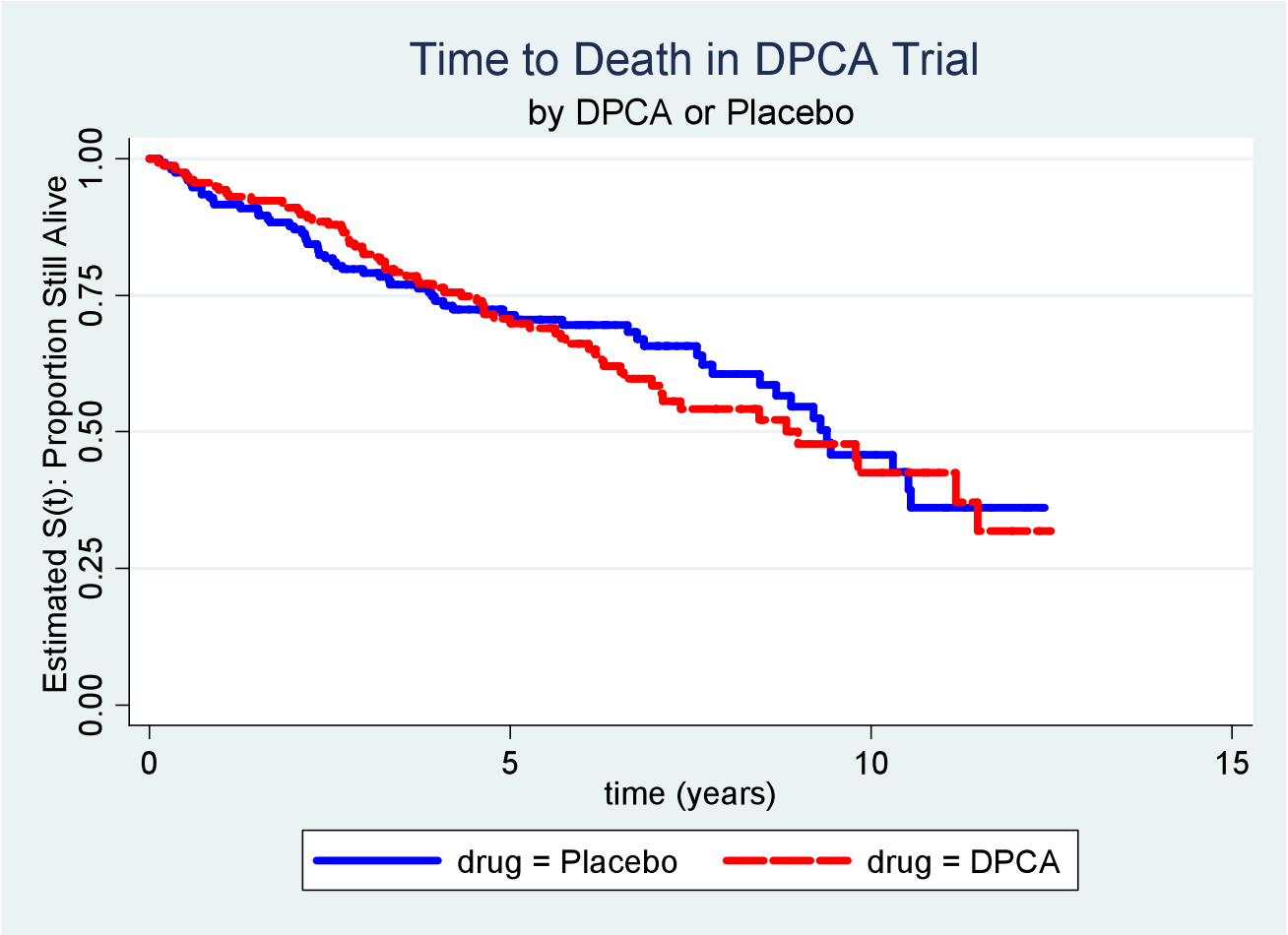
Examples of Logrank and Breslow-Gehan Test

- PBC trial
  - Kaplan-Meier curves for time to death from PBC for each group, with 95% CIs



Examples of Logrank and Breslow-Gehan Test

- PBC trial
  - Kaplan-Meier curves for time to death from PBC for each group, without 95% CIs



Testing Drug/Survival Relationship

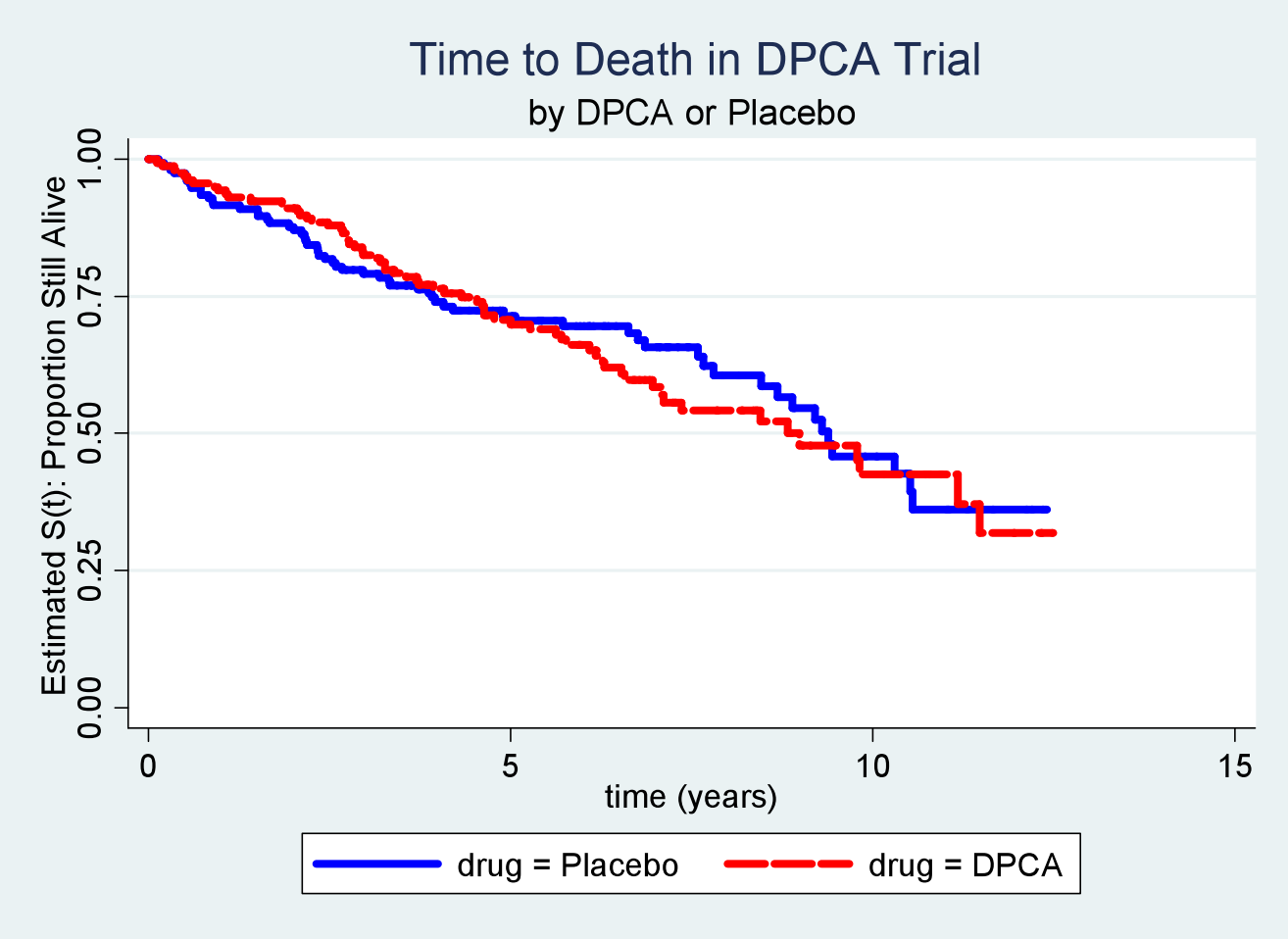
- Hypothesis test setup
  - H_o: $S_{DPCA}(t) = S_{PLACEBO}(t)$
  - H_A: $S_{DPCA}(t) \neq S_{PLACEBO}(t)$

- Log-rank results:
  - p = .75

- Breslow/Wilcoxon/Gehan results:
  - p = .96

Examples of Logrank and Breslow-Gehan Test

- PBC trial
  - Kaplan-Meier curves for time to death from PBC for each group, without 95% CIs



Testing Drug/Survival Relationship

- Hypothesis test setup
  - H_o: $S_{DPCA}(t) = S_{PLACEBO}(t)$
  - H_A: $S_{DPCA}(t) \neq S_{PLACEBO}(t)$

- Log-rank results:
  - p = .75

- Breslow/Wilcoxon/Gehan results:
  - p = .96

Examples from Literature

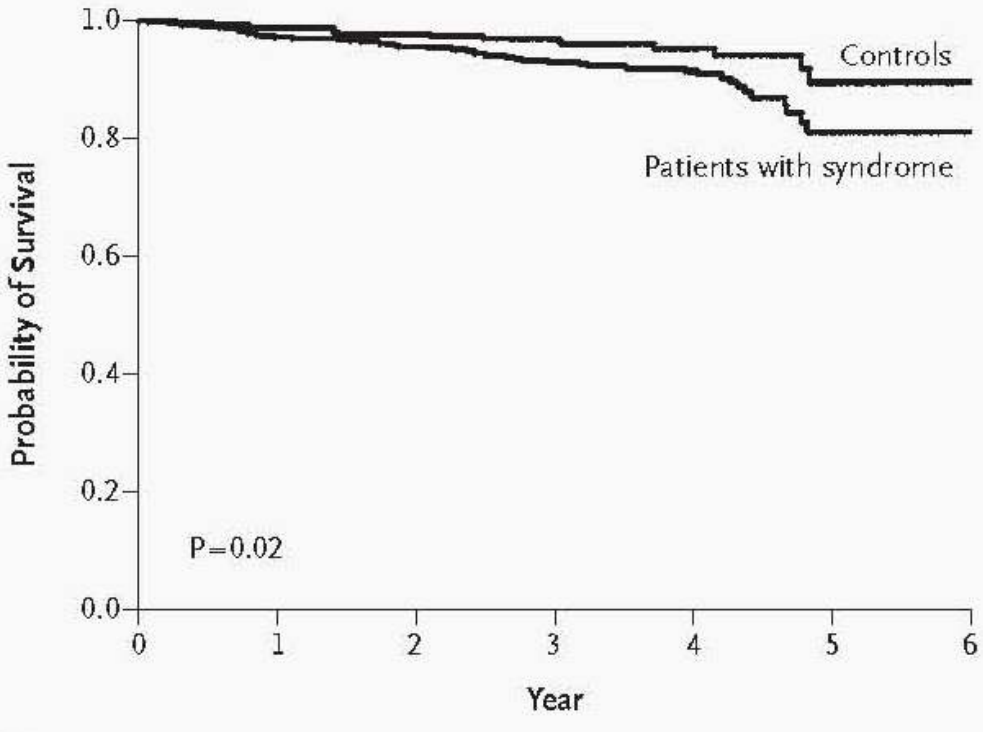
- Obstructive sleep apnea as a risk factor for stroke and death*

- Subjects were followed until death or stroke (events) or censoring
  - "In this observational cohort study, consecutive patients underwent polysomnography, and subsequent events (strokes and deaths) were verified. The diagnosis of the obstructive sleep apnea syndrome was based on an apnea–hypopnea index of five or higher (five or more events per hour); patients with an apnea–hypopnea index of less than five served as the comparison group."
  - "The Kaplan–Meier method and the log-rank test were used to compare event-free survival among patients with and those without the obstructive sleep apnea syndrome"

Notes: * Yaggi, H., et al. (2005). Obstructive sleep apnea as a risk factor for stroke and death. *New England Journal of Medicine,* 353, 19.

Examples from Literature

- Sleep apnea/death and stroke



No. at Risk							
Controls	325	266	262	228	89	23	1
Patients with syndrome	697	562	552	465	179	33	3

Figure 2. Kaplan–Meier Estimates of the Probability of Overall Survival among Patients with the Obstructive Sleep Apnea Syndrome and Controls.

Examples from Literature

- Return to work following injury: The role of economic, social, and job-related factors*

- Subjects were followed until returning to work or censoring
  - "The main dependent variable in the analysis is the time (in days) from injury to the first time the study patient returned to work. Kaplan-Meier estimates of the cumulative proportion of patients returning to work were computed. These estimates take into account how long patients were followed as well as when they returned to work. A log-rank test was used to test the association between the cumulative probability of RTW and each of the risk factors considered one at a time."

Notes: * MacKenzie, E., et al. (1998). Return to work following injury: The role of economic, social, and job-related factors. *America Journal of Public Health,* 88, 11.

Examples from Literature

- Kaplan Meier (tracking proportion HAVING event by time t, $1 - \hat{S}(t)$ as we previously defined it

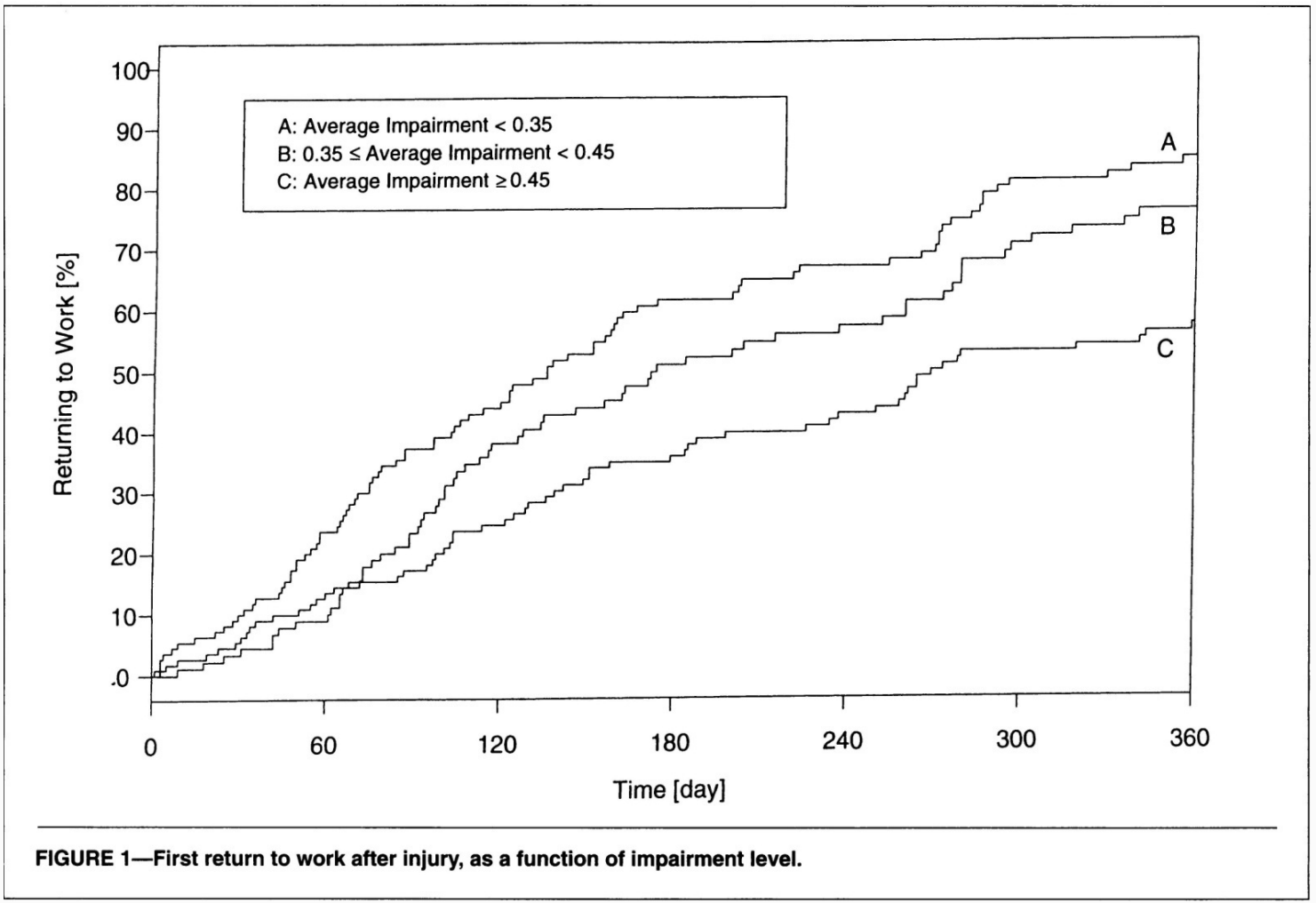


FIGURE 1—First return to work after injury, as a function of impairment level.