JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section B

Estimating the Survival Curve: The Kaplan Meier Approach

Central Problem

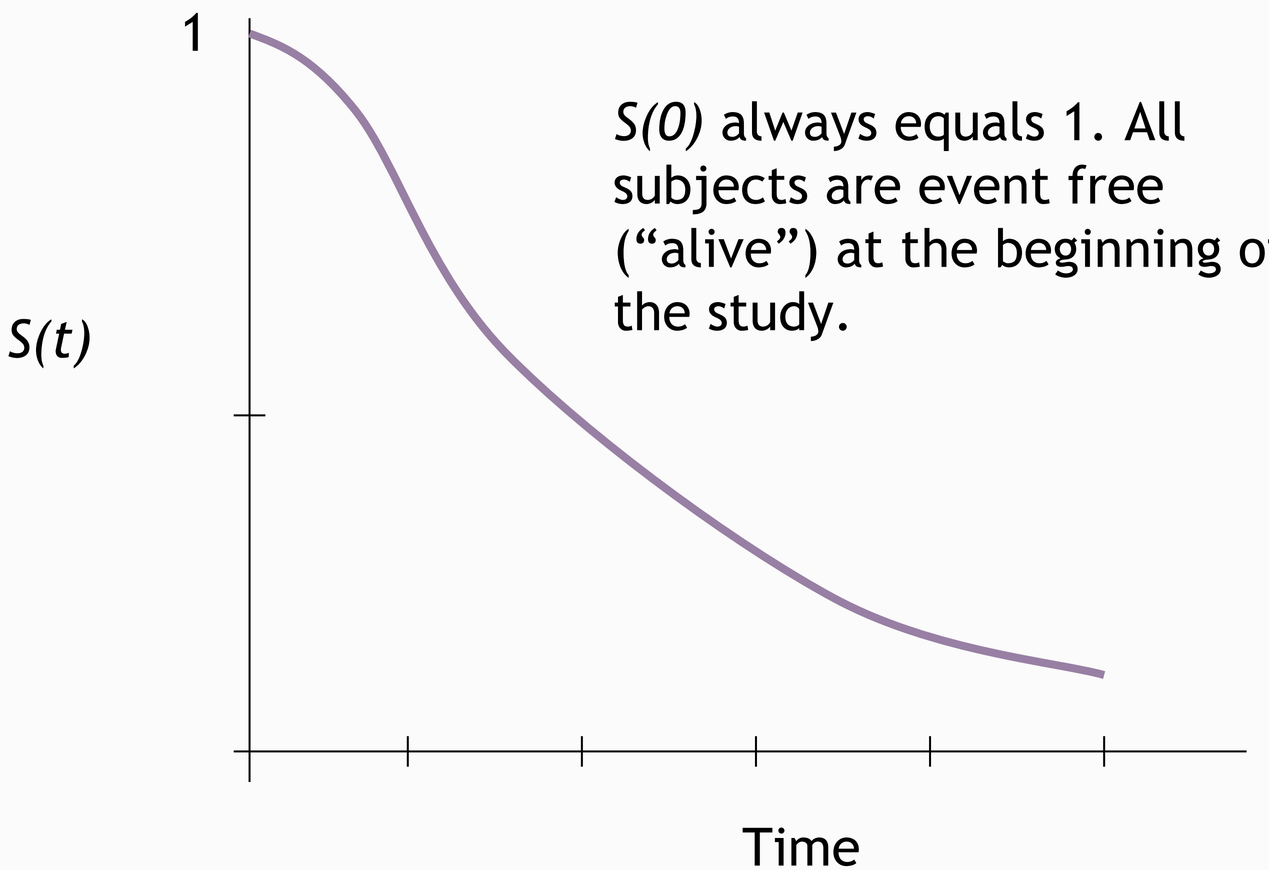
- Estimation of the "survival curve"

- *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

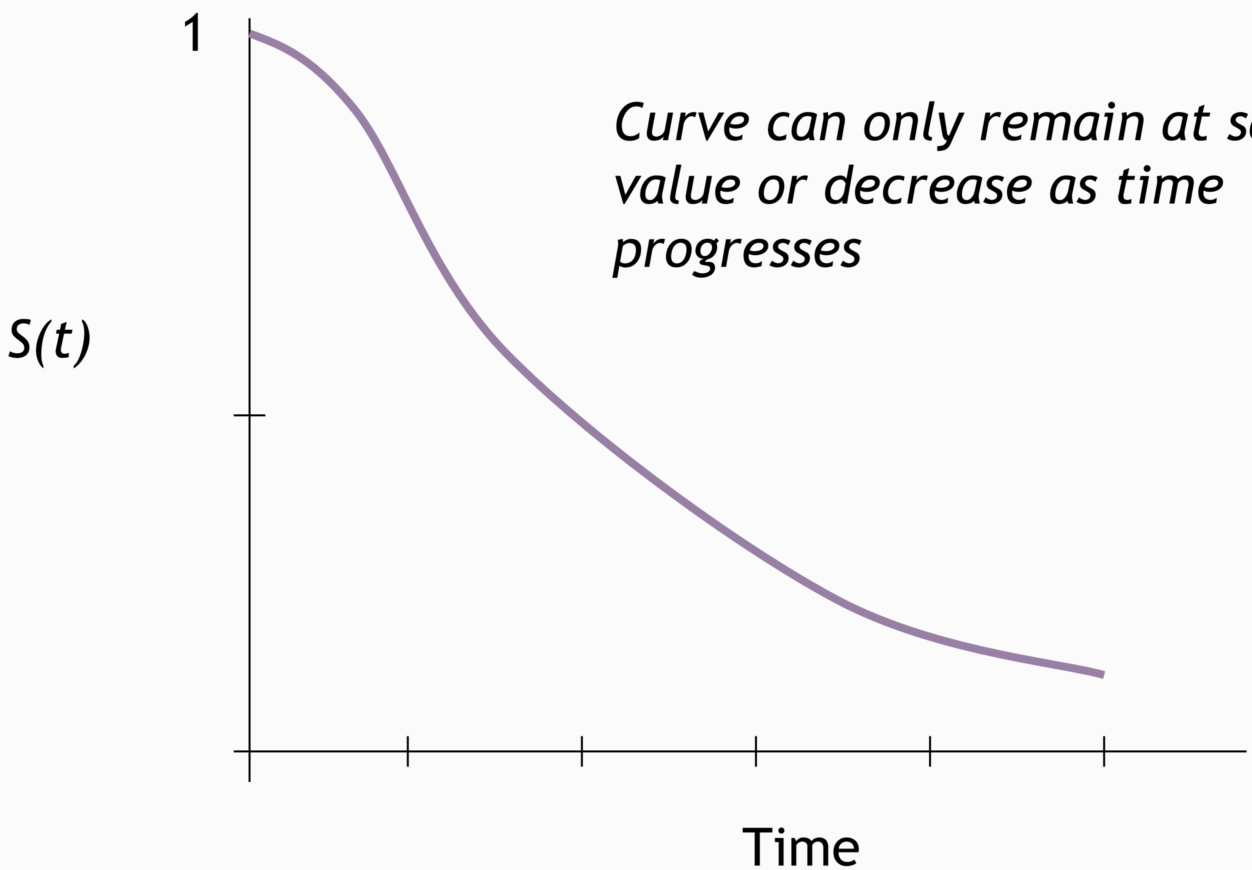


S(0) always equals 1. All subjects are event free ("alive") at the beginning of the study.

Central Problem

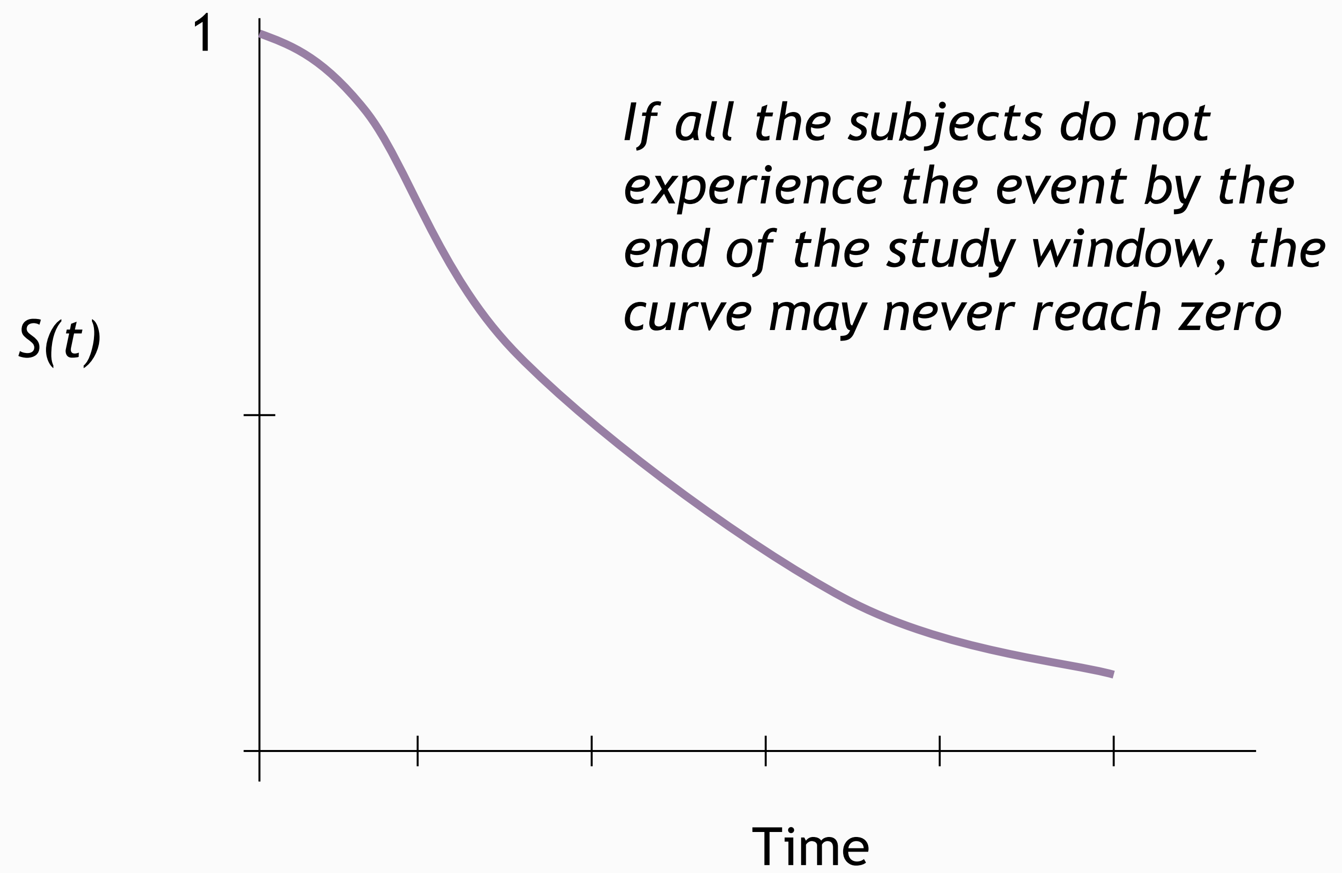
- Estimation of the “survival curve”

- *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 $\hat{S}(t)$



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

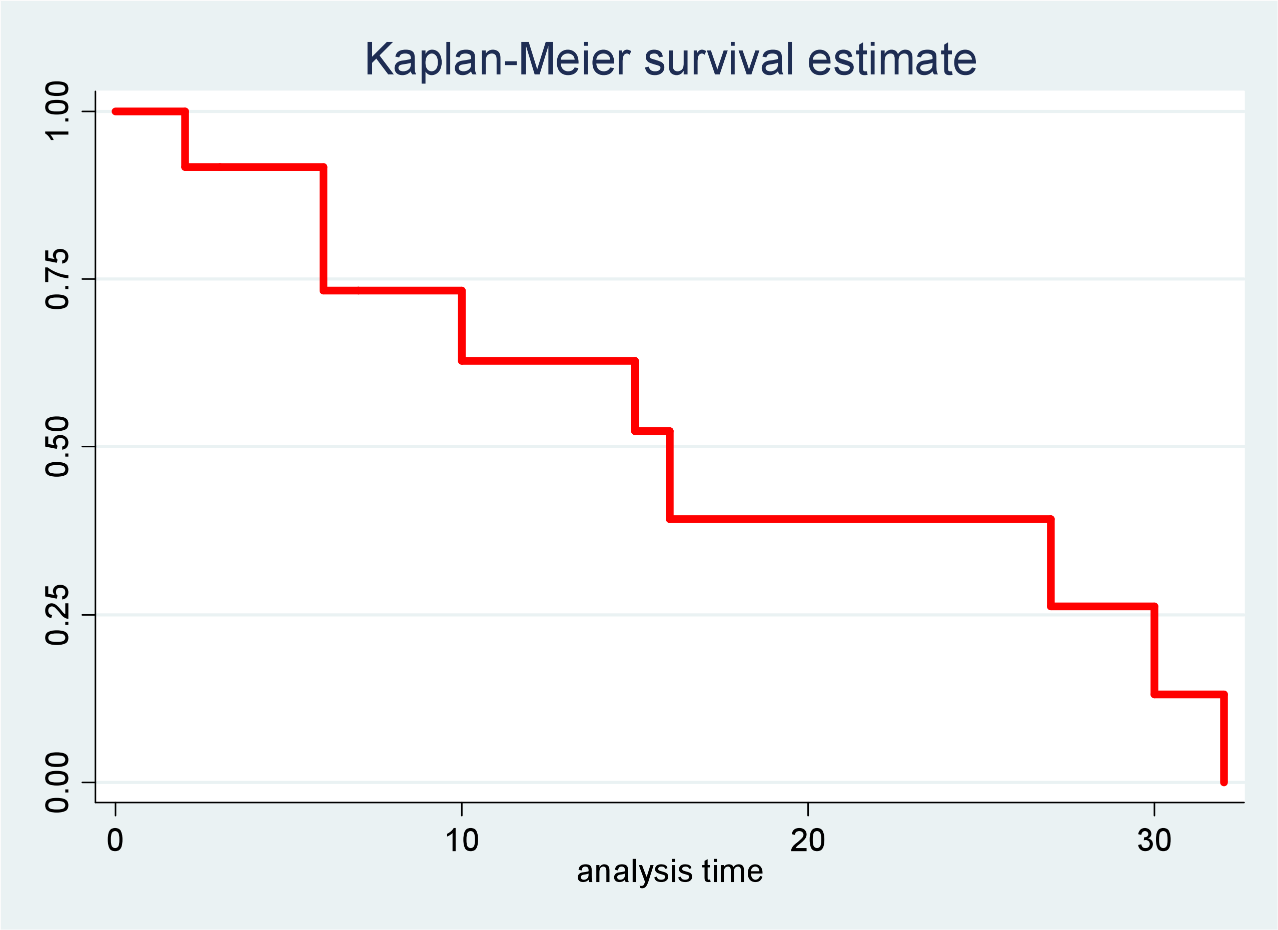
t	$\hat{S}(t)$
2	.92
6	.74
10	.64
15	.52
16	.39
27	.26
30	.13
32	0

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

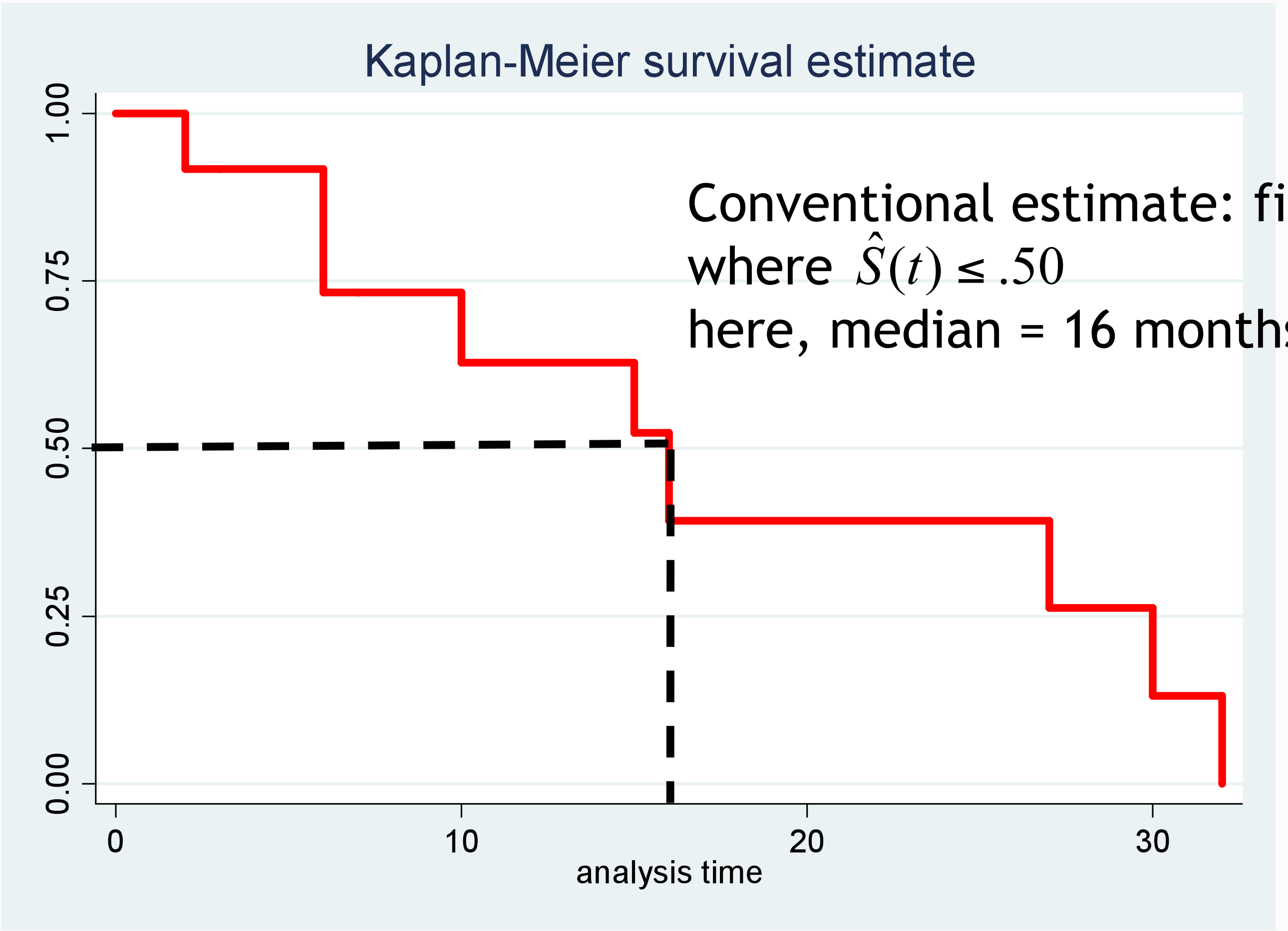
Kaplan-Meier Estimate

- 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+
  - $\hat{S}(5) = \left(\frac{N(5) - E(5)}{N(5)} \right) \times \hat{S}(1) = \left(\frac{8 - 1}{8} \right) \times .90 = .88 \times .90 = .79$

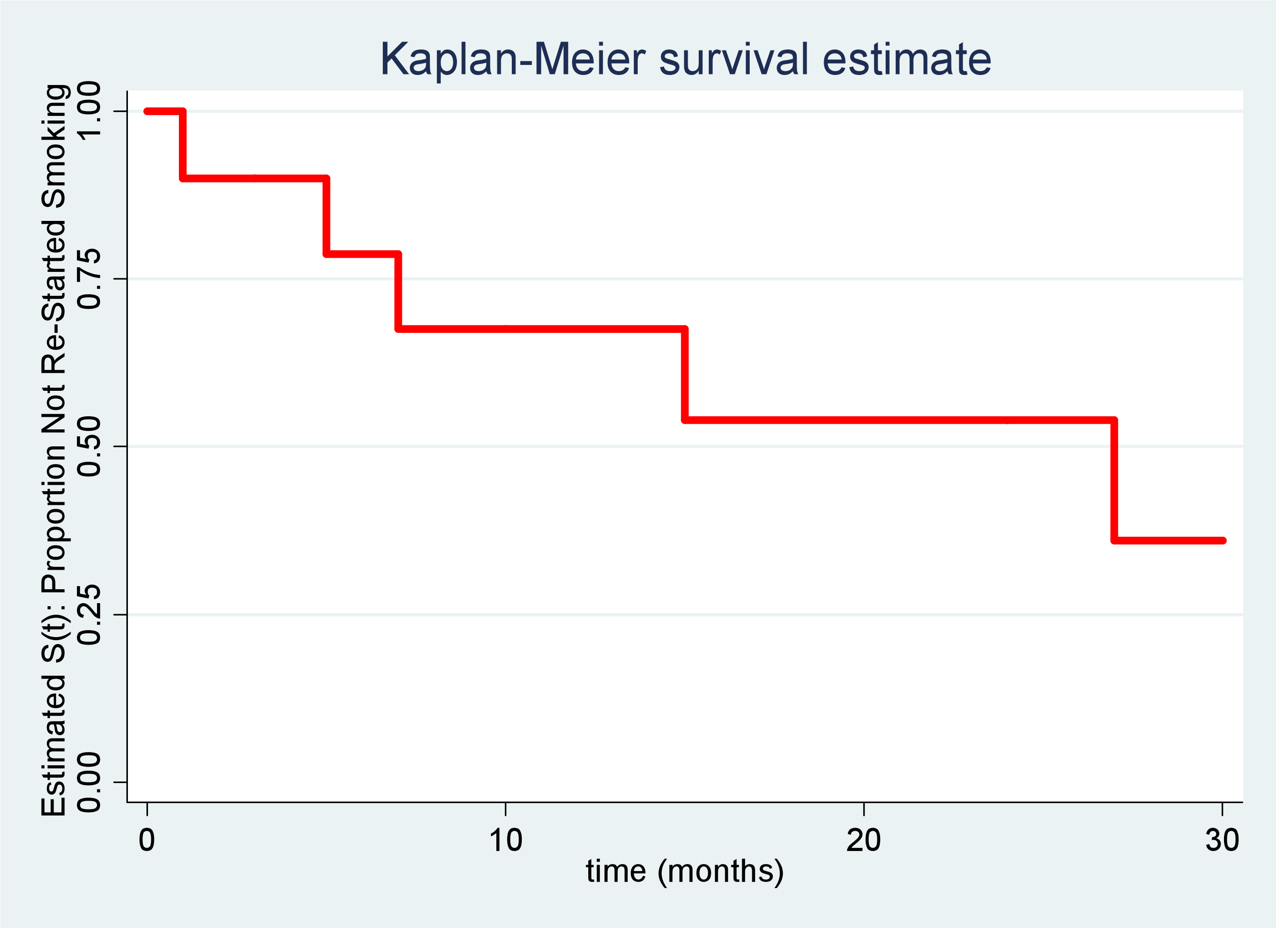
Kaplan-Meier Estimate

- 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