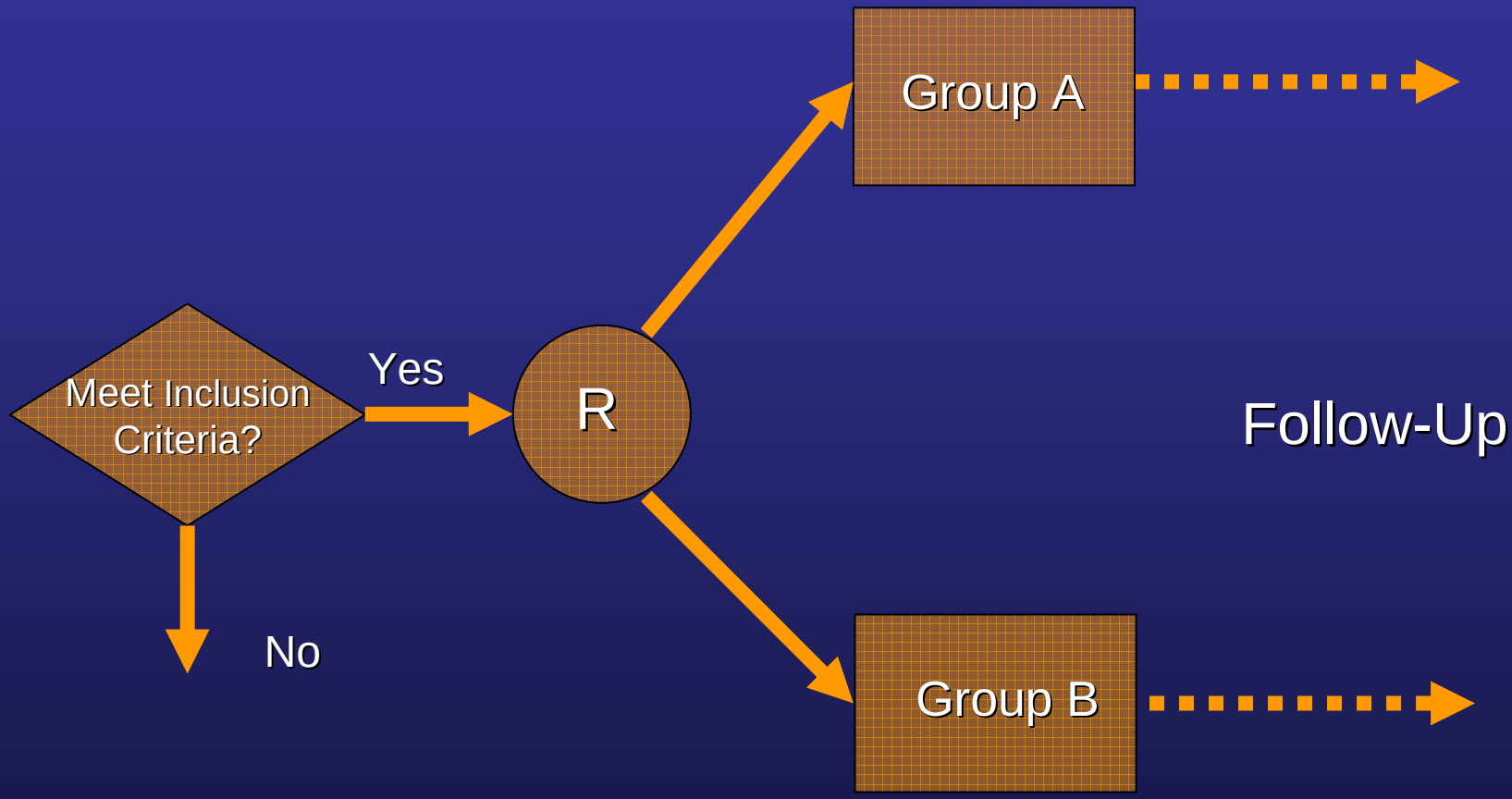
Stayin' Alive: An Introduction to Survival Analysis

David L. Streiner, Ph.D.
Director, Kunin-Lunenfeld Applied Research Unit
Assistant V.P., Research
Baycrest Centre for Geriatric Care

Professor, Department of Psychiatry
University of Toronto

A Traditional RCT:



Elements of an RCT:

- Usually, patients entered over relatively brief period
- Followed for a preset period of time
- All patients reach end-point
- Difficult to handle loss to follow-up
- Focus is on patient's *state* at follow-up

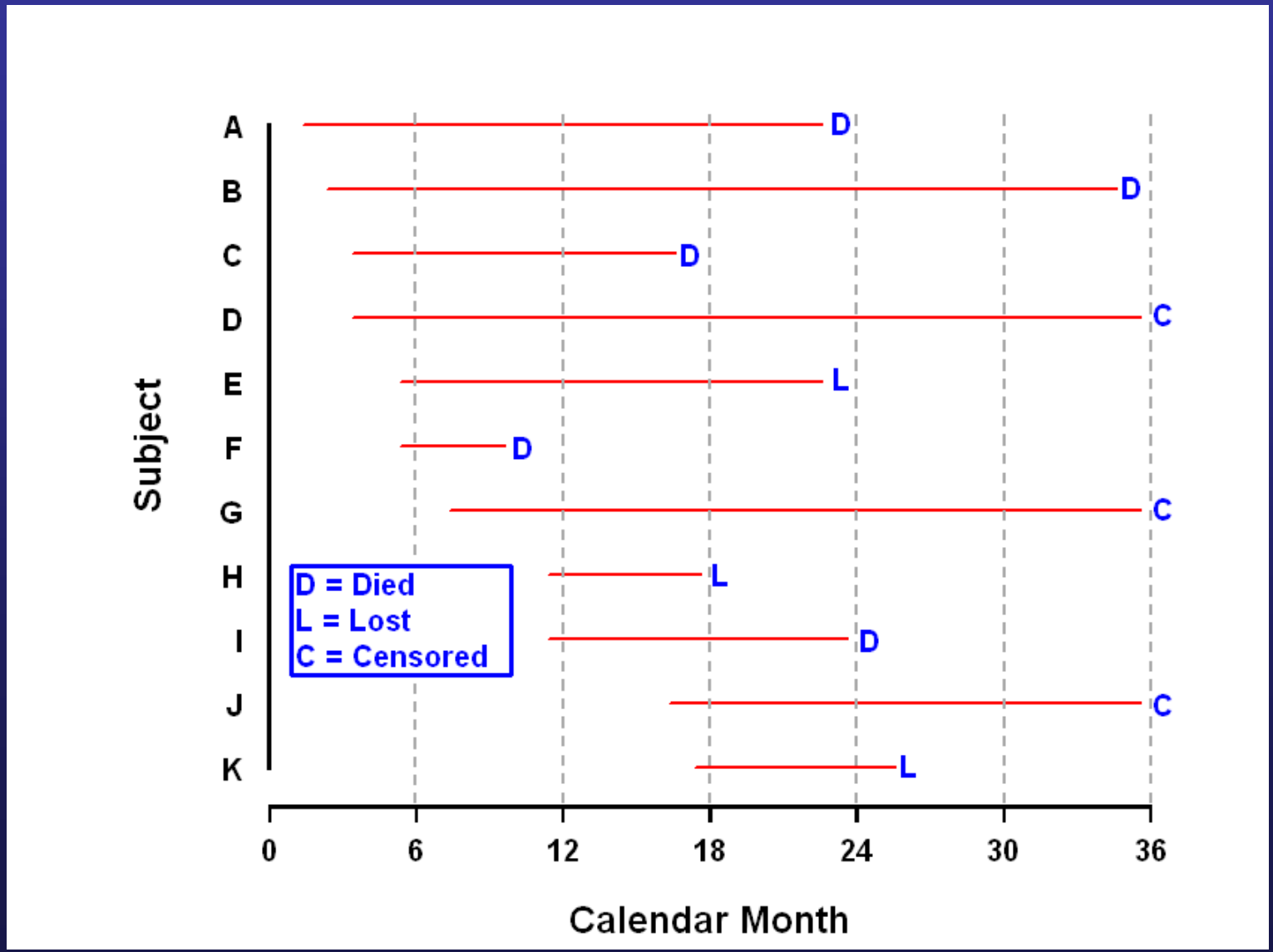
Why Survival Analysis?

- In some studies, recruitment takes 2-3 years
- Patients followed until end of study
- If minimum follow-up is 5 years, then:
  - Some patients followed for 5 years; others up to 8 years
  - Much greater chance of attrition
- Some patients don't reach end-point
- Focus is on *time* to end-point

Possible Endpoints:

- Died during trial
- Lost to follow-up
- Still alive at end of trial (“right censored”)

Results of 11 Patients:



How Not *to Analyze the Data:*

1. Mean Survival
   - Use only those for whom we have complete data
   - Looks only at those who died during study
   - Problems:
     - Throws out much of the data
     - Loads the dice against us, since it doesn't count those who survive (censored)

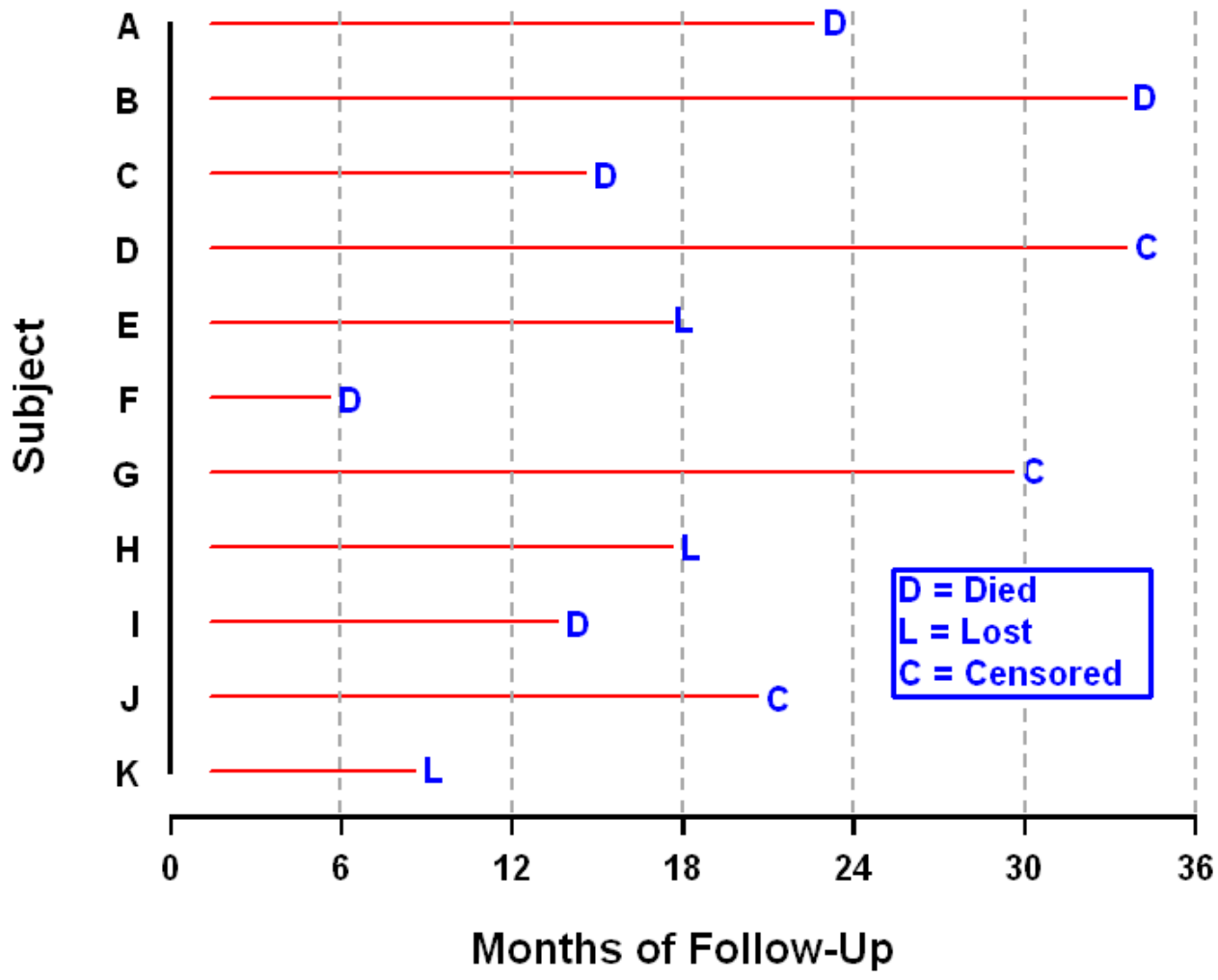
How Not *to Analyze the Data:*

2. Survival Rate
   - Counts people still alive at a fixed period (e.g., 5 years)
   - Often used in cancer trials
   - Problems:
     - Doesn't count those who were lost or censored before 5 years
     - Doesn't use most of the data
     - Time point is arbitrary

How to Analyze the Data:

3. Survival Analysis
   - Shifts from looking at *people* to looking at *time*
   - See how many people still alive at the end of multiple time points
   - Begin by shifting each line to a common starting point

Shifted to a Common Start:



How to Analyze the Data:

- For each time period, figure out:
  - Number of people at risk (still in the study at beginning of interval)
  - Number who died
  - Number who were lost to follow-up and who were censored (from our perspective, they're the same – we don't know how long they survived after we stopped following them)

How to Analyze the Data:

Number of Months in Study	Number at Risk	Number Died	Number Lost
0 - 6	11	1	0
6 - 12	10	0	1
12 - 18	9	2	2
18 – 24	5	1	1
24 – 30	3	0	1
30 - 36	2	1	1

How to Analyze the Drop-Outs:

- What do we do with people who were lost or censored?
  - If we say they're "at risk," we underestimate probability of dying in that period
  - If we don't count them, we overestimate the probability
  - Compromise by assuming they were at risk for half the period

How to Analyze the Data:

- For each time period, figure out the *hazard*
- *Hazard* = risk of dying in that period

- Hazard = $\dfrac{\text{Number of Deaths}}{\text{Number at Risk} - \text{Number Lost} / 2}$

How to Analyze the Data:

- For the period 0 – 6 months:
  - 11 people at risk
  - 1 died
  - 0 lost
- Hazard = 1/11 = .0909

How to Analyze the Data:

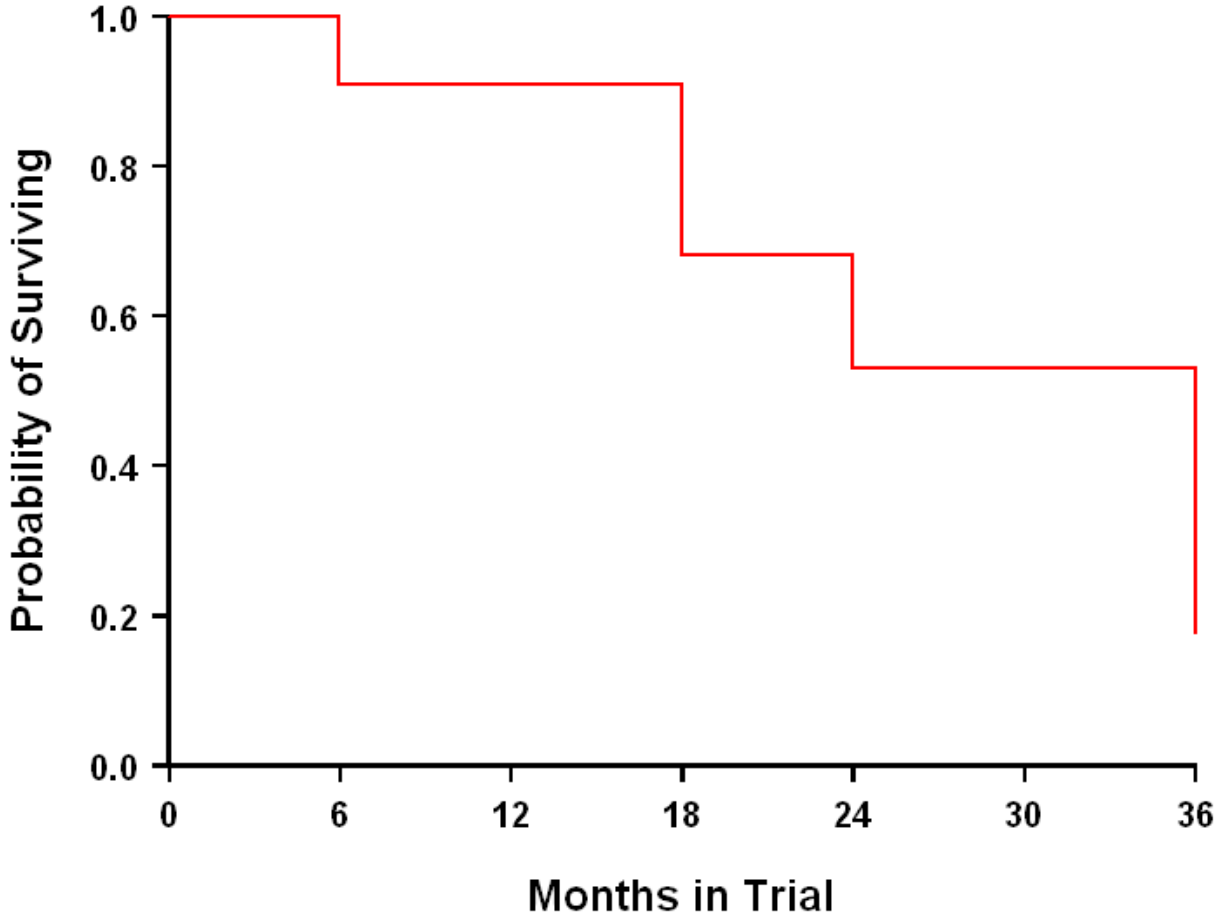
- For the period 12 – 18 months:
  - 9 people at risk
  - 2 died
  - 2 lost
- Hazard = 2/(9 – 1) = .2500

How to Analyze the Data:

Number of Months in Study	Hazard	Probability of Survival	Cumulative Probability of Survival
0 - 6	0.0909	0.9091	0.9091
6 - 12	0.0000	1.0000	.09091
12 - 18	0.2500	0.7500	0.6818
18 – 24	0.2222	0.7778	0.5303
24 – 30	0.0000	1.0000	0.5303
30 - 36	0.6667	0.3333	0.1768

Now plot the cumulative probability of surviving:

The Survival Function:



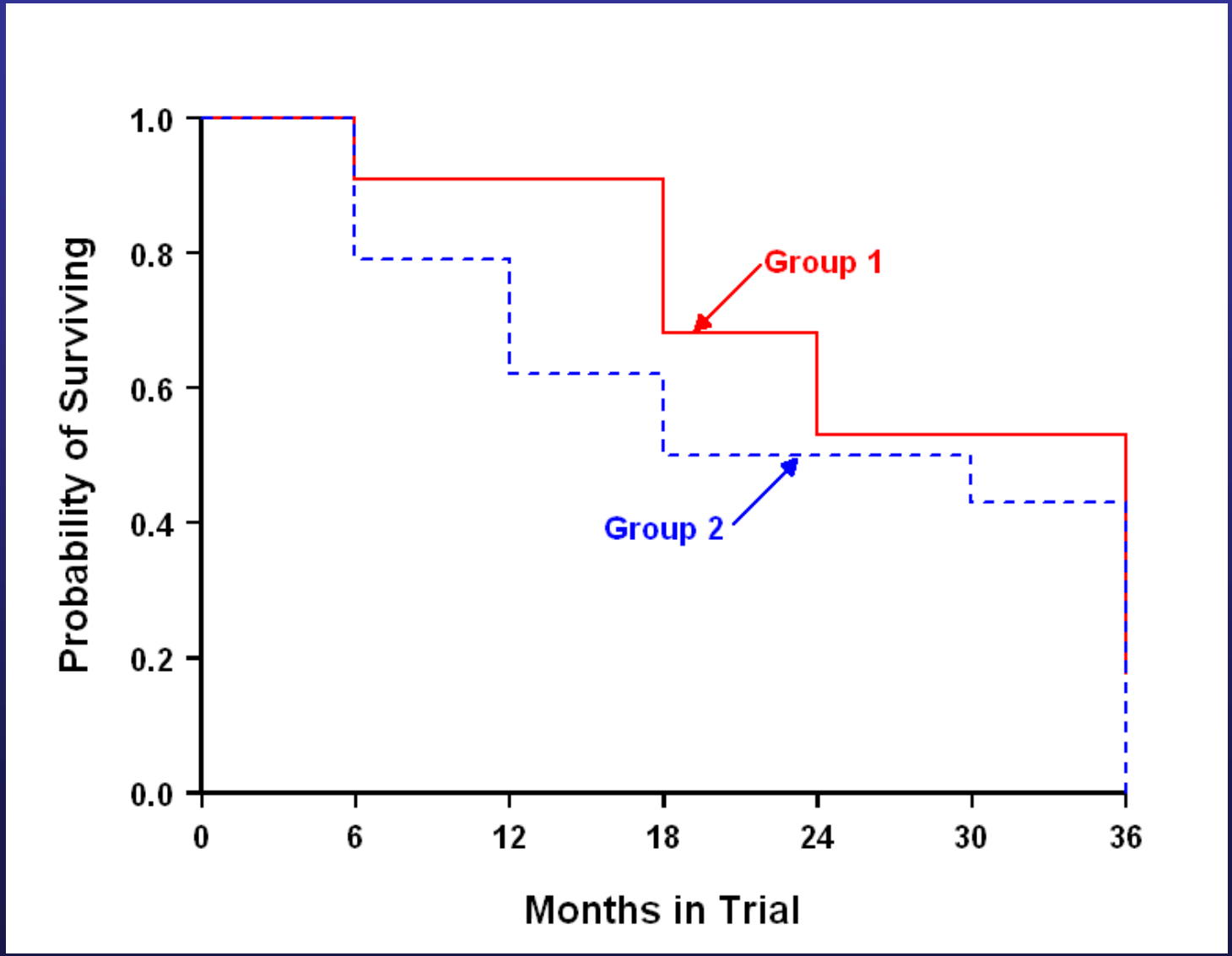
Variations on a Theme:

- Called the *actuarial* method
  - Resembles what actuaries do
  - Hazard calculated at fixed periods
  - Used if don't know exact time of death
- *Kaplan-Meier* method
  - Hazard calculated whenever death occurs
  - Need to know exact time of death
- Not much difference

More Than One Group:

- Real power of survival analysis
- Can compare two or more groups

Comparing Groups:



Comparing Groups:

- Can calculate *Relative Risk* of surviving:
- $RR = \frac{1 - P_1}{1 - P_2}$

Comparing Groups:

- In *Survival Rate*, do at the end of one fixed period (e.g., 5 year survival rate)
- In *Survival Analysis*, do at the end of every fixed period (actuarial) or every time there's a death (Kaplan – Meier)
- Sum up relative risks with *Mantel-Cox chi-squared*

Comparing Groups:

- Can also calculate *hazard ratio:*

- Hazard Ratio = $\dfrac{\text{Hazard (Group 1, Time } t\text{)}}{\text{Hazard (Group 2, Time } t\text{)}}$

Comparing Groups:

- Difference between *relative risk* and *hazard ratio*:
  - Both are a comparison of risk of living (or dying) between the groups
  - RR compares groups *from entry into trial*
  - Hazard ratio compares people at time *t who have survived up to time (t* – 1)
    - E.g., Given that you survived 18 months, what's the risk of dying by 24 months

More Variations on a Theme:

- Sometimes want to *adjust for* baseline differences (e.g., age, sex, co-morbidities), or *examine effects* of these differences

- Use *Cox Proportional Hazards Model*
  - A Kaplan – Meier survival analysis with more variables thrown into the mix

Should We Compare Groups?

- Definitive answer – it all depends
  - Was the comparison decided on *before* the study began?
  - Was the *p* level adjusted to account for many analyses?
- If the answer is *Yes*, results are kosher
- If the answer is *No*, results are traif

Assumptions:

- All patients enter trial at an equivalent point
- End-point the same for all people
- Loss to follow-up unrelated to outcome
- No changes over time regarding
  - Diagnosis
  - Treatment

Questions?