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

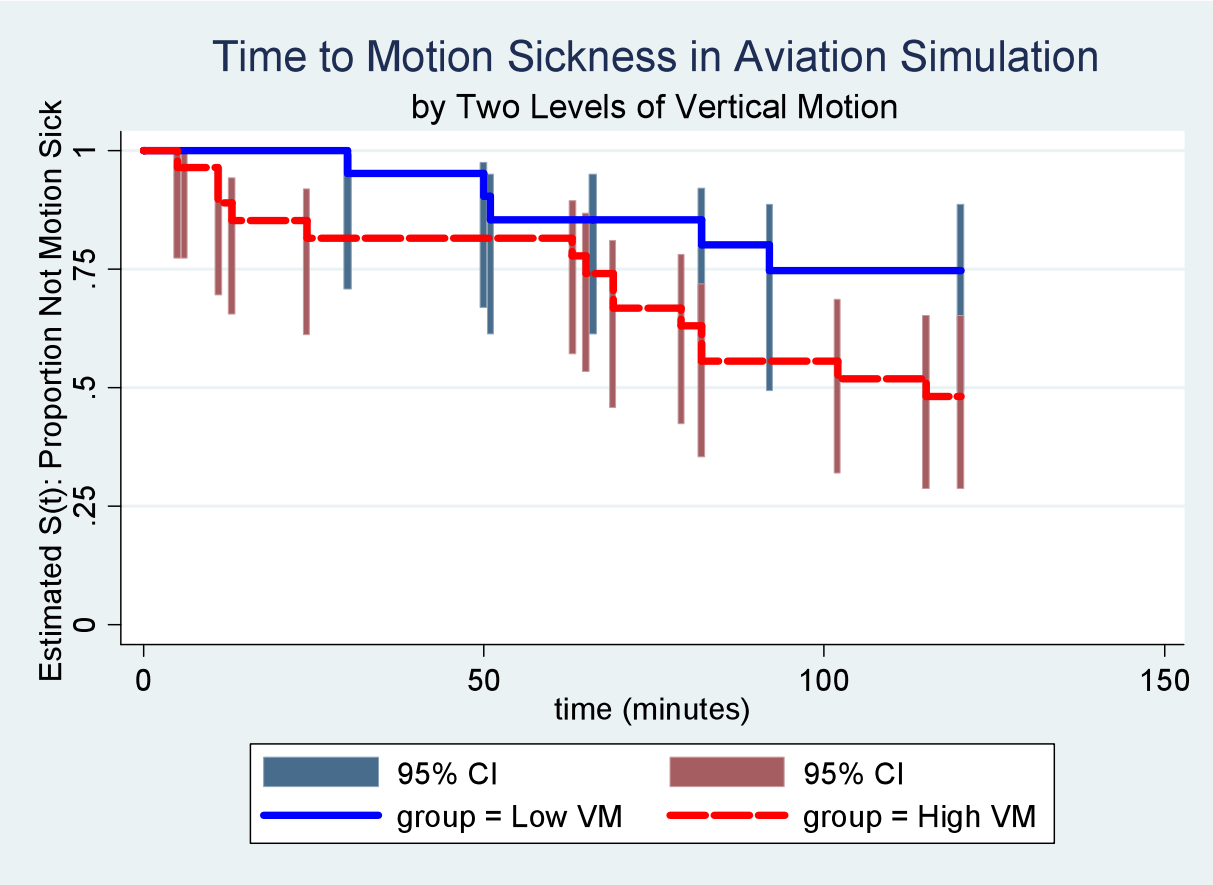
Examples of Logrank and Breslow-Gehan Test

- 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.

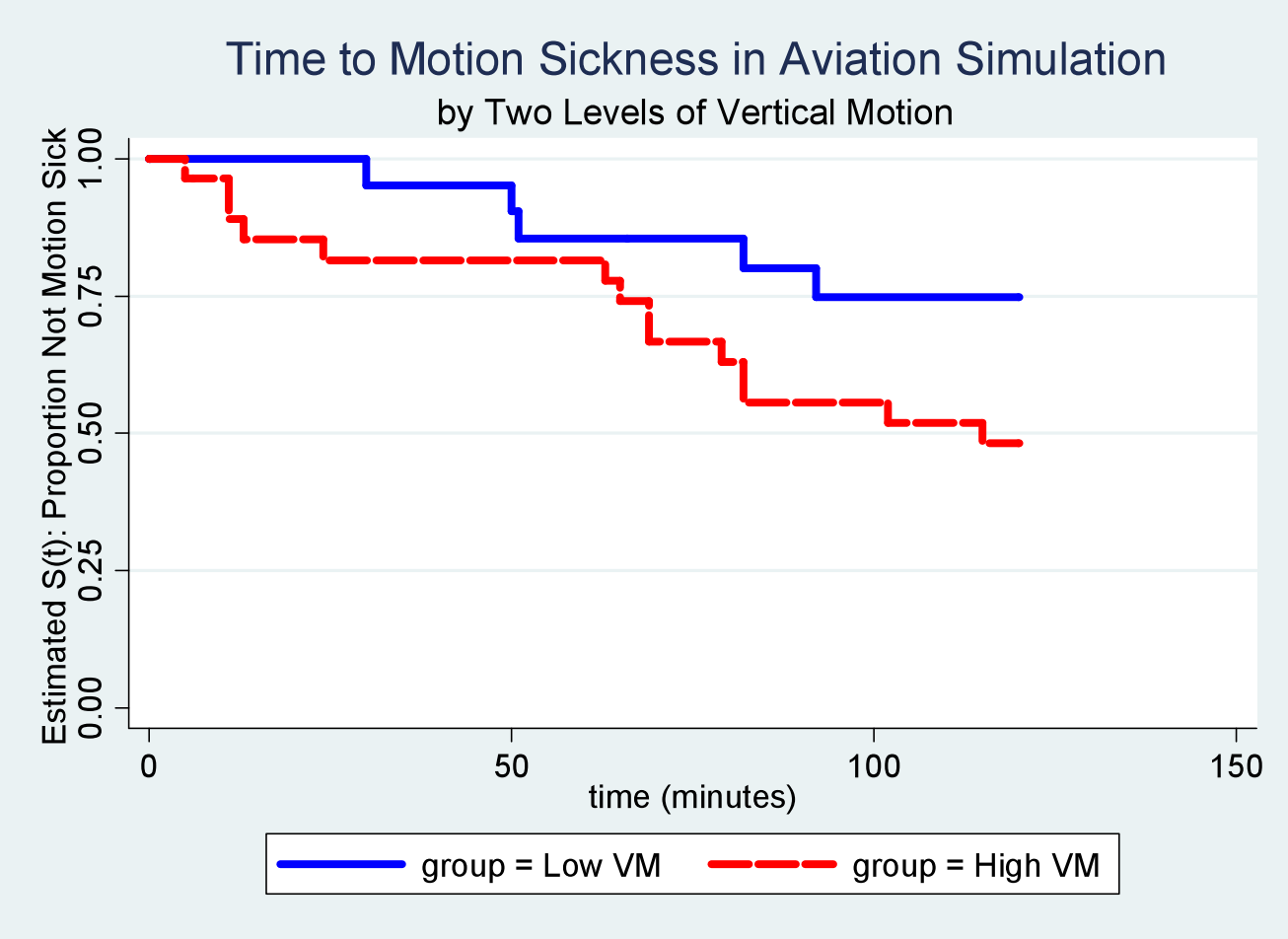
Examples of Logrank and Breslow-Gehan Test

- 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

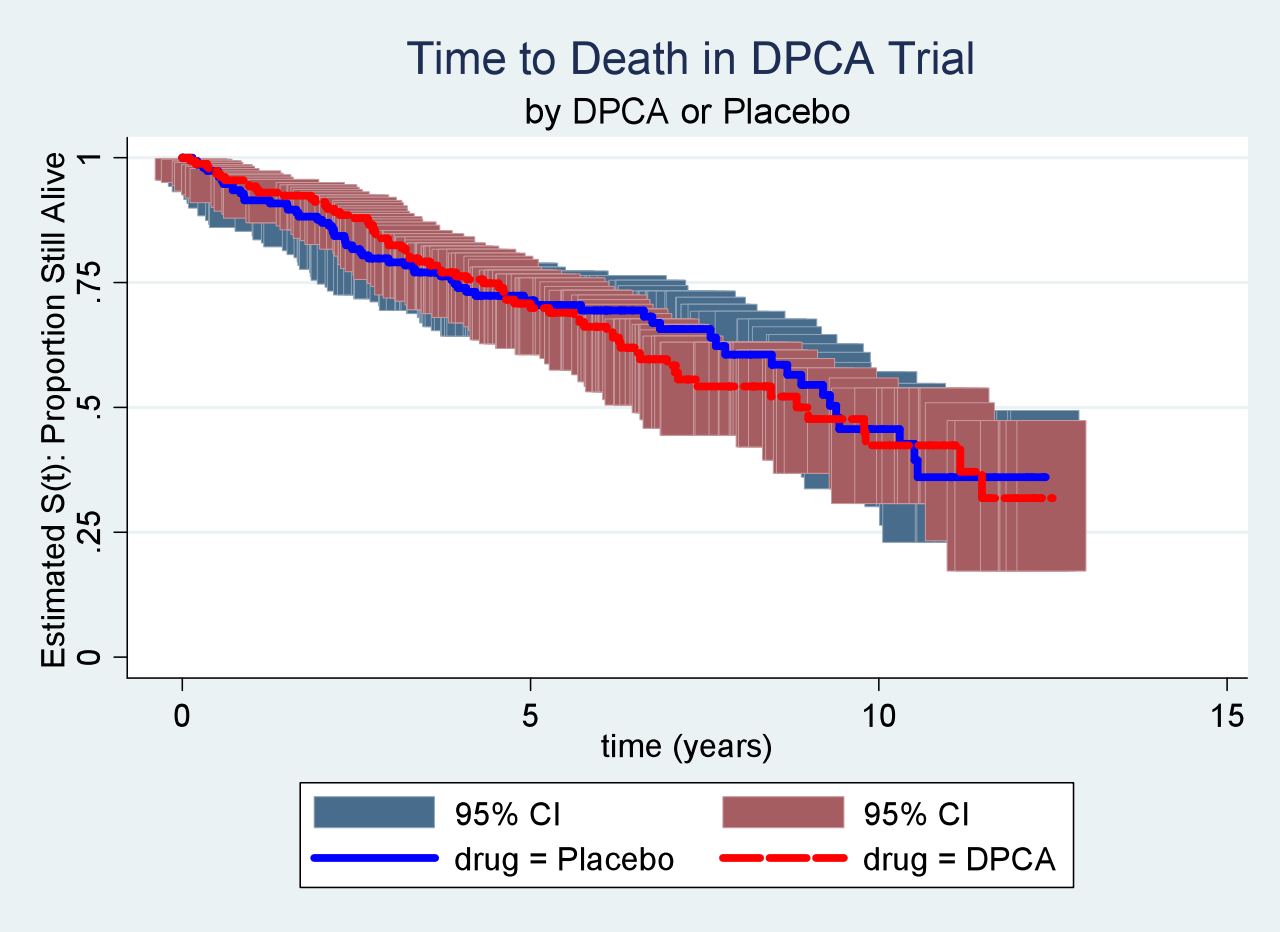
- Breslow/Wilcoxon/Gehan results:
  - p = .075

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

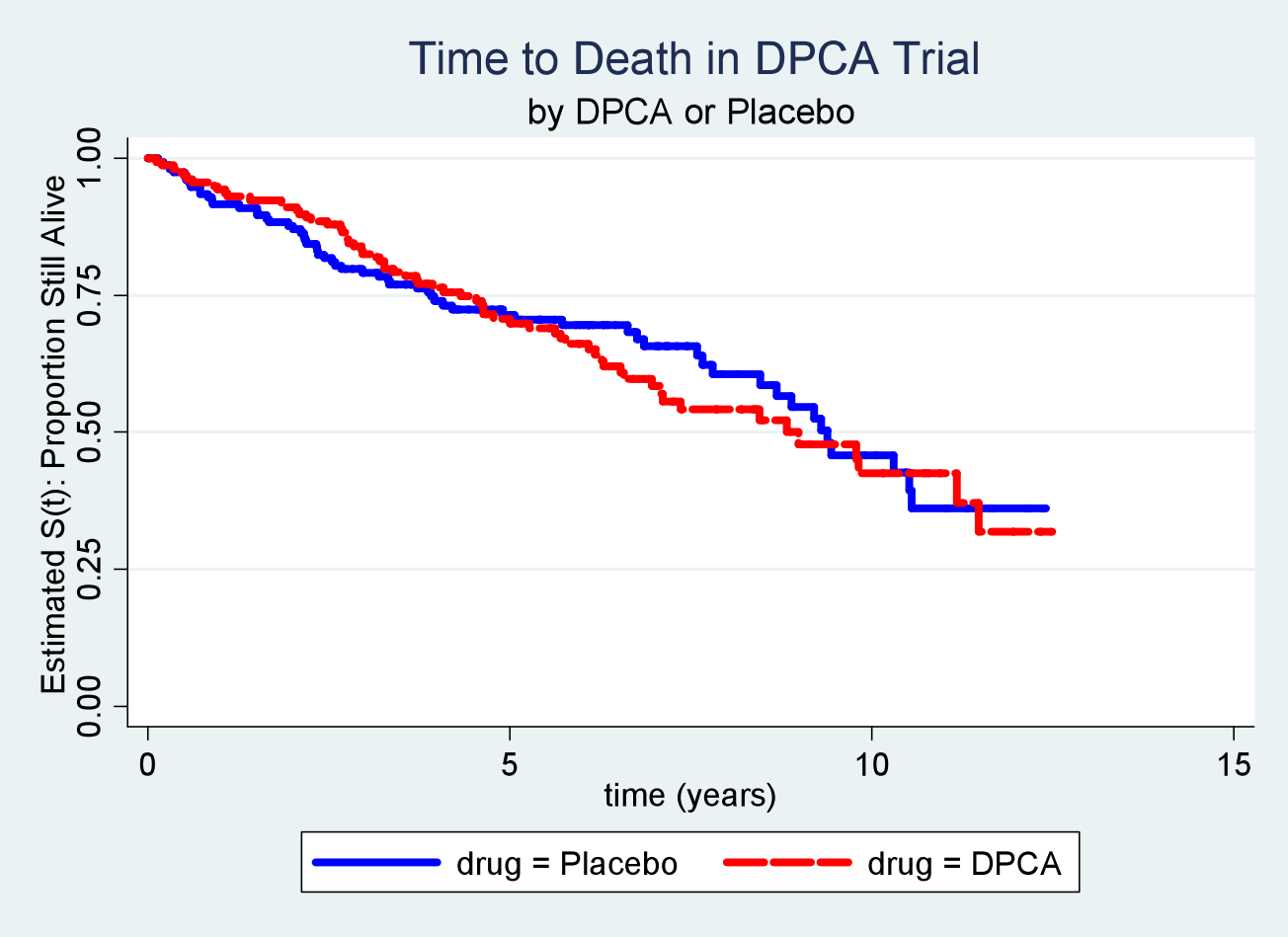
Examples of Logrank and Breslow-Gehan Test

- 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

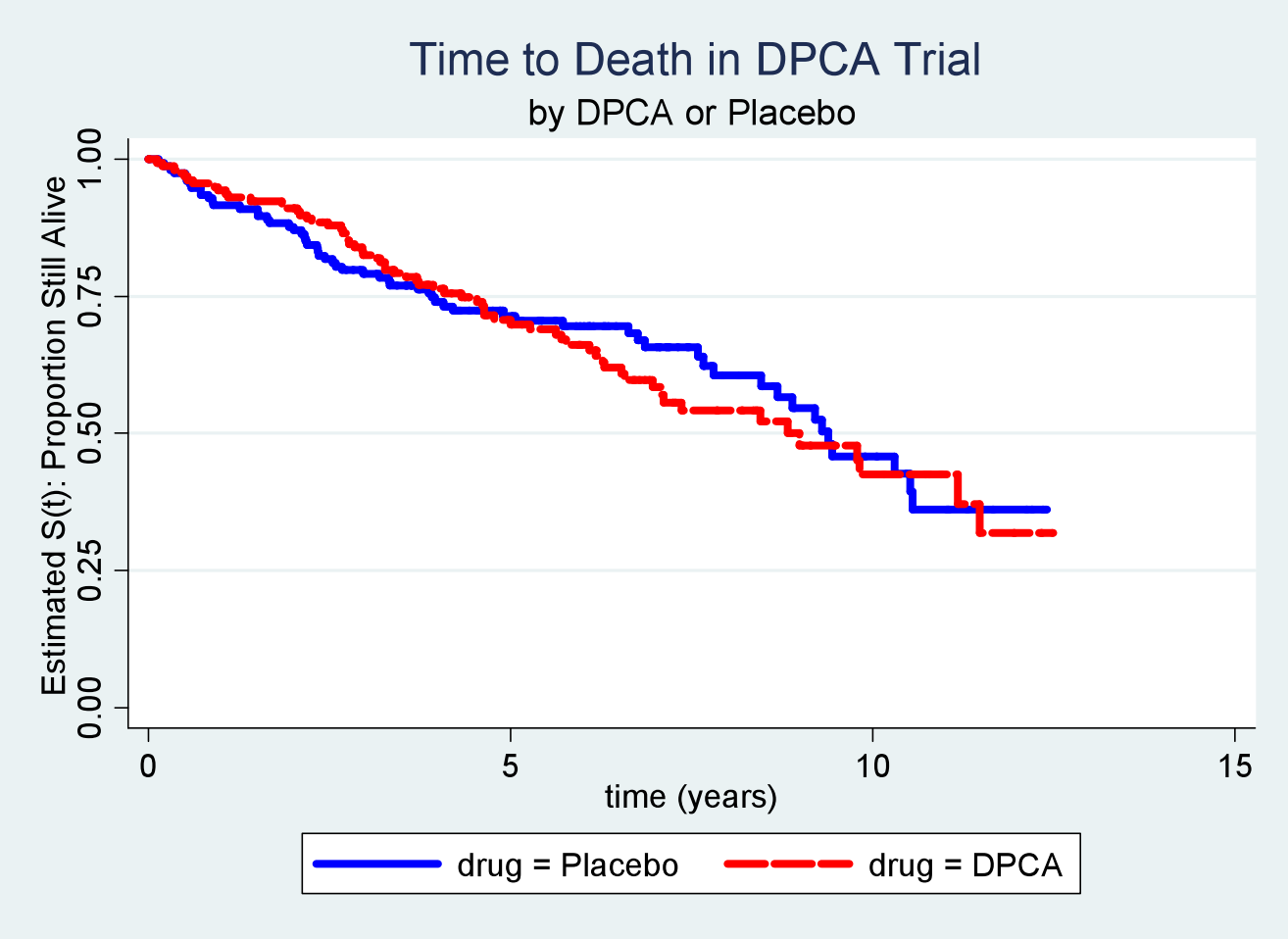
- Hypothesis test setup
  - H_o: $S_{DPCA}(t) = S_{PLACEBO}(t)$
  - H_A: $S_{DPCA}(t) \neq S_{PLACEBO}(t)$

- 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

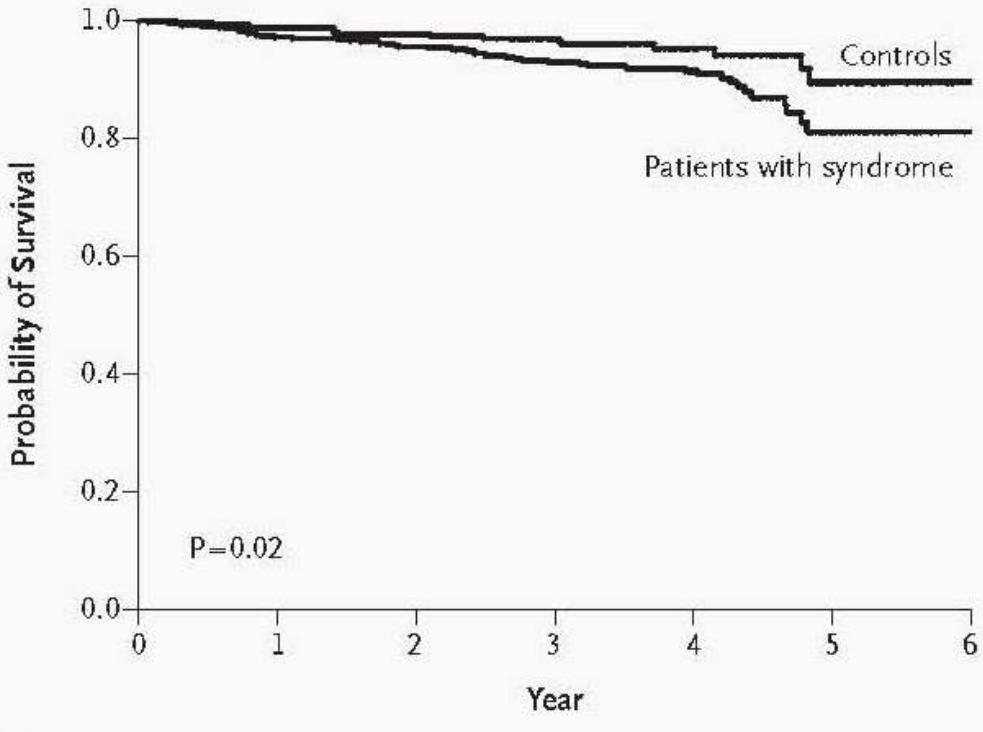
- Obstructive sleep apnea as a risk factor for stroke and death*

- 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

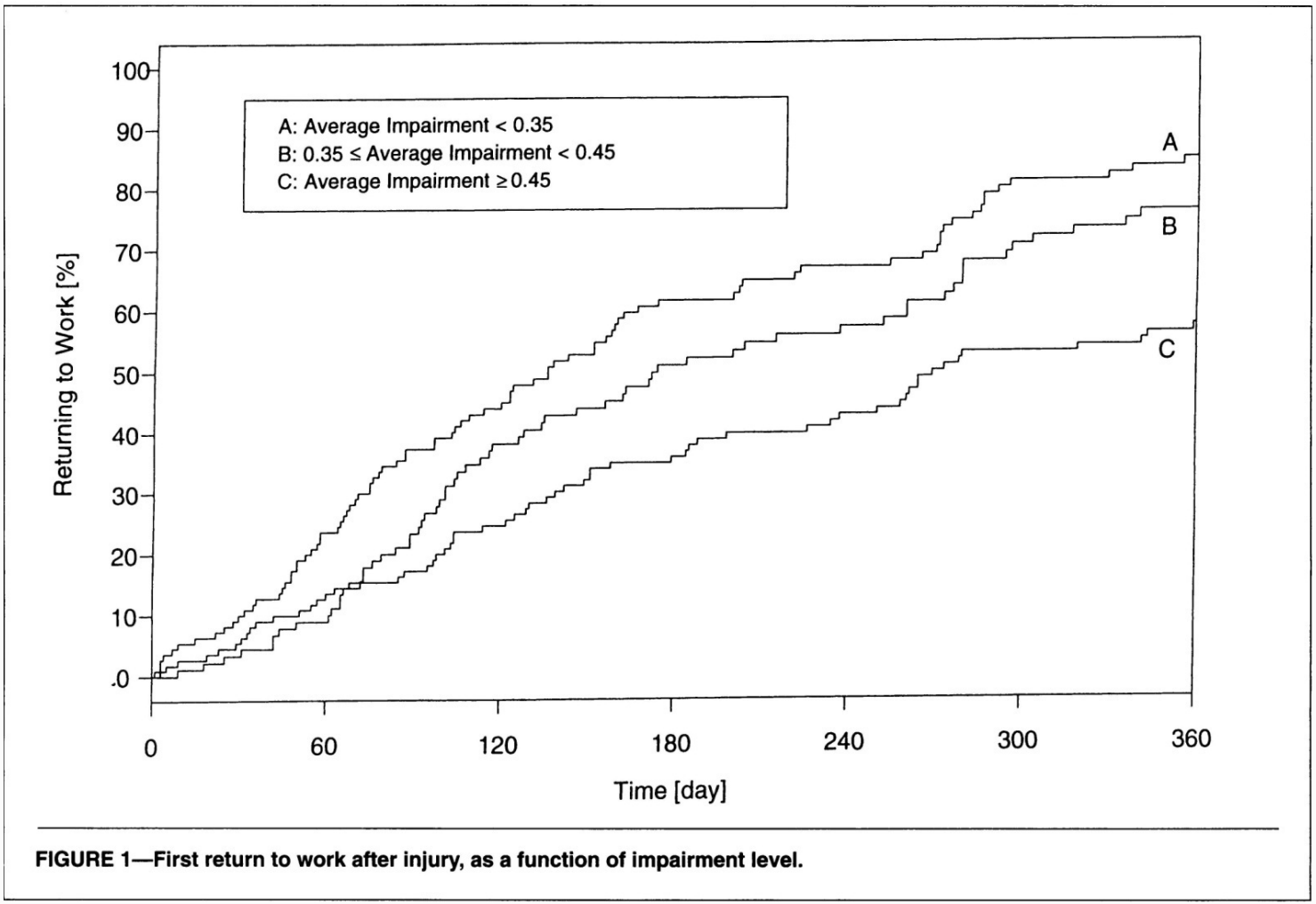


FIGURE 1—First return to work after injury, as a function of impairment level.