MODULE 6: INTRODUCTION TO SURVIVAL ANALYSIS

Elizabeth Brown, ScD
Member, Fred Hutchinson Cancer Research Center
and
Research Professor
Department of Biostatistics
University of Washington

OVERVIEW

- Session 1
  - Introductory examples
  - The survival function
  - Survival Distributions
  - Mean and Median survival time

- Session 2
  - Censored data
  - Risk sets
  - Censoring Assumptions
  - Kaplan-Meier Estimator and CI
  - Median and CI

- Session 3
  - Two-group comparisons: logrank test
  - Trend and heterogeneity tests for more than two groups

- Session 4
  - Introduction to Cox regression

OVERVIEW – MODULE 9

Module 9: Survival analysis in Clinical Trials

- Quick review of basics
- Estimating survival after Cox model fit
- More two-sample tests
  - Weighted logrank
  - Additional tests based on functionals and metrics
- Adjustment, precision and post-randomization variables
- Power
- Choice of outcome
- Information accrual in sequential monitoring

OVERVIEW – MODULE 13

Module 13: Survival analysis for Observational Data

- More complicated Cox models
  - Adjustment
  - Interaction
- Competing Risks
- Choice of time variable
- Left Entry/Truncation
- Immortal time bias
- Index event bias
- Time-dependent covariates

MODULE 6
INTRODUCTION TO SURVIVAL ANALYSIS

SESSION 1:
SURVIVAL DATA: EXAMPLES

PRELIMINARIES

- No prior knowledge of survival analysis techniques assumed
- Familiarity with standard one- and two-sample statistical methods (estimation and testing) is assumed
- Emphasis on application rather than mathematical details
- Examples

SESSIONS/BREAKS

- 8:30 – 10:00
  - Break until 10:30
- 10:30 – 12:00
  - Break until 1:30
- 1:30 – 3:00
  - Break until 3:30
- 3:30 – 5:00

WHAT IS SURVIVAL ANALYSIS ABOUT?

- Studies the occurrence of an event over time
  - Time from randomization to death (cancer RCT)
  - Time from acceptance into a heart transplant program to death
  - Time from randomization to diagnosis of Alzheimer's Disease in a prevention trial
  - Time from randomization to ovarian cancer death in a randomized screening trial
  - Time from birth to removal of supplementary oxygen therapy
  - Time from first VTE diagnosis to recurrent VTE

YOUR EXAMPLES

WHAT IS SURVIVAL ANALYSIS ABOUT?

- Explores factors that are thought to influence the chance that the event occurs
  - Treatment
  - Age
  - Gender
  - Body Mass Index
  - Diet

  - Etc.

EXAMPLE 1

- Levamisole and Fluorouracil for adjuvant therapy of resected colon carcinoma
  Moertel et a. l, 1990, 1995

- 1296 patients, enrolled 1 – 5 weeks after surgery

- Stage B_2 or C

- 3 unblinded treatment groups in stage C (2:1:1 ratio)

  – Observation only

  – Levamisole (oral, 1yr)

  – Levamisole (oral, 1yr) + fluorouracil (intravenous 1yr)

Moertel CG, Fleming TR, Macdonald JS, et al. (1990) NEJM: 322(6):352–358.

Moertel CG. et al (1995). Annals of Internal Medicine: 122(5):321.

EXAMPLE 1

- Randomization
  - Dynamic method based on accrued:
  - For B_2, extent of invasion, time since surgery
  - For C, extent of invasion, time since surgery, number of lymph nodes involved

EXAMPLE 1

- Statistical analysis
  - Survival primary outcome (recurrence secondary)
  - Kaplan-Meier survival curves
  - Log-rank statistic
  - Cox proportional-hazards model for all multivariable analysis
  - Backward regression, maximal partial-likelihood estimate statistic
  - O'Brien-Fleming boundary for sequential monitoring; stopped early for stage C

EXAMPLE 1
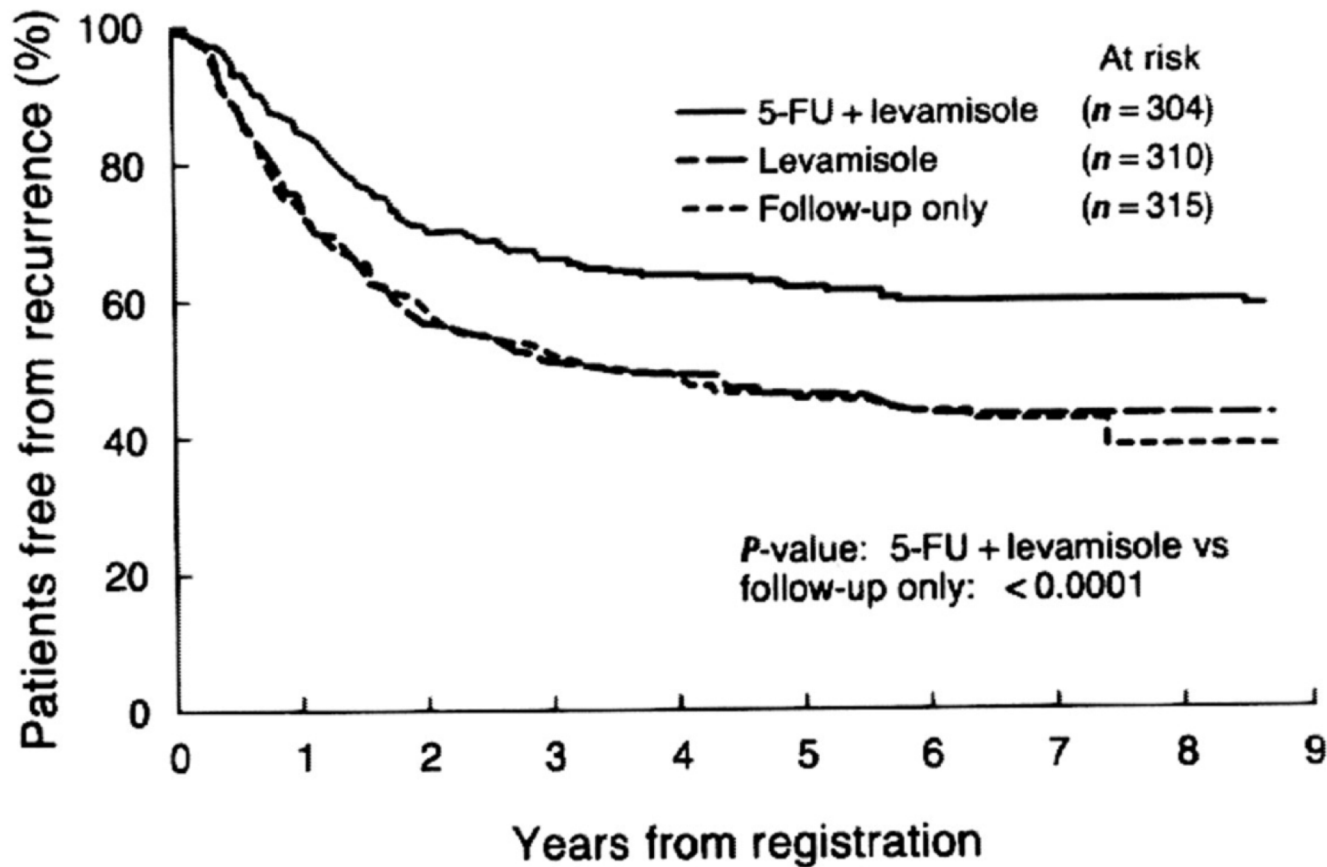
Figure 1: Recurrence-free interval according to treatment arm. Patients who died without recurrence have been censored. 5-FU = fluorouracil.

From: Fluorouracil plus Levamisole as Effective Adjuvant Therapy after Resection of Stage III Colon Carcinoma: A Final Report
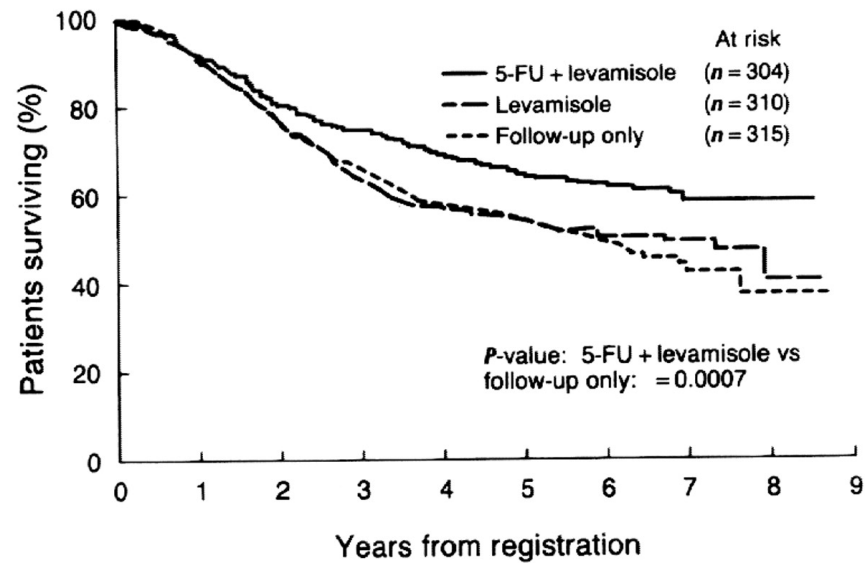
Figure Legend:

Survival according to treatment arm. 5-FU = fluorouracil.

EXAMPLE 1

- Results (stage C) after 2nd interim analysis
- Fluorouracil + Levamisole reduced the
  - Recurrence rate by 41% (95% CI 23% - 54%) (p<0.0001)
  - Death rate by 33% (95% CI 10% - 50%) (p<0.006)
- Levamisole reduced the
  - Recurrence rate by 2%
  - Death rate by 6%
- Toxicity was mild (with few exceptions)
- Patient compliance excellent

EXAMPLE 1

- R survival package data "colon"
  - 929 eligible stage C patients (971 randomized – 42 ineligible)
  - Treatment groups (rx)
  - Sex, age
  - Obstruction of colon by tumor (obstruct)
  - Perforation of colon (perfor)
  - Adherence to nearby organs (adhere)
  - Number of lymph nodes with detectable cancer (nodes)
  - Days until event or censoring (time)
  - Censoring status (status)

EXAMPLE 1

- Multivariable analysis:
  - Proportional hazards model
  - "we kept the variable of treatment in the model and used backward regression for other covariates"
  - Other covariates (P < 0.01)
    - Depth of primary tumor invasion,
    - Invasion of adjacent structures
    - Regional implants
    - Number of metastatic lymph nodes
    - Histological differentiation
    - Preoperative carcinoembryonic antigen level

EXAMPLE 1

- Multivariable results: "After correction for the influence of prognostic factors through the use of a proportional hazards model, patients receiving fluorouracil plus levamisole were again found to have a significant survival advantage when compared with patients assigned to observation only; they had a 33% reduction in mortality rate (95% CI, 16% to 47%; *P* = 0.0007). Therapy with levamisole alone showed essentially no effect (6% reduction in death rate; *P* = 0.57."

   Moertel et al (1995)

EXAMPLE 2 – ALZHEIMER'S

- Petersen et al. 2005, NEJM

- Subjects with amnestic subtype of mild cognitive impairment

- Adaptive randomization based on MMSE score, age, Apo ε4 genotype

- Three arms: Vitamin E, Donepezil, and Placebo

- Primary outcome: Time from randomization to possible or probable AD diagnosis

- Length of double-blind treatment: 3 years

Petersen RC, Thomas RG, Grundman M. et al. (2005) NEJM. 352(23):2379–2388.

EXAMPLE 2 – ALZHEIMER'S

- Primary analysis: Cox regression adjusted for randomization influencing variables MMSE score, age and Apo E genotype

- 769 enrolled: 253 donepezil, 257 vitamin E, 259 placebo

- 230 dropped out: 92 donepezil, 74 vitamin E, 66 placebo

  – Treatment related toxicity: GI complaints, muscle aches, insomnia

- Dropout was observed to be related to MMSE score

EXAMPLE 2 – ALZHEIMER'S

- 212 developed possible or probable AD

- "There were no significant differences ... during the three years of treatment"

- Vitamin E vs Placebo

  – Hazard Ratio 1.02 (95% CI, 0.74, 1.41), p-value 0.91

- Donepezil vs Placebo

  – Hazard Ratio 0.80 (95% CI, 0.57, 1.13), p-value 0.42

EXAMPLE 2 – ALZHEIMER'S

- Pre-specified analyses at 6 month intervals
  - Donepezil vs Placebo significantly reduced likelihood of progression to AD during the first 12 months (p-value 0.04)
  - Finding supported by secondary outcome measures
  - Subgroup ≥ 1 apolipoprotein E ε4 alleles significantly reduced likelihood of progression to AD over 3 years
  - Vitamin E vs Placebo: no significant differences
  - Vitamin E vs Placebo: also no significance for above subgroup
- Simulations assuming informative treatment-related dropout did not change primary conclusions

EXAMPLE 2 – RESULTS

- Overall and at 6 and 12 months
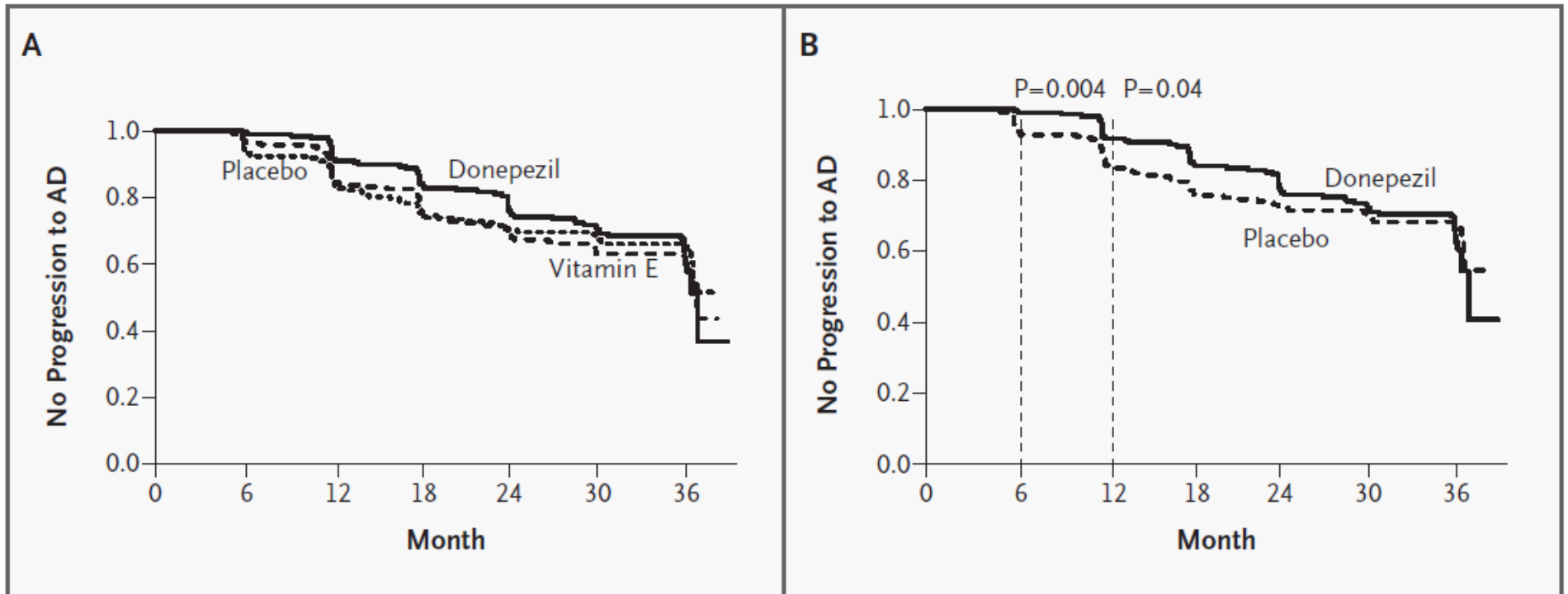
EXAMPLE 2 – RESULTS

- APOE ε4 results
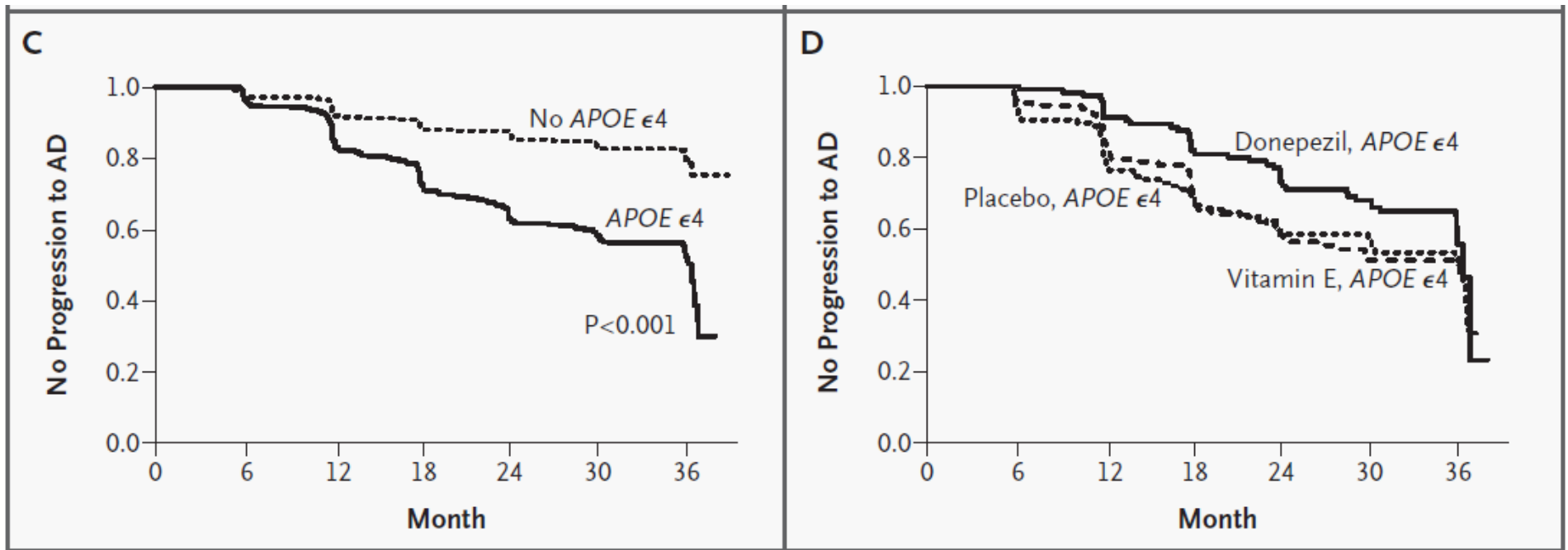
EDITORIAL

- "long-awaited results"

- Donepezil standard therapy for AD

- "Implications …. Enormous"
  - "The clear-cut negative findings for vitamin E, which is widely used despite the dearth of evidence of its efficacy, are especially noteworthy."

  - Findings for donepezil "much less clear"

EDITORIAL COMMENTS

- "rate of progression … somewhat lower in the treatment group during the first year of the study"

- "by two years, even this small effect had worn off"

- Possible explanation: "Reduced statistical power later in the study as the number of subjects at risk declined owing to death, withdrawal and development of AD

- Secondary analyses suggest… benefits wore off

EXAMPLE 2 – RESULTS

- Interesting steps…..
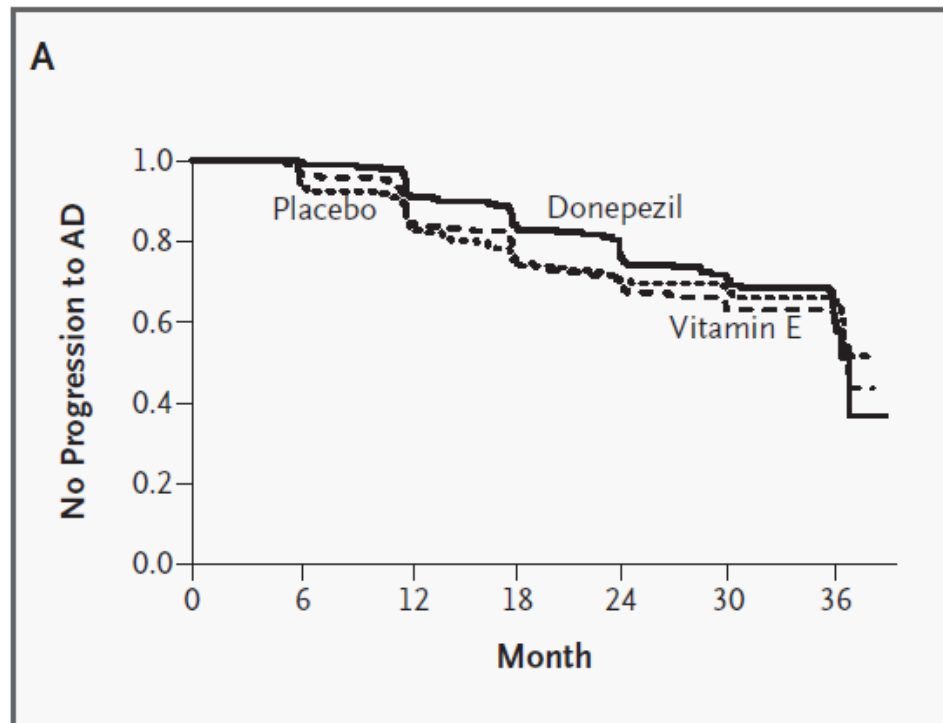
SCREENING TRIAL

- 202,546 women 50-72 years of age, England, Wales, Northern Ireland
- Randomized to one of three arms in 1:1:2 ratio between June 1, 2001 and Oct 21, 2005.
  - Annual multimodal screening (serun CA 125 + algorithm)
  - Annual transvaginal ultrasound
  - No screening
- Screening ended Dec 31, 2011.
- Not blinded
- Primary outcome: death from ovarian cancer (by end of 2014)

Jacobs IJ, Menon U, Ryan A, et al. (2016) The Lancet. 387(10022):945–956.

OVARIAN CANCER SCREENING TRIAL

- Primary analysis: Cox regression (proportional hazards)

  – MMS vs. no screening: Mortality reduction =

     $(1 - HR)100 = 15\%$ (95% CI: -1% – 33%) P = .10

  – USS vs. no screening: Mortality reduction =

     $(1 - HR)100 = 11\%$ (95% CI: -7% - 27%) P = .21

OVARIAN CANCER SCREENING TRIAL
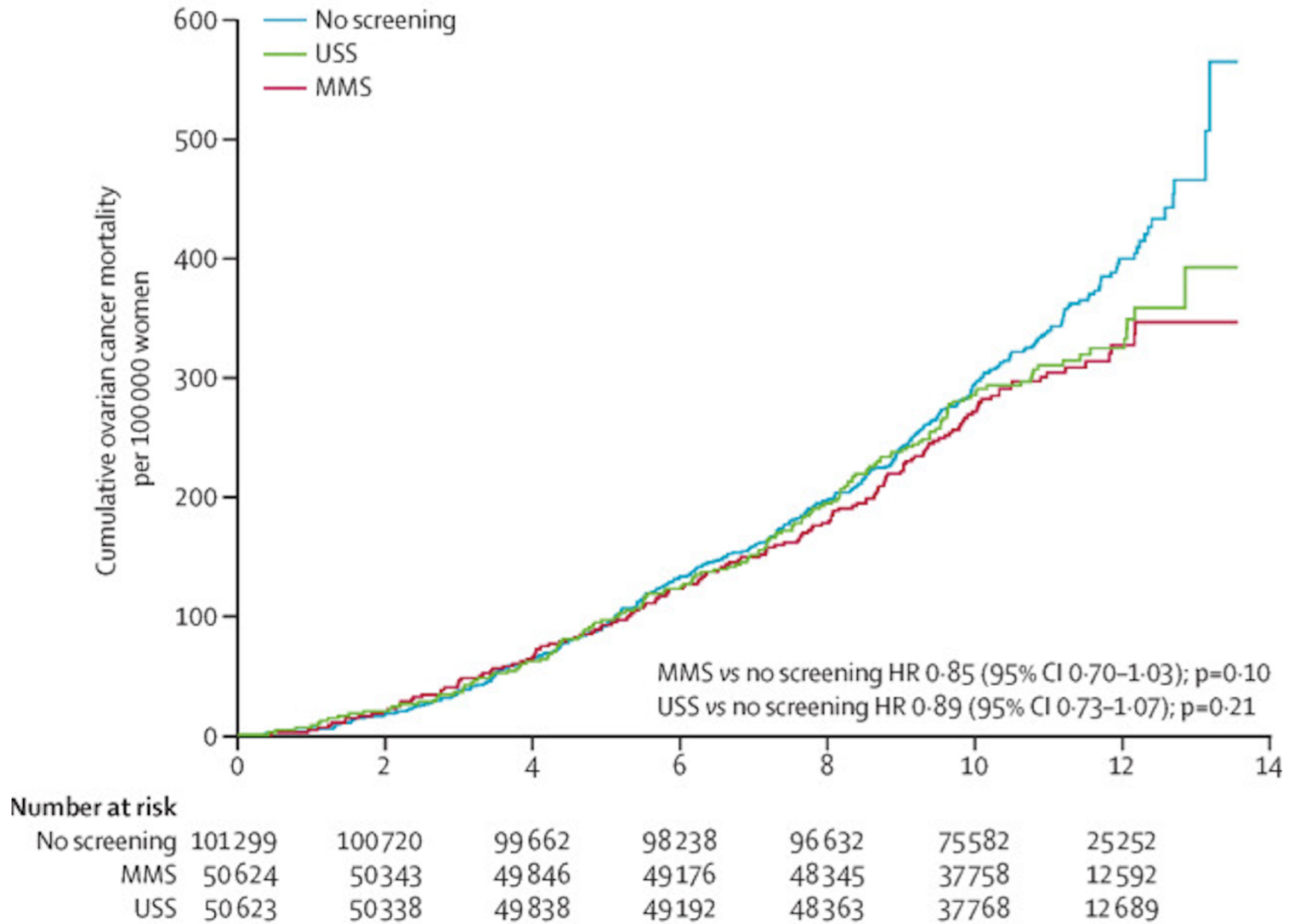
OVARIAN CANCER SCREENING TRIAL

- Why the delayed difference?

OVARIAN CANCER SCREENING TRIAL

- Secondary analyses conducted motivated by late separation of survival curves

- Post-hoc Weighted* logrank test:
  - MMS mortality reduction = 22% (3-38%) P = .023
  - USS mortality reduction = 20% (0 – 35%) P = .049

  * by pooled cumulative mortality

SURVIVAL DATA AND FUNCTION

- Original applications in biometry were to survival times in cancer clinical trials

- Many other applications in biometry: eg. disease onset ages

- Interest centers not only on average or median survival time but also on probability of surviving beyond 2 years, 5 years, 10 years, etc.

- Best described with the entire survival function S(t).

  - For T = a subject's survival time, S(t) = P[T> t].

  - Characterizes the entire distribution of survival times T.

  - Gives useful information for each t.

SURVIVAL FUNCTION
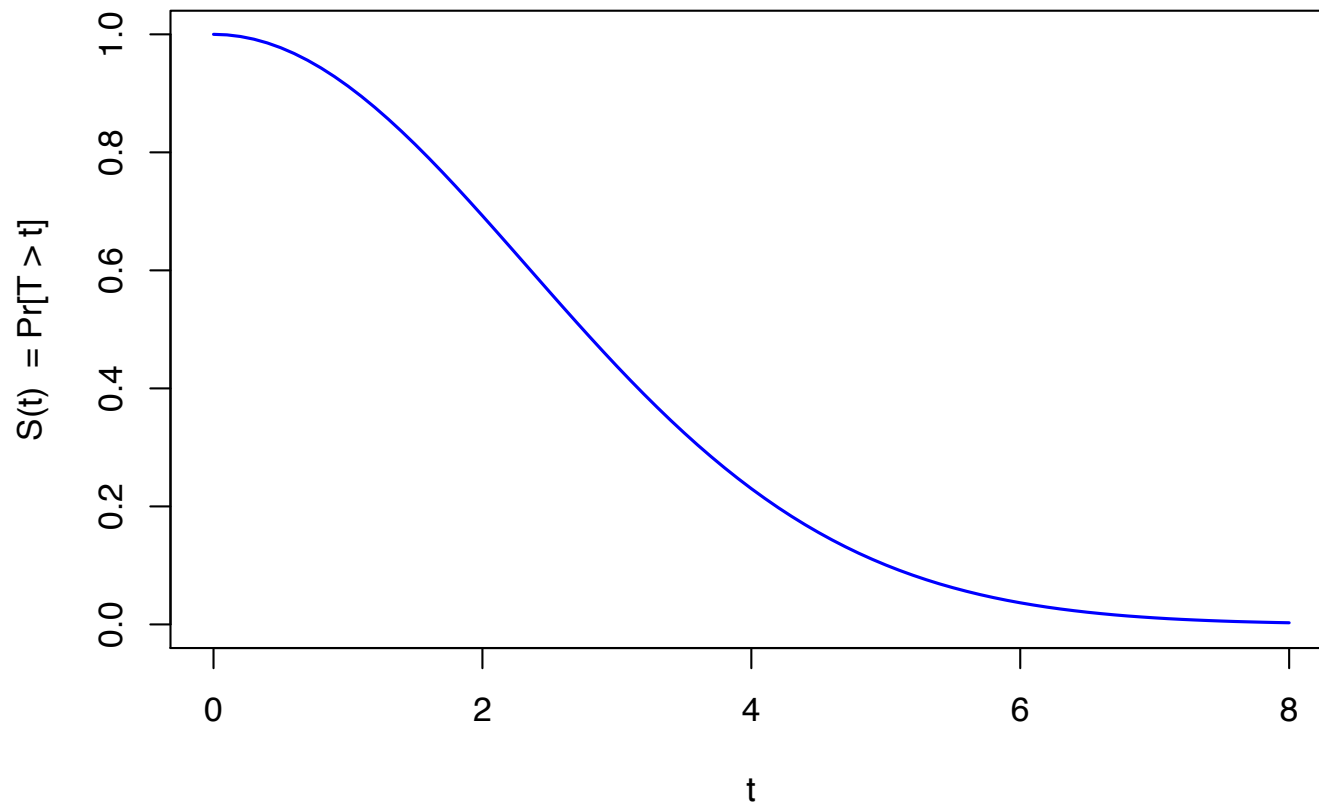
Survival Function

SURVIVAL DISTRIBUTION

- Continuous probability distribution of times T
- Only non-negative T's are possible: $\Pr(T<0)=0$
- Density function $\quad f(t) = \lim_{\Delta t \to 0} \dfrac{1}{\Delta t} \Pr(t \le T < t + \Delta t)$
- Area under the $f(t)$ curve between two points is the probability T is between the two points.

DENSITY AND SURVIVAL FUNCTIONS
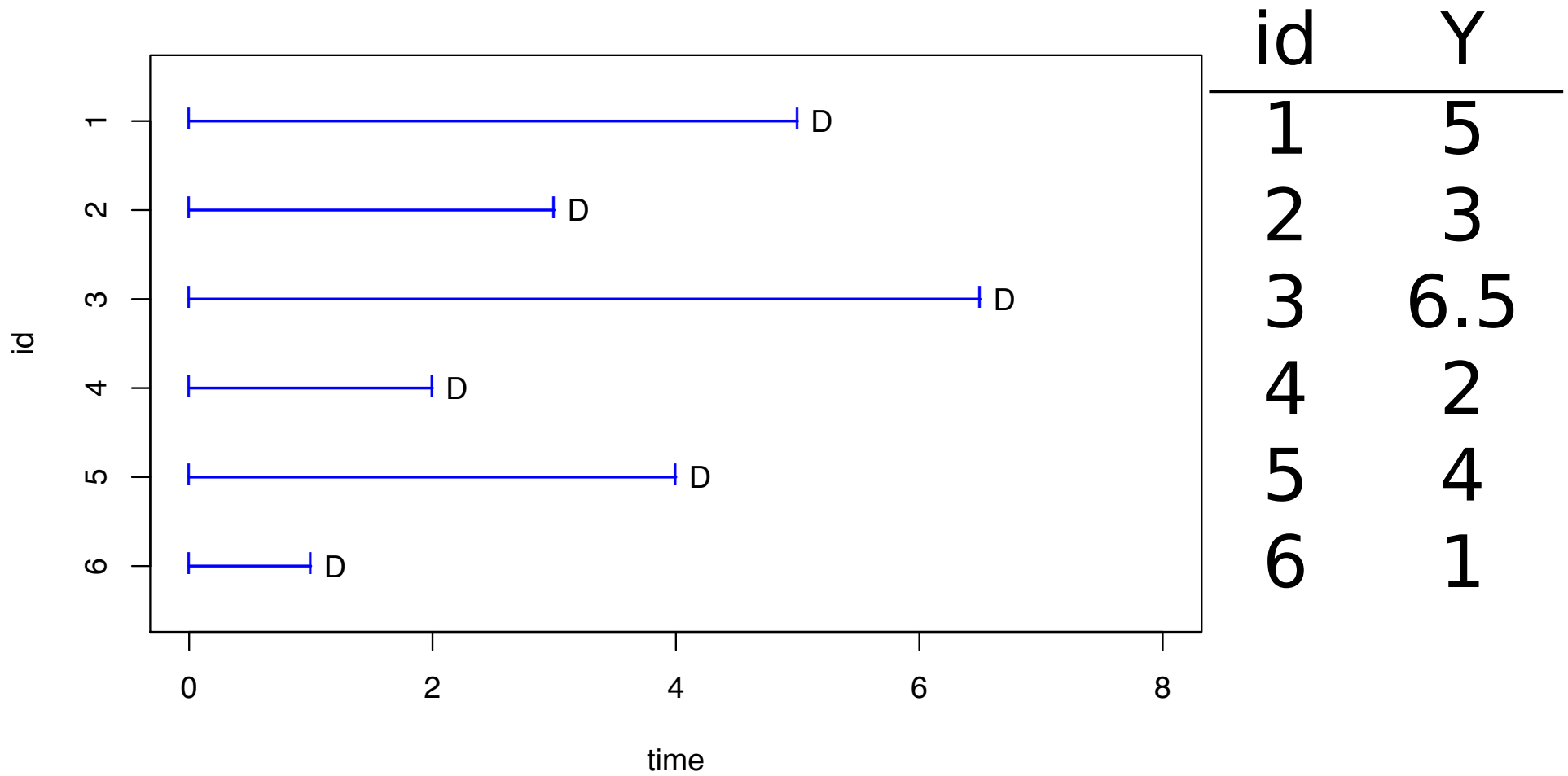

Density Function
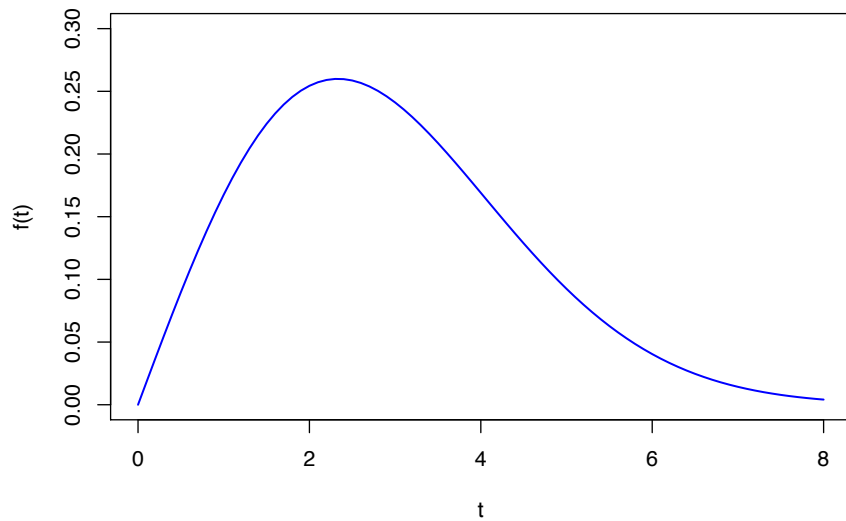
Survival Function

MEDIAN SURVIVAL TIME
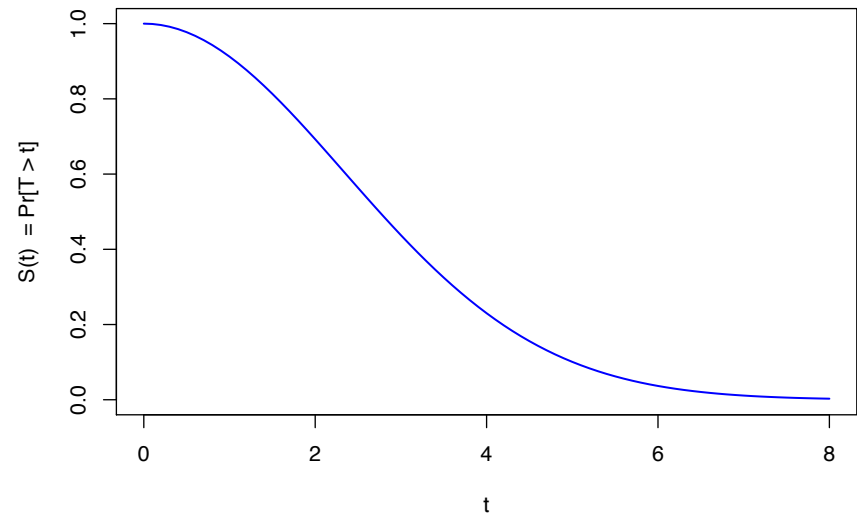
Median Survival Time
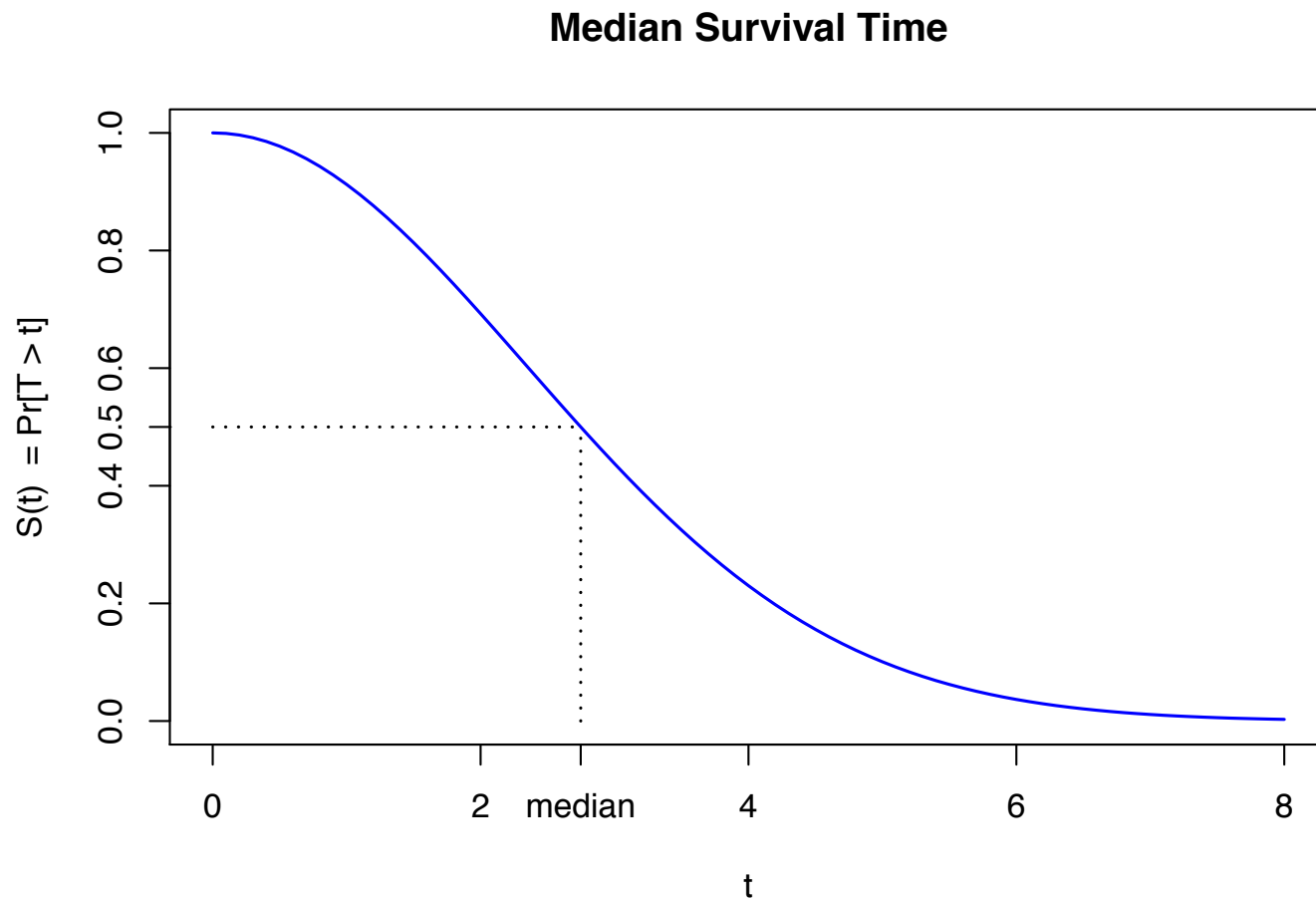
MEDIAN SURVIVAL TIME

Density Function
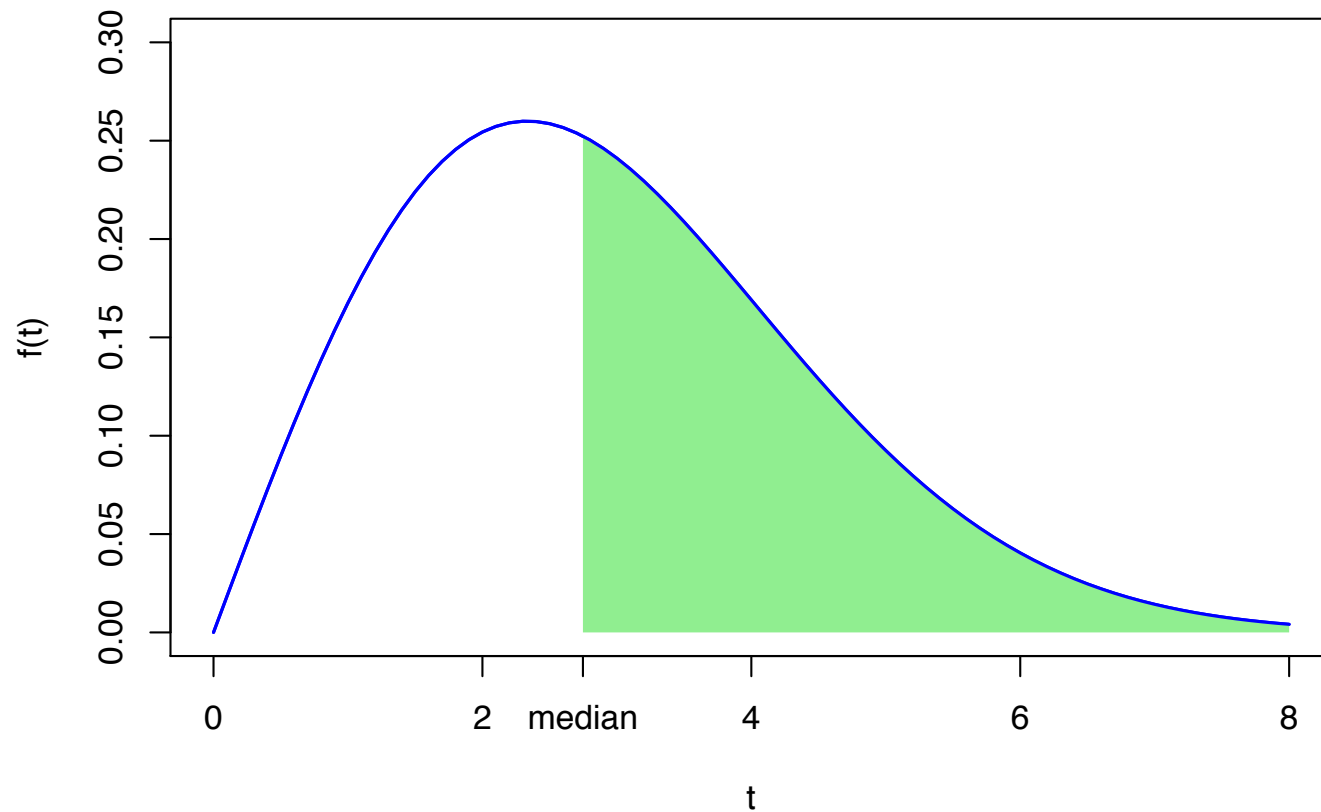
ILLUSTRATIVE DATA

id	Y
1	5
2	3
3	6.5
4	2
5	4
6	1

SURVIVAL FUNCTION ESTIMATE

- Nonparametric Estimate: reduce estimate by 1/n every time there is an event (death): Empirical survival function estimate

Survival Function Estimate
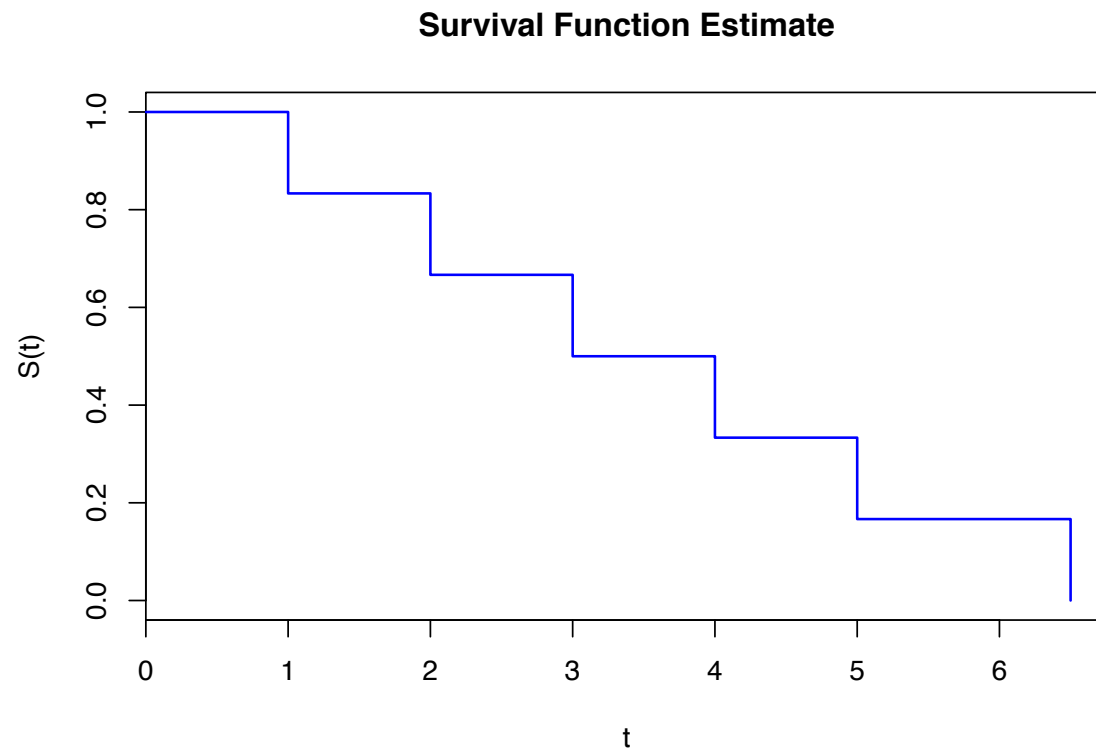
MEDIAN ESTIMATE
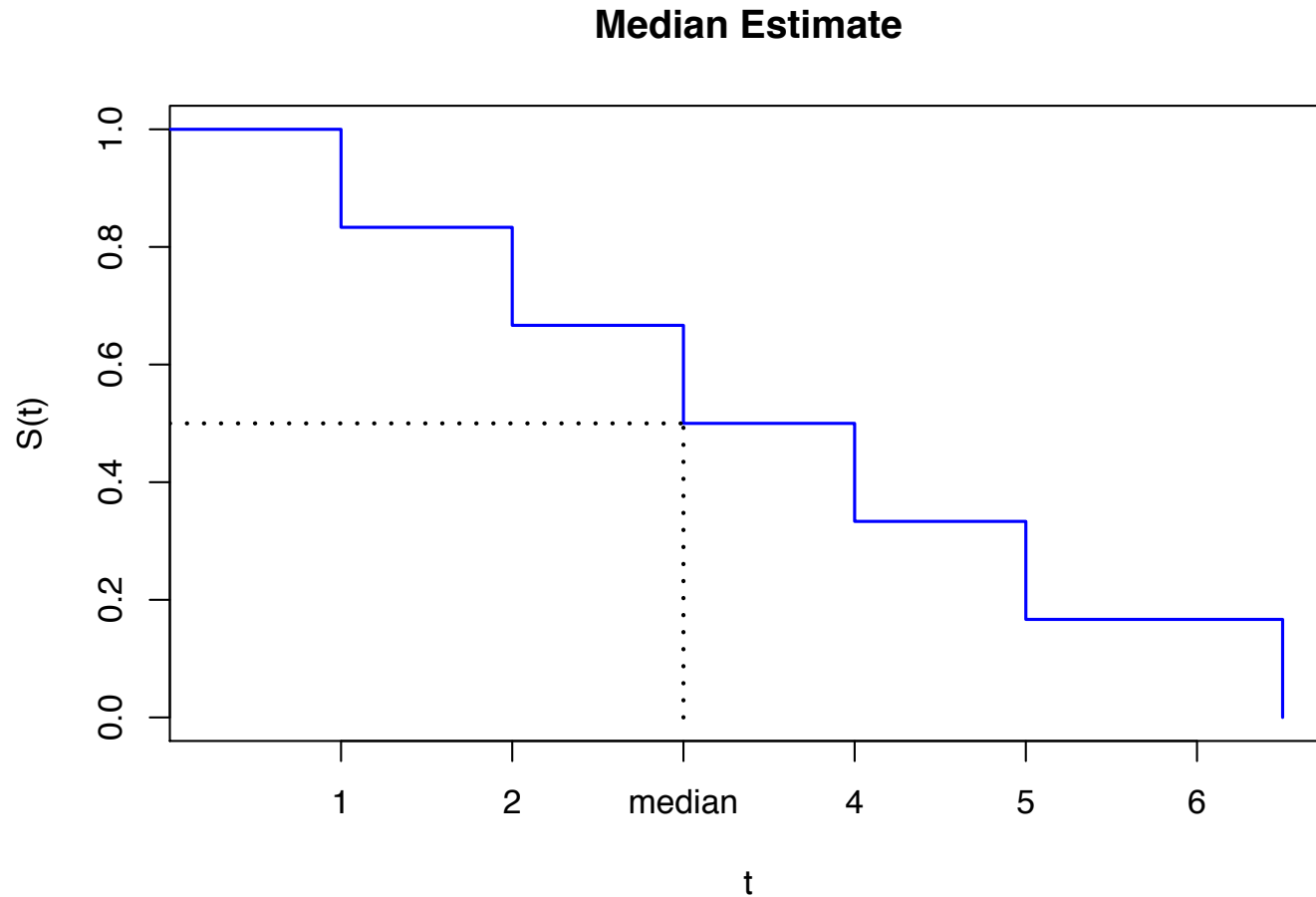
Median Estimate

By convention: median is earliest time where survival estimate ≤ .5

OTHER WAYS TO DESCRIBE A SURVIVAL DISTRIBUTION

- So far we have looked at the density function and survival function S(t).

- Also of interest: "hazard" function λ(t)

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr[t \leq T < t + \Delta t | T \geq t]$$

- Instantaneous rate at which death occurs at t in those who are alive at t

- Examples:

  – Age-specific death rate

  – Age-specific disease incidence rate

HAZARD FUNCTION FOR HUMANS

Human Mortality

$\lambda(t)$ vs age in years
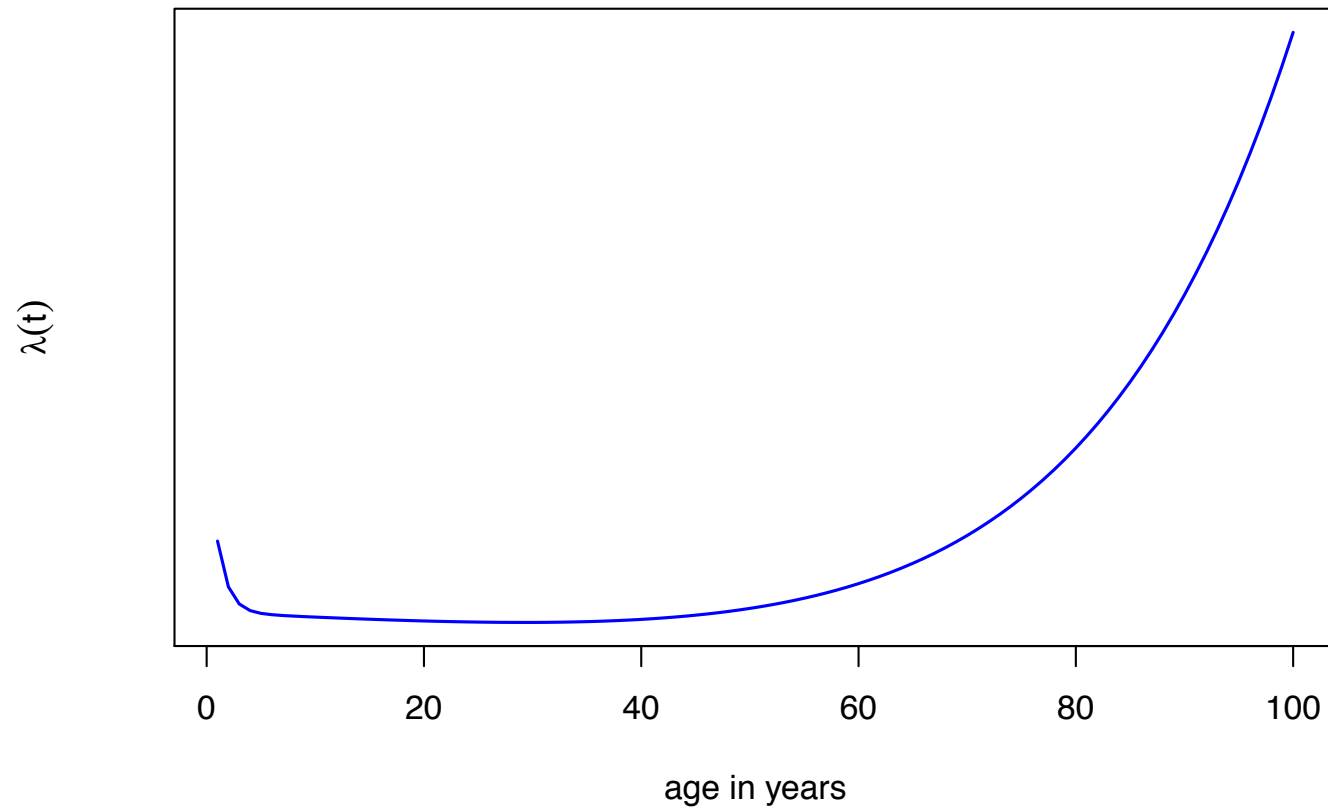
EQUIVALENT CHARACTERIZATIONS

- Any <u>one</u> of the density function(f(t)), the survival function(S(t)) or the hazard function(λ(t)) is enough to determine the survival distribution.

- They are each functions of each other:

$$\bullet \ S(t) = \int_t^\infty f(s)ds = e^{-\int_0^t \lambda(s)ds}$$

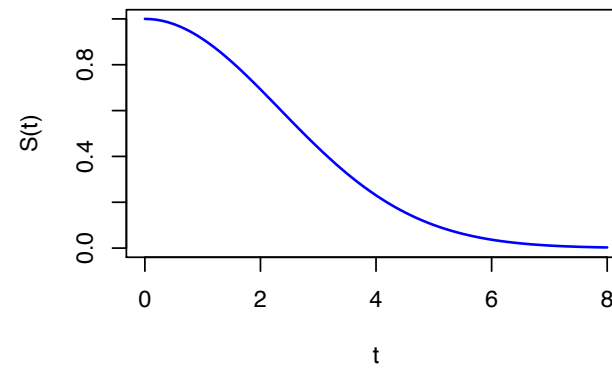$$\bullet \ f(t) = -\frac{d}{dt}S(t) = \lambda(t)e^{-\int_0^t \lambda(s)ds}$$

$$\bullet \ \lambda(t) = \frac{f(t)}{S(t)}$$

EQUIVALENT CHARACTERIZATIONS

EQUIVALENT CHARACTERIZATIONS

QUESTIONS?