Harvard-MIT Division of Health Sciences and Technology
HST.951J: Medical Decision Support, Fall 2005
Instructors: Professor Lucila Ohno-Machado and Professor Staal Vinterbo

6.873/HST.951 Medical Decision Support
Spring 2005

Survival Analysis

Lucila Ohno-Machado

Outline

- Basic concepts & distributions
  – Survival, hazard
  – Parametric models
  – Non-parametric models
- Simple models
  – Life-table
  – Product-limit
- Multivariate models
  – Cox proportional hazard
  – Neural nets

What we are trying to do

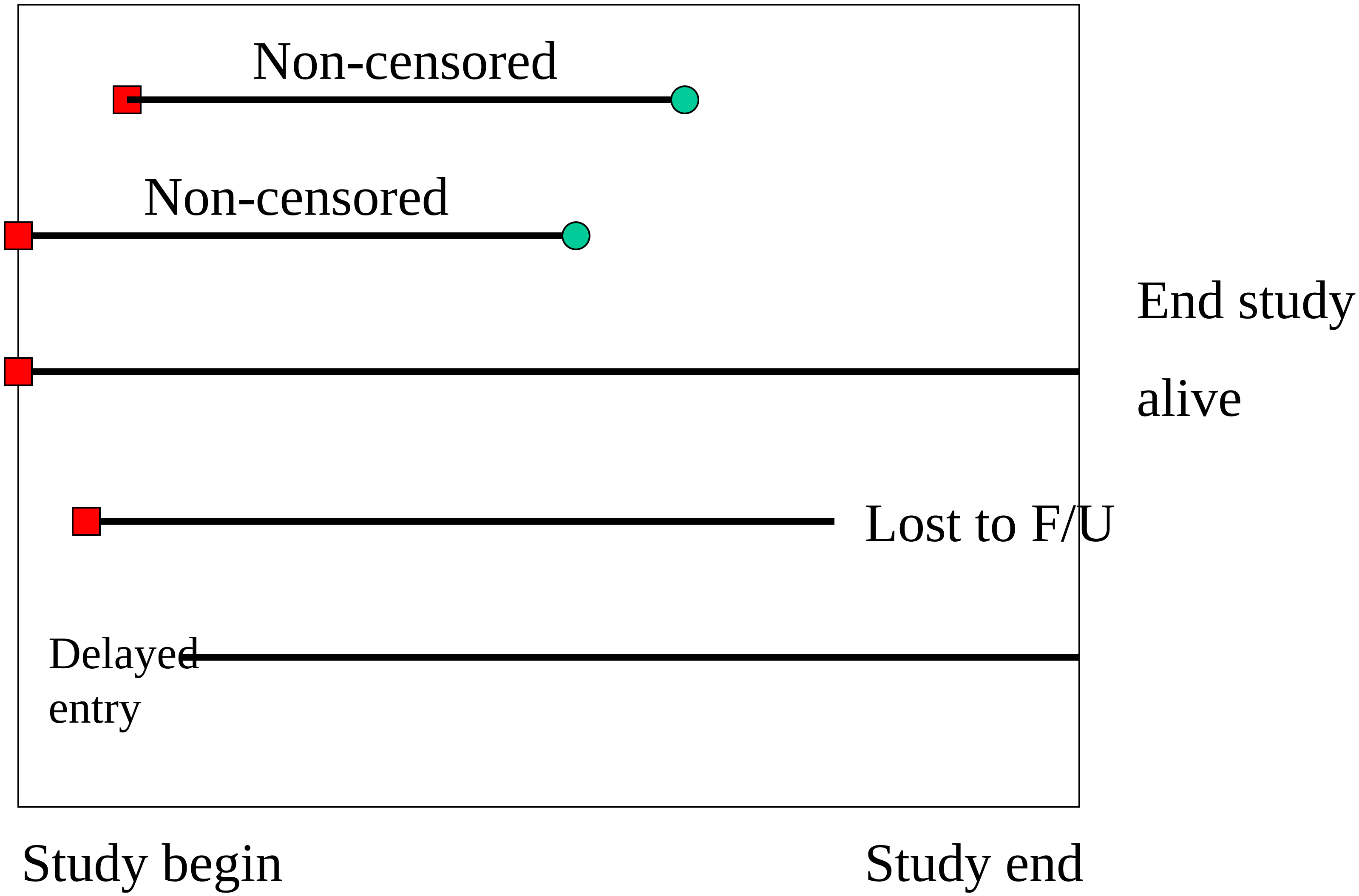
	Variable 1	Variable 2	Years of survival
Case 1	0.7	-0.2	8+
Case 2	0.6	0.5	4
…	-0.6	0.1	2
	0	-0.9	3+
	-0.4	0.4	2
	-0.8	0.6	3
	0.5	-0.7	4

Predict survival (or more frequently predict the probability of at least *n* years of survival)

Using these

- and evaluate performance on new cases
- and determine which variables are important

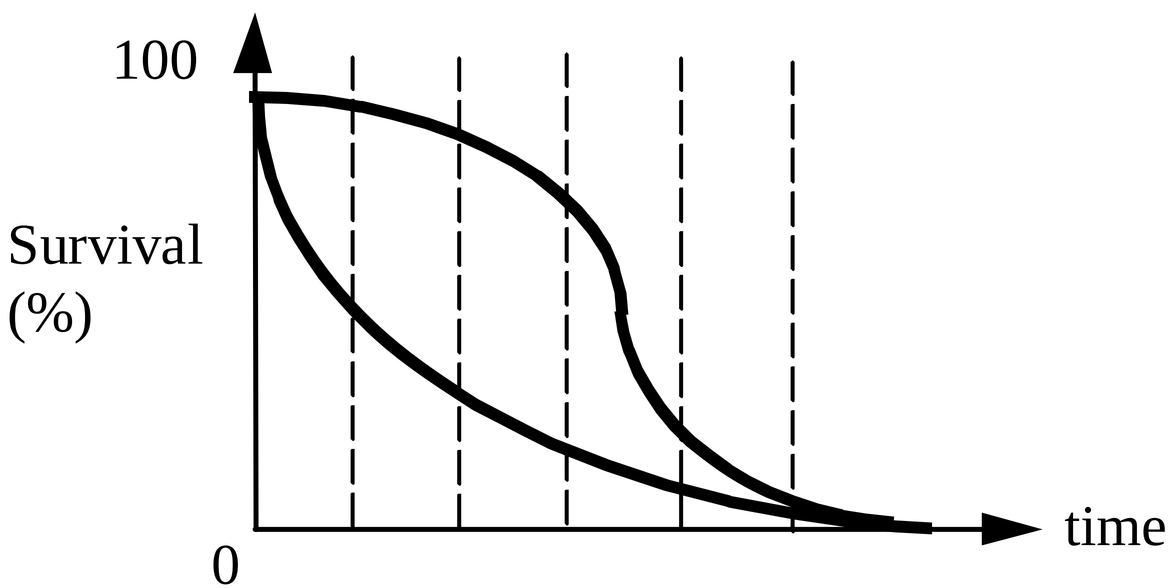
Censoring


Non-censored
Non-censored
End study
alive
Lost to F/U
Delayed
entry
Study begin
Study end

Survival function

Probability that an individual survives at least t, T is patient's survival

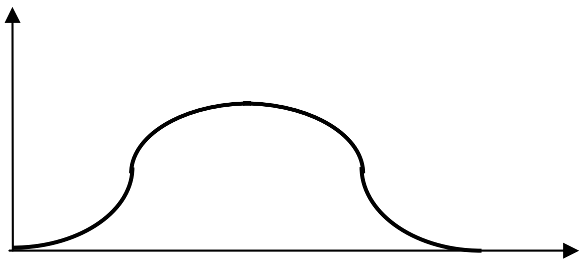
- S(t) = P(T > t) = 1 – F(t)
- Survival is cumulative, non-increasing function
- F(t) is cumulative distribution of death (failure)
- By definition, S(0) = 1 and S(∞)=0

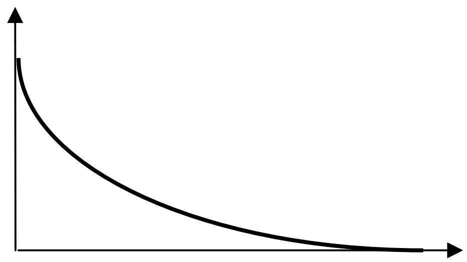


Unconditional failure rate

- pdf of T
- $f(t) = \lim_{\Delta t \to 0} P(\text{individual dies } (t, t+\Delta t)) / \Delta t$
- f(t) always non-negative
- Area below density is 1
- Estimated by

patients dying in the interval/total patients

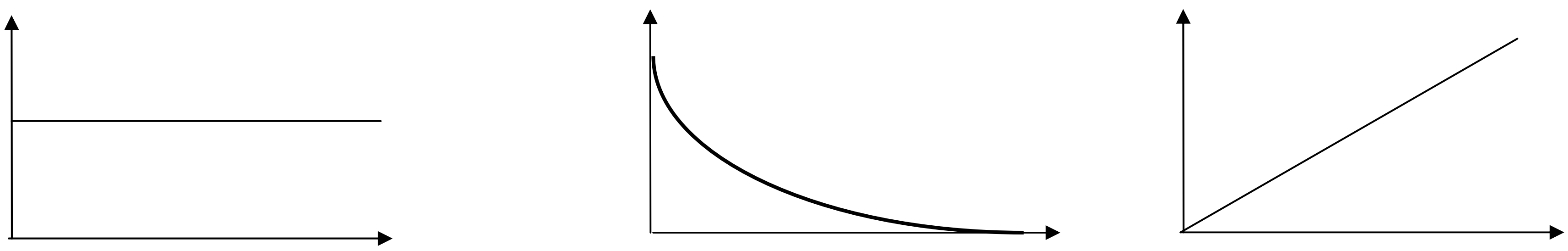




Conditional failure rate

- Hazard function
- $h(t) = \lim_{\Delta t \to 0} P($survivor until t dies $(t, t+\Delta t))/ \Delta t$
- h(t) is conditional instantaneous failure rate
- Estimated by

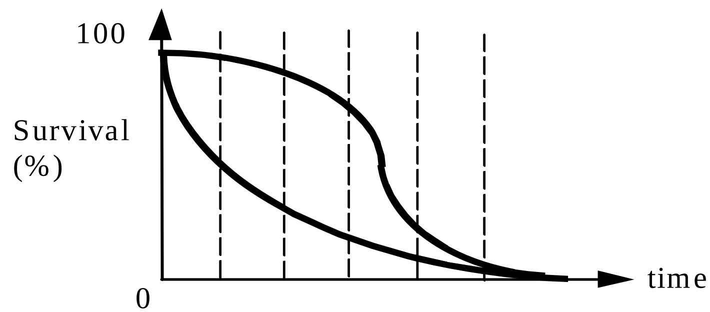
\# patients dying in the interval/survivors at t



Hazard Function

λ

$$f(t) = \partial F(t) / \partial(t)$$
$$F(t) = 1 - S(t)$$
$$f(t) = -\partial S(t) / \partial(t)$$



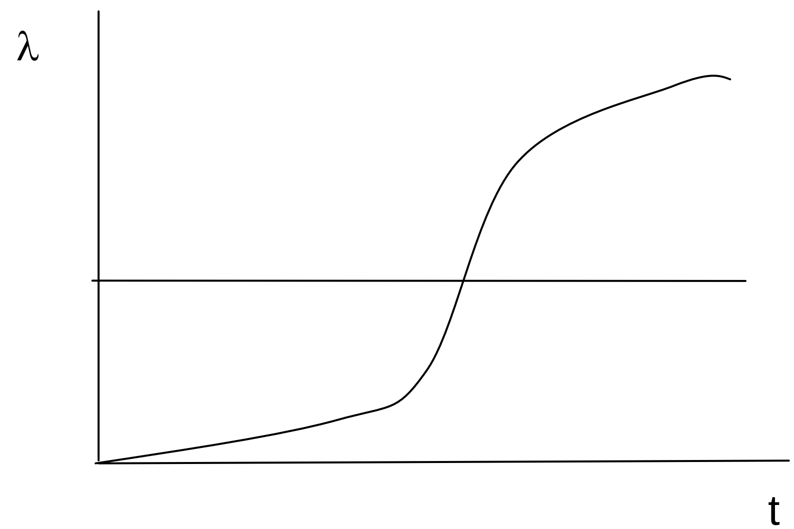
$$\lambda(t) = \lim_{u \to 0} \frac{P\{t < T \le t+u \mid T > t\}}{u}$$

$$\lambda(t) = \lim_{u \to 0} \frac{P\{t < T \le t+u\} \mid P\{T > t\}}{u}$$

$$\lambda(t) = \lim_{u \to 0} \frac{[F(t+u) - F(t)] \mid u}{S(t)}$$

$$\lambda(t) = \frac{\partial F(t) / \partial t}{S(t)}$$

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



Cumulative Hazard Function Λ

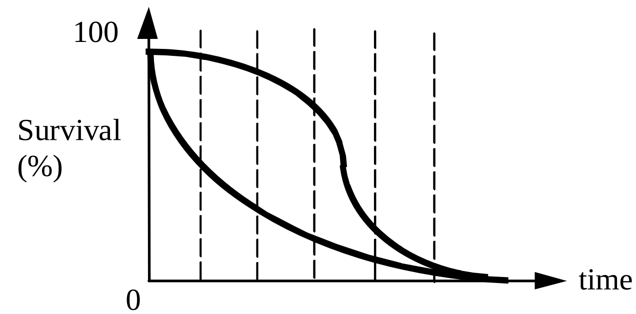
$$\lambda(t) = \frac{f(t)}{S(t)}$$

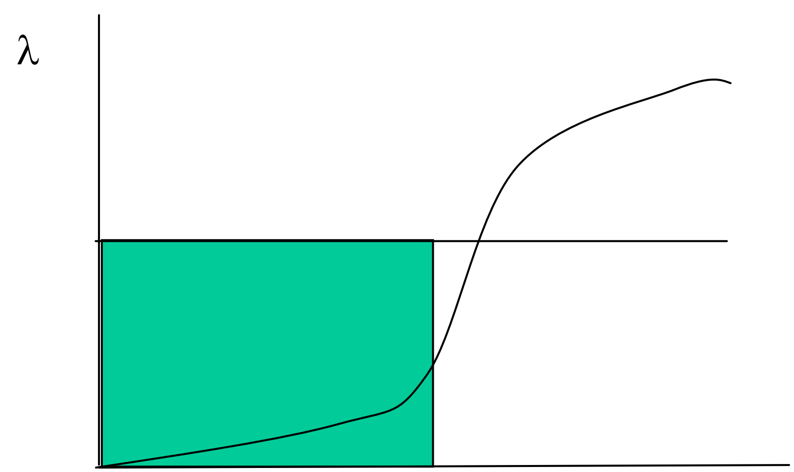
$$\frac{\partial \log S(t)}{\partial t} = \frac{\partial S(t)/\partial t}{S(t)} = -\frac{f(t)}{S(t)}$$

$$\lambda(t) = -\frac{\partial \log S(t)}{\partial t}$$

$$\int_0^t \lambda(v)dv = -\log S(t) = \Lambda(t)$$

$$S(t) = e^{-\Lambda(t)}$$

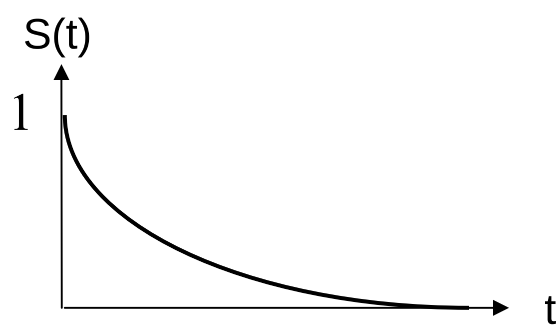


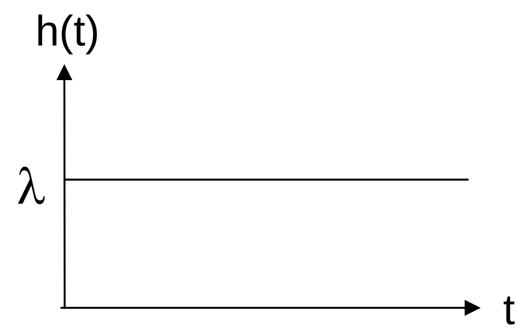


Parametric estimation

Example: Exponential

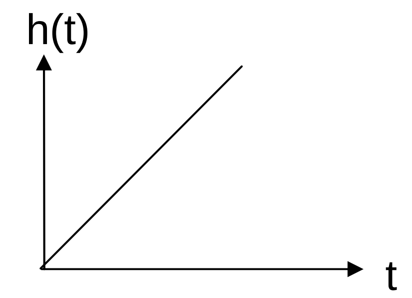
- $f(t) = \lambda e^{-\lambda t}$
- $S(t) = e^{-\lambda t}$
- $h(t) = \lambda$

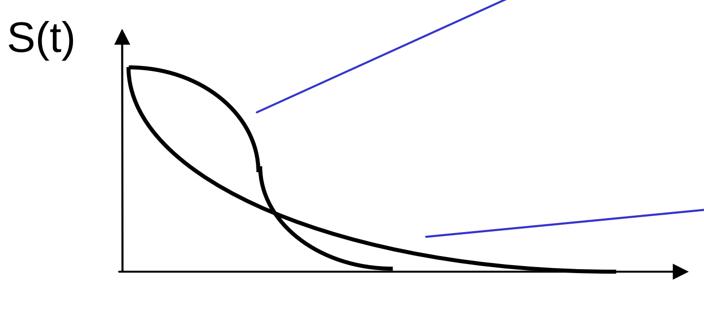


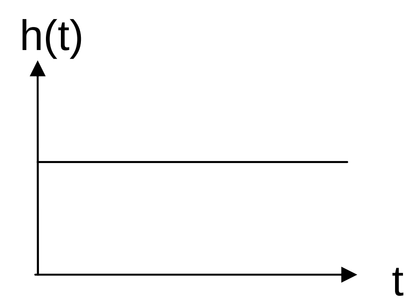


Weibull distribution

- Generalization of the exponential
- For $\lambda,\gamma > 0$
- $f(t) = \gamma\lambda(\lambda t)^{\gamma-1}\, e^{-\lambda t^{\gamma}}$
- $S(t) = e^{-\lambda t^{\gamma}}$
- $h(t) = \gamma\lambda(\lambda t)^{\gamma-1}$



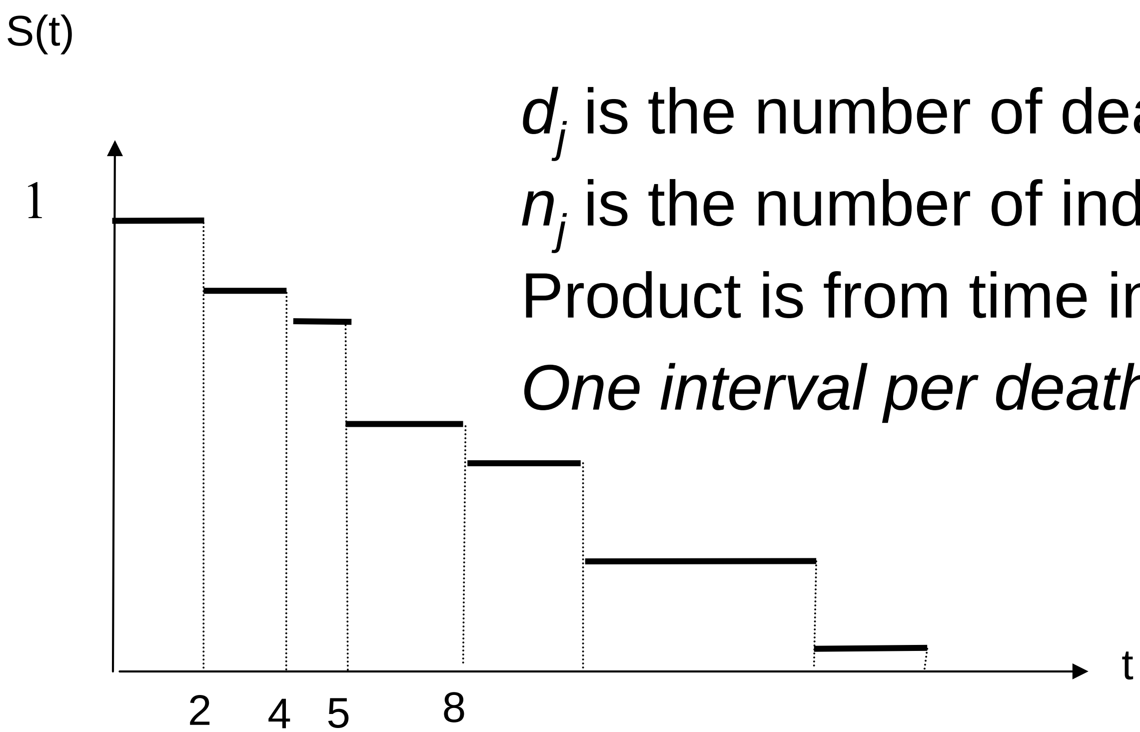




Non-Parametric estimation

Product-Limit (Kaplan-Meier)

$S(t_i) = \Pi \, (n_j - d_j)/ n_j$



d_j is the number of deaths in interval j

n_j is the number of individuals at risk

Product is from time interval 1 to j

One interval per death time

Kaplan-Meier

- Example
- Deaths: 10, 37, 40, 80, 91,143, 164, 188, 188, 190, 192, 206, …

Life-Tables

- AKA actuarial method

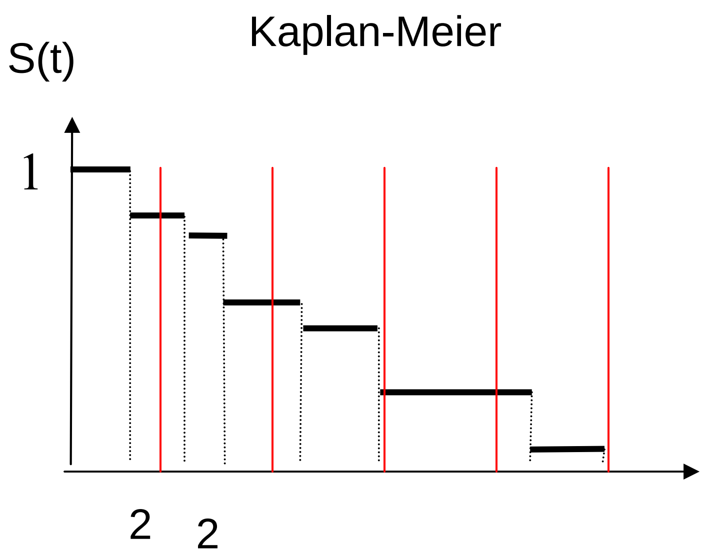
$S(t_i) = \Pi\ (n_j - d_j)/\ n_j$

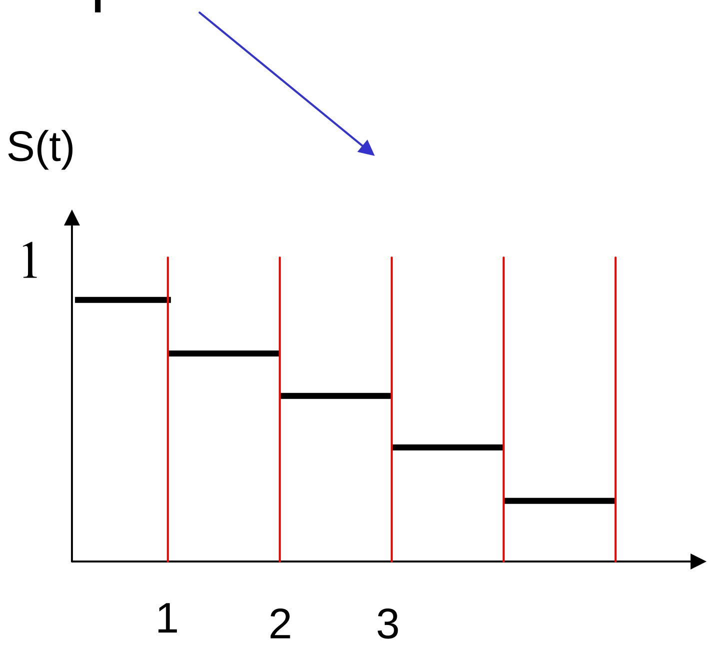
d_j is the number of deaths in interval j

n_j is the number of individuals at risk

Product is from time interval 1 to j

- Pre-defined intervals j are independent of death times





Life-Table

hazard

survival

density

Simple models

Multiple strata

Multivariate models

- Several strata, each defined by a set of variable values
- Could potentially go as far as “one stratum per case”?
- Can it do prediction for individuals?

Cox Proportional Hazards

- Regression model
- Can give estimate of hazard for a particular individual relative to baseline hazard at a particular point in time
- Baseline hazard can be estimated by, for example, by using survival from the Kaplan-Meier method or parametrically

Proportional Hazards

$$\lambda_i = \lambda\, e^{\beta x_i}$$

where λ is baseline hazard (ie, for the “baseline” – usually the most common patient) and x_i is covariate vector for a specific patient i

Cox proportional hazards

$$h_i(t) = h_0(t)\, e^{\beta x_i}$$

- Survival

$$S_i(t) = [S_0(t)]^{e^{\beta x_i}}$$

Cox Proportional Hazards

$$h_i(t) = h_0(t)\, e^{\beta x_i}$$

- From the set of m individuals at risk at time **j** (R_j), the probability of picking exactly the one who died is

$$\frac{h_0(t)\, e^{\beta x_i}}{\Sigma_m\, h_0(t)\, e^{\beta x_m}}$$

- Then likelihood function to maximize to all j is
- $L(\beta) = \Pi_j \left(e^{\beta x_i} / \Sigma_m e^{\beta x_m}\right)$
- MLE uses LogLikelihood

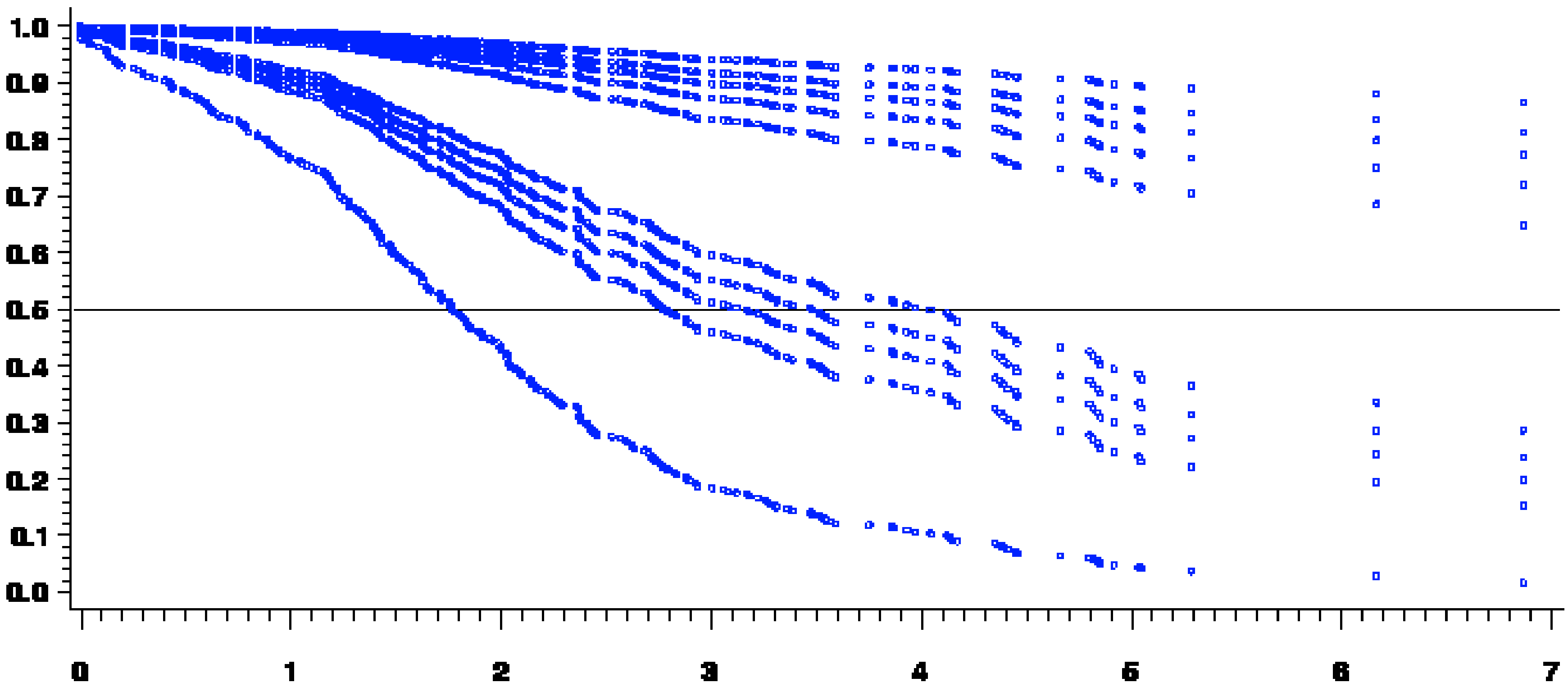
Important details

- Survival curves can't cross if hazards are proportional
- There is a common baseline h_0, but we don't need to know it to estimate the coefficients
- Ie, we don't need to know the shape of hazard function
- Cox model is commonly used to interpret importance of covariates (amenable to variable selection methods)
- It is the most popular multivariate model for survival
- Testing the proportionality assumption is difficult and hardly ever done

Estimating survival for a patient using the Cox model

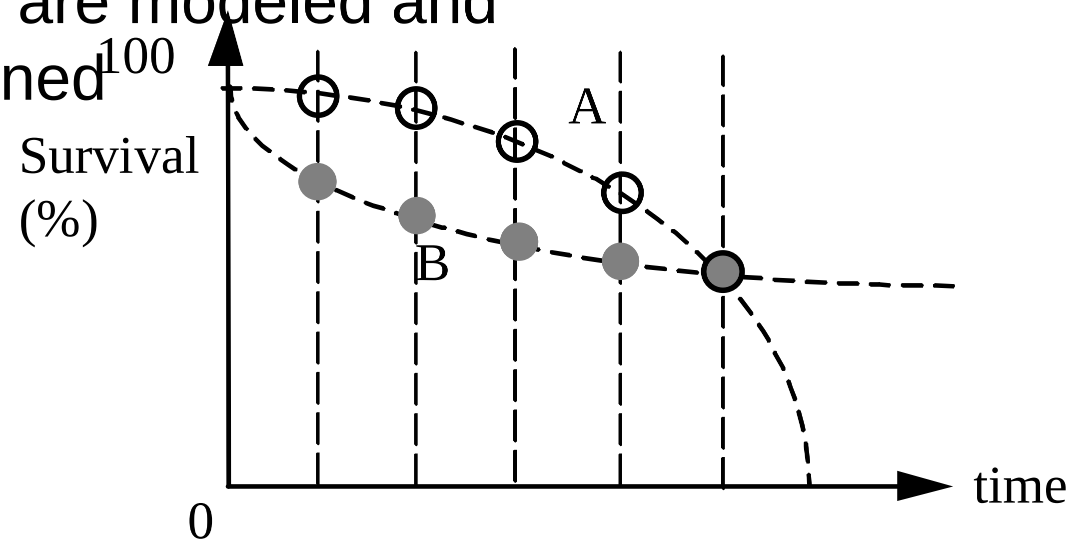
- Need to estimate the baseline
- Can use parametric or non-parametric model to estimate the baseline
- Can then create a continuous “survival curve estimate” for a patient
- Baseline survival can be, for example:
  - Kaplan-Meier estimate

Example of survival estimates



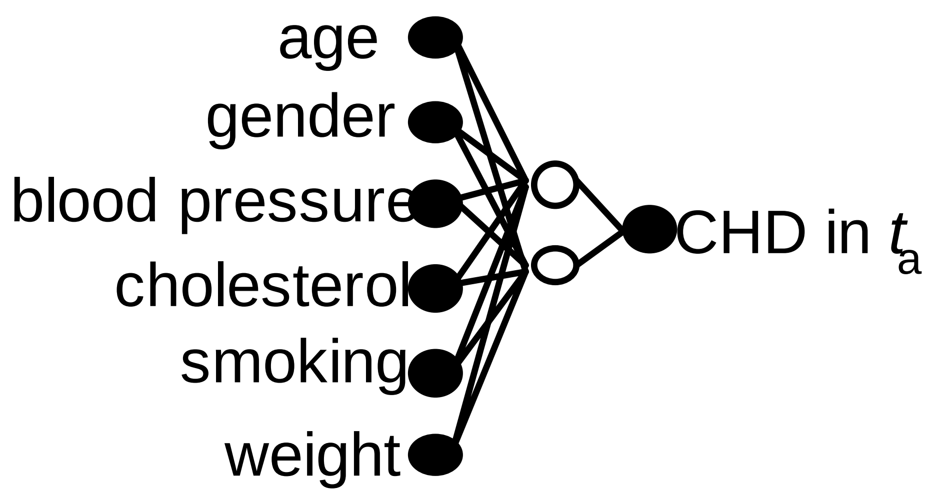
What if the proportionality assumption is not OK?

- Survival curves may cross
- Other multivariate models can be built
- Survival at certain time points are modeled and combined



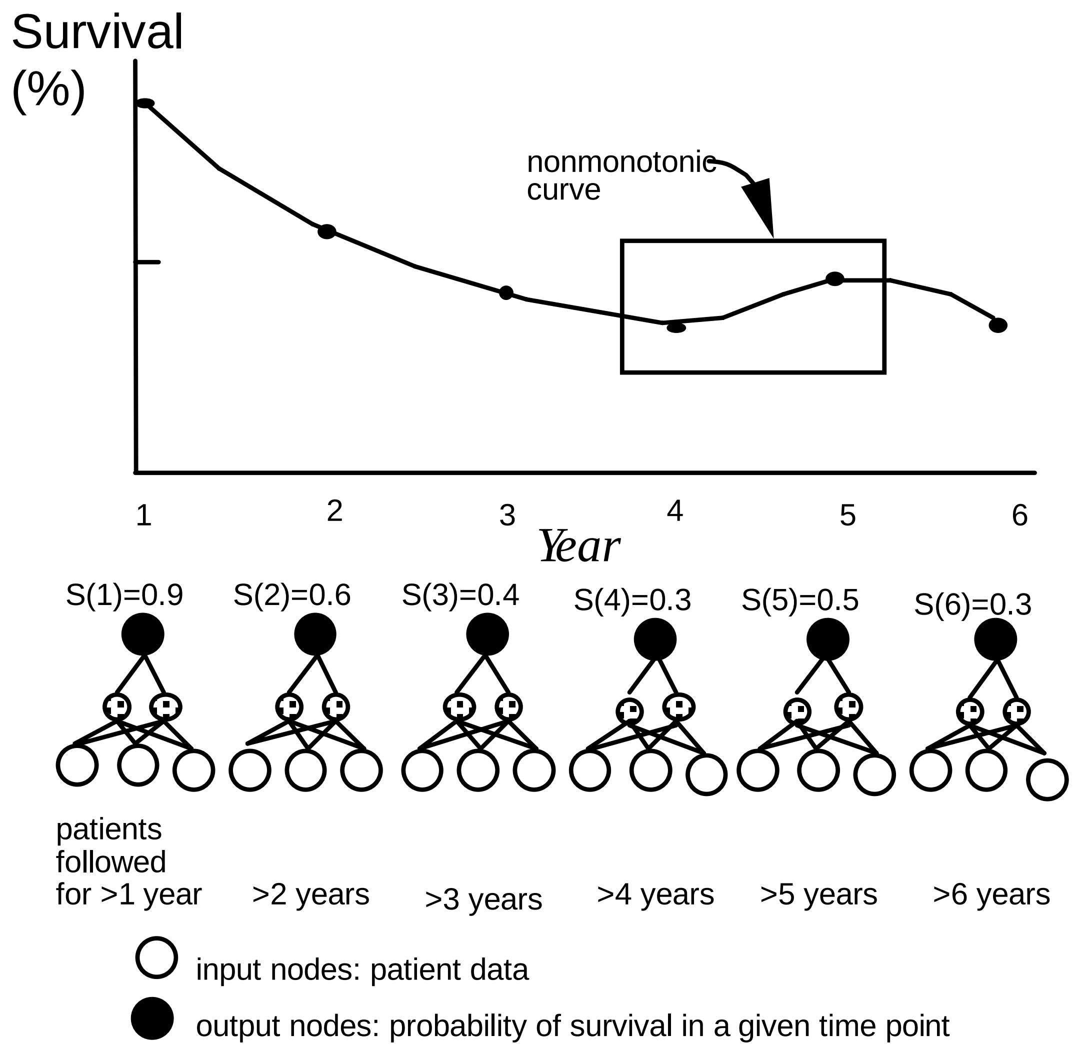
Single-point models

- Logistic regression
- Neural nets



Problems

- Dependency between intervals is not modeled (no links between networks)
- Nonmonotonic curves may appear
- How to evaluate?



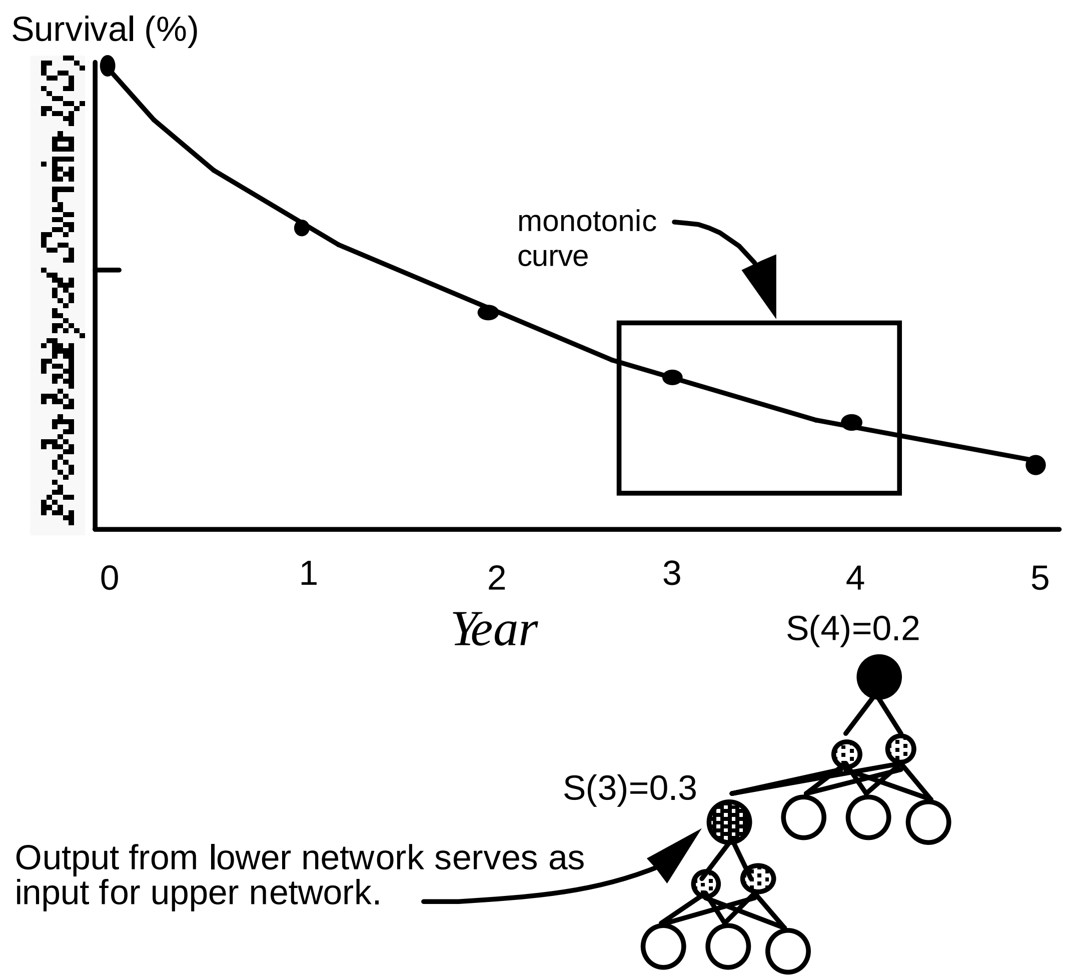
Figures removed due to copyright reasons.

Please see Tables III, V, VI and figures 6, 8, and 10 in:

Ohno-Machado, Lucila, and Mark A. Musen. "Modular Neural Networks for Medical Prognosis: Quantifying the Benefits of Combining Neural Networks for Survival Prediction." *Connection Science* 9, no. 1 (March 1997): 71-86.

Accounting for dependencies

- “Link” networks in some way to account for dependencies



Figures removed due to copyright reasons.

Please see Tables III, V, VI and figures 6, 8, and 10 in:

Ohno-Machado, Lucila, and Mark A. Musen. "Modular Neural Networks for Medical Prognosis: Quantifying the Benefits of Combining Neural Networks for Survival Prediction." *Connection Science* 9, no. 1 (March 1997): 71-86.

Survival without Coronary Disease

Year of follow-up	Test set χ^2	Test set p
2	15.0118	0.0589
4	11.1389	0.1939
6	19.6175	0.0118
8	30.3247	0.0001
10	23.6363	0.0026
12	11.6443	0.1677
14	9.3273	0.3154
16	6.7588	0.5628
18	26.1660	0.0009
20	22.3739	0.0042
22	12.7683	0.1200

Test set Area under the ROC curve	*standard error*
0.7038	0.0242
0.7117	0.0190
0.7352	0.0152
0.7337	0.0138
0.7333	0.0130
0.7448	0.0123
0.7752	0.0121
0.8059	0.0119
0.8275	0.0122
0.8374	0.0122
0.8324	0.0163

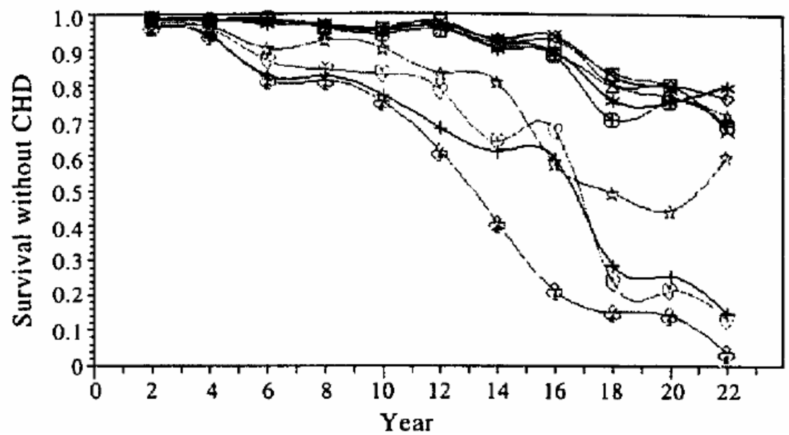


Figure removed due to copyright reasons.

Please see figure 10 in:

Ohno-Machado, Lucila, and Mark A. Musen. "Sequential versus standard neural networks for pattern recognition: an example using the domain of coronary heart disease." *Comput Biol Med* 27, no. 4 (Jul 1997): 267-81.

Summary

- Kaplan-Meier for simple descriptive analysis
- Cox Proportional for multivariate prediction if survival curves don’t cross
- Other methods for multivariate survival exist: logistic regression, neural nets, CART, etc.