Survival Analysis

One-Day Short Course

Jianghua He, Ph.D.
Associate Professor, Department of Biostatistics
University of Kansas Medical Center
jhe@kumc.edu

KU MEDICAL CENTER
The University of Kansas

Outline

Introduction

- Time to event data /Censoring and Truncation
- Basic survival functions/Group comparison of survival data

Regression Analyses

- Parametric Approach -- Accelerated failure time models
- Semiparametric Approach – Cox models

Extension to the Cox models

- Weighted Cox models
- Time-dependent covariate Cox models
- Time-varying coefficient survival models

Other topics

- Competing risk modeling
- Bayesian survival analyses

KU MEDICAL CENTER
The University of Kansas

Introduction--Definition

Survival analysis is a branch of statistics that deals with analysis of **time duration until one or more events happen**, such as death in biological organisms and failure in mechanical systems. This topic is called **reliability theory** or **reliability analysis** in engineering, **duration analysis** or **duration modelling** in economics, and **event history analysis** in **sociology**.

--- Wikipedia

KU MEDICAL CENTER
The University of Kansas

Introduction--Definition

With survival analysis, we can try to answer questions such as:

- What are the 5-year or 10-year survival rates for certain diseases after diagnosis?
- What are the risk factors that contribute to the risk of events?
- Does the new treatment reduce the risk (increase the survival rate) compared to the placebo or old treatment?
- Can multiple causes of death or failure be taken into account?

Introduction

Examples of Survival Data

- Times to infection of kidney dialysis patients since starting of dialysis
- Times to death for a breast-cancer trial from the beginning of treatment
- Times to weaning from breastfeeding for new-born babies from birth.
- Time to recovery after surgeries

Note: 1. Events are not necessary bad.

2. Baseline (time 0) should be meaningful.

Introduction

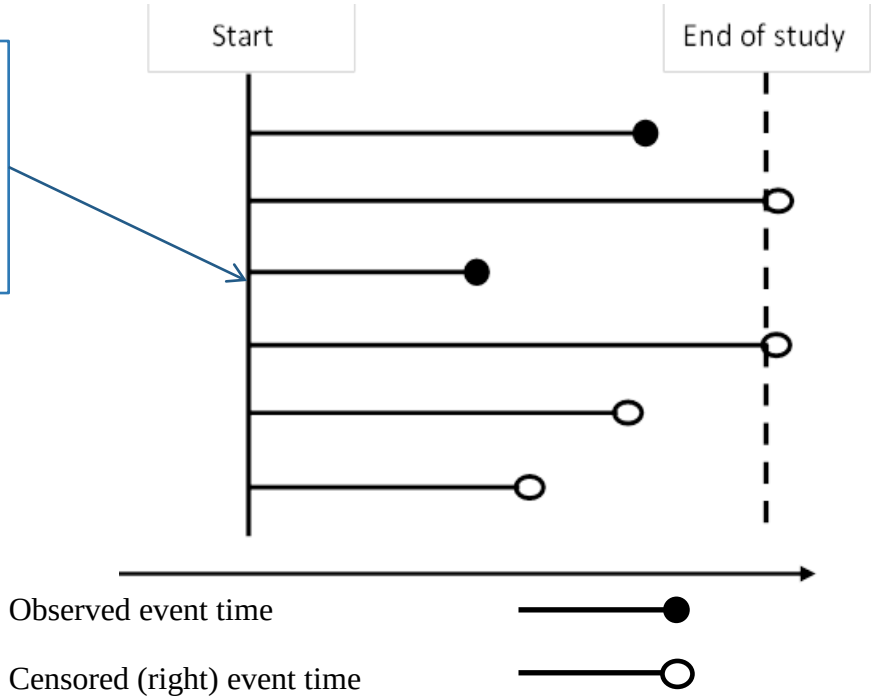
What makes time to event data different from other data?

- Censoring → Exact time of event may not be observed.
- Truncation → Subjects may be excluded when their survival times do not satisfy certain conditions.

Introduction---Censoring

Assume there is a study of cancer patients' survival after a new cancer treatment. The following plot shows a typical outcome of the study

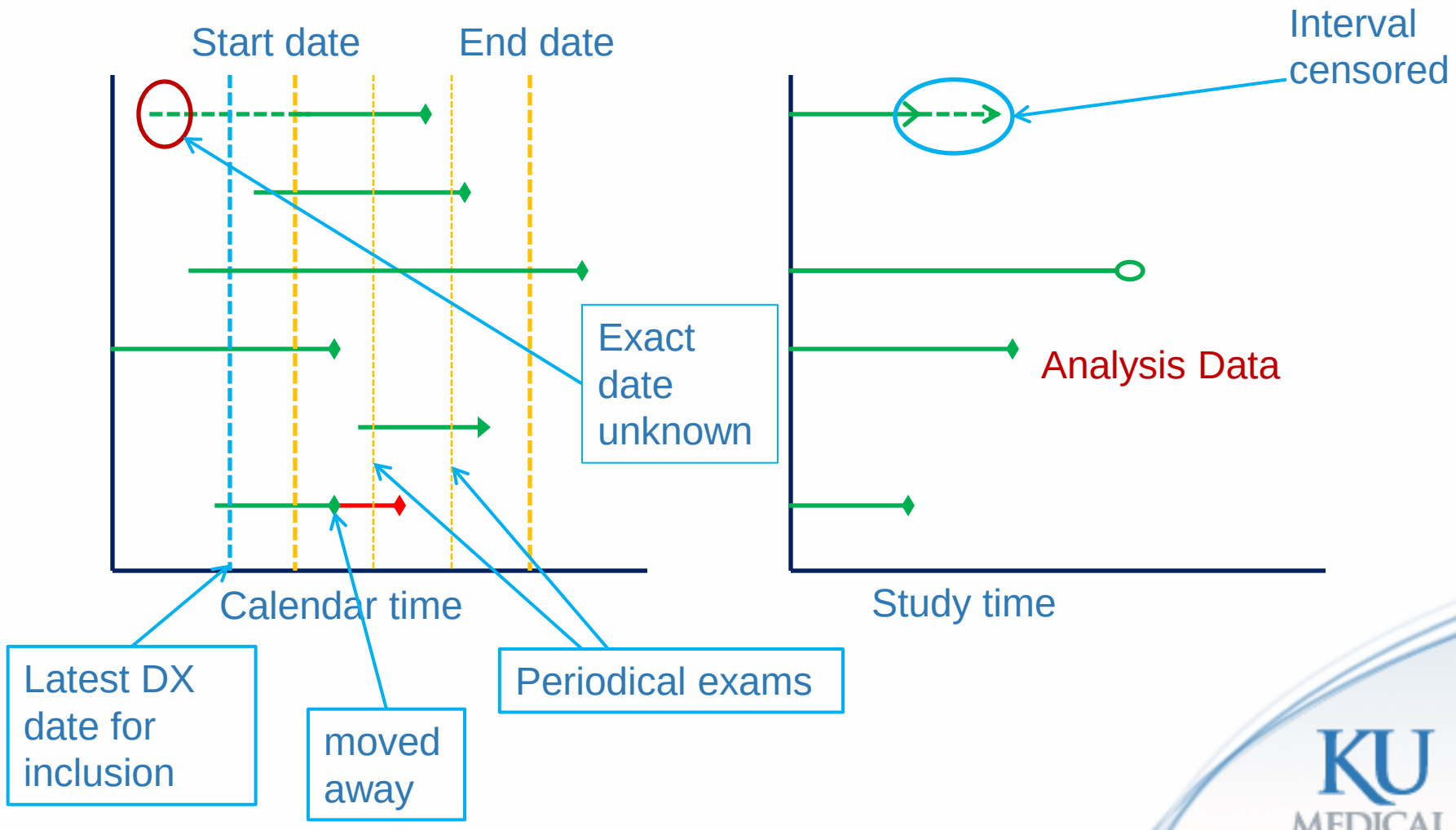
Time 0 may not be the same calendar time for subjects due to the enrollment process



Reasons of censoring could be: dropout, death unrelated to the specific cancer, end of follow-up, etc.

Introduction--- Design

Time to death from diagnosis of cancer



KU MEDICAL CENTER
The University of Kansas

Introduction---Censoring

Sometimes, an event is not observed but is known to happen after certain time point (right censored), or before certain time point (left-censored), or during a certain time period (interval-censored).

- Linear models cannot handle censored observations directly and have to exclude all unobserved observations or use computed or censored observations. The uniqueness of survival analysis is that it can handle censored observations directly.
- Logistic regression doesn't take event time into consideration and will lose some useful information. The results may be biased if we treat censored cases as non-events or exclude them from analyses.

Introduction---Censoring

In reality, right censoring is the most commonly seen situation. The observed survival time $X = \min(T, C)$, where C is the censoring time. Another important element is the censoring indicator $\delta = 1\ (\text{event})\ \text{if X} = \text{T, otherwise}\ \delta = 0(\text{censored})$.

Data: (x_1, δ_1), (x_2, δ_2), (x_3, δ_3),…, (x_n, δ_n)

For interval censored data, both the upper limit and lower limit will be provided.

Introduction---Censoring

Noninformative/Independent Censoring

Time to censoring C can be considered another time to event variable. Independent censoring means time to event of interest T and C are independent. Knowing the censoring times provide no extra information about the potential event times besides what is already known (conditional independence).

Most of the survival analyses methods are based on this assumption. However, this assumption cannot be tested within a given data (Identifiability Dilemma).

Introduction---Censoring

Noninformative/Independent Censoring

Independence assumption is often based clinical consideration. For example, time to cancer death and censoring due to traffic accidents can be considered as independent.In a clinical trial, censoring due to dropouts unrelated to disease severity can be considered independent.

Sometimes patients who are doing worse are more likely to drop out (they might think the treatment is harmful), sometimes it is the other way around (less severe patients are less motivated to follow the protocol). These types of censoring should not be considered as independent.

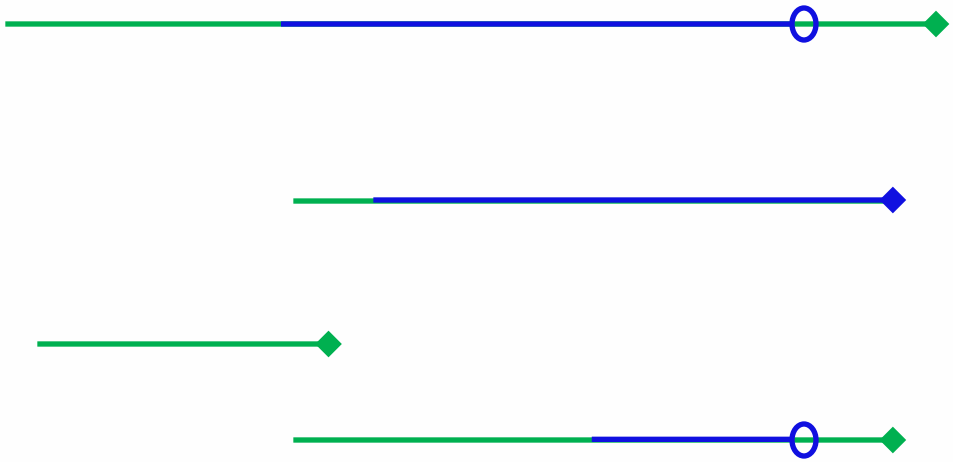
Introduction---Truncation

Truncation means a study includes only individuals with event times satisfying certain conditions. A study of survival for elderly admitted to retirement center (age has to be older enough to be retired)

- A study of first time having marihuana based on high school students (age has to be older enough to be admitted into high school).

Survival analysis have special features to handle truncation. In reality, we should be aware of the significance of truncation and interpret our results carefully.

Introduction---Truncation



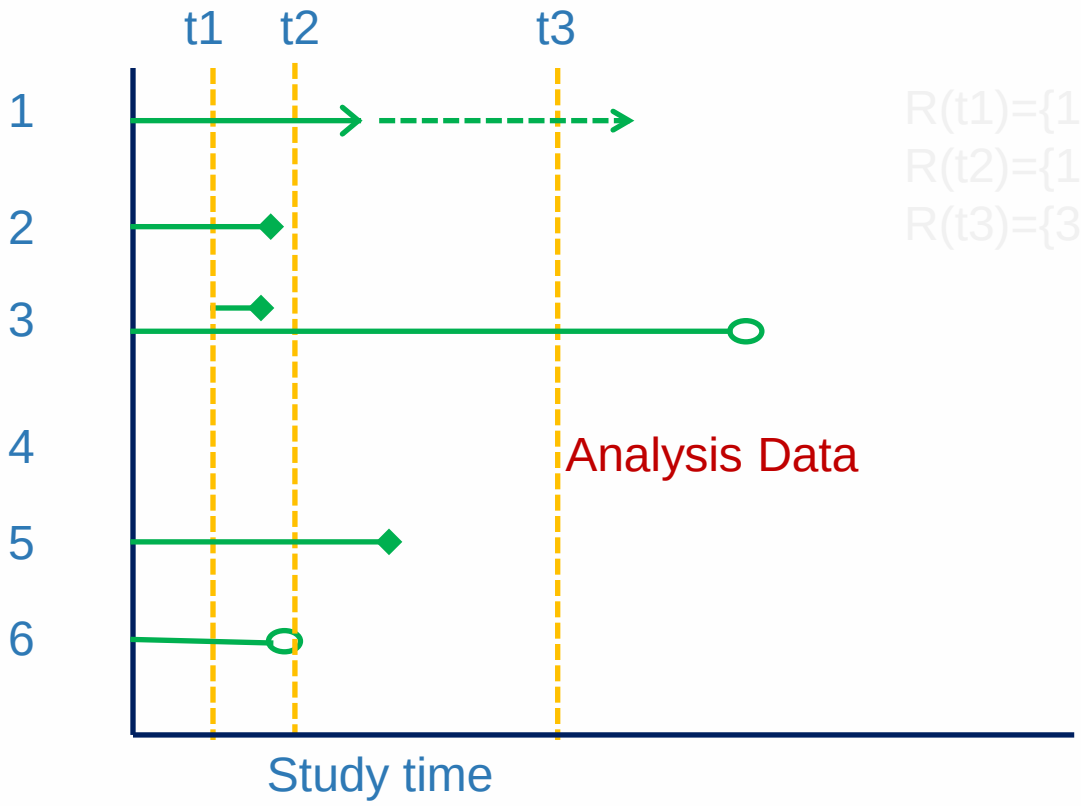
Actual process

Observed process

KU MEDICAL CENTER
The University of Kansas

Introduction--- Risk Set

Who are at risk right before time t1, t2, t3?



R(t1)={1, 2, 3, 5, 6}
R(t2)={1, 3, 5, 6}
R(t3)={3}

KU MEDICAL CENTER
The University of Kansas

Introduction---Examples

Some course materials are based on the following textbooks.

Survival Analysis: Techniques for Censored and Truncated Data, 2nd ed., by J. P. Klein and M. L. Moeschberger

Data and SAS program codes are available online

http://www.ats.ucla.edu/stat/examples/sakm/

Applied Survival Analysis 2nd ed., by D.W. Hosmer, S. Lemeshow, and S. May.

Data and program codes (SAS, STATA, SPSS)

http://www.ats.ucla.edu/stat/examples/asa2/default.htm

www.umass.edu/statdata/statdata/stat-survival.html

Introduction---Examples

Kidney Dialysis

Purpose: To assess the impact of catheter placement methods on the time to first exit-site infection (in months) in patients with renal insufficiency.

Treatment: Surgical placement (n=43) vs. Percutaneous placement (76).

Many right censored cases.

Introduction---Examples

Male Laryngeal Cancer

Purpose: To assess the impact of catheter placement methods on the time to first exit-site infection (in months) in patients with renal insufficiency.

Outcome: Intervals (in years) between first treatment and either death or the end of the study (Jan 1, 1983)

Subjects: A total of 90 males diagnosed with cancer of the larynx during the period 1970-1978 at a Dutch hospital were included in the study. There were 33, 17, 27, and 13 patients diagnosed with stage I, II, III, or IV larynx cancer, respectively.

Introduction---Examples

Bronchopulmonary Dysplasia (BPD)

Purpose: To determine factors that predict the length of time to getting off oxygen treatment for very low birth weight infants(<1500 grams) with BPD.

Outcome: the total number of days that the baby required supplemental oxygen therapy.

Treatment: surfactant replacement therapy was used starting in Aug 1989.

Design: Retrospective data collected for study period from December 1987 to March1991. A total 78 infants met the inclusion criteria, and 35 infants received surfactant replacement therapy (43 didn't). By the end of study follow-up, 5 infants remained on oxygen.

Introduction---Examples

Framingham Heart Study

A joint project of the National Heart, Lung and Blood Institute and Boston University to study risk factors of heart diseases.

The researchers recruited 5,209 men and women between the ages of 30 and 62 from the town of Framingham, Massachusetts. Since 1948, the subjects return to the study every two years for a detailed medical history, physical examination, and laboratory tests.

Introduction---Basic Functions

Continuous Survival Time X

Cumulative distribution function $F(x) = P(X \leq x)$

Probability density function

$$f(x) = \frac{dF(x)}{dx} = \lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x}$$

Survival function: $0 \leq S(x) \leq 1$

$$S(x) = P(X > x) = 1 - F(x)$$

$$= \int_x^{\infty} f(x)dx$$

Thus

$$f(x) = -\frac{dS(x)}{dx}$$

KU MEDICAL CENTER
The University of Kansas

Introduction---Basic Functions

Hazard function: $h(x) \geq 0$

$$h(x) = \lim_{\Delta x \to 0} \frac{P(x \leq X < x + \Delta x | X \geq x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\frac{P(x \leq X < x + \Delta x)}{S(x)}}{\Delta x} = \frac{f(x)}{S(x)} = \frac{-\frac{dS(x)}{dx}}{S(x)}$$

$\boldsymbol{h(x)dx}$ may be thought as the "approximate" probability for an individual “alive” at x to “die” immediately. $\boldsymbol{h(x)}$ is also called the instantaneous risk. Or we can consider the hazard function as the rate of losing the remaining life.

Cumulative Hazard function:

$$\mathrm{H}(x) = \int_x^\infty h(t)dt = -\ln(S(x))$$

KU MEDICAL CENTER
The University of Kansas

Introduction---Basic Functions

Discrete Survival Time X

X can take value $x_1, x_2, x_3, \ldots$ where $x_1 < x_2 < x_3 \ldots$

Probability mass function $p(x_j) = P(X = x_j), j = 1, 2, 3, \ldots$

Cumulative distribution function $F(x_j) = P(X \leq x_j) = \sum_{k \leq j} P(X = x_k)$

Survival function $S(x_j) = 1 - F(x_j) = \sum_{k > j} P(X = x_k)$

Hazard function $h(x_j) = p(X = x_j | X \geq x_j) = \frac{p(x_j)}{S(x_{j-1})}$

Cumulative hazard function $H(x_j) = \sum_{k \leq j} \frac{p(x_k)}{S(x_{k-1})}$

Note: $H(x_j) \neq \ln(S(x_j))$ for discrete case.

KU MEDICAL CENTER
The University of Kansas

Introduction-Some Parametric Distributions

Distribution	Hazard Rate $h(x)$	Survival Function $S(x)$	$h(x)$ Properties
Exponential $\lambda > 0, x \geq 0$	λ	$e^{-\lambda x}$	Constant hazard
Weibull $\alpha, \lambda > 0, x \geq 0$	$\alpha\lambda x^{\alpha-1}$	$e^{-\lambda x^{\alpha}}$	$\alpha = 1$ constant $\alpha > 1$ increasing $\alpha < 1$ decreasing
Log logistic $\alpha, \lambda > 0, x \geq 0$	$\frac{\alpha\lambda x^{\alpha-1}}{1+\lambda x^{\alpha}}$	$\frac{1}{1+\lambda x^{\alpha}}$	hump-shaped
Gompertz $\alpha, \theta > 0, x \geq 0$	$\theta e^{\alpha x}$	$\exp[\frac{\theta}{\alpha}(1-e^{\alpha x})]$	Increasing often used for aging
Pareto $\lambda > 0, \theta > 0, x \geq \lambda$	$\frac{\theta}{x}$	$\frac{\lambda^{\theta}}{x^{\theta}}$	Deceasing $x \geq \lambda > 0$

Others: Gamma, log normal, normal, exponential power, Inverse Gaussian, generalized Gamma, etc.

Introduction---Likelihood

Type of observations	Likelihood components
Exact lifetimes ($X = x$)	$f(x)$
Right-censored ($X > C_r$)	$S(C_r)$
Left-censored ($X < C_l$)	$1 - S(C_l)$
Interval-censored ($C_r < X < C_l$)	$S(C_l) - S(C_r)$

$$L \propto \prod_{i \in D} f_i(x_i) \prod_{i \in R} S_i(C_r) \prod_{i \in L} (1 - S_i(C_l)) \prod_{i \in I} (S_i(C_l) - S_i(C_r))$$

When there is right censoring only,

$$L \propto \prod_{i=1}^{n} [f_i(t)]^{\delta} [S_i(t)]^{1-\delta} = \prod_{i=1}^{n} [h_i(t)]^{\delta} [S_i(t)]$$

where δ is the censoring indicator that is 1 if the event time is observed, otherwise, 0.

Introduction---Likelihood

When there is truncation, denote the likelihood component based on censoring as l

Type of observations	Likelihood components
Right-truncation ($X > Y_R$ excluded)	$l/[1 - S(Y_R)]$
Left-truncation ($X < Y_L$ excluded)	$l/S(Y_L)$
Interval-censored ($X > Y_R \mid X < Y_L$ excluded)	$l/[S(Y_L) - S(Y_R)]$

For population follow-up study using age as the survival time, all individuals would be left truncated at the entry age as individuals who died before the beginning of the study would not be enrolled in the study. There will be no truncation if study time is used as survival time and the entry age is used as a covariate.

Introduction---Kaplan Meier $\hat{S}(t)$

Given a dataset with n individuals

- events occur at D distinct times $t_1 < t_2 < \cdots < t_D$
- d_i events at t_i and Y_i individuals at risk at t_i^-

Kaplan Meier (1958) (Product limit estimator) estimator of survival function

$$\hat{S}(t) = \begin{cases} 1 & if\ t < t_1 \\ \prod_{t_i \le t} \left[1 - \frac{d_i}{Y_i}\right] & if\ t \ge t_1 \end{cases}$$

Here, $\frac{d_i}{Y_i}$ can be considered as the conditional probability of having an immediate event at t_i, then $\left(1 - \frac{d_i}{Y_i}\right)$ is a natural estimator of the probability of survival at t_i conditional on survival up to t_i^-.

Introduction--- KM $\hat{S}(t)$

Time to relapse: 6, 6, 6, 6+, 7, 9+, 10, 11+, 11+, 13, 16, 17+, 19+, 20+ , 22, 23, 25+, 32+, 34+, 34+, 35+ (+ means right-censored)

6, 6, 6, 6+

7, 9+

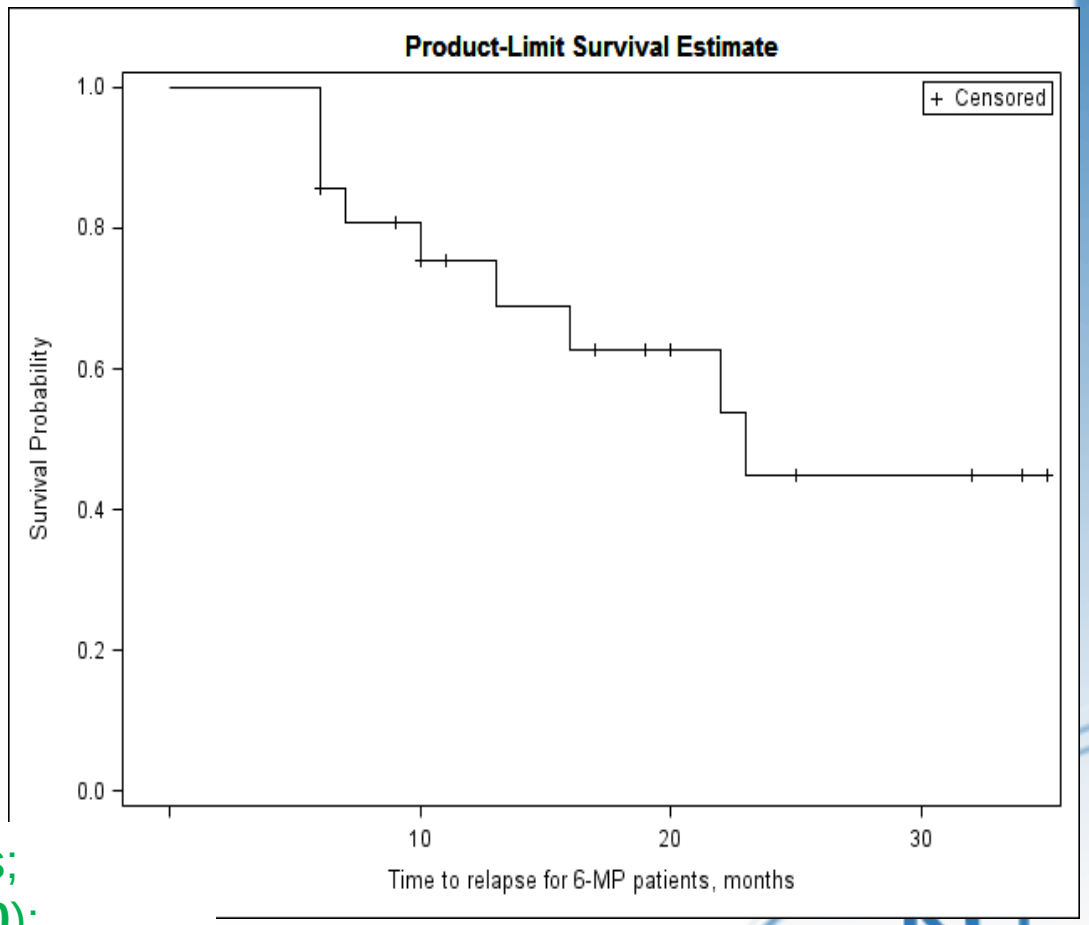
10,11+,11+

13

16,17+,19+,20+

22

23,25+,32+,34+,34+, 35+




```
proc lifetest data=relapse plots=s;
              time r_time*censor(0);
run;
```

Introduction---KM $\hat{S}(t)$

The Greenwood's formula for the variance of KM survival function:

$$\hat{V}\left[\hat{S}(t)\right] = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{Y_i(Y_i - d_i)}$$

Can be used to construct 95% pointwise confidence intervals for the survival function. However, the bounds could be outside of [0,1].

Introduction---Survival Function

Other point-wise confidence intervals for survival functions and these CIs are bounded within 0-1.

Log-log transformation: $\ln(-\ln[\hat{S}(t_0)])$

Log transformation:$\ln[\hat{S}(t_0)]$

Logit transformation:$\ln\{\frac{\hat{S}(t_0)}{1-\hat{S}(t_0)}\}$

Arcsine square-root transformation: $\arcsin[\sqrt{\hat{S}(t_0)}]$

PROC LIFETEST < options > ;

CONFTYPE= LINEAR/ASINSQRT/LOGLOG/LOG/LOGIT;

Confidence Bands (apply to the entire time range)

CONFBAND=ALL (both EP and HW)

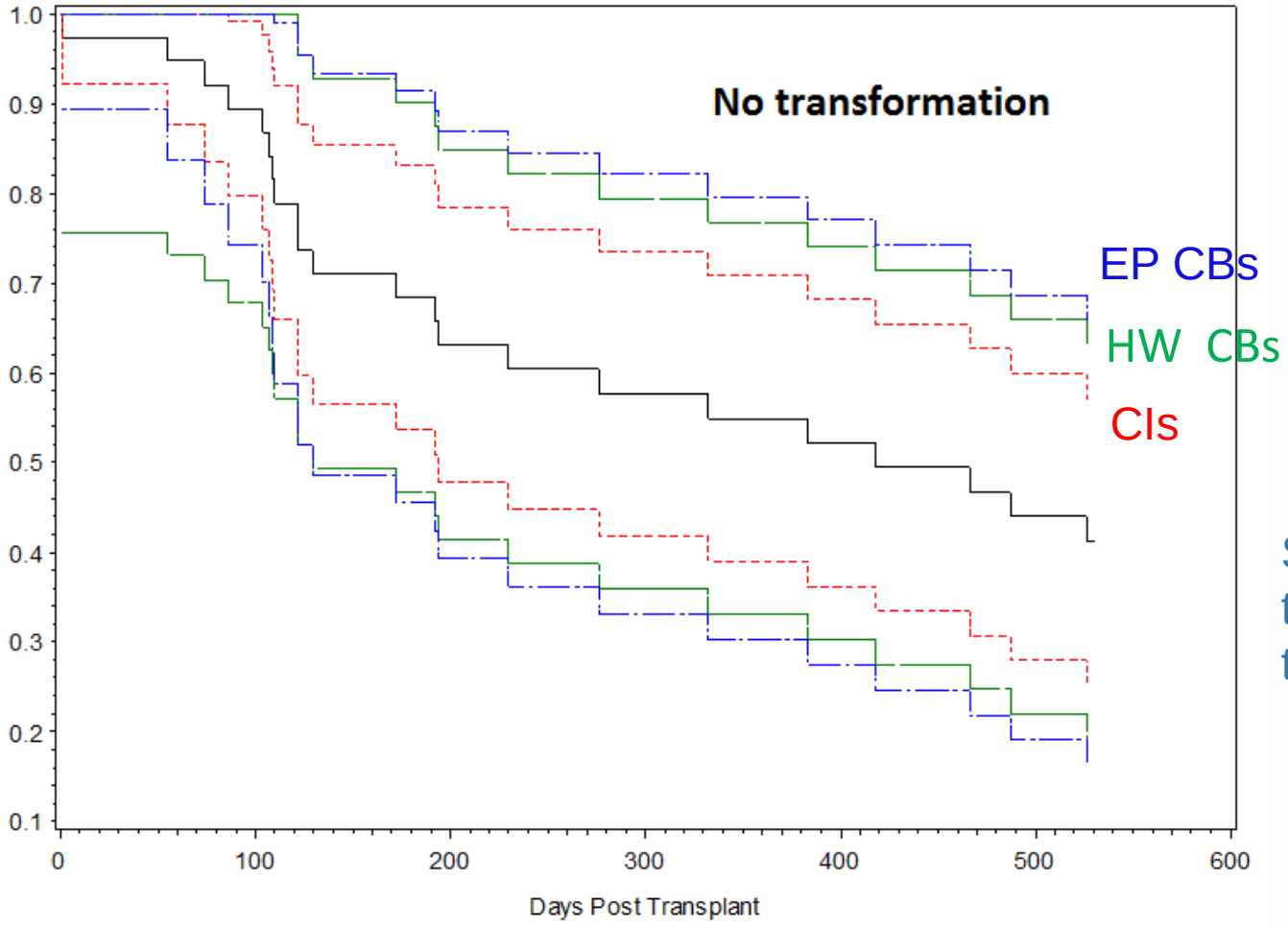
EP (equal-precision CBs)

HW (Hall-Wellner CBs)

KU MEDICAL CENTER
The University of Kansas

Introduction---Survival Function

Pointwise Confidence Intervals and Confidence Bands

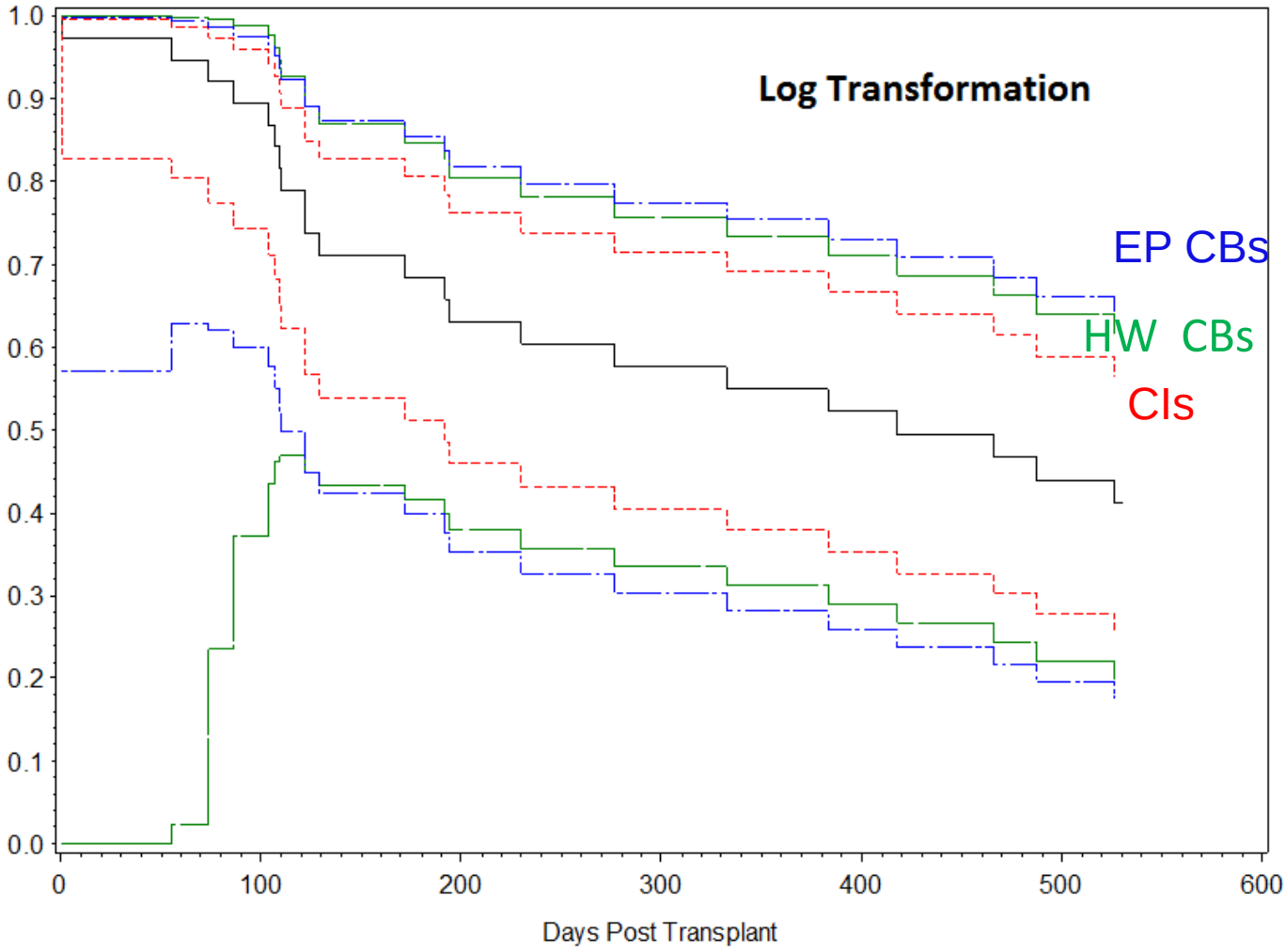


SAS automatically truncate all bounds to be within 0-1

KU MEDICAL CENTER
The University of Kansas

Introduction---Survival Function

Pointwise Confidence Intervals and Confidence Bands

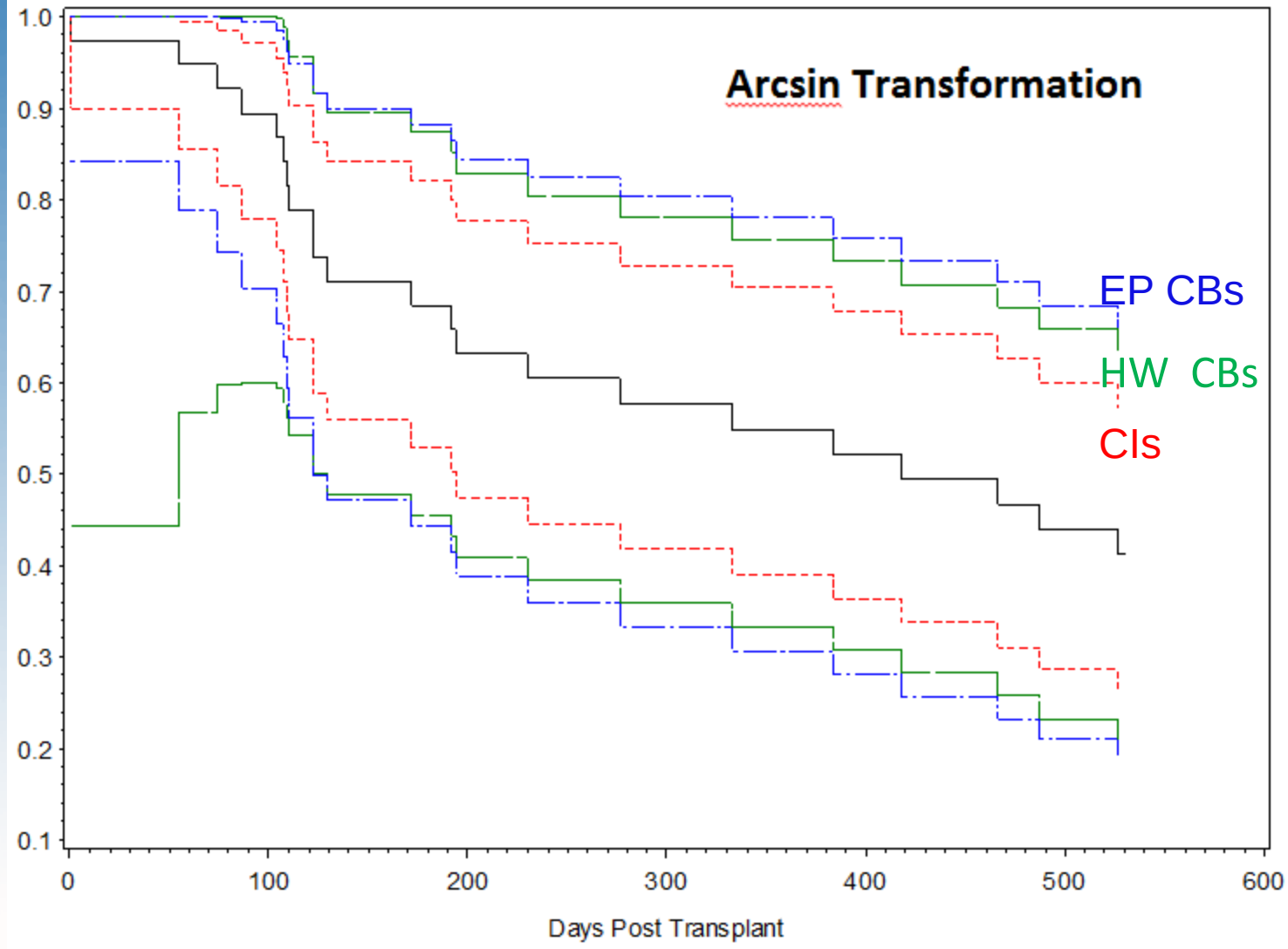


Note: Lower Confidence Bands are not monotonic

KU MEDICAL CENTER
The University of Kansas

Introduction---Survival Function

Pointwise Confidence Intervals and Confidence Bands



Note: Lower Confidence Bands are not monotonic

KU MEDICAL CENTER
The University of Kansas

Introduction---Survival Function

Life table

A life table is a summary of the survival data grouped into convenient intervals. Life table is often used in actuarial for life insurance. This approach is much more convenient than K-M estimator when the sample size is very large.

Intervals $I_1, I_2, \ldots, I_k$ and the corresponding knots: $t_0, t_1, t_2 \ldots, t_k$.

Introduction---Survival Function

Life table

An example of cost of group life insurance for adults of AIChE Group

Monthly Premium Contributions (Current as of June 1, 2015)

Monthly Rates per unit of coverage Including 20% Discount Effective through October 31, 2015						
$10,000 per unit	**$10,000 to $290,000**		**$300,000 to $540,000**		**$550,000 to $1,000,000**	
Member/Spouse Issue Age	Non-Smoker	Smoker	Non-Smoker	Smoker	Non-Smoker	Smoker
Under 30	$0.32	$0.40	$0.29	$0.36	$0.28	$0.34
30-34	0.36	0.46	0.33	0.42	0.31	0.40
35-39	0.49	0.64	0.44	0.57	0.42	0.54
40-44	0.84	0.98	0.76	0.89	0.71	0.84
45-49	1.34	1.58	1.21	1.42	1.14	1.34
50-54	2.08	2.44	1.87	2.20	1.77	2.08
55-59	3.18	3.74	2.87	3.37	2.71	3.19

Source: http://www.aicheinsurance.com/sites/aiche/Pages/GroupTerm-CoverageLimitsandRates.aspx

Life table

$I_i=(t_{i-1}, t_i]$	Risk set at t_{i-1}^-	Events in I_i	Censored cases in I_i	$P(t>t_i \mid t> t_{i-1})$
I_1	$n_0 = n$	d_1	c_1	$1 - \frac{d_1}{(n_0 - \frac{c_1}{2})}$
I_2	$n_1 = n - d_1 - c_1$	d_2	c_2	$1 - \frac{d_2}{(n_1 - \frac{c_2}{2})}$
…	…	…	…	…
I_k	$n_{k-1} = n - \sum_{j=1}^{k-1}(d_j + c_j)$	d_k	c_k	$1 - \frac{d_k}{(n_{k-1} - \frac{c_k}{2})}$

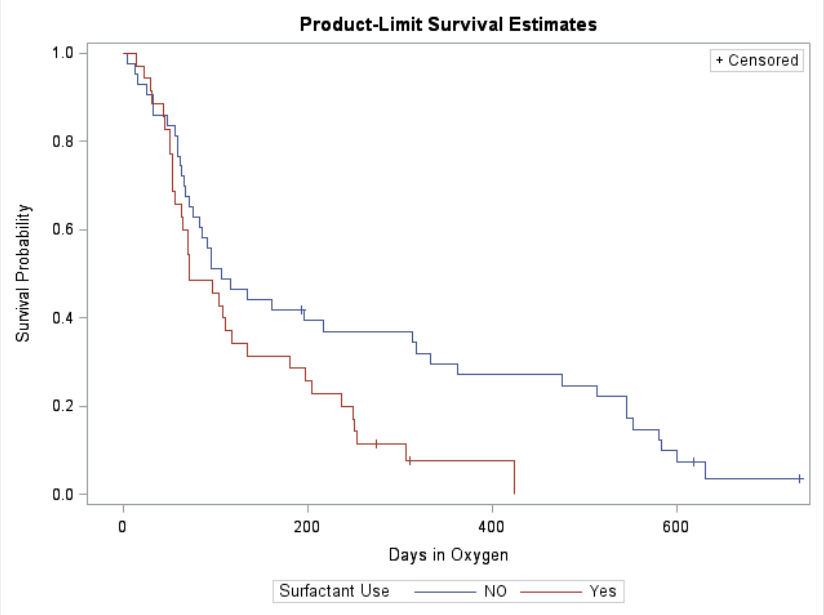
Survival Function is in a similar form of KM estimator

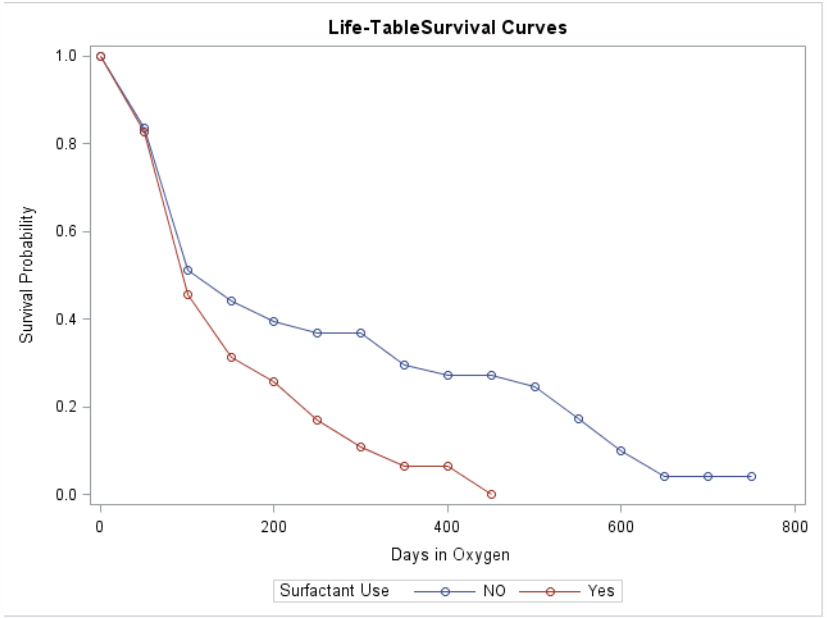
$$\check{S}(t) = \prod_{t_i \le t} \left[1 - \frac{d_i}{(n_{i-1} - \frac{c_i}{2})} \right]$$

KU MEDICAL CENTER
The University of Kansas

Introduction---Survival Function

Life table






```
proc lifetest data=bpd method=lt intervals=(0 to 750 by 50) plots=s;
time ondays*censor(0) ;
strata surfact;
run;
```

KU MEDICAL CENTER
The University of Kansas

Introduction

Nelson-Aalen estimator of cumulative hazard function

$$\tilde{H}(t) = \begin{cases} 0 & if\ t < t_1 \\ \sum_{t_i \le t} \frac{d_i}{Y_i} & if\ t \ge t_1 \end{cases}$$

Based on this estimator, an alternative estimator of the survival function is given by

$$\tilde{S}(t) = \exp\left[-\tilde{H}(t)\right]$$

This estimate is equal or greater than the K-M estimate.

Note: The N-A estimator $\tilde{H}(t)$ is the first term in a Taylor series expansion of $-ln\left[\hat{S}(t)\right] = \hat{H}(t)$.

KU MEDICAL CENTER
The University of Kansas

Introduction

Estimate of mean

$$\mu = \int_0^\infty x f(x) dx = \int_0^\infty -x dS(x) = -xS(x)|_0^\infty + \int_0^\infty S(x) dx$$

$$= \int_0^\infty S(x) dx \text{ , we have } \hat{\mu}_\tau = \int_0^\tau \hat{S}(t) dt$$

where τ is the longest observed event time or a pre-specified value by study investigators. This is equivalent to assuming all individuals survival after τ die immediately at τ.

$$\hat{V}^*[\hat{\mu}_\tau] = \frac{n_d}{n_d - 1} \sum_{i=1}^{D} \left[\int_{t_i}^{\tau} \hat{S}(t) dt \right]^2 \frac{d_i}{Y_i(Y_i - d_i)}$$

where n_d is the total # of observed events and $\frac{n_d}{n_d - 1}$ is a adjusting factor. (SAS does so)

Introduction

Estimates of mean

$$\hat{\mu}_\tau = \int_0^\tau \hat{S}(t)dt$$

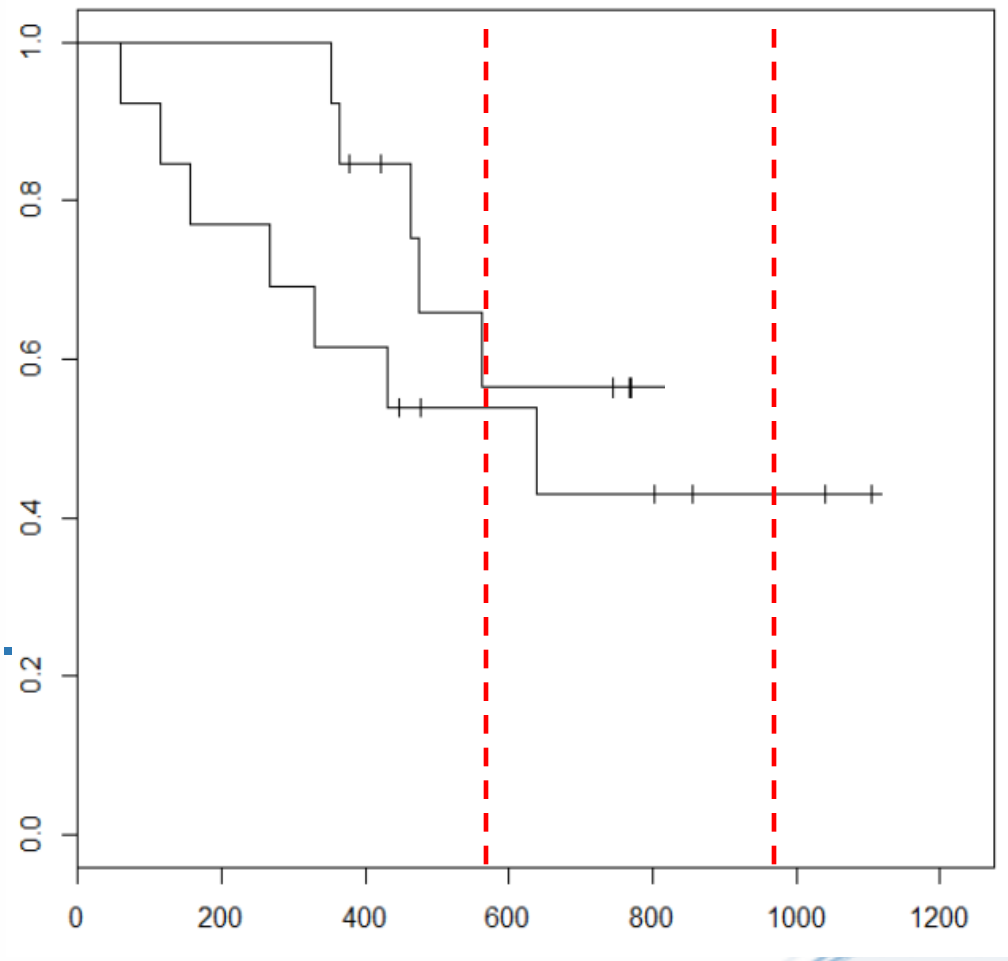
this is equivalent to the area under the survival curve up to τ.

When we compare means among groups, make sure the same τ is applied to all groups for a fair comparison, and τ is appropriate for all groups ($\tau < \max_i(t_{ij})$ for $\forall\, i$).

Introduction

Estimate of mean

Note: SAS calculates mean to the Maximum event time (not including censored cases) without specifying the time limit.




```
Proc lifetest data=data TIMELIM=500;
     time time*censor(0);
run;
```

KU MEDICAL CENTER
The University of Kansas

Introduction

Median or quantiles

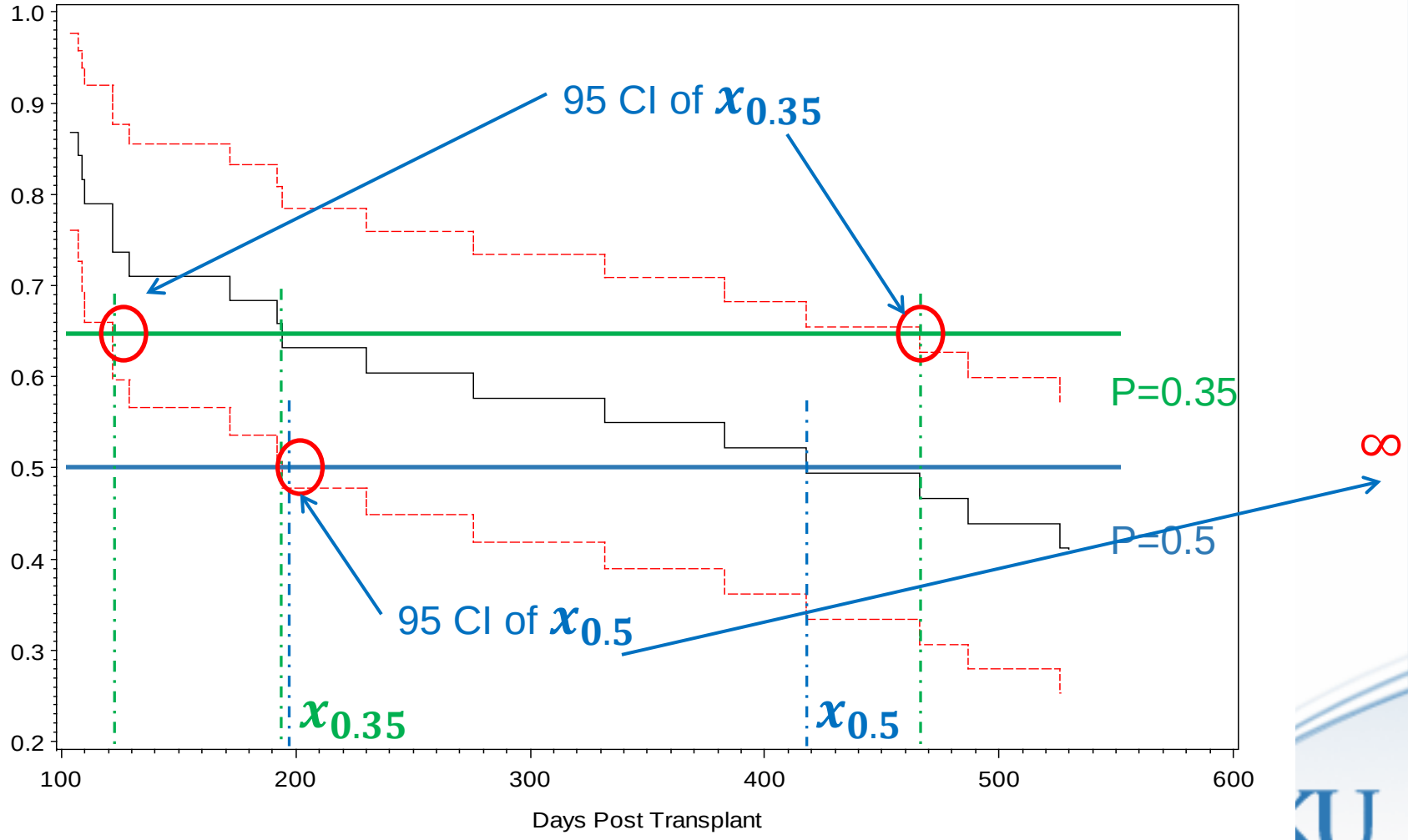
$$x_p = \inf\{t: S(t) \leq 1 - p\}$$

It is easy to identify quantiles based on a plot of the survival function.

1. Draw a horizontal line at $1 - p$
2. Look for the time value where survival function intersect with the horizontal line. If there are many such values, choose the smallest value.

Confidence intervals for x_p can be identified visually.

Introduction



Introduction

Group Comparison---Log Rank Test

Assume there are K samples ($K \geq 2$)

Null hypothesis: All the groups have the same hazard rate for all $t \leq \tau$.

$$H_0: h_1(t) = h_2(t) = \cdots = h_K(t)$$

Alternative hypothesis: At least one group has a different hazard rate for some $t \leq \tau$

$$H_A: at\ least\ one\ of\ the\ h_j(t)'s\ is\ different\ for\ some\ t \leq \tau.$$

Note: the key factor for power is the number of events rather than the total sample size.

KU MEDICAL CENTER
The University of Kansas

Introduction

Group Comparison---Log Rank Test

Let $t_1 < t_2 < ... < t_D$ be the distinct deaths times in the pooled sample.

d_{ij}—# of events at t_i from the jth group. $\sum_{j=1}^{K} d_{ij} = d_i$ (# of events at t_i from the pooled sample)

Y_{ij}—# of individuals at risk right before t_i from the jth group. $\sum_{j=1}^{K} Y_{ij} = Y_i$ (risk set of the pooled sample)

$i = 1,2 \ldots D$ and $j = 1,2 \ldots K$

$$Z_j(\tau) = \sum_{i=1}^{D} W(t_i) \left\{ d_{ij} - Y_{ij} \frac{d_i}{Y_i} \right\}, j = 1,2 \ldots K$$

where $W(t_i)$ is a weight function defined by the pooled data. There are many options for the weight function.

Introduction

Group Comparison---Log Rank Test

The variance for $Z_j(\tau)$ is given by

$$\hat{\sigma}_{jj}(\tau) = \sum_{i=1}^{D} W^2(t_i)\ d_i \frac{Y_{ij}}{Y_i}\left(1 - \frac{Y_{ij}}{Y_i}\right)\left(\frac{Y_i - d_i}{Y_i - 1}\right),$$

$$j = 1,2 \ldots K$$

and the covariance of $Z_j(\tau), Z_g(\tau)$ is expressed by

$$\hat{\sigma}_{jg}(\tau) = -\sum_{i=1}^{D} W^2(t_i)\, d_i \frac{Y_{ij}}{Y_i}\frac{Y_{ig}}{Y_i}\left(\frac{Y_i - d_i}{Y_i - 1}\right), j \neq g.$$

The test statistic for H0 is a quadratic form

$$(Z_1(\tau), \ldots Z_{K-1}(\tau))\boldsymbol{\Sigma}^{-1}(Z_1(\tau), \ldots Z_{K-1}(\tau))^T \sim \chi^2_{K-1}\ under\ H0$$

for large samples.

Introduction

Test	$W(t_i)$
Log-Rank	1
Gehan (called Wilcoxon in SAS)	Y_i
Tarone-Ware	$\sqrt{Y_i}$
Peto-Peto	$\tilde{S}(t_i)$
Modified Peto-Peto	$\tilde{S}(t_i)Y_i/(Y_i+1)$
Fleming-Harrington (define $0^0 = 1$)	$[\hat{S}(t_{i-1})]^p[1-\hat{S}(t_{i-1})]^q$ $p \geq 0, q \geq 0$

$\tilde{S}(t_i)$ is similar to KM $\hat{S}(t_i)$ with Y_i replaced by $Y_i + 1$;

For FH weight, p and q adjust weight at different t.

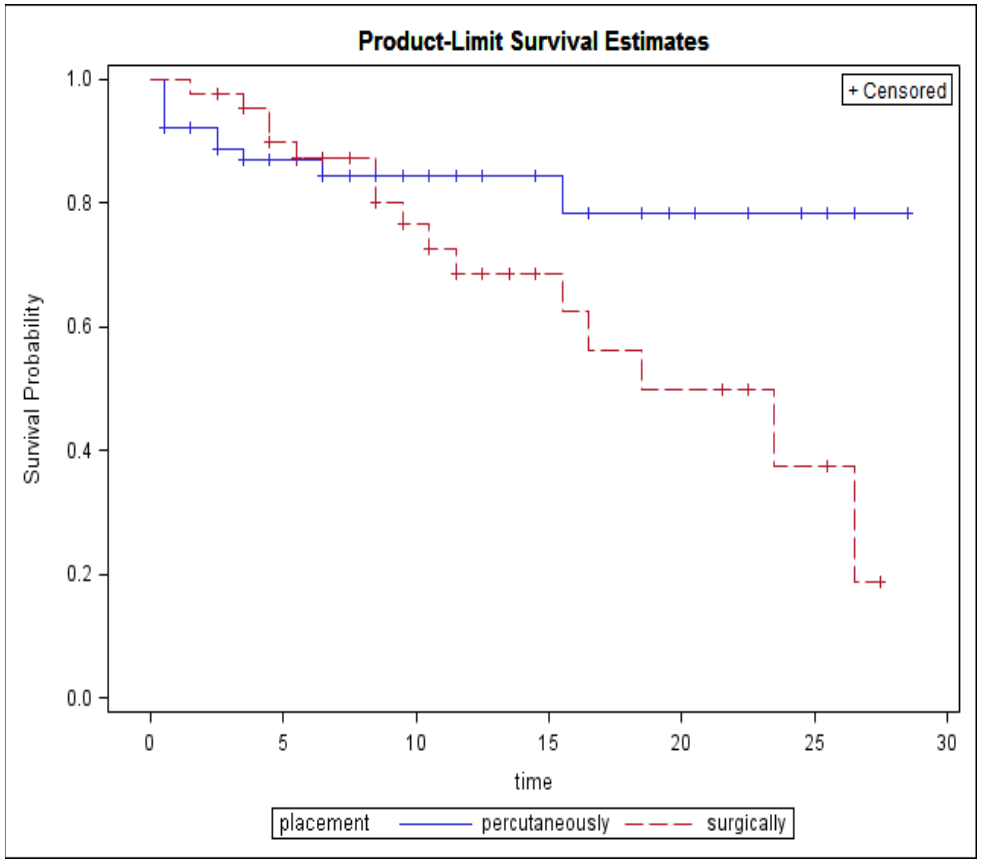
p=q=0 → log rank p=1, q=0 → close to Peto-Peto

p>q → relatively more weight for small t, otherwise, relatively more weight for large t

Introduction

Two methods for placing catheters in kidney dialysis patients: surgical placement vs. percutaneous placement

Is there a difference in terms of infection risk between the two methods?



KU MEDICAL CENTER
The University of Kansas

Introduction

Test of Equality over Group

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	2.5295	1	0.1117
Wilcoxon	0.0021	1	0.9636
Tarone	0.4027	1	0.5257
Peto	1.3992	1	0.2369
Modified Peto	1.2759	1	0.2587
Fleming(0,1)	9.6680	1	0.0019

```
proc lifetest data=kidney outsurv=a;
time time*infect(0);
strata /test= all group=placement FLEMING(0, 1);
run;
```

*All include FLEMING(1,0) (default). When FLEMING(**0**, **1**) is specified, FLEMING(1,0) will be replaced;

KU MEDICAL CENTER
The University of Kansas

Introduction

With different weight functions, different conclusions can be made. In reality, which weight to use should be pre-specified based on clinical considerations. For example, sometimes it may be more relevant to consider the difference between two treatments immediately after baseline.

Note: these tests average differences across time that they are not appropriate for survival functions crossing each other.

Introduction

When survival functions cross each other, Renyi-type tests are recommended.

SAS macro for this test is available at the following paper online: http://pharmasug.org/proceedings/2011/SP/PharmaSUG-2011-SP06.pdf

These tests are based on the supreme of the weighted (Observed-Expected) difference over time instead of the total weighted (Observed-Expected) difference.

%_***renyi***(kidney,placement,time,infect,**0**,LOGRANK,**0** ,**0**);

Renyi Test Statistic: 2.1421242939

Approximate P-value: 0.0643669627

KU MEDICAL CENTER
The University of Kansas

Introduction

Test for trend

Assume there are K samples

Null hypothesis: For $t \leq \tau$,

$$H_0: h_1(t) = h_2(t) = \cdots = h_K(t)$$

Alternative hypothesis: For $t \leq \tau$

$$H_A: h_1(t) \leq h_2(t) \leq \cdots \leq h_K(t)$$

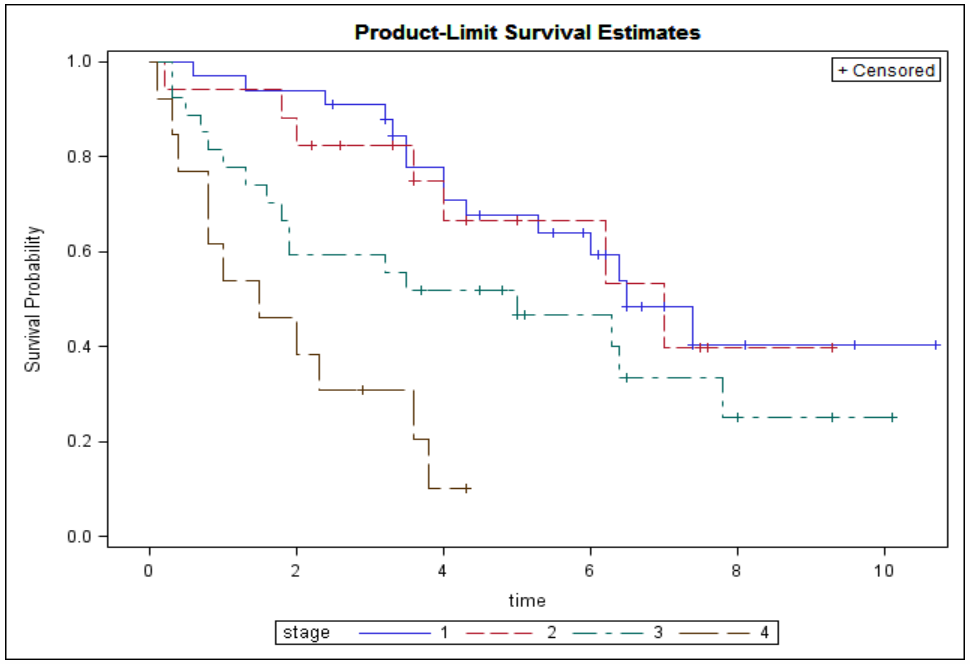
with at least one strict inequality

```
proc lifetest data=Larynx;
time time*death(0);
strata /trend group=stage all;
run;
```

KU MEDICAL CENTER
The University of Kansas

Introduction

Test for trend--Larynx cancer patients at different stages



Test	z-Score	Pr > |z|
Log-Rank	3.7190	0.0002
Wilcoxon	4.2248	<.0001
Tarone	4.0580	<.0001
Peto	4.1293	<.0001
Modified Peto	4.1363	<.0001
Fleming(1)	4.1201	<.0001

When there is no censoring, the test using Gehan or Peto-Peto weights reduces to the Jonckheere-Terpstra test (trend test for categorical data).

KU MEDICAL CENTER
The University of Kansas

Introduction

Summary

Survival is a process

Survival Function/Hazard Function

Censoring and truncation

Log-rank test

- Which weight function should we use?
- What if survival functions cross each other

KU MEDICAL CENTER
The University of Kansas

outline

Introduction

- Time to event data /Censoring and truncation
- Basic survival functions/Group comparison of survival data

Regression Analyses

- **Parametric Models -- Accelerate failure time models**
- **Semiparametric Models – Cox models**

Extension to the Cox models

- Weighted Cox models
- Time-dependent covariate Cox models
- Time-varying coefficient survival models

Other topics

- Competing risk modeling
- Bayesian survival analyses

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Accelerated Failure Time (AFT) models

Loglinear model of survival time

$$\ln(X) = \mu + \boldsymbol{\beta}^T \boldsymbol{Z} + \sigma W$$

where $\beta^T = (\beta_1, \ldots, \beta_p)$ is a vector of regression coefficients and W is the error distribution.

Different error distributions provide different AFT models.

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Accelerated Failure Time (AFT) models

Multiplicative effect of risk factors on failure time X

$$E(X|\boldsymbol{Z}) = \mathrm{E}[\exp(\mu + \boldsymbol{\beta}^T\boldsymbol{Z} + \sigma W)]$$
$$= \exp(\boldsymbol{\beta}^T\boldsymbol{Z})E(X_0)$$

where $E(X_0)$ is the expected survival time when the values of all risk factors are 0s.

Here, $e^{-\boldsymbol{\beta}^T\boldsymbol{Z}}$ is called the acceleration factor (sometimes $e^{\boldsymbol{\beta}^T\boldsymbol{Z}}$ is called that), denoting the accelerating effects on the survival processes, such as aging, cancer advancing, or recovering after surgeries etc.

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Accelerated Failure Time (AFT) models

$S(x|\boldsymbol{Z}) = S_0\left(xe^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)$, where $S_0(x)$is the survival function for $X_0(\boldsymbol{Z}=0)$.

The implication is that the median survival time for an individual with a covariate $\boldsymbol{Z}$ is the median survival time of X_0 times $e^{\boldsymbol{\beta}^T\boldsymbol{Z}}$ (this is true for all quantiles).

$$h(x|\boldsymbol{Z}) = e^{-\boldsymbol{\beta}^T\boldsymbol{Z}}h_0(xe^{-\boldsymbol{\beta}^T\boldsymbol{Z}}) \ for\ all\ x$$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Weibull AFT models: $\ln(X) = \mu + \boldsymbol{\beta}^T \boldsymbol{Z} + \sigma W$

$f_W(w) = \exp(w - e^w)$ (extreme value distribution)

$Y = e^w \sim exponential(1)$

$$X = e^{\mu+\boldsymbol{\beta}^T \boldsymbol{Z}+\sigma W} = [e^{\frac{\mu+\boldsymbol{\beta}^T \boldsymbol{Z}}{\sigma}} e^w]^\sigma = [e^{\frac{\mu+\boldsymbol{\beta}^T \boldsymbol{Z}}{\sigma}} Y]^\sigma$$

$$\sim Weibull(1/\sigma, e^{-\frac{\mu+\boldsymbol{\beta}^T \boldsymbol{Z}}{\sigma}})$$

$X \sim exponential[e^{-(\mu+\boldsymbol{\beta}^T \boldsymbol{Z})}]$ if σ=1, which is a memoryless process, that is,

$$\mathrm{E}(X - x | X \geq x) = E(X) \; for \; \forall x > 0.$$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Weibull AFT models: $\ln(X) = \mu + \boldsymbol{\beta}^T \boldsymbol{Z} + \sigma W$

$$X \sim Weibull(1/\sigma, e^{-\frac{\mu+\boldsymbol{\beta}^T \boldsymbol{Z}}{\sigma}})$$

This AFT model is flexible in modeling constant (σ=1), increasing (σ<1), or decreasing (σ >1) hazard rate.

This is the AFT model with a proportional hazards representation

$$h(x|\boldsymbol{Z} = \boldsymbol{z_1})/h(x|\boldsymbol{Z} = \boldsymbol{z_2}) = e^{-\boldsymbol{\beta}^T(\boldsymbol{z_1}-\boldsymbol{z_2})\alpha} = e^{-\boldsymbol{\beta}^T \Delta \boldsymbol{z}\frac{1}{\sigma}}$$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Log Logistic AFT models: $\ln(X) = \mu + \boldsymbol{\beta}^T \boldsymbol{Z} + \sigma W$

$f_W(w) = e^w/(1+e^w)^2$

$Y = e^W \sim log - logistic(1,1)\ \ (f_Y(y) = \frac{1}{(1+y)^2})$

$X_0 = e^{\mu+\sigma W} = [e^{\frac{\mu}{\sigma}} e^W]^\sigma \sim llog(1/\sigma, e^{-\mu/\sigma})$

$X = e^{\mu+\boldsymbol{\beta}^T \boldsymbol{Z}+\sigma W} = [e^{\frac{\mu+\boldsymbol{\beta}^T \boldsymbol{Z}}{\sigma}} e^W]^\sigma \sim llog(1/\sigma, e^{-\frac{\mu+\boldsymbol{\beta}^T \boldsymbol{Z}}{\sigma}})$

Hump-shaped hazard rate. The only AFT model with a proportional odds representation

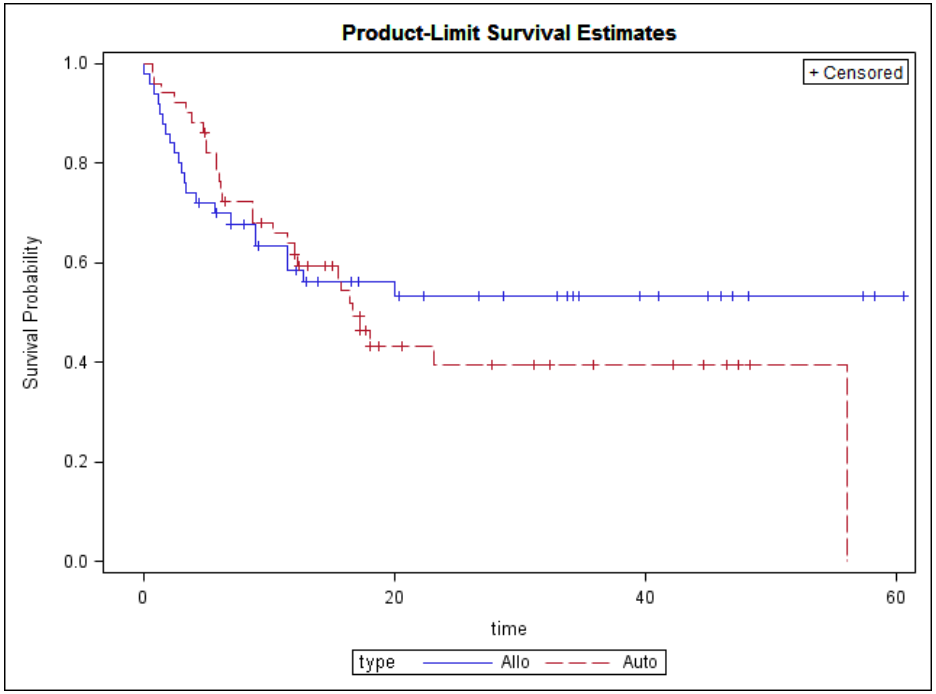
$$\frac{S(x|\boldsymbol{Z})}{1-S(x|\boldsymbol{Z})} = e^{\alpha\, \boldsymbol{\beta}^T \boldsymbol{Z}} \frac{S_0(x)}{1-S_0(x)} = e^{\frac{1}{\sigma} \boldsymbol{\beta}^T \boldsymbol{Z}} \frac{S_0(x)}{1-S_0(x)}$$

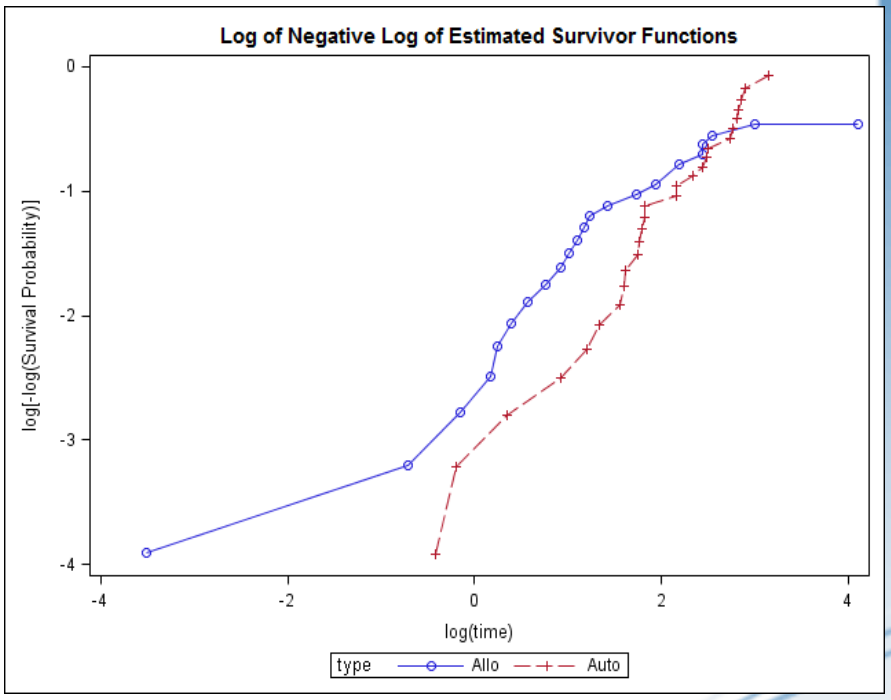
Odds ratio $(\boldsymbol{Z_1}/\boldsymbol{Z_2})$=$e^{\frac{1}{\sigma} \boldsymbol{\beta}^T (\boldsymbol{Z_1}-\boldsymbol{Z_2})}$ for all x.

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Allo and Auto bone marrow transplants for leukemia patients





KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Allo and Auto bone marrow transplants for leukemia patients

```
proc lifereg data = transplant;
class type;
model time*ind(0)= type /covb distribution=weibull;
run;
```

	Parameter		DF	Estimate	Standard Error	95% Confidence Limits	
$\hat{\mu}$	Intercept		1	3.5947	0.2896	3.0272	4.1622
$\hat{\beta}_{allo}$	Type	Allo	1	0.3736	0.4204	-0.4504	1.1977
	Type	Auto	0	0.0000	.	.	.
$\hat{\sigma}$	Scale		1	1.4737	0.1811	1.1582	1.8751
$\hat{\alpha}$	Weibull Shape		1	0.6786	0.0834	0.5333	0.8634

Other distribution options:
loglogistic/exponential/ lognormal/ gamma

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Allo and Auto bone marrow transplants for leukemia patients

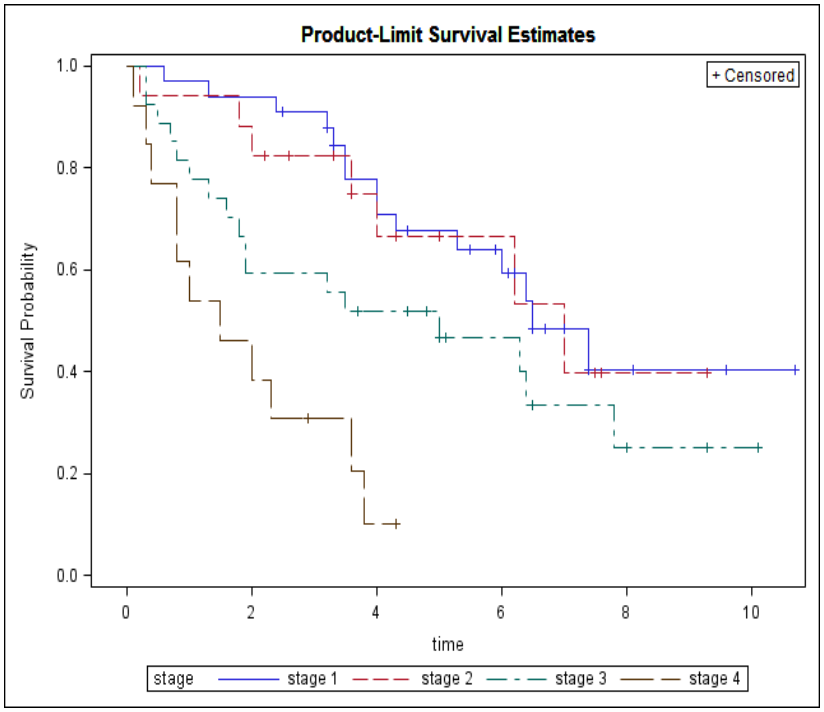
Weibull AIC (smaller is better)	**450.084**
Exponential AIC (smaller is better)	**459.976**
LLogistic AIC (smaller is better)	**446.458**
Lognormal AIC (smaller is better)	**445.987**
Gamma AIC (smaller is better)	**447.965**

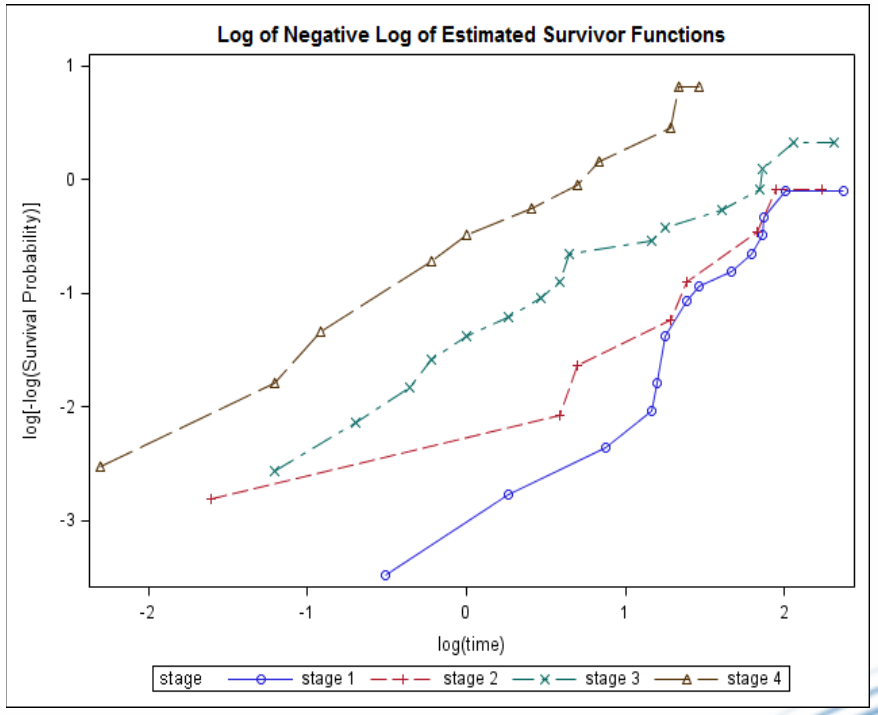
For this data, log normal seems the best fit.

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Patients with larynx cancer at different stages





KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Patients with larynx cancer at different stages

Exponential AIC (smaller is better) **293.796**

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.7550	0.9902	1.8143	5.6957	14.38	0.0001
Stage II	1	-0.1456	0.4602	-1.0475	0.7563	0.10	0.7517
Stage III	1	-0.6483	0.3552	-1.3444	0.0478	3.33	0.0680
Stage IV	1	-1.6350	0.3985	-2.4161	-0.8540	16.83	<.0001
age	1	-0.0197	0.0142	-0.0476	0.0081	1.93	0.1651
Scale	0	1.0000	0.0000	1.0000	1.0000		
Weibull Shape	0	1.0000	0.0000	1.0000	1.0000		

The model setting is

$$\ln(X_i) = \mu + \boldsymbol{\beta}^T \boldsymbol{Z_i} + w_i$$

$$\{w_i, i = 1,2, \ldots n\} \; are \; independent \; with \; f_W(w)$$

$$= \exp[-\exp(w)]$$

The estimated model is (ignore the significance of the tests)

$$\ln(X_i) = 3.755 - 0.1456 * stageII - 0.6483 * stageIII$$

$$- 1.6350 * stageIV - 0.0197 * age \; + w_i$$

Regression Analyses

Patients with larynx cancer at different stages

So the estimated model is (ignore the significance of the tests)

$$\ln(X_i) = 3.755 - 0.1456 * stageII - 0.6483 * stageIII - 1.6350 * stageIV - 0.0197 * age + w_i$$

Q1. With age fixed, the expected survival time changes by how much when the cancer stage changes from stage I to stage II, from stage I to stage IV?

$$\exp(-0.1456) = 0.86$$

The expected survival time decreases by 14% when the cancer is stage II vs. stage I.

Q2. With cancer stage fixed, the expected survival time changes by how much when age is 5 years older.

$$\exp(-0.0197 * 5) = 0.91$$

The expected survival time decreases by 9% when the patient is 5 years older.

Regression Analyses

Patients with larynx cancer at different stages

So the estimated model is (ignore the significance of the tests)

$$\ln(X_i) = 3.755 - 0.1456 * stageII - 0.6483 * stageIII - 1.6350 * stageIV - 0.0197 * age + w_i$$

Q3. What is the median survival time for an individual of age 65 with Larynx cancer of stage I, stage II, stage III, and Stage IV?

$$Y = \exp(w) \sim exponential(1)$$

$$S(y) = exp(-y) = 0.5 \Rightarrow M_y = \ln(2)$$

$$M_{(stgI,age=65)} = \exp(3.755 - 0.0197 * 65) * M_y = 8.2 \text{ yrs}$$

$$M_{(stgII,age=65)} = \exp(3.755 - 0.1456 - 0.0197 * 65) * M_y = 7.1 \text{ yrs}$$

$$M_{(stgIII,age=65)} = \exp(3.755 - 0.6483 - 0.0197 * 65) * M_y = 4.3 \text{ yrs}$$

$$M_{(stgV,age=65)} = \exp(3.755 - 1.6350 - 0.0197 * 65) * M_y = 1.6 \text{ yrs}$$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Summary---AFT models

Parametric models with options for different hazard function forms (constant, increasing, decreasing, hump-shaped)

Multiplicative effects of factors on survival times

This type of models are not used as commonly as general linear models for continuous variables in reality. Why?

A more flexible model that is easy to implement: The Cox model

Regression Analyses

Semiparametric approach: the Cox Model

Data $[(T_j, \delta_j, Z_j], j = 1, 2, \ldots, n]$.

$$\lambda(t|Z) = \lambda_0(t)\exp(\boldsymbol{\beta}^T \boldsymbol{Z}) = \lambda_0(t)\exp\left(\sum_{k=1}^{p} \beta_k Z_k\right)$$

$h_0(t)$: the arbitrary baseline hazard rate for individuals with $\boldsymbol{Z} = \mathbf{0}$. The functional form of $h_0(t)$ is not specified (**nonparametric**).

$\exp(\boldsymbol{\beta}^T \boldsymbol{Z})$: the effects of covariates are multiplicative on hazard function. The function is specified (**parametric**).

***(Cox, Regression Models and Life Tables, 1972)**

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: proportional hazards model

Given two individuals with covariates Z and Z^*

$$HR = \frac{H(t|\boldsymbol{Z})}{H(t|\boldsymbol{Z}^*)} = \frac{\lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z})}{\lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}^*)} = \exp[\boldsymbol{\beta}^T(\boldsymbol{Z}-\boldsymbol{Z}^*)]$$

The hazard rates at any time point are proportional with a fixed ratio independent of t (constant hazard ratio with given Z and Z^*).→ survival functions should not cross each other.

Given a parameter β of covariate Z, the interpretation of β is the log(HR) of 1 unit change in Z.

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model

As the baseline hazard is not specified, the full likelihood is unknown. how can the parameters be estimated?

Regression Analyses

Cox Model: partial likelihood (no ties)

Given covariates Z_j, the event and censoring times for the jth individual are independent (**Independent/noninformative censoring)**.

D events at D distinct time points (no ties).

$$t_1 < t_2 < \cdots < t_D$$

$Z_{(i)k}$ denotes the kth covariates of the individual with an event at t_i, then the partial likelihood is

$$l_p(\boldsymbol{\beta}) = \prod_{\boldsymbol{i=1}}^{\boldsymbol{D}} \frac{\exp\left(\sum_{k=1}^{p} \beta_k Z_{(i)k}\right)}{\sum_{\boldsymbol{j \in R(t_i)}} \exp\left(\sum_{k=1}^{p} \beta_k Z_{jk}\right)}$$

where $R(t_i)$ is the risk set at t_i, including all individuals at risk at t_i.

Note, the log partial likelihood doesn't depend on $h_0(t)$.

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: maximum partial likelihood estimators

The log partial likelihood is

$$L_p(\boldsymbol{\beta}) = \boldsymbol{ln}\left[l_p(\boldsymbol{\beta})\right] = \sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} \sum_{k=1}^{p} \beta_k Z_{(i)k} - \sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} ln[\sum_{\boldsymbol{j \in R(t_i)}} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)]$$

$$\frac{\partial L_p(\boldsymbol{\beta})}{\partial \beta_l} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} Z_{(i)l} - \sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} \frac{\sum_{\boldsymbol{j \in R(t_i)}} Z_{jl} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)}{\sum_{\boldsymbol{j \in R(t_i)}} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)} = \mathbf{0}$$

$$\sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} [Z_{(i)l} - \frac{\sum_{\boldsymbol{j \in R(t_i)}} Z_{jl} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)}{\sum_{\boldsymbol{j \in R(t_i)}} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)}] = \mathbf{0}$$

$$\sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} (Z_{(i)l} - \sum_{\boldsymbol{j \in R(t_i)}} Z_{jl} W_{jt_i}) = 0, \; where \; W_{jt_i} = \frac{\exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)}{\sum_{\boldsymbol{j \in R(t_i)}} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)})$$

$$\sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} (Z_{(i)l} - \bar{Z}_{wi}) = 0$$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: maximum partial likelihood estimators

The information matrix $I(\boldsymbol{\beta})$ is the negative of the matrix of second derivatives of log likelihood.

$$I(\boldsymbol{\beta}) = [I_{gl}(\boldsymbol{\beta})]_{p\times p}$$

With

$$I_{gl}(\boldsymbol{\beta}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} \frac{\sum_{\boldsymbol{j\in R(t_i)}} Z_{jg} Z_{jl} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)}{\sum_{\boldsymbol{j\in R(t_i)}} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)})$$

$$-\sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} [\frac{\sum_{\boldsymbol{j\in R(t_i)}} Z_{jg} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)}{\sum_{\boldsymbol{j\in R(t_i)}} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)} * \frac{\sum_{\boldsymbol{j\in R(t_i)}} Z_{jl} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)}{\sum_{\boldsymbol{j\in R(t_i)}} \exp\left(\sum_{k=1}^{p} \beta_j Z_{jk}\right)}]$$

The empirical information matrix: $I(\boldsymbol{b})$

The MPLE: $\boldsymbol{b} \sim \boldsymbol{MN}(\boldsymbol{\beta}, I^{-1}(\boldsymbol{b}))$ when D is large.

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: maximum partial likelihood estimators

Hypothesis Tests: Global Test: $\boldsymbol{\beta = \beta_0}$

Wald Test → based on the normality of $\boldsymbol{b}$.

$$\mathrm{X}_{\mathrm{w}}^2 = (\boldsymbol{b - \beta_0})^{\mathrm{T}} I(\boldsymbol{b})(\boldsymbol{b - \beta_0}) \sim \mathrm{X}_{\mathrm{p}}^2$$

Log partial likelihood ratio test

$$\mathrm{X}_{\mathrm{LR}}^2 = \mathbf{2}[\mathbf{L_p}(\boldsymbol{b}) - \boldsymbol{L_p}(\boldsymbol{\beta_0})] \sim \mathrm{X}_{\mathrm{p}}^2$$

Score Test → based on the normality of the score function

Let $\mathrm{U_k}(\boldsymbol{\beta}) = \frac{\partial L_p(\boldsymbol{\beta})}{\partial \beta_k}$

$$\boldsymbol{U}(\boldsymbol{\beta}) = [U_1(\boldsymbol{\beta}), \ldots, U_p(\boldsymbol{\beta})]^{\boldsymbol{T}} \sim \boldsymbol{MN}[\mathbf{0}, \boldsymbol{I}(\boldsymbol{b})]$$

$$X_{SC}^2 = \boldsymbol{U}(\boldsymbol{\beta_0}) I^{-1}(\boldsymbol{b}) \boldsymbol{U}(\boldsymbol{\beta_0}) \sim X_p^2$$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: maximum partial likelihood estimators

Hypothesis Tests: Local Test: $\boldsymbol{\beta_1 = \beta_{10}}$

$\boldsymbol{\beta = (\beta'_1, \beta'_2)'}$, $\boldsymbol{I\begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}}$, and $\boldsymbol{\beta_{MPLE} = b = (b'_1, b'_2)'}$

Wald Test → based on the normality of $\boldsymbol{b_1}$.

$$\mathrm{X}^2_{\mathrm{w}} = (\boldsymbol{b_1 - \beta_{10}})^{\mathrm{T}}[I^{11}(\boldsymbol{b})]^{-1}(\boldsymbol{b_1 - \beta_{10}}) \sim \mathrm{X}^2_{\mathrm{q}}$$

$I^{11}(\boldsymbol{b})$ is the upper qxq submatrix of $\boldsymbol{I}^{-1}(\boldsymbol{b})$.

Log partial likelihood ratio test

$$\mathrm{X}^2_{\mathrm{LR}} = \mathbf{2}[\mathbf{L_p}(\boldsymbol{b}) - \boldsymbol{L_p(\beta_{10}, b_2(\beta_{10}))}] \sim \mathrm{X}^2_{\mathrm{q}}$$

$\boldsymbol{b_2(\beta_{10})}$ is the MPLE of $\boldsymbol{b_2}$.

Regression Analyses

Cox Model: maximum partial likelihood estimators

Hypothesis Tests: Local Test: $\boldsymbol{\beta_1 = \beta_{10}}$

$\boldsymbol{\beta} = (\boldsymbol{\beta'_1}, \boldsymbol{\beta'_2})'$, $\boldsymbol{I}\begin{pmatrix} \boldsymbol{I_{11}} & \boldsymbol{I_{12}} \\ \boldsymbol{I_{21}} & \boldsymbol{I_{22}} \end{pmatrix}$, and $\boldsymbol{\beta_{MPLE} = b = (b'_1, b'_2)'}$

Score Test → based on the normality of the score function

Let $\boldsymbol{U_1[\beta_{10}, b_2(\beta_{10})]}$ be the qx1 vector of scores for $\boldsymbol{\beta_1}$evaluated at $\boldsymbol{\beta_{10}}$ and $\boldsymbol{b_2(\beta_{10})}$.

$X^2_{SC} =$

$\boldsymbol{U_1[\beta_{10}, b_2(\beta_{10})]'[I^{11}(\beta_{10}, b_2(\beta_{10}))]U_1[\beta_{10}, b_2(\beta_{10})]} \sim X^2_q$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: maximum partial likelihood estimators

Assume there is only one binary variable Z

$\lambda(t|Z=0)=\lambda_0(t);\ \lambda(t|Z=1)=\lambda_0(t)\exp(\beta)$

Hypothesis Test: Local Test: $\beta=0$

$$U(\beta=0)=\frac{\partial L_p(\beta)}{\partial\beta}_{\beta=0}=\sum_{i=1}^{D} Z_{(i)}-\sum_{i=1}^{D}\frac{\sum_{j\in R(t_i)} Z_j\exp(\beta Z_j)}{\sum_{j\in R(t_i)}\exp(\beta Z_j)}$$

$$=d_1-\sum_{i=1}^{D}\frac{Y_{1i}}{Y_{0i}+Y_{1i}}$$

Y_{0i}, Y_{1i} are # of subjects at risk at t_i

$$I(0)=\sum_{i=1}^{D}\frac{Y_{0i}Y_{1i}}{(Y_{0i}+Y_{1i})^2}$$

When there are no ties, the score test is exactly the same as the two sample log rank test.

Regression Analyses

Cox Model: partial likelihood (with ties)

$t_1 < t_2 < \cdots < t_D$; d_i: # of events at t_i; $\boldsymbol{R}_i$: risk set at t_i.

D_i: the set of all individuals who had events at t_i

$\boldsymbol{S}_i = \sum_{j \in D_i} \boldsymbol{Z}_j$: the sum of the vectors $\boldsymbol{Z}_j$over the event set D_i.

Method 1: Breslow (1974)

$$l_p^B(\boldsymbol{\beta}) = \prod_{\boldsymbol{i=1}}^{\boldsymbol{D}} \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{S_i})}{[\sum_{\boldsymbol{j} \in \boldsymbol{R}(\boldsymbol{t_i})} \exp(\boldsymbol{\beta}^T \boldsymbol{Z_j})]^{\boldsymbol{d_i}}}$$

$$= \prod_{\boldsymbol{i=1}}^{\boldsymbol{D}} \prod_{\boldsymbol{k} \in D_i} \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{Z_k})}{[\sum_{\boldsymbol{j} \in \boldsymbol{R}(\boldsymbol{t_i})} \exp(\boldsymbol{\beta}^T \boldsymbol{Z_j})]^{\boldsymbol{d_i}}} = \prod_{\boldsymbol{i=1}}^{\boldsymbol{D}} \prod_{\boldsymbol{k} \in D_i} \hat{p}_{ik}$$

This partial likelihood considers each of the d_i events at t_i as distinct, so the conditional probability is the product of the individual conditional probabilities. This approximation works well when there are few ties.

Regression Analyses

Cox Model: partial likelihood (with ties)

Method 2: Efron (1977)

$$l_p^E(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{S_i})}{\prod_{k=1}^{d_i}\left(\sum_{j\in R(t_i)} \exp(\boldsymbol{\beta}^T \boldsymbol{Z_j}) - \frac{\boldsymbol{k-1}}{\boldsymbol{d_i}} \sum_{j\in D_i} \exp(\boldsymbol{\beta}^T \boldsymbol{Z_j})\right)}$$

$$= \prod_{i=1}^{D} \left\{ \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{Z_{i1}})}{\sum_{j\in R(t_i)} \exp(\boldsymbol{\beta}^T \boldsymbol{Z_j})} \times \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{Z_{i2}})\ I(d_i \geq 2)}{\sum_{j\in R(t_i)} \exp(\boldsymbol{\beta}^T \boldsymbol{Z_j}) - \frac{1}{d_i}\sum_{\boldsymbol{j}\in D_i} \exp(\boldsymbol{\beta}^T \boldsymbol{Z_j})} \right.$$

$$\left. \ldots \times \frac{\exp\left(\boldsymbol{\beta}^T \boldsymbol{Z_{id_i}}\right) I(d_i \geq 2)}{\sum_{j\in R(t_i)} \exp(\boldsymbol{\beta}^T \boldsymbol{Z_j}) - \frac{d_i - 1}{d_i}\sum_{j\in D_i} \exp(\boldsymbol{\beta}^T \boldsymbol{Z_j})} \right\}$$

$$= \prod_{i=1}^{D} \hat{p}_{i1}\hat{p}_{i2|i1} \cdots \hat{p}_{id_i|i1,\ldots,i(d_i-1)}$$

$l_p^E(\boldsymbol{\beta})$ takes sequential perspective, differentiate the conditional probability of multiple events at the same time. Thus, $l_p^E(\boldsymbol{\beta})$ works better when there are more than a few ties.

Regression Analyses

Cox Model: partial likelihood (with ties)

Method 3: Exact permutation (more intensive computations)

Let $\boldsymbol{Q_i}$ denote the set with all possible subsets with d^i events

$$l_p^{EX}(\boldsymbol{\beta}) = \prod_{\boldsymbol{i=1}}^{\boldsymbol{D}} \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{S_i})}{\sum_{\boldsymbol{q}\in \boldsymbol{Q_i}} \exp\left(\boldsymbol{\beta}^T \boldsymbol{S_q^*}\right)}$$

where $\boldsymbol{S_q^*} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{D_i}} \boldsymbol{Z_{qj}}$ and q is the set of individuals.

When there are no ties, $l_p^B(\boldsymbol{\beta}) = l_p^E(\boldsymbol{\beta}) = l_p^{EX}(\boldsymbol{\beta}) = l_p(\boldsymbol{\beta})$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Estimation of the Survival Functions

For the proportional hazards model,

$$\lambda(t|Z) = \lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z})$$

$$H(t|Z) = \int_0^t \lambda_0(\mu)\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\,d\mu$$

$$= \exp(\boldsymbol{\beta}^T\boldsymbol{Z})\int_0^t \lambda_0(\mu)d\mu \quad = \exp(\boldsymbol{\beta}^T\boldsymbol{Z})H_0(t|Z)$$

$$S(t|Z) = \exp[-H(t|Z)] = \exp[-\exp(\boldsymbol{\beta}^T\boldsymbol{Z})H_0(t)]$$

$$= \{\exp[-H_0(t)]\}^{\exp(\boldsymbol{\beta}^T\boldsymbol{Z})} = [S_0(t)]^{\exp(\boldsymbol{\beta}^T\boldsymbol{Z})}$$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Estimation of the Survival Functions

With $\boldsymbol{\beta}$ estimated as $\boldsymbol{b}$,

$$\left\{\boldsymbol{b}, \widehat{\boldsymbol{V}}(\boldsymbol{b})\right\}, t_1 < t_2 < \cdots < t_D$$

d_i: # of events at t_i

The Breslow estimator of baseline cumulative hazard rate:

$$\widehat{H}_0(t) = \sum_{t_i \le t} \frac{d_i}{\sum_{\boldsymbol{j} \in \boldsymbol{R}(\boldsymbol{t_i})} \exp\left(\boldsymbol{b}^T \boldsymbol{Z_j}\right)}$$

$\widehat{H}_0(t)$ is a step function with jumps at $\{t_i\}$.

When $\boldsymbol{Z_j} = \boldsymbol{0}$, $\widehat{H}_0(t)$ is the N-A estimator.

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Estimation of the Survival Functions

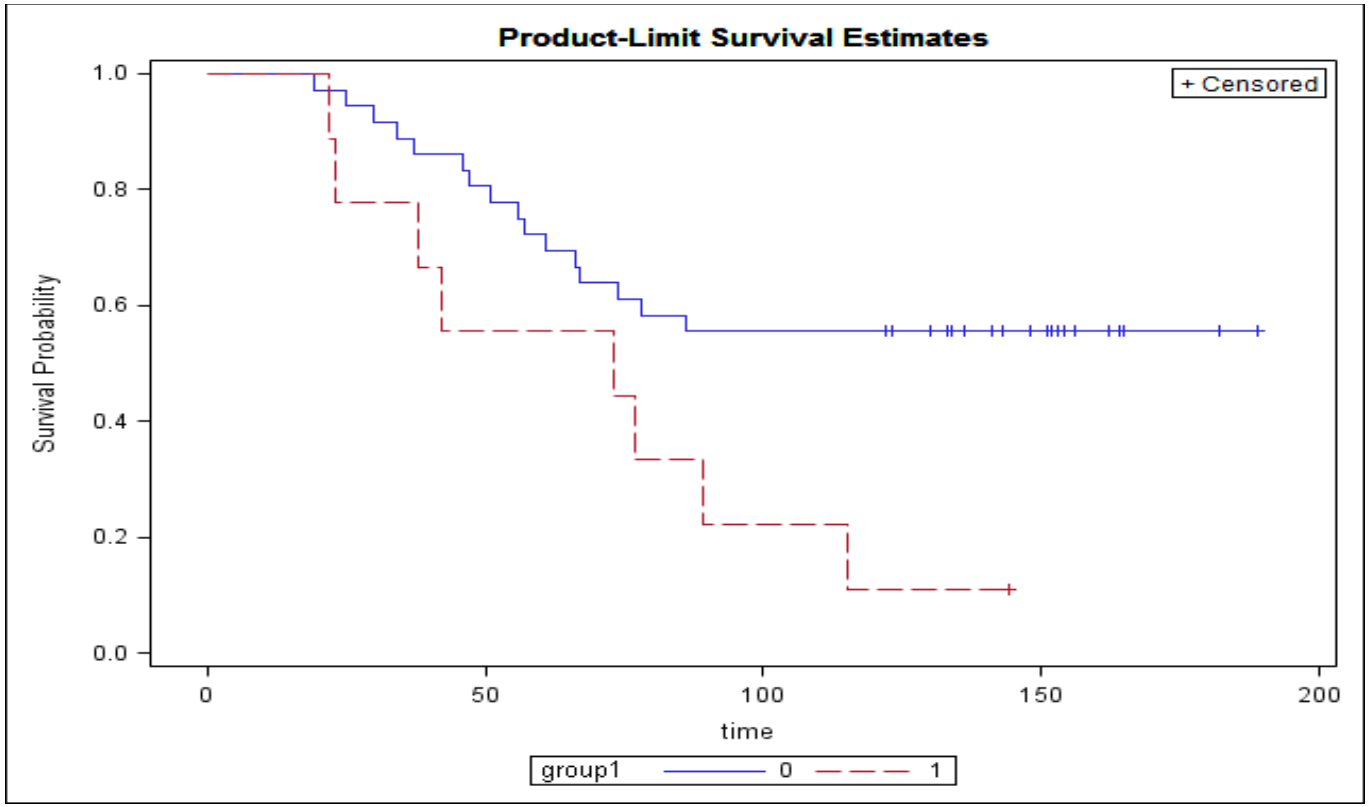
With $\boldsymbol{\beta}$ estimated as $\boldsymbol{b}$, we may want to estimate the survival function for a new individual with covariate vector $\boldsymbol{Z_{new}}$.

$$\hat{S}_{new}(t|\boldsymbol{Z} = \boldsymbol{Z_{new}}) = [\hat{S}_0(t)]^{\exp(\boldsymbol{b}^T \boldsymbol{Z_{new}})}$$

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Time to death for a breast cancer trail



Group1: 1—immunoperoxidase positive 0---otherwise

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Time to death for a breast cancer trail

```
proc phreg data = breast_cancer;
        model time*death(0)=group1;
run;
```

Total	Event	Censored	Percent Censored
45	24	21	46.67

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	167.488	163.041
AIC	167.488	165.041
SBC	167.488	166.219

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.4463	1	0.0350
Score	5.4943	1	0.0191
Wald	5.0804	1	0.0242

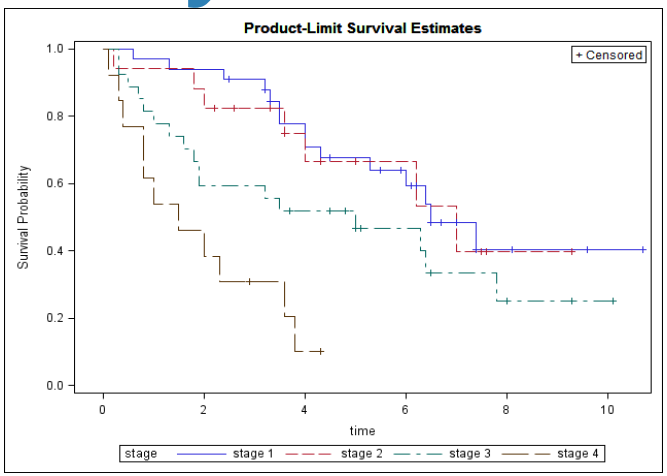
Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
group1	1	0.98022	0.43489	5.0804	0.0242	2.665

Regression Analyses

Cox Model: Larynx Cancer

proc phreg data=larynx;
model time*death(**0**)=z1 z2 z3 age;
run;



Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
z1(stg II)	1	0.13842	0.46232	0.0896	0.7646	1.148
z2(stg III)	1	0.63815	0.35609	3.2116	0.0731	1.893
z3(stg IV)	1	1.69333	0.42218	16.0876	<.0001	5.438
age	1	0.01890	0.01425	1.7589	0.1848	1.019

The estimates of hazard ratios are similar to those from the exponential AFT model (HR=1.16, 1.91, and 5.12 for z1, z2, and z3, respectively).

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Ties

```
proc phreg data=kidney ;
  model time*infect(0)= z1 /itprint; *default is Breslow's method.
run;
```

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
z1	1	-0.61817	0.39813	2.4108	0.1205	0.539

```
proc phreg data=kidney;
  model time*infect(0)= z1 /ties = efron itprint;
run;
```

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
z1	1	-0.61257	0.39791	2.3699	0.1237	0.542

```
proc phreg data=kidney;
  model time*infect(0)= z1 /ties = discrete itprint; *use this is times are discrete
run;
```

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
z1	1	-0.62944	0.40190	2.4529	0.1173	0.533

```
proc phreg data=kidney;
  model time*infect(0)= z1 /ties = exact itprint;
run;
```

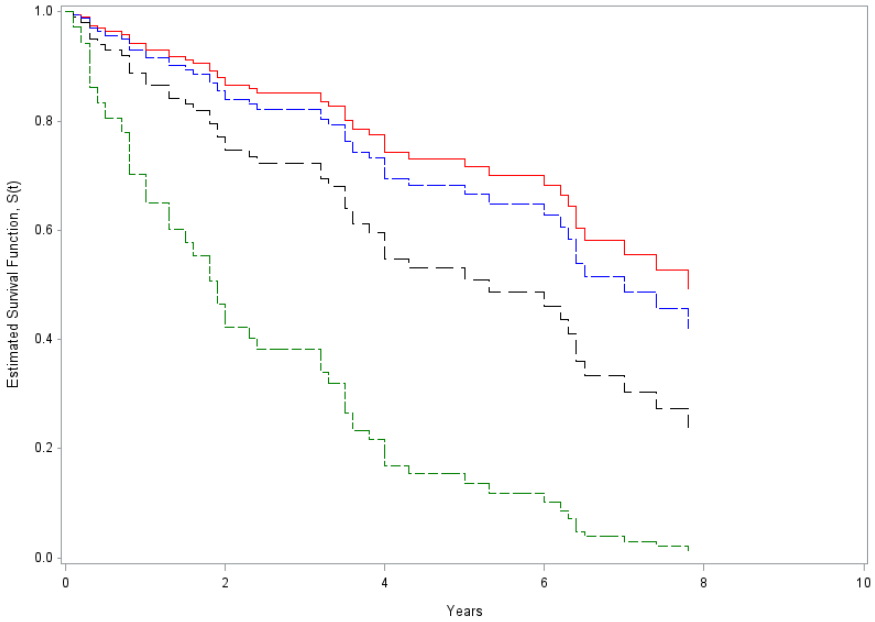
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
z1	1	-0.61265	0.39794	2.3703	0.1237	0.542

Regression Analyses

Cox Model: estimate survival functions

Larynx cancer of stage I-IV at age 60

```
data pred;
 input z1 z2 z3 age;
cards;
0 0 0 60
1 0 0 60
0 1 0 60
0 0 1 60
;
run;
```




```
proc phreg data =larynx;
 model time*delta(0) = z1 z2 z3 age;
 baseline out = surv60 survival = survival lower = slower upper = supper
        covariates = pred /method = ch nomean  cltype = loglog ;
run;
*ch: Breslow estimate for baseline survival function
```

Regression Analyses

Cox Model: Diagnosis

Independent censoring—not testable with the given data (identifiability dilemma)

Proportional hazards assumption

Regression Analyses

Cox Model: Diagnosis

1. **Assess the overall fit** using Cox-Snell residuals
2. **Identify the functional form** of independent variable Z using Martingale residuals
3. **Assess the proportional hazards assumption**

Regression Analyses

Cox Model: Diagnosis of overall fit

Given data $(T_j, \delta_j,\ Z_j(t))$ and a proportional hazards model

$$h\left(t\middle|Z_j(t)\right) = h_0(t)\exp\left[\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{Z}_j(t)\right],$$

the distribution of $S\left(t\middle|Z_j(t)\right)$ is uniform on [0,1] and the distribution of $H\left(t\middle|Z_j(t)\right) = -\ln(S\left(t\middle|Z_j(t)\right)]$ is censored exponential (1) if the model is correct.

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Diagnosis of overall fit

Cox-Snell residuals

$$r_j = \hat{H}_0(T_j)\exp(\boldsymbol{b}^T\boldsymbol{Z}_j),$$

Where $\hat{H}_0(T_j)$is Breslow's estimator of cumulative baseline hazard.

Do these residuals behave like a sample from EXP(1)?

The plot of r_j vs the $\tilde{H}_{NA}(r_j)$ should be a straight line through the origin with a slope of 1, where $\tilde{H}_{NA}(r_j)$ is the Nelson-Aalen estimator of the cumulative hazard rate of using r_j's as survival times.

Regression Analyses

Cox Model: Diagnosis of functional form

Martingale residuals

$$\widehat{M}_j = N_j(\infty) - \int_0^{\infty} Y_j(t)\exp[\boldsymbol{b}^{\boldsymbol{T}} Z_j(t)]d\widehat{H}_0(t),$$

$$j = 1,2,\ldots,n$$

When the data is right censored and all covariates are fixed at the baseline, the martingale residual reduces to

$$\widehat{M}_j = \delta_j - r_j$$

The martingale residuals are estimates of the excess number of events seen in the data but not predicted by the model.

Regression Analyses

Cox Model: Diagnosis of functional form

To find the functional form of a single variable Z', we can fit a Cox model to the data based on other variables with known functional form and compute the martingale residuals, then plot the martingale residuals against Z'. The smoothed-fitted curve gives an indication of the functional form for Z'.

Regression Analyses

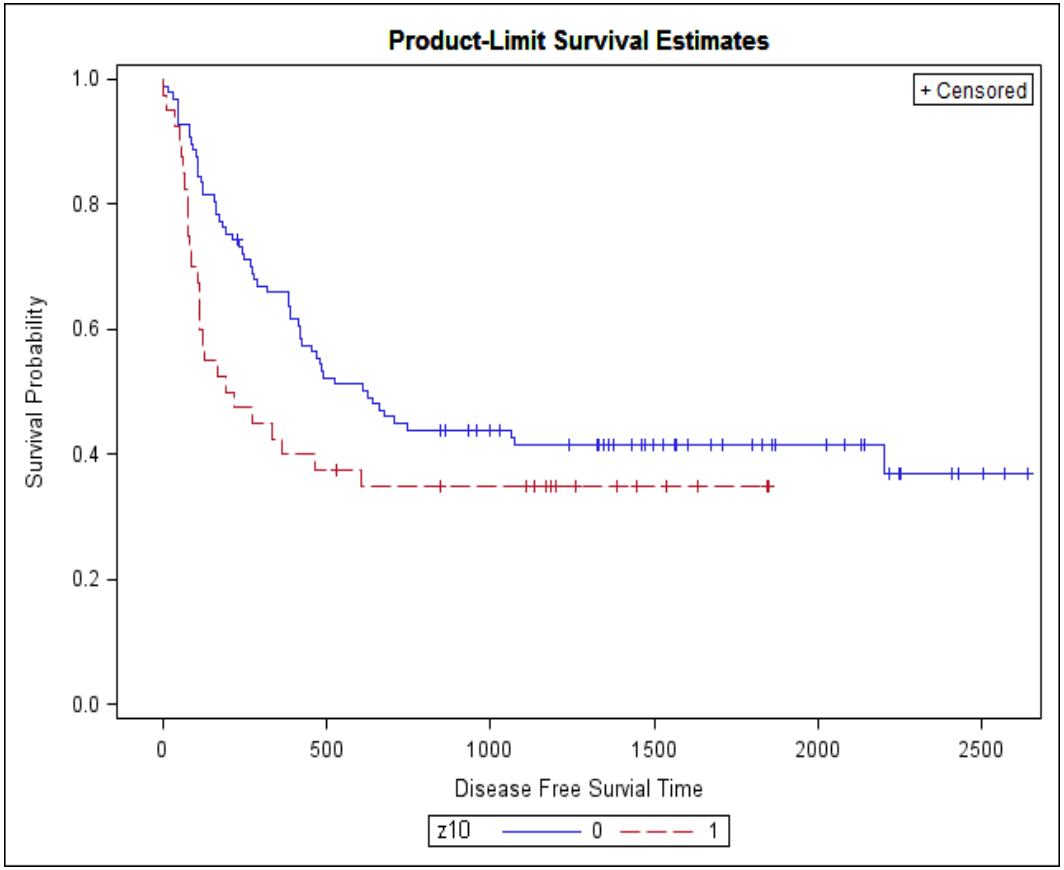
Cox Model: Diagnosis of proportional hazards

a. Plot log cumulative hazards of all groups vs t. (When the PH assumption is satisfied, the vertical distances of any two curves should be constant.
b. Plot the difference of log cumulative hazard between each group vs the control group, then all the curves should be roughly flat.
c. Plot cumulative hazard rate of each group vs that of the control group. All the curves should be roughly straight lines.

Regression Analyses

Cox Model: Diagnosis of proportional hazards

Bone marrow transplants



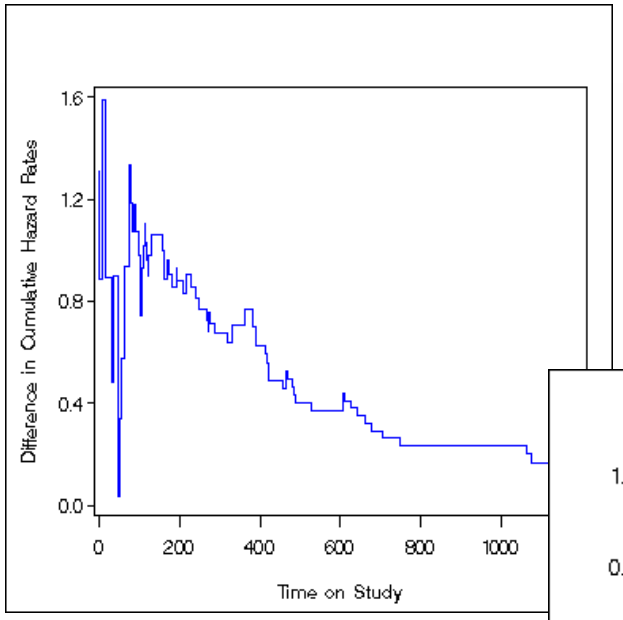
Z10: Use methotrexate or not(1-Yes, 0-No)

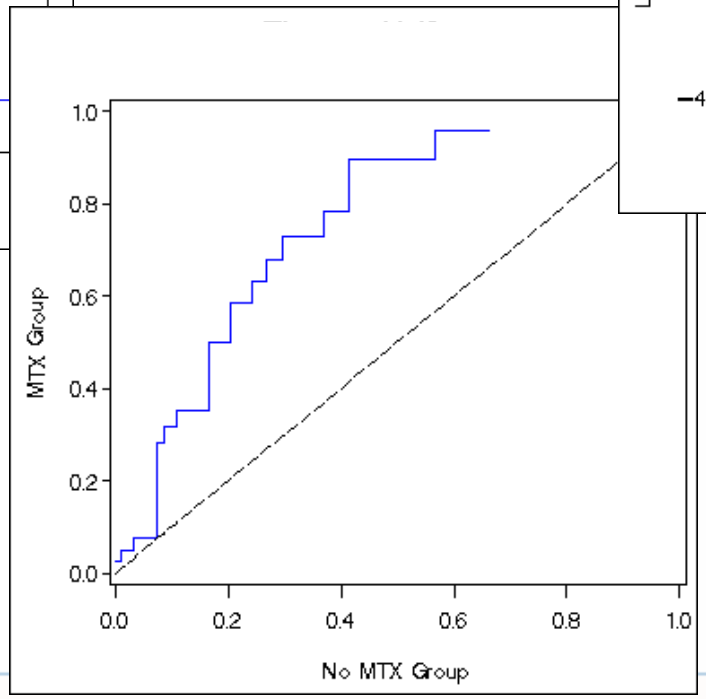
KU MEDICAL CENTER
The University of Kansas

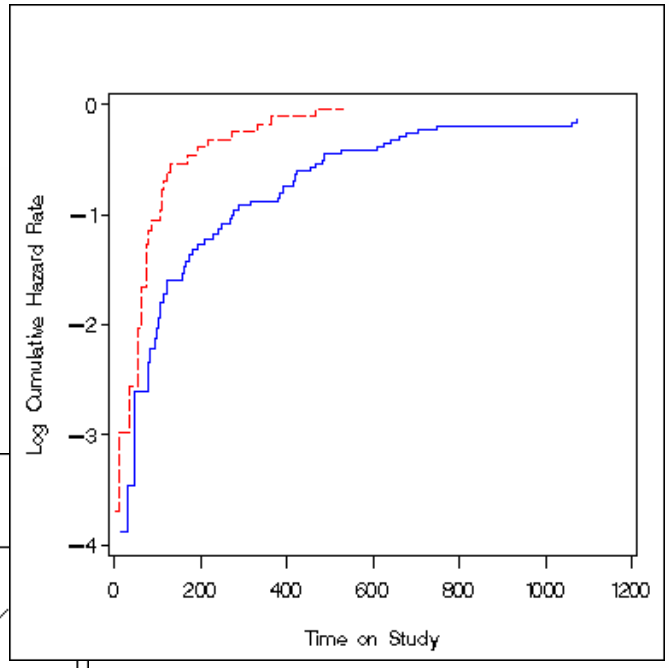
Regression Analyses

Cox Model: Diagnosis of proportional hazards

Bone marrow transplants





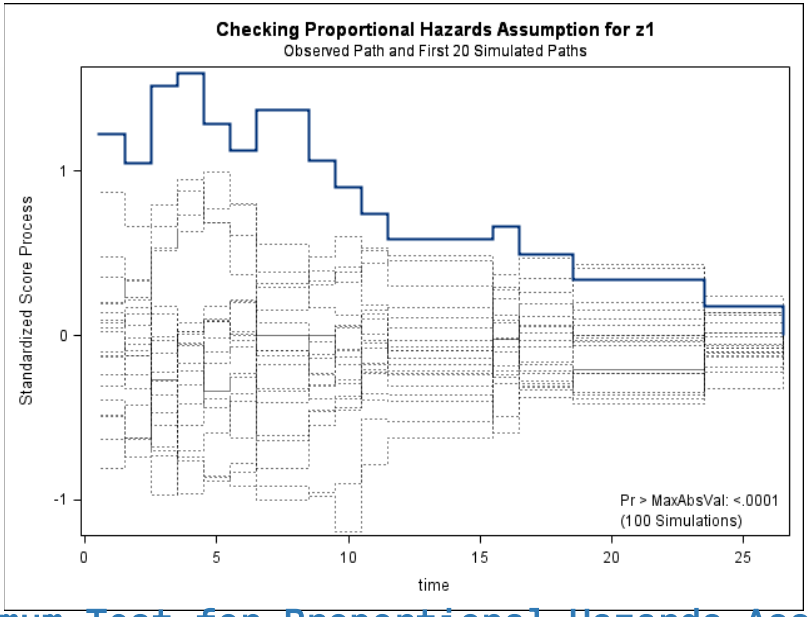


KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Diagnosis of proportional hazards

```
proc phreg data=kidney;
*Kidney Catheter Placements;
model time*infect(0)= z1;
assess var=(z1) PH/ seed=1737 resample=100;
run;
```



Supremum Test for Proportional Hazards Assumption

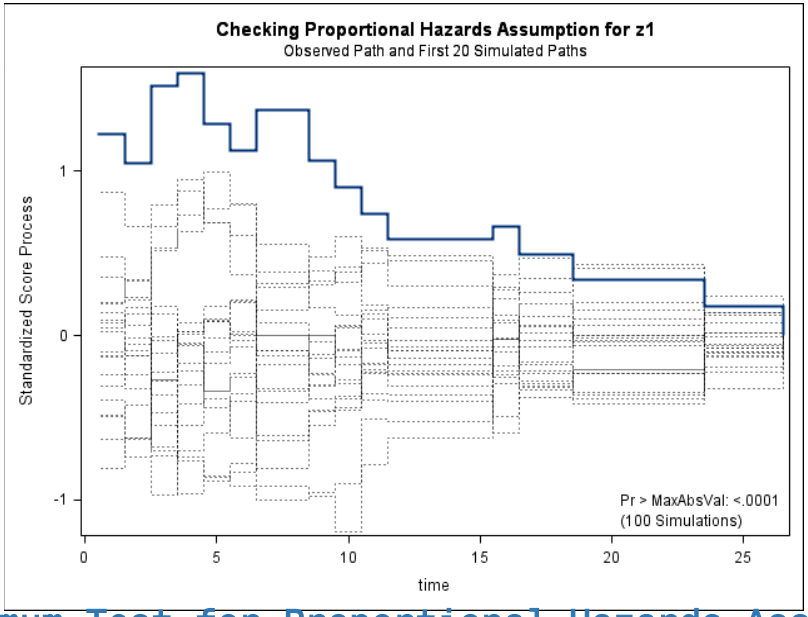
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
z1	1.5973	100	1737	<.0001

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Diagnosis of proportional hazards

```
proc phreg data=kidney;
*Kidney Catheter Placements;
model time*infect(0)= z1;
assess var=(z1) PH/ seed=1737 resample=100;
run;
```



Supremum Test for Proportional Hazards Assumption

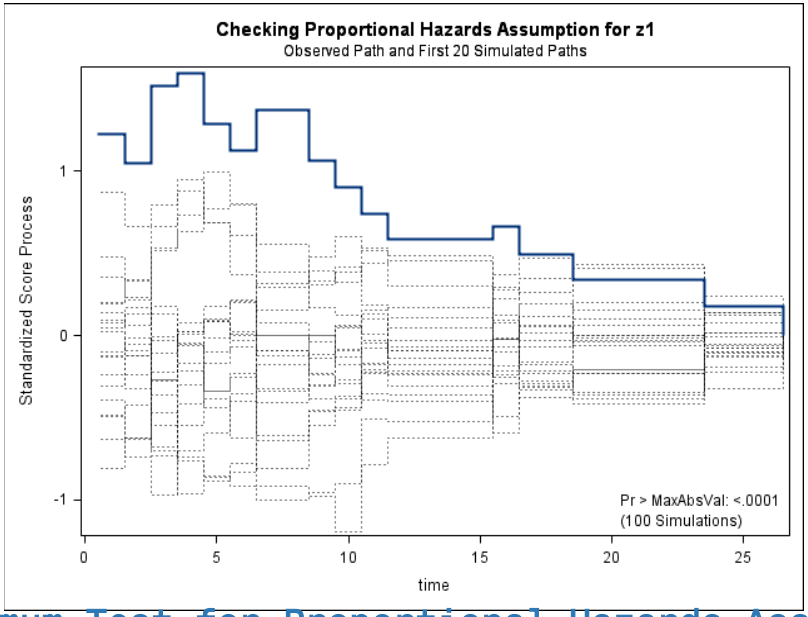
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
z1	1.5973	100	1737	<.0001

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Diagnosis

```
proc phreg data=kidney;
*Kidney Catheter Placements;
model time*infect(0)= z1;
assess var=(z1) PH/ seed=1737 resample=100;
run;
```



Supremum Test for Proportional Hazards Assumption

Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
z1	1.5973	100	1737	<.0001

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Cox Model: Diagnosis

Outliers and Influential Observations by SAS

DFBETA specifies the approximate changes in the parameter estimates $\hat{\beta} - \hat{\beta}_{(j)}$ when the j th observation is omitted.

RESDEV specifies the deviance residual $\widehat{D}_j$.This is a transform of the martingale residual to achieve a more symmetric distribution.

RESMART specifies the martingale residual $\widehat{M}_j$.

XBETA specifies the estimate of the linear predictor $X\hat{\beta}$.

$$\widehat{D}_j = sign\left[\widehat{M}_j\right]\{-2\left[\widehat{M}_j + \delta_j \log\left(\delta_j - \widehat{M}_j\right)\right]\}^{\frac{1}{2}}, \quad \widehat{D}_j = 0 \; if \; \widehat{M}_j = 0$$

To assess the effect of a given individual on the model, a plot of the deviance $\widehat{D}_j$ vs risk score $X\hat{\beta}$ .When there is light to moderate censoring, the $\widehat{D}_j$ should look like a sample of normally distributed noise. When there is heavy censoring, a large collections of points near zero will distort the normal approximation. In either case, potential outliers will have deviance residuals whose absolute values are too large.

This part is similar to those for linear regression.

Regression Analyses

Cox Model: Diagnosis

Outliers and Influential Observations by SAS

```
proc phreg data = sec1_10;
  model time*ind(0) = dtype kscore;
   output out = sect1_10out resmart=M resdev =D xbeta=xb dfbeta =dfbeta1
dfbeta12  /method = ch;
*dfbeta1 : Difference in the parameter for dtype;
*dfbeta12 : Difference in the parameter for kscore;
run;
```

KU MEDICAL CENTER
The University of Kansas

Regression Analyses

Which time scale should we use, age or study time? It depends on the research interest and underlying survival functions.

Some references

Pencina, M. J., Larson, M. G., & D'Agostino, R. B. (2007). Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in medicine*, *26*(6), 1343-1359.

Kom, E. L., Graubard, B. I., & Midthune, D. (1997). Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *American journal of epidemiology*, *145*(1), 72-80.

Thiébaut, A., & Bénichou, J. (2004). Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. Statistics in medicine, 23(24), 3803-3820.

KU MEDICAL CENTER
The University of Kansas

outline

Introduction

- Time to event data /Censoring and truncation
- Basic survival functions/Group comparison of survival data

Regression Analyses

- Parametric Approach -- Accelerated failure time models
- Semiparametric Approach – Cox models

Extension to the Cox models

- Weighted Cox models: nonproportional hazards
- Time-dependent covariate Cox models
- Time-varying coefficient survival models

Other topics

- Competing risk modeling
- Bayesian survival analyses

KU MEDICAL CENTER
The University of Kansas

Extension to the Cox models

Cox Model: Non-Proportional Hazards

$\beta(t)$ is a function of t rather than a constant. Hazard ratio $exp[\beta(t)]$ is a function of t.

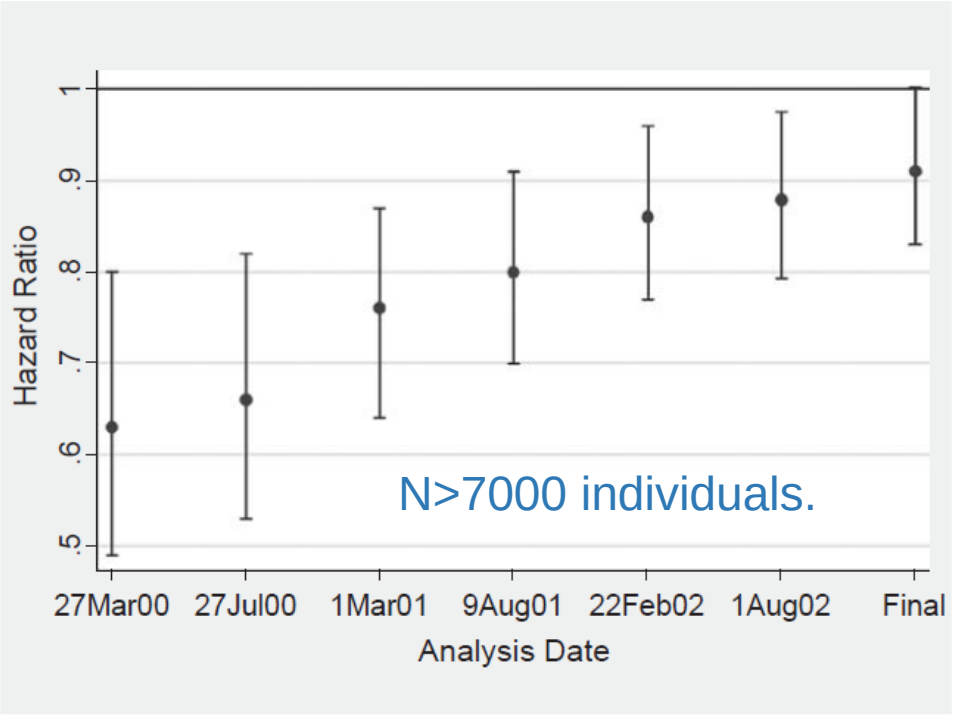
Note: the interpretation of $\beta(t)$ is different from a regular time-varying function, such as height, weight, and blood pressure, as the number of survivors keep decreasing over time.

For a aging process with age as t in years, $\beta(200)$ carries much less weight compared with $\beta(40)$.

Extension to the Cox models

Cox Model: Non-Proportional Hazards

In reality, Cox models are often used even when the assumption is not satisfied, because of the simplicity of the model.



Hazard ratio and 95% CI for all-cause mortality (candesartan vs placebo) at each interim analysis and at study closeout.

Source: Pocock et al. 2005. American Heart Journal. The data monitoring experience in the Candesartan in Heart Failure Assessment of Reduction in Mortality and morbidity (CHARM) program.

Extension to the Cox models

Can we use the Cox model when the proportional hazards assumption is violated? What is the interpretation? Any issue we should be aware of?

Extension to the Cox models

Cox Model: Non-Proportional Hazards

A Cox model estimate β_{PL} under non-proportional hazards situation can be considered as an estimate of the average effect across the time period for fixed covariates.

$$\beta^* = \int_0^\infty \beta(x) dF(x)$$

The estimate β_{PL} depends heavily on the censoring that may be biased in estimating β^* in reality.

KU MEDICAL CENTER
The University of Kansas

Extension to the Cox models

Cox Model: Non-Proportional Hazards

Cox estimates adjusted for censoring

Weighted partial likelihood estimator

$$l_p(\boldsymbol{\beta}) = \prod_{i=1}^{D} \left\{ \frac{\exp\left(\sum_{k=1}^{p} \beta_k Z_{(i)k}\right)}{\sum_{j \in R(t_i)} \exp\left(\sum_{k=1}^{p} \beta_k Z_{jk}\right)} \right\}^{w_i}$$

where $w_i = \frac{1}{\hat{s}_i}$ and $\hat{s}_i$ is the K-M estimate of survival function for censoring. This estimate give more weight to values of $\beta(x)$ at larger x values when censoring rate is high. w_i=1 when there is no censoring.

Van Houwelingen, H.C., van de Velde, C.J.H., Stijnen, T. (2005). Interim analysis on survival data: Its potential bias and how to repair it. Statistics in Medicine 24:2823-2835.

Extension to the Cox models

Cox Model: Non-Proportional Hazards

Cox estimates adjusted for censoring

Weighted partial likelihood estimator

$$l_p(\boldsymbol{\beta}) = \prod_{i=1}^{D} \left\{ \frac{\exp\left(\sum_{k=1}^{p} \beta_k Z_{(i)k}\right)}{\sum_{j \in R(t_i)} \exp\left(\sum_{k=1}^{p} \beta_k Z_{jk}\right)} \right\}^{w_i}$$

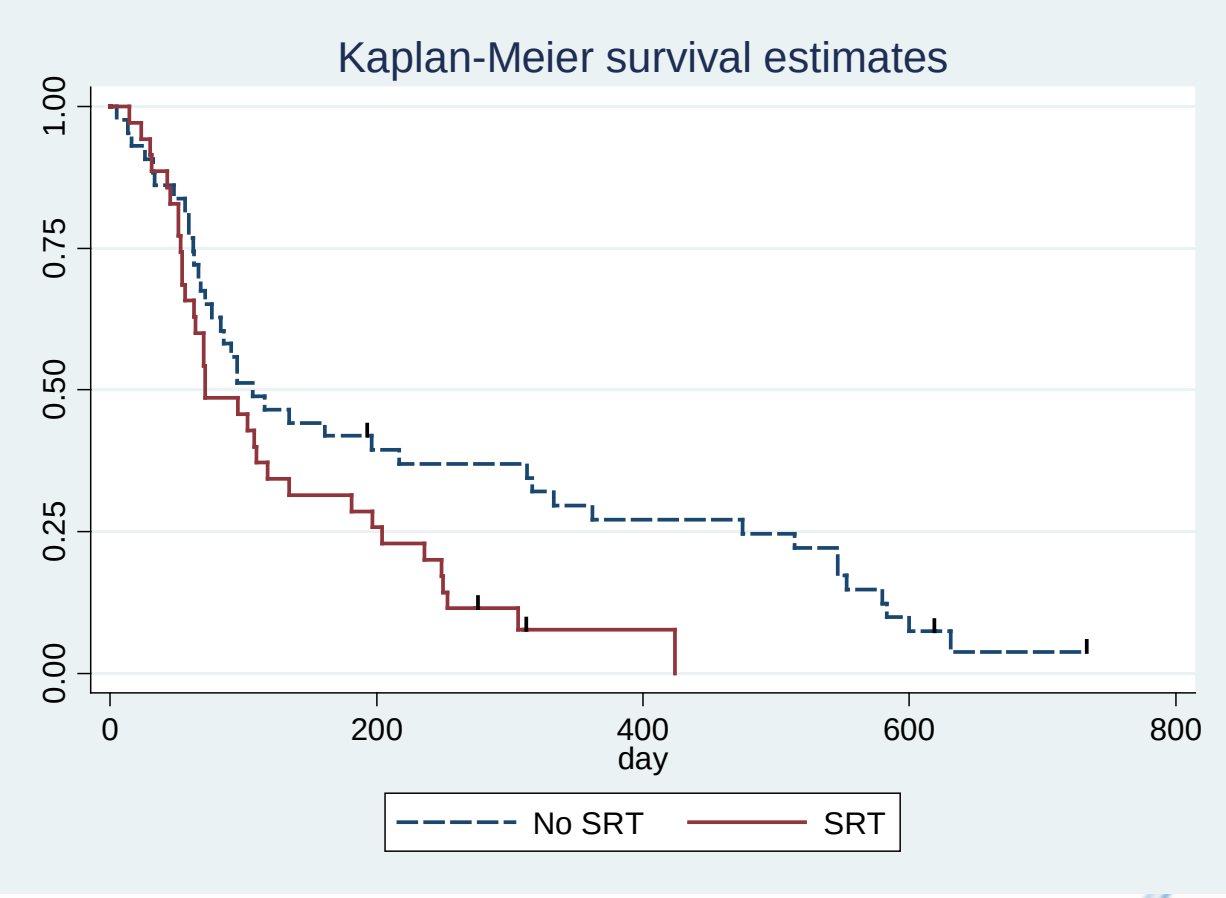
where $w_i = \frac{1}{\hat{s}_i}$ and $\hat{s}_i$ is the K-M estimate of survival function for censoring. This estimate give more weight to values of $\beta(x)$ at larger x values when censoring rate is high. w_i=1 when there is no censoring.

Van Houwelingen, H.C., van de Velde, C.J.H., Stijnen, T. (2005). Interim analysis on survival data: Its potential bias and how to repair it. Statistics in Medicine 24:2823-2835.

Extension to the Cox models

Cox Model: Non-Proportional Hazards

BDP data



KU MEDICAL CENTER
The University of Kansas

Extension to the Cox models

Cox Model: Non-Proportional Hazards

R package coxphw

Unadjusted estimates

coxphw(Surv(ondays, censor) ~ surfact, data = bpd, template = "PH")

	coef	se(coef)	exp(coef)	z	p
Surfact	0.60	0.24	1.83	2.52	0.012

The same as coxph(Surv(*ondays, censor*) ~ *surfact*, *data = bpd*)

Average regression effects

coxphw(Surv(ondays, censor) ~ surfact, data = bpd, template = "ARE")

	coef	se(coef)	exp(coef)	z	p
Surfact	0.62	0.24	1.85	2.59	0.0096

Xu R, O'Quigley J. Estimating average regression effect under non-proportional hazards. *Biostatistics* 2000; **1**:423–439. DOI: 10.1093/biostatistics/1.4.423.[Similar to the weight partial likelihood discussed in previous slides. This method is also applicable to time-dependent covariates.]

The two estimates are similar as there are only 5 censored cases.

Extension to the Cox models

Cox Model: Non-Proportional Hazards

R package coxphw

Average hazard ratios (default)

coxphw(formula = Surv(ondays, censor) ~ surfact, data = data)

	coef	se(coef)	exp(coef)	z	p
surfact	0.44	0.26	1.56	1.68	0.094

Schemper M, Wakounig S, Heinze G. The estimation of average hazard ratios by weighted Cox regression. Statistics in Medicine 2009;28(19):2473--2489.

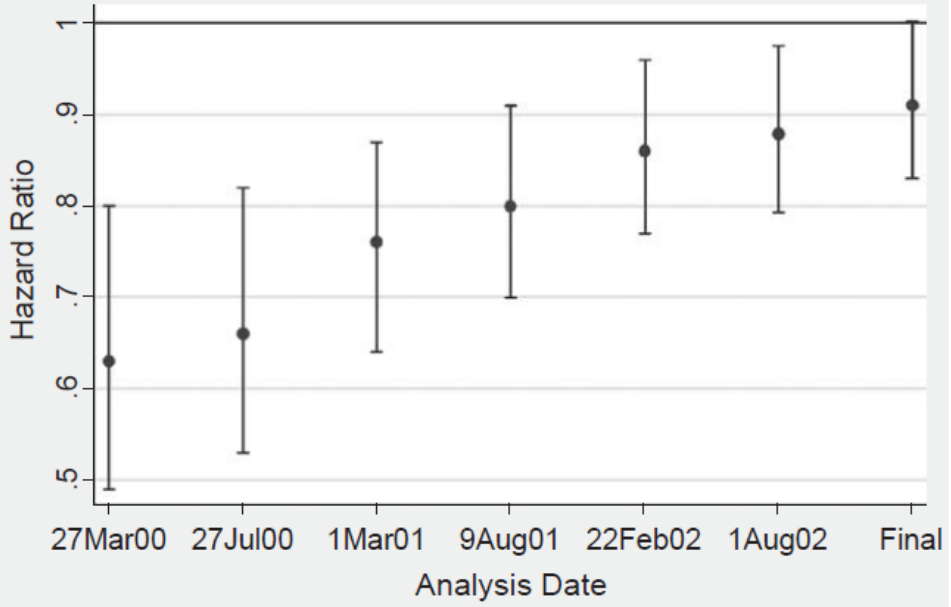
Note: this estimate is very different from Cox estimate without censoring.

KU MEDICAL CENTER
The University of Kansas

Extension to the Cox models

Cox Model: Non-Proportional Hazards

The hazard ratios with adjustment for censoring for this study are expected to be greater or less than those in the figure?



Hazard ratio and 95% CI for all-cause mortality (candesartan vs placebo) at each interim analysis and at study closeout.

Source: Pocock et al. 2005. American Heart Journal. The data monitoring experience in the Candesartan in Heart Failure Assessment of Reduction in Mortality and morbidity (CHARM) program

Regression Analyses

Cox Model: time-dependent covariates

Covariate Z may change with time →Z(t)

Using time-dependent covariates Cox model to

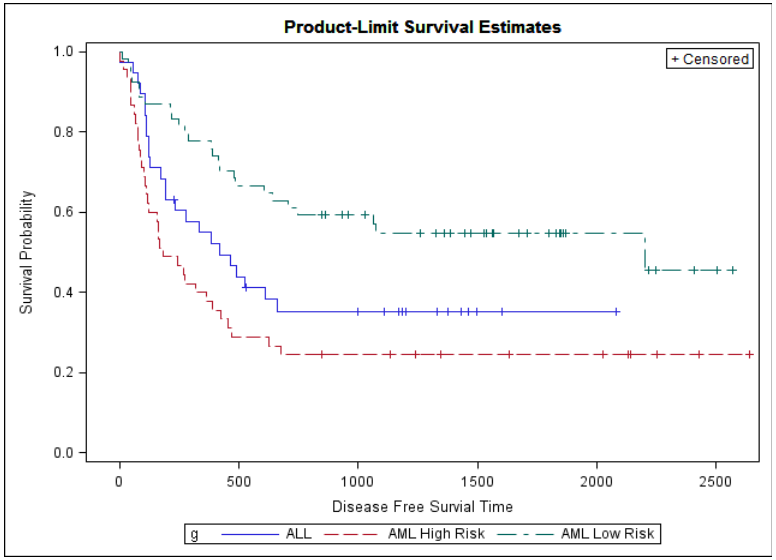
A: test the impact of intermediate events

B: test the proportional hazards assumption

Extension to the Cox models

Time-dependent covariates Cox model

Time to relapse after bone marrow transplants (n=137)



Three intermediate events may be related to the disease-free survival time.

A. Development of acute graft-versus-host disease (aGVHD) (ta—time to aGVHD)

B. Development of chronic graft-versus-host disease (cGVHD) (tc—time to cGVHD)

C. Return of the patient's platelet count to a self-sustaining level (platelet recovery) (tp—time to platelet recovery)

ALL -- acute lymphoblastic leukemia
AML-- acute myeloctic leukemia

It turns out event A and B are not related to risk of relapse while event C is an indicator of better recovery.

KU MEDICAL CENTER
The University of Kansas

Extension to the Cox models

Time-dependent covariates Cox model

This model provides a simple way of modeling time-varying effect of covariates with their proportional hazards assumptions violated.

Extension to the Cox models

Time-dependent covariates Cox model

Time to relapse/death after bone marrow transplants

```
proc phreg data= bone_marrow;
  model t2*dfree(0)= zp g2 g3;
  if (t2 >= tp & p=1) then  zp = 1;
  else  zp = 0;
*p: indicator of having platelet recovery during follow-up;
run;
```

zp is a binary variable that is 0 when platelet is not recovered and 1 when platelet is recovered, which may or may not happen during the recovery process.

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
zp	1	-1.12986	0.32800	11.8658	0.0006	0.323
g2	1	-0.49624	0.28924	2.9435	0.0862	0.609
g3	1	0.38134	0.26761	2.0306	0.1542	1.464

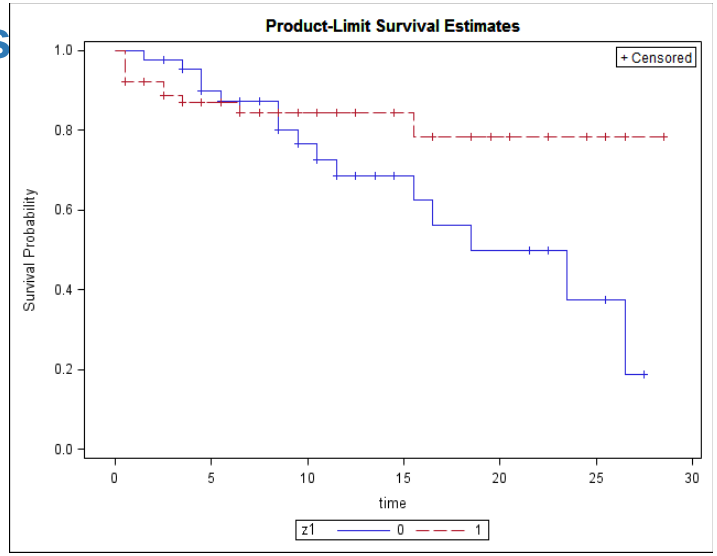
The risk of disease free decreases by 68% (1-0.32) when the platelet is recovered.

Extension to the Cox models

Time-dependent covariates Cox model

Time to infection of kidney dialysis patients

The survival functions of the two treatment cross each other, suggesting the infection risk for surgical patients maybe higher at first, but it becomes lower after about 5 months.
This is an indicator of possible violation of the proportional hazards assumption.




```
proc phreg data=kidney1;
  model time*infect(0)= z1 Z1t;
  z1t=z1*time;
run;
```

0= "surgical" **1** = "percutaneous"

Parameter	Standard			Hazard		
Parameter	DF	Estimate	Error	Chi-Square	Pr > ChiSq	Ratio
z1	1	0.94335	0.75154	1.5756	0.2094	2.569
Z1t	1	-0.25343	0.11695	4.6955	**0.0302**	0.776

The hazard ratio of percutaneous/surgical decreases with time.

b (t) =ln[hr(t)] =$0.94335-0.25343t$

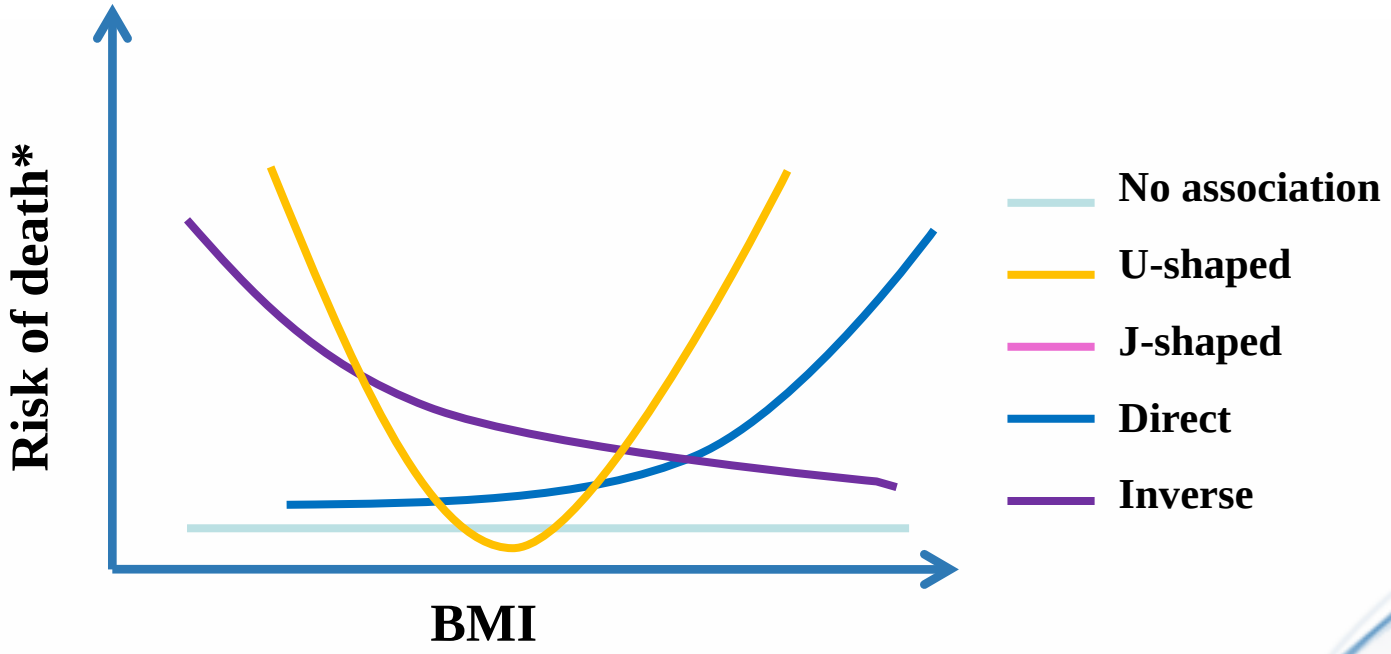
HR(t)=$2.569*(0.776)^t$. HR(t)=1 when t=3.7 months.

Extension to the Cox models

Time-dependent covariates Cox model

Application: Body weight and mortality

The association of body weight measured by body mass index (BMI=weight (kg)/height^2 (m^2)) and all-cause mortality has been reported to have different shapes. Traditional Cox model or logistic regression models were used.



***Risk can be measured as relative risk, odds ratio, and hazard ratio etc.**

KU MEDICAL CENTER
The University of Kansas

Extension to the Cox models

Time-dependent covariates Cox model

Application: Body weight and mortality

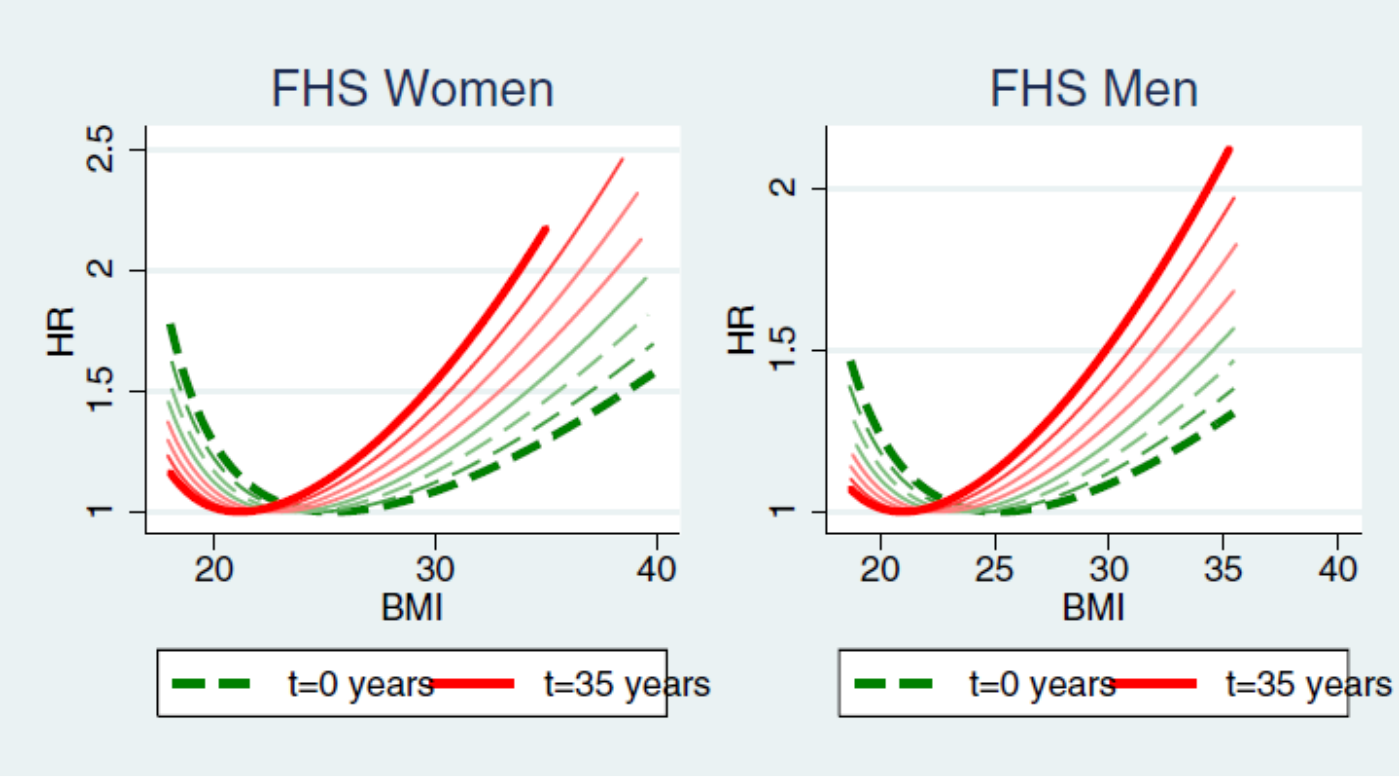
The association depends nonlinearly on BMI (using quadratic term). Using time-dependent covariate Cox model, the hazard functions can be estimated as

$$h(t) \propto \exp[(\beta_1+\beta_2 t)(\frac{1}{bmi})^2 + (\beta_3+\beta_4 t)\frac{1}{bmi}]$$

Extension to the Cox models

Time-dependent covariates Cox model

Application: Body weight and mortality



Similar results were found in some other data.

Source: **He J**, Yu Q, Zhang H, Mahnken JD. The dynamic association of body mass index and all-cause mortality in multiple cohorts and its impacts. Emerging Themes in Epidemiology 2014 11:17.

KU MEDICAL CENTER
The University of Kansas

Extension to the Cox models

Time-dependent covariates Cox model

Application: Body weight and mortality

Conclusion:

Because the association changes with time that studies with different lengths of follow-up may naturally obtain different results as their estimates using traditional Cox model were averages effects within different time period. Short studies tend to obtained inverse associations. This finding is consistent with the fact that most studies obtained inverse associations have follow-up lengths shorter than 10 years.

KU MEDICAL CENTER
The University of Kansas

Extension to the Cox models

Time-varying coefficient survival model

Time-dependent covariates Cox model can be used to estimate time-varying coefficients survival models with pre-specified functional forms.

$$\beta X(t) = \beta X f(t) = \beta f(t) X = \beta(t) X$$

Commonly used functional forms of $f(t)$: t, $\ln(t)$, and $\ln(t+1)$.

This method is restricted due to the pre-specified functional forms. A more flexible approach is to use fractional polynomials to estimate the function.

A even more flexible approach is to estimate the function by smoothing techniques.

Extension to the Cox models

Time-varying coefficient survival model

Nonparametric approach (smoothing) of modeling time-varying coefficients

Penalized partial likelihood

$$\boldsymbol{J(\beta) = ln\left[l_p(\beta)\right] - panelty}$$

where $\boldsymbol{panelty}$ usually depend on the smoothness of $\boldsymbol{\beta(t)}$. For a piece-wise linear $\beta_k(t)$ with knots at $t_1, t_2, \ldots t_M$, the penalty term could be

$$\frac{1}{2}\lambda_k \sum_{j=2}^{M+1} (\beta_{jk} - \beta_{jk-1})^2$$

Gray, R. J. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. Journal of the American Statistical Association, 1992, 87:942-951.

Extension to the Cox models

Time-varying coefficient survival model

Nonparametric approach (smoothing) of modeling time-varying coefficients

R package *coxspline* by Gray (panelized B-splines)
http://biowww.dfci.harvard.edu/~gray/
May not be compatible with the latest R.

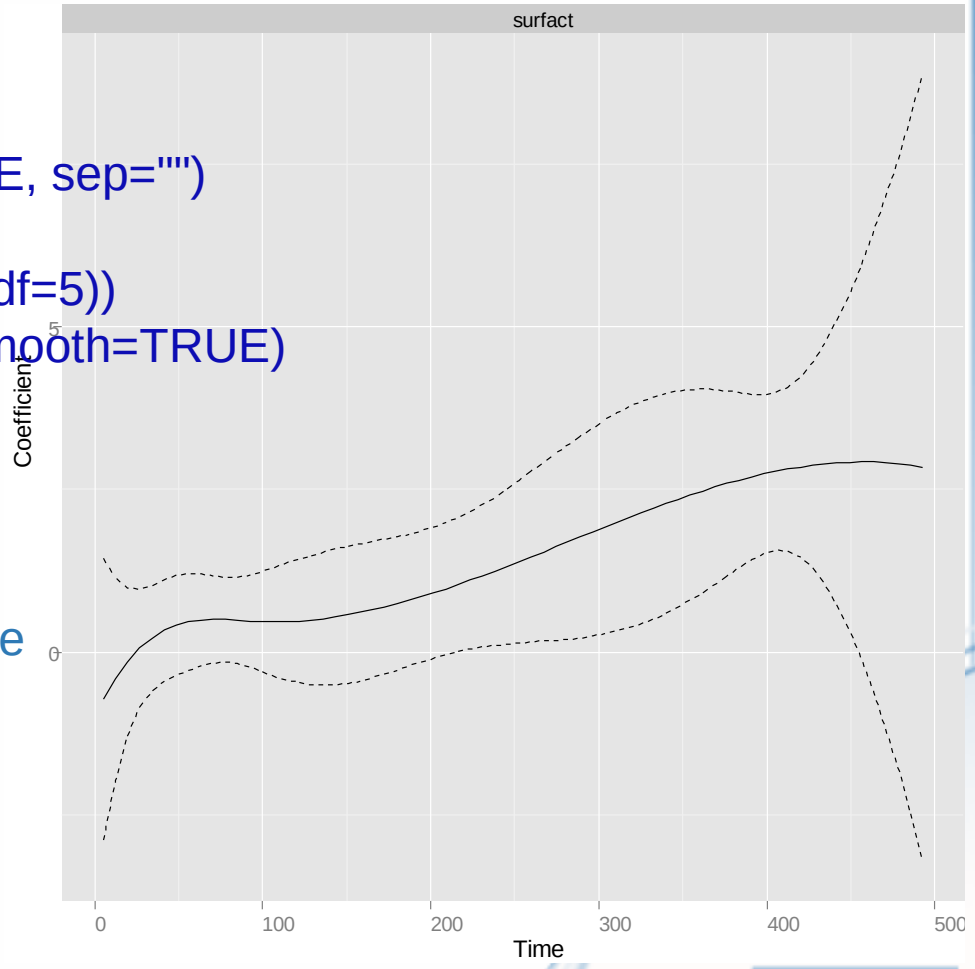
R function *splineCox* in package *dynsurv* (CRAN) by Wang, Yan, and Chen.
A tutorial is available at
http://cran.r-project.org/web/packages/dynsurv/dynsurv.pdf

Extension to the Cox models

Time-varying coefficient survival model

DBP data: using splineCox in R

```
library(dynsurv)
library(splines)
data= “~/bpd.dat"
bpd <- read.delim(data, header = FALSE, sep="")
bpdtvc<-Surv(ondays, censor)~surfact
fit<-splineCox(bpdtvc, bpd, control=list(df=5))
plotCoef(subset(coef(fit), Time<500), smooth=TRUE)
```



Not working when df=3 or lower.
Coxspline by Gray allows estimates to be piece-wise constant functions that are more robust but less smooth.

Extension to the Cox models

Time-varying coefficient survival model

Bayesian approach can be use to estimate time-varying coefficient models using Bayesian smoothing techniques, see the next section about Bayesian survival models.

Outline

Introduction

- Time to event data /Censoring and truncation
- Basic survival functions/Group comparison of survival data

Regression Analyses

- Parametric Approach -- Accelerated failure time models
- Semiparametric Approach – Cox models

Extension to the Cox models

- Weighted Cox models
- Time-dependent covariate Cox models
- Time-varying coefficient survival models

Other topics

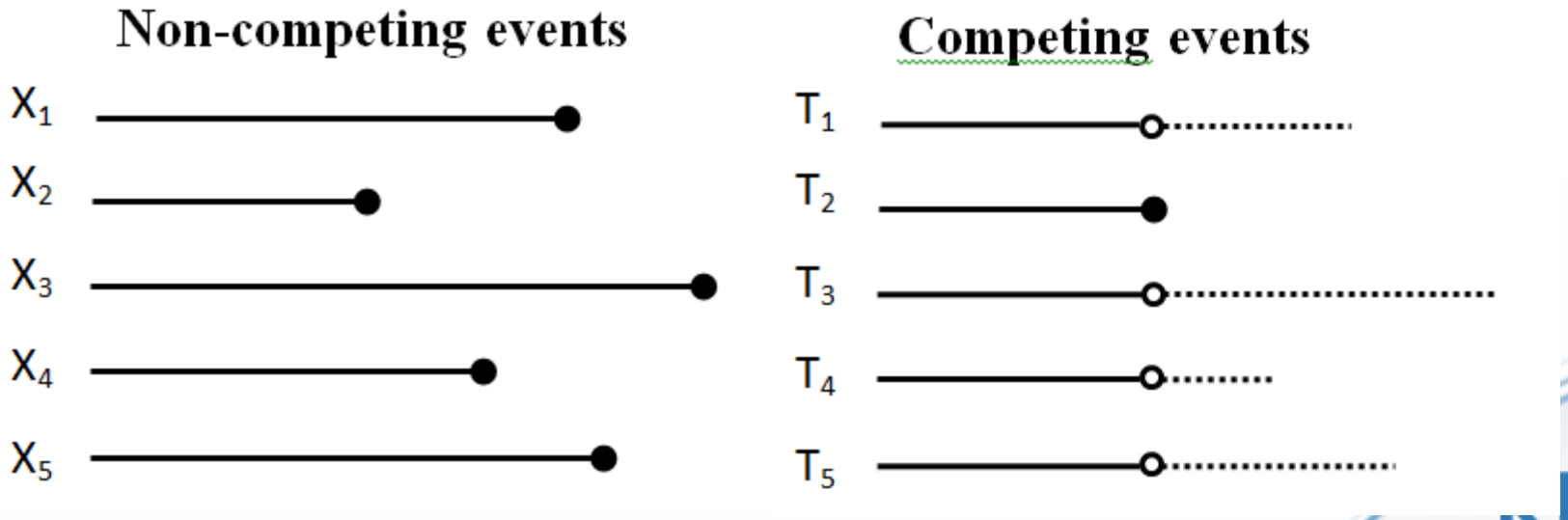
- Competing risk modeling
- Bayesian survival analyses

KU MEDICAL CENTER
The University of Kansas

Other Topics

Competing Risks Survival Models

In reality, there may be multiple events of interest. Sometimes, all event times can be observed (non-competing events), but sometimes only one of the event can be observed (competing events).



Other Topics

Competing Risks Survival Models

Let $X_i, i = 1, \ldots, K$ be the competing event times. $T = \mathrm{Min}(\mathrm{X}_1, \ldots, \mathrm{X}_K)$ is observed event time and δ indicates which of the K events happens first. $\delta = \mathrm{i}$ if $\mathrm{T} = \mathrm{X}_i$.

Cause-specific (crude) hazard rate

$$h_i^c(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, \delta = \mathrm{i} \,|T \geq t)}{\Delta t}$$

$h_i^c(t)$ measures the **instantaneous risk of having an i event** for a subject **who hasn't experienced any of the K events at time t**, that is, the risk of observing event i happening first.

Other Topics

Competing Risks Survival Models

The **overall hazard rate** of the time to failure(any of the competing events happening), T, is

$$h_T(t) = \sum_{i=1}^{K} h_i^c(t)$$

For example, all-cause mortality can be considered as the combination of mortalities of multiple mutually exclusive causes, such as, cancer-related, car accidents, murder,…etc.

Other Topics

Competing Risks Survival Models

Cumulative Incidence Function (CIF)

$$CIF_i(t) = P[T \le t, \delta = i] = \int_0^t h_i^c(\mu)\exp\{-H_T(\mu)\}d\mu$$

When all competing events are independent, $h_i^c(\mu)$ is the same as marginal $h_i(\mu)$ that is estimable using the regular Cox model assuming a single event of interest and other events are censoring. However, the independent assumption cannot be tested statistically and may not be reasonable from the clinical aspect. Further more, the effect of covariates is not directly associated with CIF.

Other Topics

Competing Risks Survival Models

Cumulative Incidence Function (CIF)

$$CIF_i(t) = P[T \le t, \delta = i] = \int_0^t h_i^c(\mu)\exp\{-H_T(\mu)\}d\mu$$

When all competing events are independent, $h_i^c(\mu) = h_i(\mu)$ that is estimable using the regular Cox model assuming a single event of interest and other events as censoring. In a hypothetic world without competing events,

$$F_i(t) = P[T \le t, \delta_i = 1] = \int_0^t h_i(\mu)\exp\{-H_i(\mu)\}d\mu$$

$h_i^c(\mu) = h_i(\mu)$, but $H_T(t) \ge H_i(t)$, so $CIF_i(t) \le F_i(t)$

Other Topics

Competing Risks Survival Models

Cause-specific hazard regression model

$$h_i^c(\mu) = h_i(\mu) = h_{i0}(\mu)\exp(\boldsymbol{\beta}'\boldsymbol{X})$$

$$CIF_i = \int_0^t h_i^c(\mu)\exp\{-H_T(\mu)\}d\mu$$

This model is based on the independence assumption that cannot be tested statistically and may not be reasonable from the clinical aspect. For example, death due to coronary heart disease and death due to stroke are likely to be related. Further more, the effect of covariates is not directly associated with CIF (observable outcomes).

Other Topics

Competing Risks Survival Models

Fine and Gray Model (Sub-distribution regression model)

$$h_i^s(\mu) = h_{i0}^s(\mu)\exp(\boldsymbol{\beta}'\boldsymbol{X})$$

$$CIF_i = \int_0^t h_i^s(\mu)\exp\{-H_i^s(\mu)\}d\mu$$

This model doesn't depend on the independence assumption among competing events. The effects of covariates are directly associate with the CIF and easy to implement.

Other Topics

Competing Risks Survival Models

SAS macro (estimate CIFs, but doesn't allow modeling)

http://www.mcw.edu/FileLibrary/Groups/Biostatistics/Software/SAS_Macro_For_Cumulative_Incidence_Functions.txt

R package: *cmprsk by Grey* (CRAN)

crr (ftime, fstatus,…)

regression modeling of subdistribution functions

cuminc (ftime, fstatus,…)

estimates CIFs

…

Other Topics

Competing Risks Survival Models

Cardiovascular disease (CVD) and other mortalities

http://www.umass.edu/statdata/statdata/data/comprisk.dat

Variable	Codes/Values
id	1 - 453
age	Years
gender	0 = Male, 1 = Female
BMI (body mass index)	kg/m^2
time	Days
ev_typ	1 = CVD, 2 = Other Cause,0 = Censor

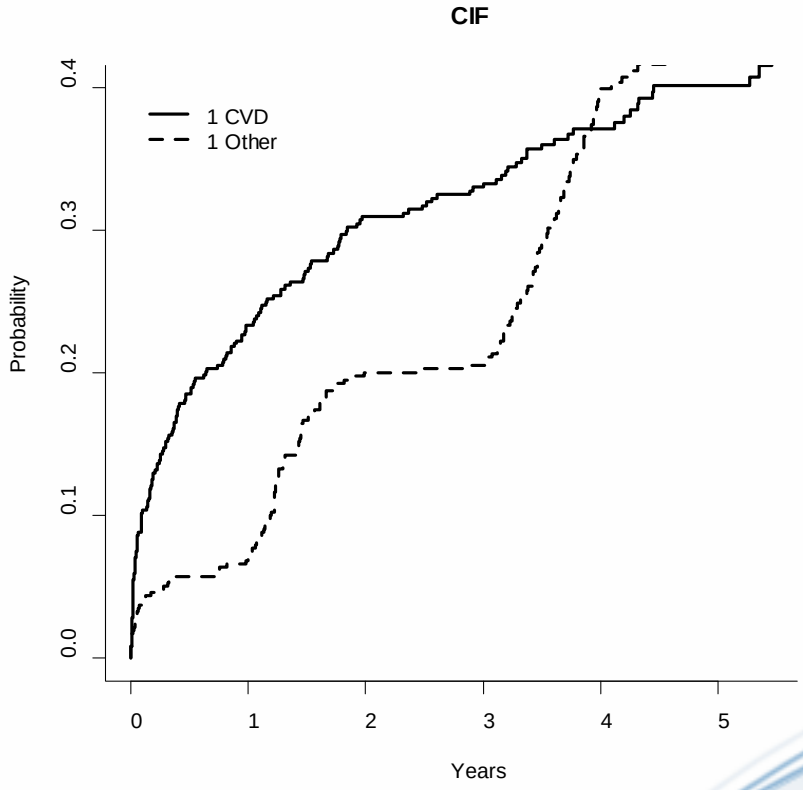
Other Topics

Competing Risks Survival Models

Cardiovascular disease (CVD) and other mortalities

The sum of CIF(cancer) and CIF(other) equals the P(total mortality) at any time point.

The sum of KM survival functions of CVD (other deaths as censoring) and cancer (CVD as censoring) is greater than the P(total mortality) at any time point.



Are deaths by CVD and other causes independent?

```
cif <- cuminc(ftime =cvd$time,fstatus  = cvd$ev_typ, rho = 0,cencode = "Censored")
plot(cif[1:2],lwd = 3,col=1,lty=1:2,ylim = c(0,0.4), main="CIF")
```

KU MEDICAL CENTER
The University of Kansas

Other Topics

Competing Risks Survival Models

CVD Mortality-- Fine and Gray model

	coef	exp(coef)	se(coef)	z	p-value
Female	-0.0752	0.928	0.1640	-0.459	0.650
Underweight	-1.3686	0.254	1.0364	-1.320	0.190
Overweight	0.4465	1.563	0.2387	1.870	0.061
Obese	0.7078	2.029	0.2290	3.091	0.002
age	0.0890	1.093	0.0069	12.886	0.000

```
finegray <- crr(ftime =cvd$time,fstatus  = cvd$ev_typ,
cov1=x,failcode ="Cancer", cencode  = "Censored")
```

KU MEDICAL CENTER
The University of Kansas

Other Topics

Competing Risks Survival Models

CVD Mortality -- Cox model

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
Female	-0.095139	0.909247	0.157705	-0.603	0.54633
Underweight	-0.801316	0.448738	1.019170	-0.786	0.43172
Overweight	0.415033	1.514420	0.217353	1.909	0.05620
Obese	0.582594	1.790677	0.215747	2.700	0.00693
age	0.083464	1.087046	0.005838	14.297	< 2e-16

```
cox <- coxph(formula =
Surv(time,dth_cvd==0)~gender+underwt+overwt+obese+age,
data=cvd)
```

There are lots of censoring cases in the data. The default method of handling ties for *COXPH* is Efron's method.

KU MEDICAL CENTER
The University of Kansas

Other Topics

Competing Risks Survival Models

Other Mortality-- Fine and Gray model

	coef	exp(coef)	se(coef)	z	p-value
Female	-0.0103	0.990	0.17757	-0.0582	0.95
Underweight	0.7597	2.138	0.35957	2.1127	0.035
Overweight	-0.5262	0.591	0.18213	-2.8891	0.004
Obese	-0.8401	0.432	0.20890	-4.0216	5.8e-05
age	-0.0464	0.955	0.00422	-11.0097	<0

```
finegray <- crr(ftime =cvd$time,fstatus  = cvd$ev_typ,
cov1=x,failcode = "Other", cencode  = "Censored")
```

KU MEDICAL CENTER
The University of Kansas

Other Topics

Competing Risks Survival Models

Other Mortality -- Cox model

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
Female	-0.008966	0.991074	0.178794	-0.050	0.96000
underwt	0.718968	2.052315	0.308726	2.329	0.01987
overwt	-0.335500	0.714980	0.182532	-1.838	0.06606
obese	-0.621722	0.537019	0.220110	-2.825	0.00473
age	-0.027848	0.972536	0.004893	-5.691	1.26e-08

```
cox <- coxph(formula =
Surv(time,dth_cvd==0)~gender+underwt+overwt+obese+age,
data=cvd)
```

KU MEDICAL CENTER
The University of Kansas

Other Topics

Bayesian Survival Analysis

With the development of computation techniques, Bayesian methods are used more and more common.

For Bayesian analysis, all parameters are considered as random variables rather than constant (frequentist). Given prior distributions of parameters, posterior distributions parameters can be obtained.

Bayes' theorem

$$p(\theta|T) = \frac{p(T|\theta)p(\theta)}{p(T)} \sim p(T|\theta)p(\theta)$$

Other Topics

Bayesian Survival Analysis

Bayesian approach is very useful

- When there are strong prior distributions based on expert opinions or data from previous studies.
- When the model is very complex that numeric solutions based on likelihoods are difficult. Bayesian models can use sampling methods (MCMC) to obtain samples from the posterior distribution of parameter.

KU MEDICAL CENTER
The University of Kansas

Other Topics

Bayesian Survival Analysis

In SAS, both PROC LIFEREG and PROC PHREG have the Bayes option that provide Bayesian estimates. Unless informative priors are provided, the results are similar to those of frequentist's approaches.

```
ODS graphics on;
proc phreg data=bpd;
model ondays*censor(0)=surfact;
bayes seed=23855 NBI=10000 NMC=50000  thinning=10 outpost=Post
DIAGNOSTICS=ALL PLOTS=(trace autocorr);
run;
ods graphics off;
```

Other Topics

Bayesian Survival Analysis

Bayesian survival analysis is more useful in estimating more complicated models than regular AFT and Cox model. Researchers often need to write their own codes.

Reference book: Bayesian Survival Analysis
By **Ibrahim**, Joseph G., **Chen**, Ming-Hui, **Sinha**, Debajyoti.

Other Topics

Frailty model using proc PHREG (Random Effects)
https://www.youtube.com/watch?v=ZfgRBuM4u3U

Multi-state models
R package: mstate