ANALYSIS OF CENSORED DATA – SURVIVAL ANALYSIS

DANYU LIN

Dennis Gillings Distinguished Professor
Department of Biostatistics
University of North Carolina
Chapel Hill, NC 27599-7420
email: lin@bios.unc.edu
website: http://dlin.web.unc.edu/

June 30 – July 1, 2014
East China Normal University

Google: The Sexiest Jobs?

For Today's Graduate, Just One Word: Statistics

— *The New York Times*, August 5, 2009

"I keep saying that the sexy job in the next 10 years will be statisticians. And I'm not kidding."

— Hal Varian, Google's Chief Economist

Data Scientist: The Sexiest Job of the 21st Century

— *Harvard Business Review*, October 2012

The Sexiest Job of the 21st Century? Data Analyst

— *CNBC*, June 5, 2013

Why Survival Analysis?

- Many studies are concerned with times to events (e.g., death, disease, machine breakdown, bankrupt, employment).
- Event times are censored if subjects are not followed long enough.
- Standard statistical methods cannot handle censoring.
- Special statistical methods have been developed to provide valid and efficient analysis of censored data.

PART I

ANALYSIS OF SURVIVAL DATA

I. INTRODUCTION

A. Survival Data

1. Survival times (failure times): times to the occurrence of a given event (failure) measured from a well-defined starting point (randomization)

- death
- physical symptoms/diseases
- machine failure
- bankruptcy
- purchase of product

2. Censoring: some subjects are not observed for the full time to failure as a result of

- loss to follow-up
- end of study

3. Latent variables and observable data:

$(X_i, \Delta_i, Z_i) \quad (i = 1, \ldots, n)$

$X =$ observation time (last contact date)

$\Delta =$ failure (censoring) indicator

$Z =$ covariates (treatment, etc.)

$T =$ survival (failure) time

$C =$ censoring time

$X = \min(T, C)$

$\Delta = I(T \leq C)$

- $I(\mathcal{A}) =$ indicator function for $\mathcal{A}$

$Z(t) = \{Z_1(t), \ldots, Z_p(t)\}'$

B. Real Examples:

1. Leukemia Study. In a clinical trial, the drug 6-MP was compared to placebo with respect to the ability to maintain remission in leukemia patients.

Lengths of remission (in weeks) for two groups of patients

6-MP	6,6,6,6*,7,9*,10,10*,11*,13,16,17*,19*,
	20*,22,23,25*,32*,32*,34*,35*
Placebo	1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,
	17,22,23

* denotes censored observations

2. Mayo PBC Study

In a double-blind clinical trial, 312 PBC patients were randomized to DPCA and placebo. Additional 106 patients were also followed.

ID	X	Δ	Z_1	Z_2	Z_3	...
1	400	1	0	58	14.5	...
2	1504	0	1	38	3.4	...
			$\vdots$			

3. Colon Cancer Study. A national intergroup trial was conducted in the 1980's to study the drugs Lev and 5-FU for adjuvant therapy of resected colon carcinoma. Patients with Stage C disease were randomly assigned to observation, Lev alone, or Lev+5-FU.

Group	Patients	Deaths
observation	315	114
Lev	310	109
Lev+5-FU	304	78

C. Scientific Questions:

1. Estimating survival distribution
2. Testing equality of two or more survival distributions
3. Estimating effects of covariates (e.g., treatment) on survival time

D. Characterizing Survival Distribution:

Distribution function: $F(t) = \Pr(T \leq t)$

Density function: $f(t) = dF(t)/dt$

Survival function:

$$S(t) = \Pr(T > t)$$
$$= 1 - F(t)$$

Hazard function:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$
$$= f(t)/S(t)$$

Cumulative hazard function: $\Lambda(t) = \int_0^t \lambda(u)du$

$$\lambda(t) = -d \log S(t)/dt \implies$$
$$S(t) = e^{-\Lambda(t)}$$

E. Independent Censoring: The subjects censored at time t should be "representative" of the subjects under observation at t. Subjects cannot be censored because they are at unusually high or low risk of failure.

One-sample case: T and C are independent.

Two-sample case: T and C are independent within each group.

Regression case: T and C are independent conditional on Z.

F. Naive Methods

Ignore censoring status: failure times associated with censored observations are underrepresented.

Delete censored cases: estimates are biased towards shorter failure times because larger failure times are more likely to be censored.

Example:

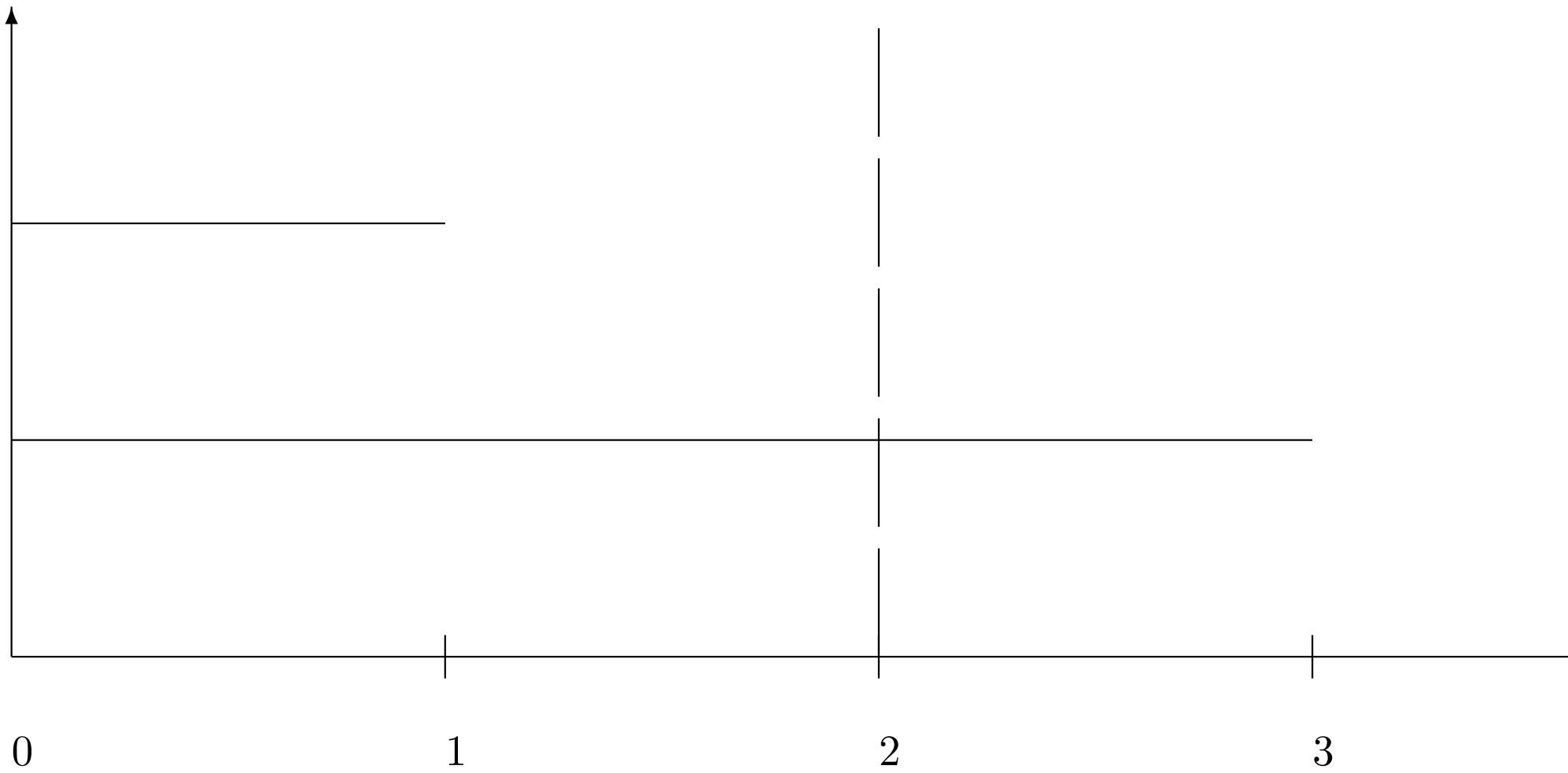
$$\Pr(T = 1) = \Pr(T = 3) = 0.5$$

$$E(T) = 2$$

$$\Pr(C = 2) = 1$$

$$E(X) = 1.5$$

$$E(X|\Delta = 1) = 1$$


0
1
2
3

G. Nelson-Aalen and Kaplan-Meier Estimators

Nelson-Aalen estimator

$$\begin{aligned}\Lambda(t+\Delta t)-\Lambda(t) &\approx \lambda(t)\Delta t \\ &\approx \Pr(t \le T < t+\Delta t | T \ge t) \\ &= \Pr(t \le T < t+\Delta t | \underbrace{T \ge t, C \ge t}_{X \ge t})\end{aligned}$$

$$\Lambda(t_l)-\Lambda(t_{l-1}) \approx \Pr(t_{l-1} \le T < t_l | X \ge t_{l-1}) \approx d_l/y_l$$

$$\widehat{\Lambda}(t) = \sum_{l:t_l \le t} d_l/y_l \longrightarrow \sum_{k:T_k^0 \le t} D_k/\overline{Y}_k$$

$$as\ m \to \infty, \quad \max_{1 \le l \le m} |t_l - t_{l-1}| \to 0$$

$T_1^0 < T_2^0 < \ldots =$ distinct observed failure times

$D_k =$ # failures at T_k^0

$\overline{Y}_k =$ #subjects at risk at T_k^0

Kaplan-Meier estimator

$$\Pr(T > t) = \Pr(T > t_0)\Pr(T > t_1|T > t_0)\Pr(T > t_2|T > t_1)\ldots$$

$$S(t) \approx \prod_{l:t_l \le t} \Pr(T \ge t_l|T \ge t_{l-1})$$

$$= \prod_{l:t_l \le t} \{1 - \Pr(T < t_l|T \ge t_{l-1})\}$$

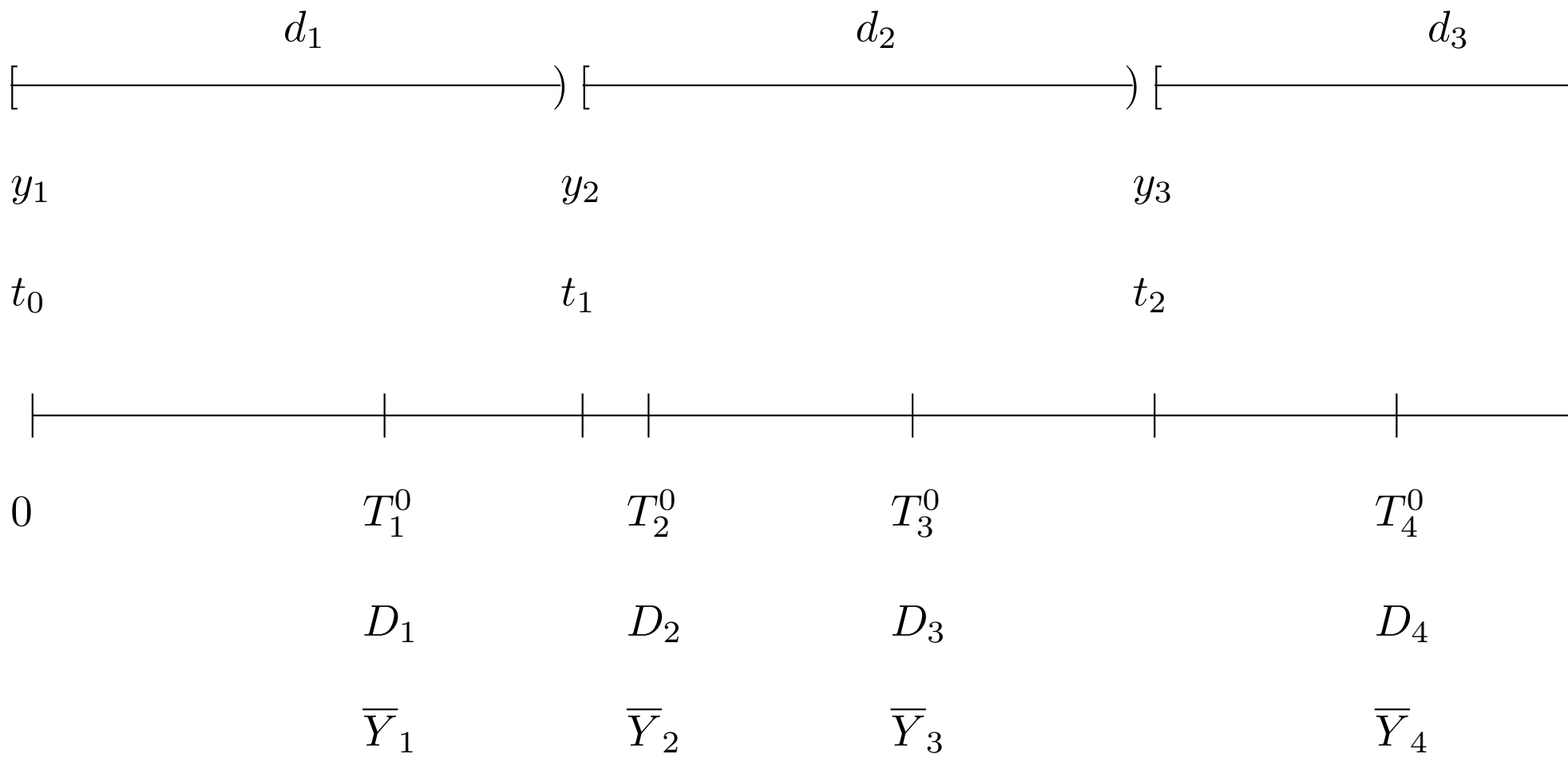
$$\widehat{S}(t) = \prod_{l:t_l \le t} (1 - d_l/y_l) \longrightarrow \prod_{k:T_k^0 \le t} (1 - D_k/\overline{Y}_k)$$

$$as\ m \to \infty, \quad \max_{1 \le l \le m} |t_l - t_{l-1}| \to 0$$

$$\widehat{S}(T_k^0) = (1 - D_1/\overline{Y}_1)\ldots(1 - D_{k-1}/\overline{Y}_{k-1})(1 - D_k/\overline{Y}_k)$$

$$= \widehat{S}(T_{k-1}^0)(1 - D_k/\overline{Y}_k)$$

For uncensored data, $\widehat{S}$ reduces to one minus the empirical distribution function.


d_1
d_2
d_3
y_1
y_2
y_3
t_0
t_1
t_2
0
T_1^0
T_2^0
T_3^0
T_4^0
D_1
D_2
D_3
D_4
$\overline{Y}_1$
$\overline{Y}_2$
$\overline{Y}_3$
$\overline{Y}_4$

$T_1^0 < T_2^0 < \ldots < T_L^0 \equiv$ distinct time points of observed failures

$D_k = \#$ failures at T_k^0

$0 = t_0 < t_1 < \ldots < t_m = t \equiv$ partition of the interval [0,t]

$d_l \equiv \#$ failures in $[t_{l-1}, t_l)$

$y_l \equiv \#$ subjects at risk at t_{l-1}

$\equiv \#$ subjects under observation just prior to t_{l-1}

$\overline{Y}_k \equiv \#$ subjects at risk at T_k^0

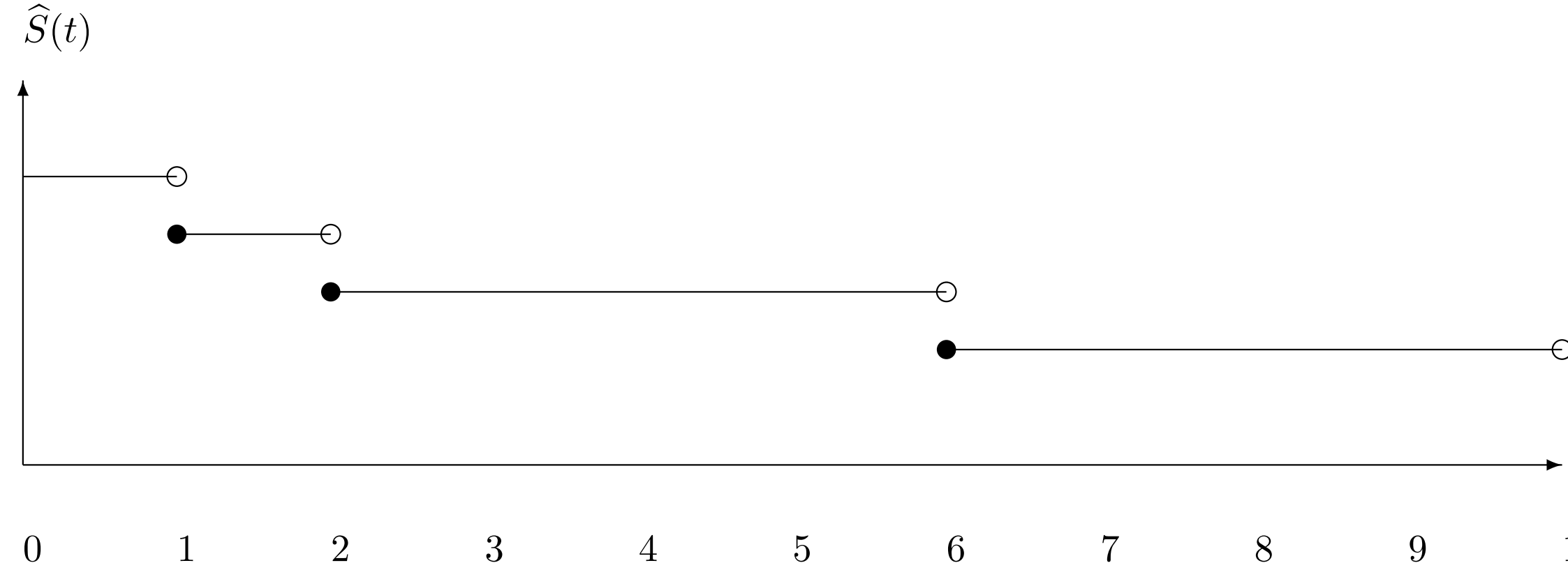
Illustration of Kaplan-Meier estimator

Data : $1 \quad 2 \quad 2 \quad 4^* \quad 5^* \quad 6 \quad 7^* \quad 8^* \quad 9^* \quad 10^*$

X_i	1	2	2	4	5	6	7	8	9	10
δ_i	1	1	1	0	0	1	0	0	0	0

Calculations:

T_k^0	D_k	$\overline{Y}_k$	$1 - \frac{D_k}{\overline{Y}_k}$	$\widehat{S}(T_k^0)$
1	1	10	$1 - \frac{1}{10} = \frac{9}{10}$	$\frac{9}{10}$
2	2	9	$1 - \frac{2}{9} = \frac{7}{9}$	$\frac{9}{10} \times \frac{7}{9} = \frac{7}{10}$
6	1	5	$1 - \frac{1}{5} = \frac{4}{5}$	$\frac{7}{10} \times \frac{4}{5} = \frac{14}{25}$


$\widehat{S}(t)$
0
1
2
3
4
5
6
7
8
9

Relationship between Nelson-Aalen and Kaplan-Meier estimators

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

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

$$= \prod_{k:T_k^0 \le t} e^{-D_k/\overline{Y}_k}$$

$$\approx \prod_{k:T_k^0 \le t} (1 - D_k/\overline{Y}_k) \quad (\text{if} \quad D_k/\overline{Y_k} \approx 0)$$

$$= \widehat{S}(t) \qquad (\text{Kaplan-Meier Estimator})$$

PBC data: Figs. 1-2

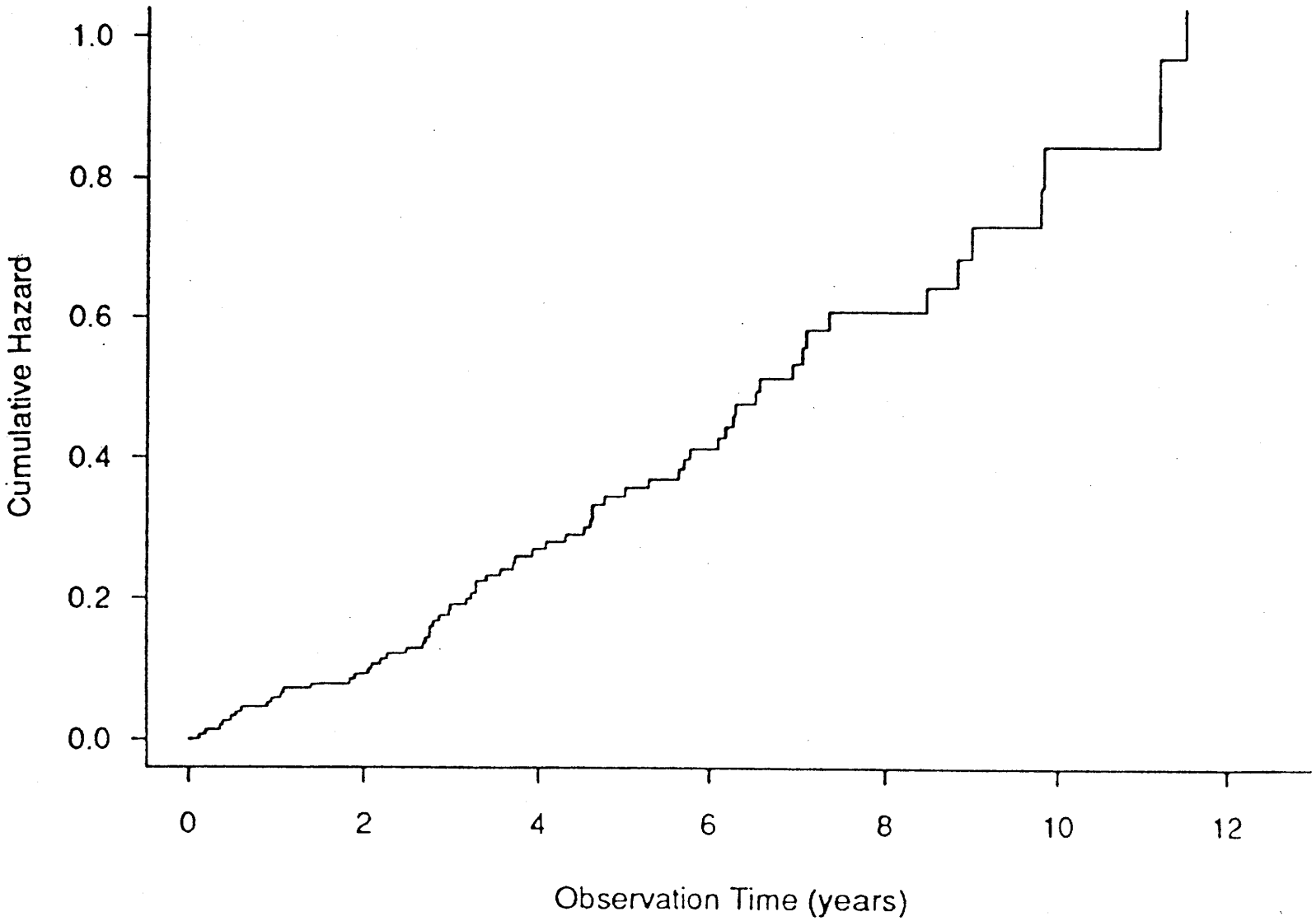


Figure 1 Nelson cumulative hazard estimate for DPCA group, PBC data.

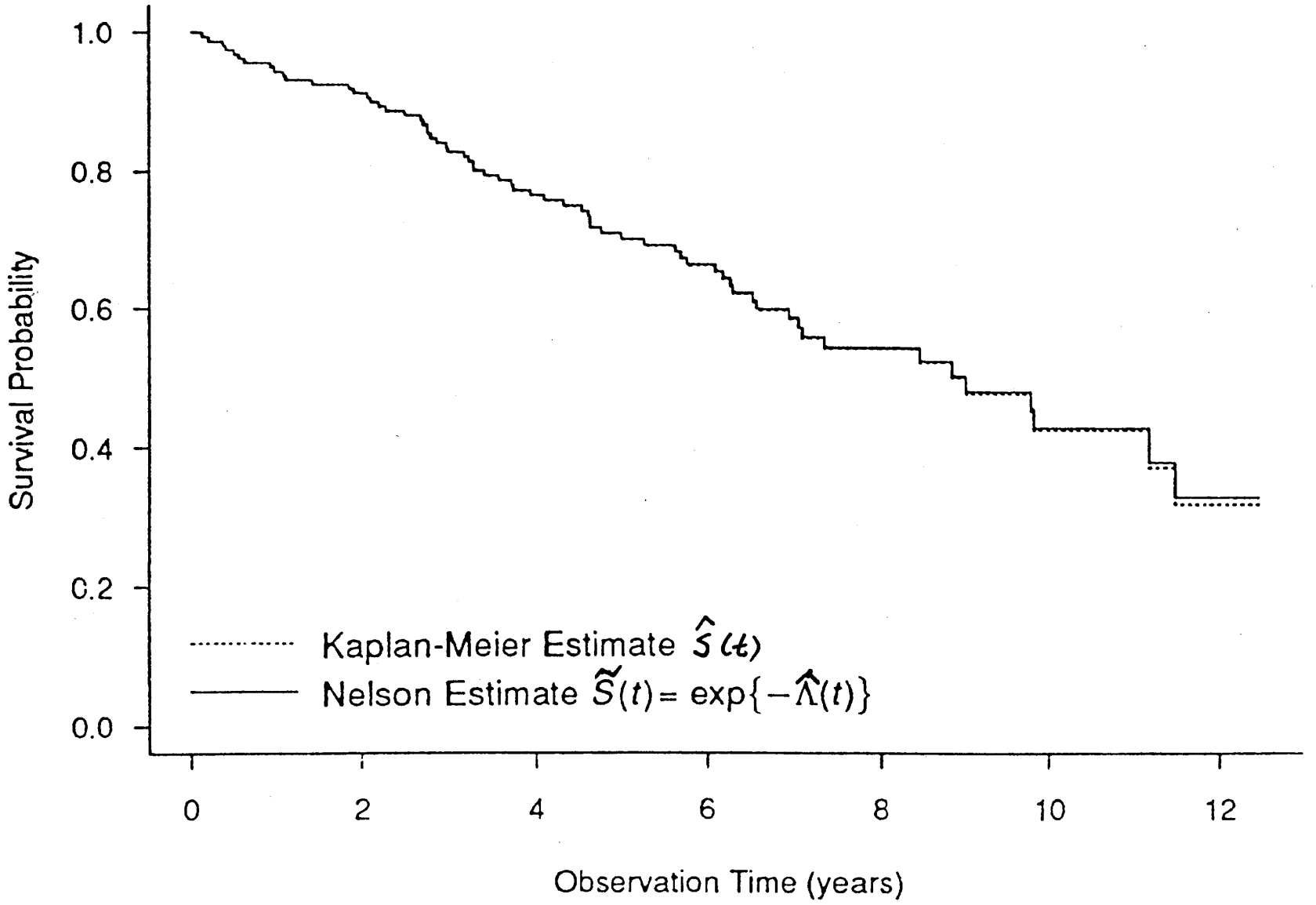


Figure 2 Estimated survival curves for DPCA group, PBC data.

H. Log-Rank Statistic:

$T_1^0 < T_2^0 < \ldots < T_L^0 =$ distinct observed failure times in the combined sample

At T_k^0	Failures	Non-failures	Risk Size
Group 1	D_{1k}	$\overline{Y}_{1k} - D_{1k}$	$\overline{Y}_{1k}$
Group 0	D_{0k}	$\overline{Y}_{0k} - D_{0k}$	$\overline{Y}_{0k}$
	D_k	$\overline{Y}_k - D_k$	$\overline{Y}_k$

$$E_{1k} = \frac{D_k \overline{Y}_{1k}}{\overline{Y}_k}, \quad V_{1k} = \frac{D_k(\overline{Y}_k - D_k)\overline{Y}_{1k}\overline{Y}_{0k}}{\overline{Y}_k^2(\overline{Y}_k - 1)}$$

$$Q = \frac{\sum_{k=1}^{L}(D_{1k} - E_{1k})}{(\sum_{k=1}^{L} V_{1k})^{1/2}} \sim N(0,1)$$

Leukemia Study: $Q^2 = 14.5$, p-value=0.00014, Fig. 3

PBC Study: $Q^2 = 0.32$, p-value=0.76, Fig. 4

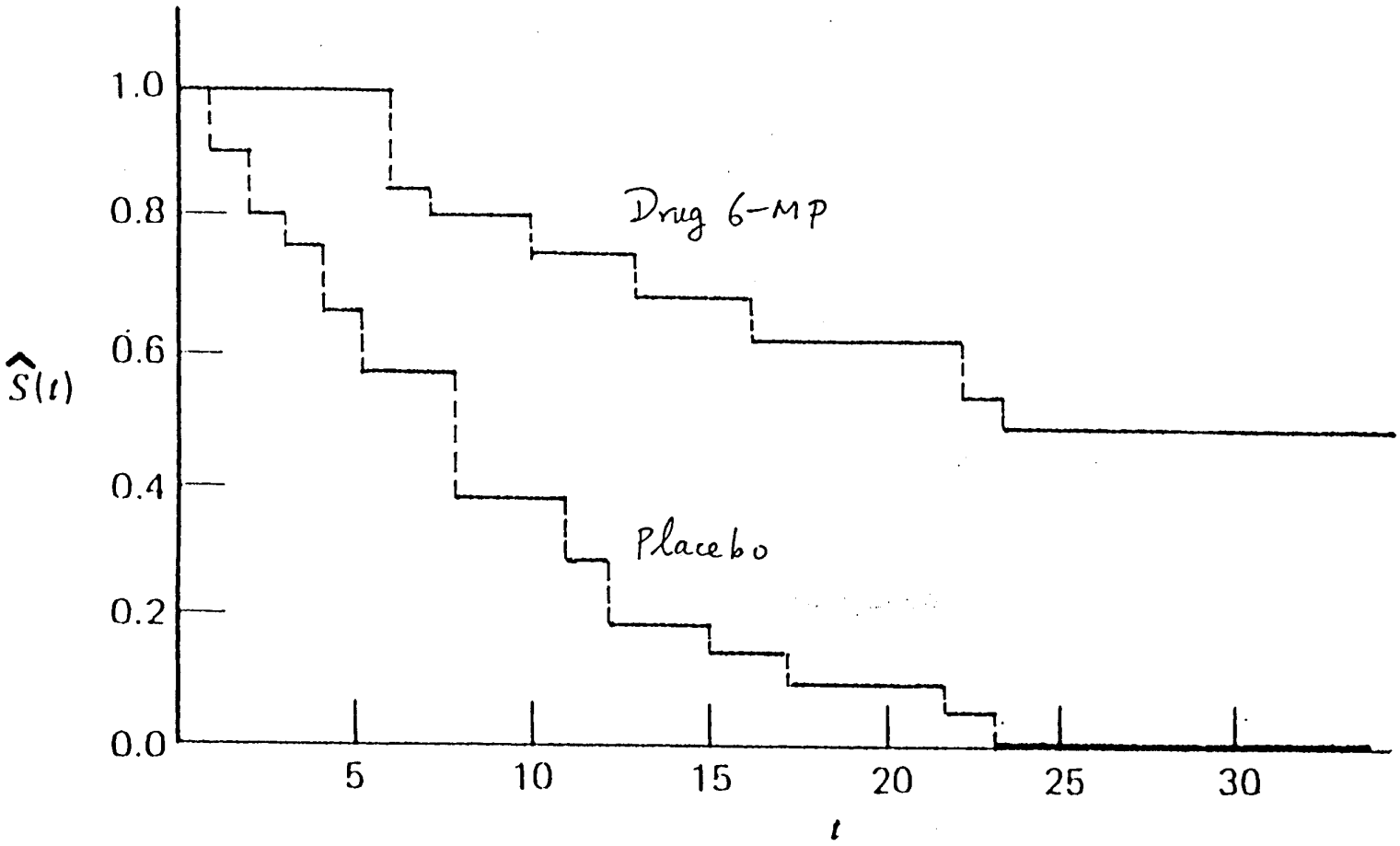
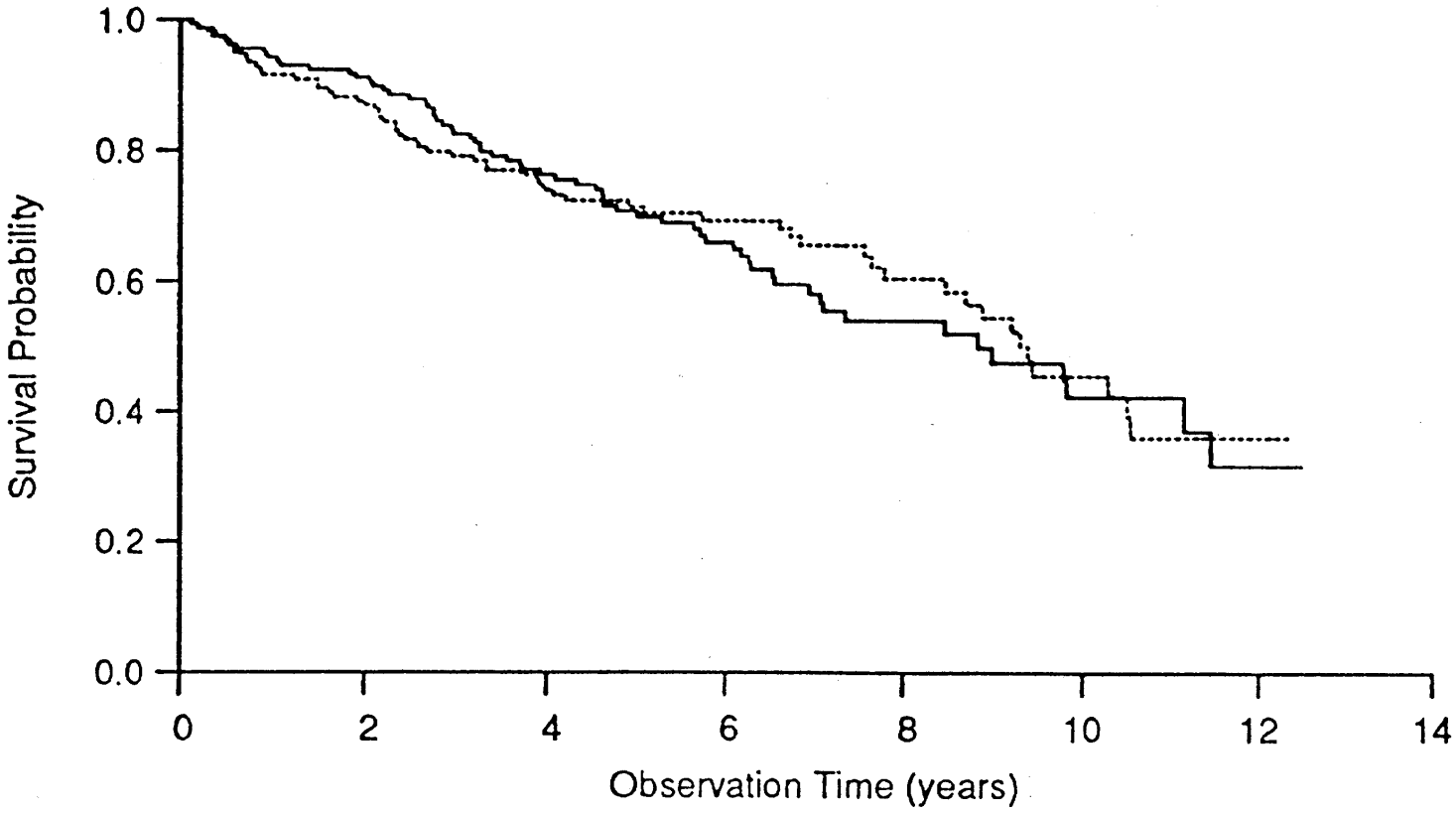


Figure 3 Product-limit estimates from two samples in Leukemia Study



Group	Time Interval 0-2	2-4	4-6	6-8	8-10	10-12	12-14
——— DPCA	14/158	22/143	12/101	10/65	5/34	2/16	0/4
····· Placebo	19/154	20/135	5/93	6/65	7/33	3/16	0/4

(# events/# at risk)

Figure 4 Estimated survival curves for the DPCA and placebo groups, PBC Data. The table below the curves gives the number of failures in each time interval, and the number of cases at risk at the beginning of the interval.

I. Cox Regression:

1. Model:

$$\lambda(t|Z) \equiv \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \Pr(t \leq T < t + \Delta t | T \geq t, Z)$$

$$= \lambda_0(t) e^{\beta_0' Z(t)}$$

- $\lambda_0(t) \equiv \lambda(t|Z = 0) =$ arbitrary baseline hazard function
- $\beta_0 =$ unknown regression parameters

2. Partial Likelihood Inference:

Partial likelihood:

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{e^{\beta' Z_i(X_i)}}{\sum_{j \in \mathcal{R}_i} e^{\beta' Z_j(X_i)}} \right\}^{\Delta_i}$$

- $\mathcal{R}_i = \{j : X_j \geq X_i\}$

Score function:

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta}$$

$$= \sum_{i=1}^{n} \Delta_i \left\{ Z_i(X_i) - \frac{\sum_{j \in \mathcal{R}_i} e^{\beta' Z_j(X_i)} Z_j(X_i)}{\sum_{j \in \mathcal{R}_i} e^{\beta' Z_j(X_i)}} \right\}$$

Information matrix:

$$\mathcal{I}(\beta) = -\frac{\partial^2 \log L(\beta)}{\partial \beta^2}$$

MPLE $\widehat{\beta}$: $\{U(\beta) = 0\}$

Properties:

$$U(\beta_0) \sim N(0, \mathcal{I}(\beta_0))$$

$$\widehat{\beta} \sim N(\beta_0, \mathcal{I}^{-1}(\widehat{\beta}))$$

3. PBC Study:

Covariate	Est	SE	Est/SE
Treatment	0.136	0.185	0.73
Age	0.035	0.009	3.89
log (Albumin)	–3.078	0.729	–4.28
log (Bilirubin)	0.884	0.099	8.96
Edema	0.786	0.296	2.65
log (Protime)	2.971	1.016	2.92

J. Citations:

Ryan, T. P. and Woodall, W. H. (2005). The most-cited statistical papers. *J. App. Stat.*, 32, 461–474

(1) With 25,869 citations (currently cited 1,984 times per year):

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, 53, 457–481.

(2) With 18,193 citations (1,342 per year):

Cox, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. B*, 34, 187–220.

Garfield, E. (1990). The most-cited papers of all time, SCI 19451988. Part 1A. The SCI top 100 will the Lowry method ever be obliterated? *Current Comments*, 7, 3–14.

Kaplan & Meier (1958) and Cox (1972) ranked among the 100 most-cited papers in the 1945-1988 *Science Citation Index*.

II. COUNTING PROCESSES AND MARTINGALES

1. Counting Process Fomulation:

$\{N_i(t), Y_i(t), Z_i(t)\}$ $(i = 1, \ldots, n)$

General case:

Counting process: $N_i(t) = \#$ observed events on subject i by time t

At-risk process: $Y_i(t) = 1$ iff subject i is at risk at time t

Covariate process: $Z_i(t) =$ covariate value at time t on subject i

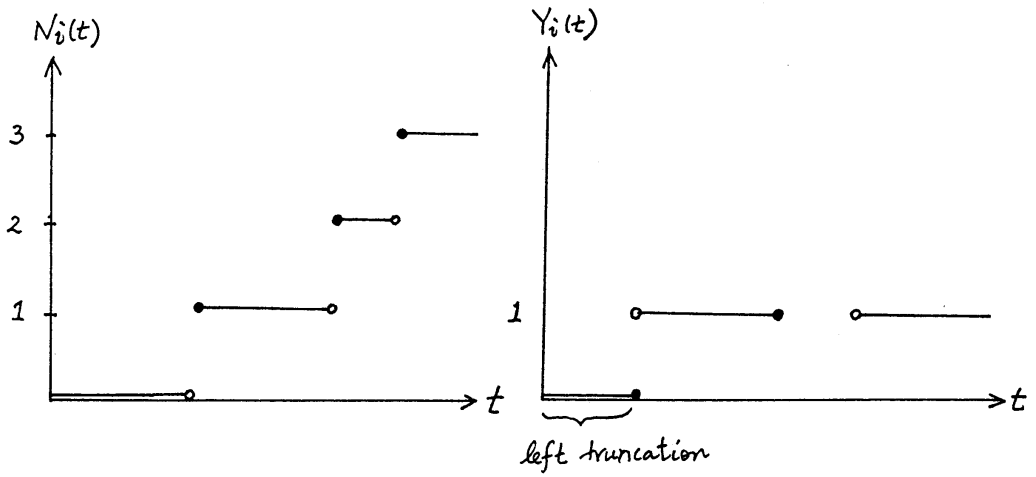
Classical survival data:

$N_i(t) = \Delta_i I(X_i \leq t)$

$Y_i(t) = I(X_i \geq t)$

Fig. 5

(a) General Case



(b) Survival Data

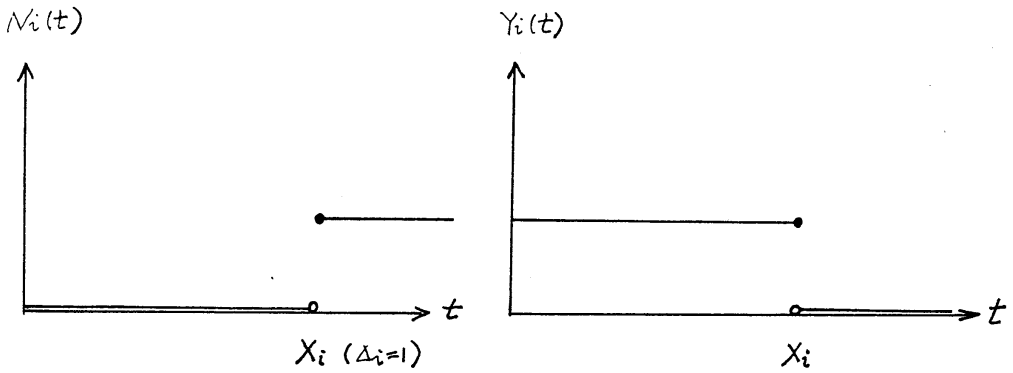


Fig. 5 Counting process and at-risk process illustration

2. Martingales:

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)\lambda_i(u)du$$

- A martingale is a pure random noise process, which has no systematic behavior in the mean: the conditional expectation of the increment given the past history (failure, censoring and covariates) is zero.
- Each $M_i(t)$ is a martingale. This fact provides the basis for studying many censored-data statistics by the martingale theory.

$$dM_i(t) = dN_i(t) - Y_i(t)\lambda_i(t)dt$$

3. Stieltjes Integration

$$S = \int_0^t f(x) dG(x)$$

- If $dG(x)/dx = g(x)$ exists, then

$$S = \int_0^t f(x) g(x) dx$$

- If G is a step function with jumps $\Delta G(x_k) = G(x_k) - G(x_k-)$, then

$$S = \sum_{k: x_k \leq t} f(x_k) \Delta G(x_k)$$

- $\int_0^t H(u) d\Lambda(u) = \int_0^t H(u) \lambda(u) du$
- $\int_0^t H(u) dN_i(u) = I(X_i \leq t) \Delta_i H(X_i)$
- $\int_0^\infty H(u) dN_i(u) = \Delta_i H(X_i)$

4. Counting-Process Martingale Representation

Nelson-Aalen estimator:

$$\widehat{\Lambda}(t) = \sum_{k:T_k^0 \le t} D_k/\overline{Y}_k = \sum_{i=1}^{n} \frac{I(X_i \le t)\Delta_i}{\overline{Y}(X_i)} = \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{\overline{Y}(u)}$$

$$\overline{Y}(t) = \sum_{i=1}^{n} Y_i(t)$$

$$\widehat{\Lambda}(t) - \Lambda(t) \approx \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{\overline{Y}(u)} - \int_0^t \frac{\sum_{i=1}^{n} Y_i(u)}{\overline{Y}(u)} \lambda(u) du$$

$$= \sum_{i=1}^{n} \int_0^t \frac{dN_i(u) - Y_i(u)\lambda(u)du}{\overline{Y}(u)}$$

$$= \sum_{i=1}^{n} \int_0^t \frac{dM_i(u)}{\overline{Y}(u)}$$

Log-rank statistic:

$$U = \sum_{k=1}^{L} \left(D_{1k} - \frac{D_k \overline{Y}_{1k}}{\overline{Y}_k} \right)$$

$$= \sum_{i=1}^{n} \Delta_i \left\{ Z_i - \overline{Z}(X_i) \right\}$$

$$= \sum_{i=1}^{n} \int_0^\infty \left\{ Z_i - \overline{Z}(t) \right\} dN_i(t)$$

$$= \sum_{i=1}^{n} \int_0^\infty \left\{ Z_i - \overline{Z}(t) \right\} \left\{ dN_i(t) - Y_i(t)\lambda(t)dt \right\}$$

$$= \sum_{i=1}^{n} \int_0^\infty \left\{ Z_i - \overline{Z}(t) \right\} dM_i(t)$$

$$Z_i = \begin{cases} 1 & \text{if patient } i \text{ was on treatment 1} \\ 0 & \text{if patient } i \text{ was on treatment 0} \end{cases}$$

$$\overline{Z}(t) = \frac{\sum_{i=1}^{n} Y_i(t) Z_i}{\sum_{i=1}^{n} Y_i(t)} = \frac{\overline{Y}_1(t)}{\overline{Y}(t)}$$

Partial likelihood score function:

$$U(\beta) = \sum_{i=1}^{n} \Delta_i \left\{ Z_i(X_i) - \frac{\sum_{j \in \mathcal{R}_i} e^{\beta' Z_j(X_i)} Z_j(X_i)}{\sum_{j \in \mathcal{R}_i} e^{\beta' Z_j(X_i)}} \right\}$$

$$= \sum_{i=1}^{n} \int_0^\infty \left\{ Z_i(t) - \frac{\sum_{j=1}^{n} Y_j(t) e^{\beta' Z_j(t)} Z_j(t)}{\sum_{j=1}^{n} Y_j(t) e^{\beta' Z_j(t)}} \right\} dN_i(t)$$

$$= \sum_{i=1}^{n} \int_0^\infty \left\{ Z_i(t) - \frac{\sum_{j=1}^{n} Y_j(t) e^{\beta' Z_j(t)} Z_j(t)}{\sum_{j=1}^{n} Y_j(t) e^{\beta' Z_j(t)}} \right\} \{ dN_i(t) - Y_i(t) e^{\beta' Z_i(t)} \lambda_0(t) dt \}$$

$$U(\beta_0) = \sum_{i=1}^{n} \int_0^\infty \left\{ Z_i(t) - \frac{\sum_{j=1}^{n} Y_j(t) e^{\beta_0' Z_j(t)} Z_j(t)}{\sum_{j=1}^{n} Y_j(t) e^{\beta_0' Z_j(t)}} \right\} dM_i(t)$$

Common theme: Censored-data statistics (when properly centered) can be written as

$$\sum_{i=1}^{n} \int_0^t H_i(u) dM_i(u)$$

5. Martingale Theory

$$U(t) = \sum_{i=1}^{n} \int_0^t H_i(u) dM_i(u)$$

$U(t)$ is a martingale

$E\{U(t)\} = 0$

$\text{Var}\{U(t)\} = E\left\{\sum_{i=1}^{n} \int_0^t H_i^2(u) Y_i(u) \lambda_i(u) du\right\}$

$V(t) = \sum_{i=1}^{n} \int_0^t H_i^2(u) Y_i(u) \lambda_i(u) du$

$U(t) \sim N(0, V(t))$

III. NELSON-AALEN AND KAPLAN-MEIER ESTIMATORS

1. Properties

$$\widehat{\Lambda}(t) - \Lambda(t) \approx \sum_{i=1}^{n} \int_0^t \frac{dM_i(u)}{\overline{Y}(u)}$$

$$E\left\{\widehat{\Lambda}(t)\right\} \approx \Lambda(t)$$

$$\mathrm{Var}\left\{\widehat{\Lambda}(t)\right\} \approx E\left\{\sum_{i=1}^{n} \int_0^t \frac{Y_i(u)d\Lambda(u)}{\overline{Y}^2(u)}\right\} = E\left\{\int_0^t \frac{d\Lambda(u)}{\overline{Y}(u)}\right\}$$

$$V(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{\overline{Y}^2(u)} = \sum_{i=1}^{n} \frac{I(X_i \le t)\Delta_i}{\overline{Y}^2(X_i)} = \sum_{k:T_k^0 \le t} D_k/\overline{Y}_k^2$$

$$\widehat{\Lambda}(t) \sim N\left(\Lambda(t), V(t)\right)$$

$$\widehat{S}(t) \approx e^{-\widehat{\Lambda}(t)} \sim N\left(S(t), \widehat{S}^2(t)V(t)\right)$$

2. Confidence Intervals for $S(t)$

Untransformaed: $\widehat{S}(t) \pm z_{1-\alpha/2}\widehat{S}(t)V^{1/2}(t)$

Log transformation: $\widehat{S}(t)e^{\pm z_{1-\alpha/2}V^{1/2}(t)}$

Log-log transformation: $\widehat{S}(t)^{\exp\left\{\pm z_{1-\alpha/2}V^{1/2}(t)/\widehat{\Lambda}(t)\right\}}$

$$z_\gamma = \Phi^{-1}(\gamma)$$

3. S function

survfit (Surv(time, status) $\sim$ group, conf.type=c('log-log'))

4. Leukemia Study: Figs. 6-8

Fig. 6 Kaplan-Meier Estimates for Leukemia Study

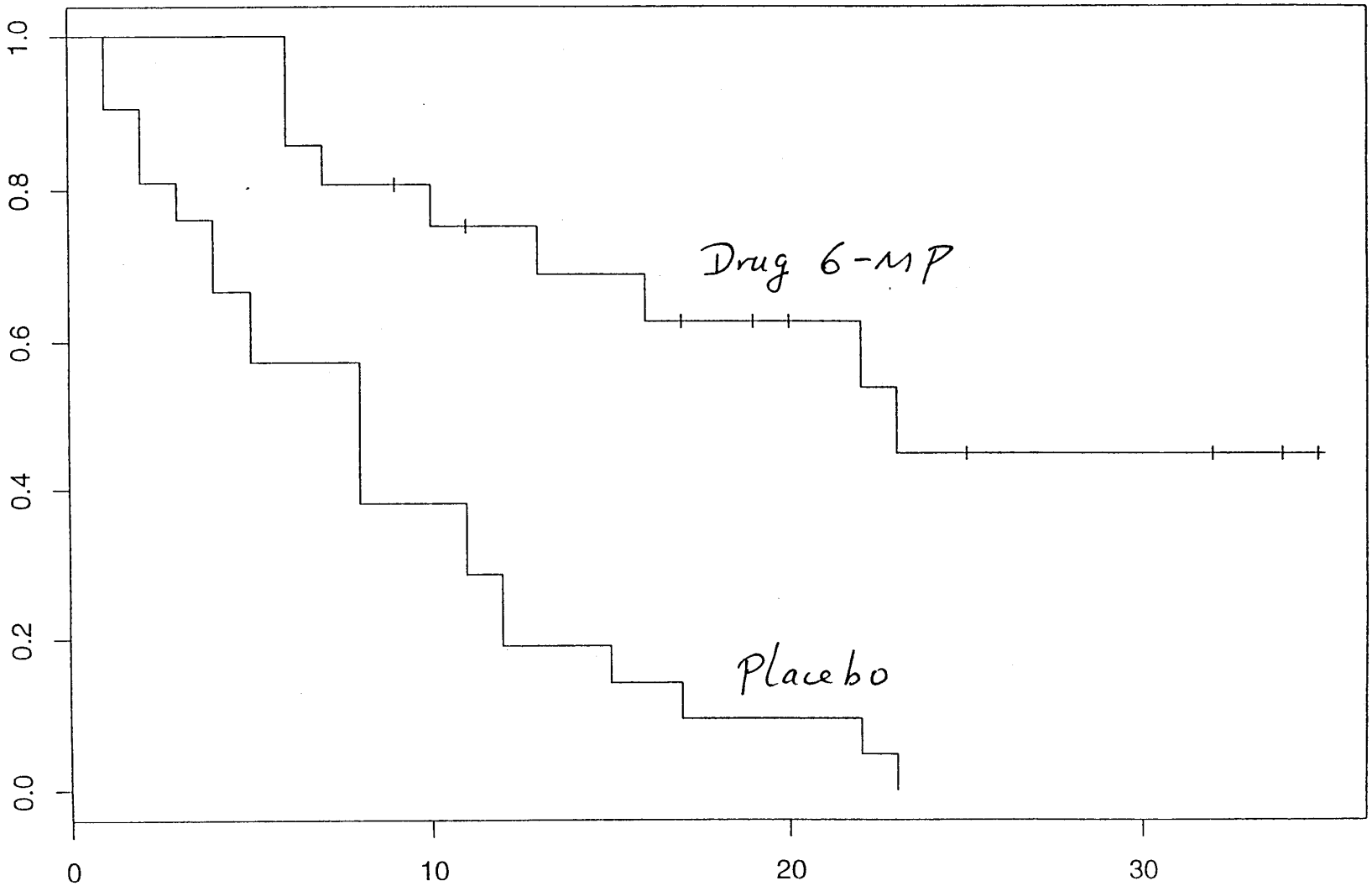


Fig. 7 95% Confidence Intervals for Placebo (Untransformed)

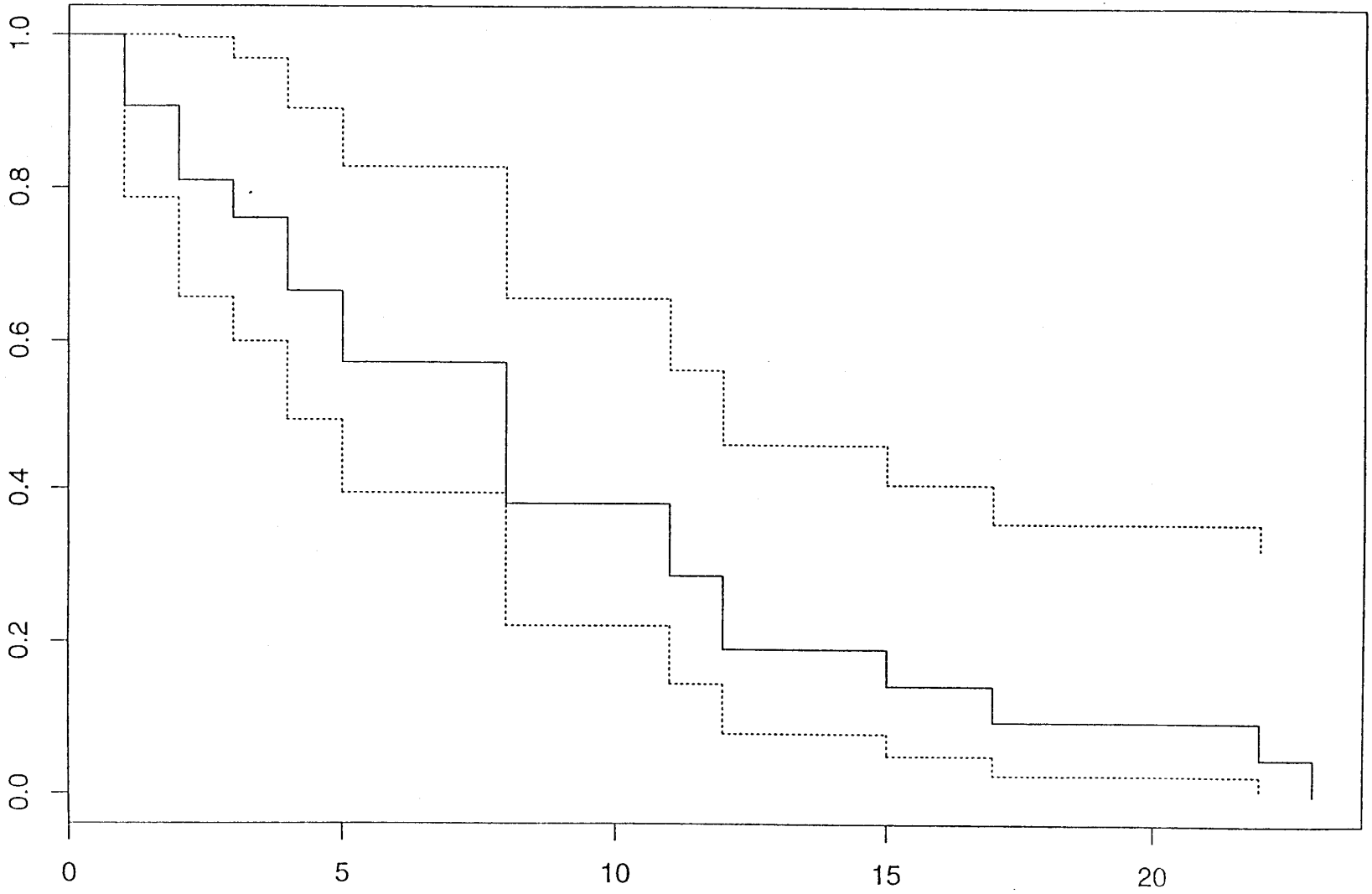
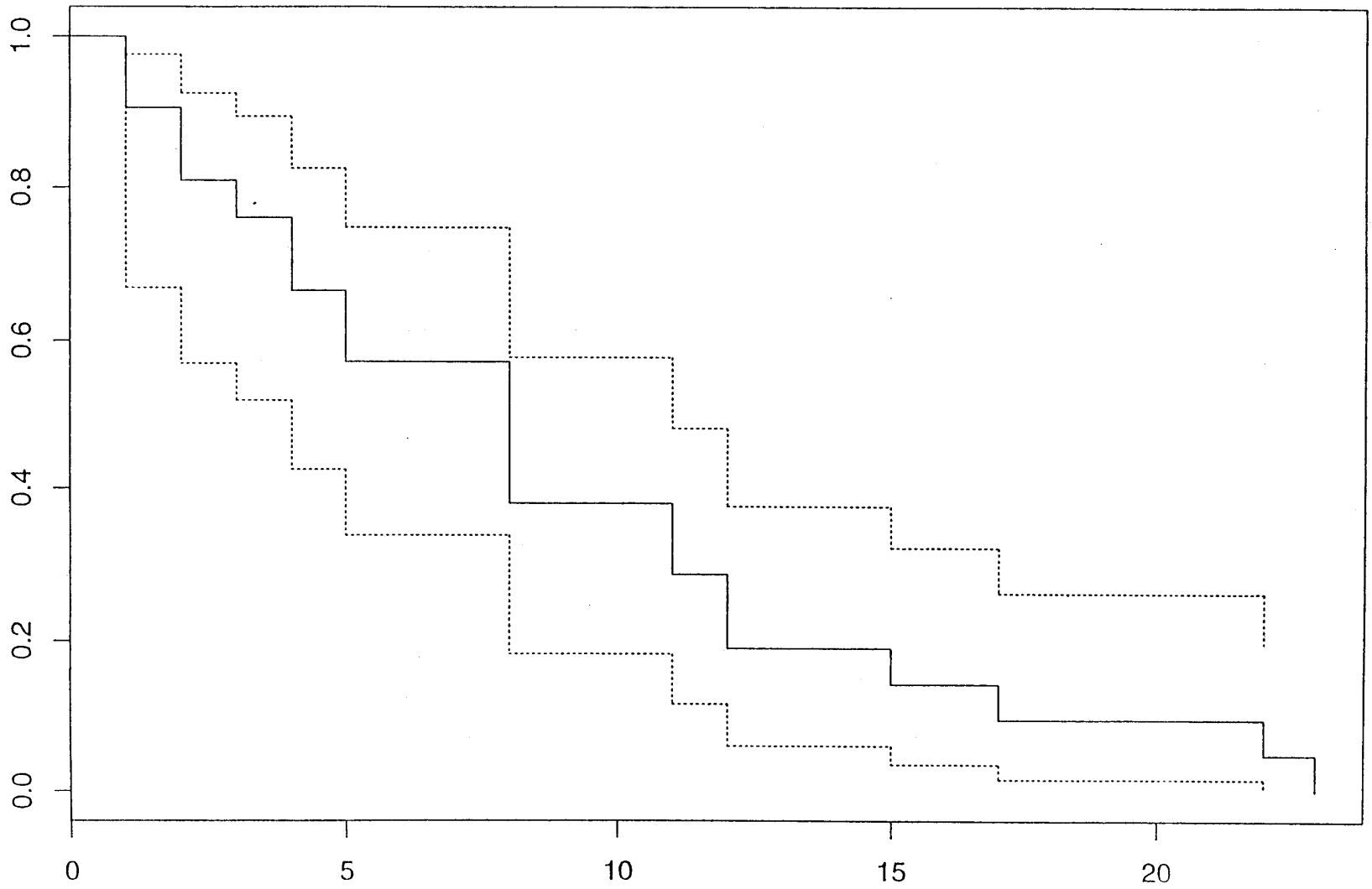


Fig. 8 95% Confidence Intervals for Placebo (Log-Log transformation)



IV. WEIGHTED LOG-RANK STATISTICS

1. Definition

$$U_w = \sum_{i=1}^{n} W(X_i)\Delta_i \left\{Z_i - \overline{Z}(X_i)\right\}$$

$$= \sum_{i=1}^{n} \int_0^{\infty} W(t) \left\{Z_i - \overline{Z}(t)\right\} dN_i(t)$$

- $W(t)$ assings weights according to timing of events
- $W(t) \xrightarrow{P} w(t)$

Statistic	$W(t)$
log-rank	$1 \quad (G^0)$
Prentice-Wilcoxon	$\widehat{S}(t) \quad (G^1)$
Harrington-Fleming G^ρ	$\widehat{S}^\rho(t)$
Gehan-Wilcoxon	$\overline{Y}(t)/n$
Tarone-Ware	$\{\overline{Y}(t)/n\}^\rho$

2. Null Distribution

$$H_0 : \lambda_1(t) = \lambda_0(t) \quad \text{for all } t$$

$$U_w = \sum_{i=1}^{n} \int_0^{\infty} W(t) \left\{ Z_i - \overline{Z}(t) \right\} dM_i(t)$$

$$E(U_w) = 0$$

$$\mathrm{Var}(U_w) = E \left\{ \sum_{i=1}^{n} \int_0^{\infty} W^2(t) \left\{ Z_i - \overline{Z}(t) \right\}^2 Y_i(t) d\Lambda_0(t) \right\}$$

$$V_w = \sum_{i=1}^{n} \int_0^{\infty} W^2(t) \overline{Z}(t) \left\{ 1 - \overline{Z}(t) \right\} dN_i(t)$$

$$= \sum_{i=1}^{n} W^2(X_i) \Delta_i \overline{Z}(X_i) \left\{ 1 - \overline{Z}(X_i) \right\}$$

$$U_w \sim N(0, V_w)$$

$$Q_w \equiv U_w / V_w^{1/2} \sim N(0,1) \Rightarrow Q_w^2 \sim \chi_1^2$$

Special case: $W = 1$

$V = \sum_{i=1}^{n} \Delta_i \mathrm{Var}(Z)$ (equal censoring)

$= D/4$ (equal allocation: $n_1 = n_0 = n/2$)

3. Software:

Data input:

time status group

S-Plus:

Survdiff (Surv(time, status) $\sim$ group, rho= ρ)

SAS:

```
proc lifetest;
        time time*status(0);
        test group;
```

4. Power

$$H_1 : \lambda_1(t) = \lambda_0(t)e^{\beta_0 w_0(t)}$$

$$U_w = \sum_{i=1}^{n} \int_0^{\infty} W(t) \left\{ Z_i - \overline{Z}(t) \right\} dN_i(t)$$

$$= \sum_{i=1}^{n} \int_0^{\infty} W(t) \left\{ Z_i - \overline{Z}(t) \right\} \left\{ dN_i(t) - Y_i(t) e^{\beta_0 w_0(t) Z_i} \lambda_0(t) dt \right\}$$

$$+ \sum_{i=1}^{n} \int_0^{\infty} W(t) \left\{ Z_i - \overline{Z}(t) \right\} Y_i(t) e^{\beta_0 w_0(t) Z_i} \lambda_0(t) dt$$

$$= \sum_{i=1}^{n} \int_0^{\infty} W(t) \left\{ Z_i - \overline{Z}(t) \right\} dM_i(t) + \mu_w$$

$$\mu_w \approx \beta_0 \sum_{i=1}^{n} \int_0^{\infty} W(t) w_0(t) \left\{ Z_i - \overline{Z}(t) \right\} Y_i(t) Z_i \lambda_0(t) dt$$

$$\approx \beta_0 \sum_{i=1}^{n} \int_0^{\infty} W(t) w_0(t) \overline{Z}(t) \left\{ 1 - \overline{Z}(t) \right\} dN_i(t)$$

$$\mu_w \approx \beta_0 \sum_{i=1}^{n} \int_0^{\infty} W(t) w_0(t) \overline{Z}(t) \left\{1 - \overline{Z}(t)\right\} dN_i(t)$$

$$\mathrm{Var}(U_w) \approx V_w = \sum_{i=1}^{n} \int_0^{\infty} W^2(t) \overline{Z}(t) \left\{1 - \overline{Z}(t)\right\} dN_i(t)$$

$$U_w \sim N(\mu_w, V_w)$$

$$Q_w \sim N(\mu_w / V_w^{1/2}, 1)$$

$w(t) \propto w_0(t) \Rightarrow \mu_w / V_w^{1/2}$ is maximized at $\beta_0 V_w^{1/2}$

$$\begin{aligned}
\text{Power} &= \Pr_{H_1}\left(Q_w > z_{1-\alpha/2}\right) \\
&= \Pr_{H_1}\left(Q_w - \mu_w / V_w^{1/2} > z_{1-\alpha/2} - \mu_w / V_w^{1/2}\right) \\
&= 1 - \Phi\left(z_{1-\alpha/2} - \mu_w / V_w^{1/2}\right)
\end{aligned}$$

5. Sample Size Determination

General: $w(t) \propto w_0(t)$

$$1 - \beta = 1 - \Phi\left(z_{1-\alpha/2} - \beta_0 V_w^{1/2}\right)$$

$$V_w = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\beta_0^2}$$

- V_w depends on n, and distributions of entry times, failure times & censoring times

Special case: $w = w_0 = 1$, equal censoring and equal allocation

$$D = \frac{4\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\beta_0^2}$$

- for one-sided test, replace $z_{1-\alpha/2}$ by $z_{1-\alpha}$
- convert D to n by specifying distributions of entry times, failure times & censoring times.

V. COX REGRESSION

A. Basic Analysis

1. Model: $\lambda(t|Z) = \lambda_0(t)e^{\beta_0' Z(t)}$

2. Data: $\{X_i, \Delta_i, Z_i(t)\} \quad (i = 1, \ldots, n)$

$$\{N_i(t), Y_i(t), Z_i(t)\} \quad (i = 1, \ldots, n)$$

3. Notation:

- $S^{(0)}(\beta, t) = \sum_{i=1}^n Y_i(t)e^{\beta' Z_i(t)}$
- $S^{(1)}(\beta, t) = \sum_{i=1}^n Y_i(t)e^{\beta' Z_i(t)} Z_i(t)$
- $S^{(2)}(\beta, t) = \sum_{i=1}^n Y_i(t)e^{\beta' Z_i(t)} Z_i(t)^{\otimes 2}$
- $a^{\otimes 2} = aa'$
- $E(\beta, t) = \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}$, $V(\beta, t) = \frac{S^{(2)}(\beta,t)}{S^{(0)}(\beta,t)} - E(\beta, t)^{\otimes 2}$

4. Inferences

Partial likelihood:

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{e^{\beta' Z_i(X_i)}}{S^{(0)}(\beta, X_i)} \right\}^{\Delta_i}$$

Score function:

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta}$$

$$= \sum_{i=1}^{n} \Delta_i \left\{ Z_i(X_i) - \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right\}$$

$$= \sum_{i=1}^{n} \int_0^{\infty} \{Z_i(t) - E(\beta, t)\} \, dN_i(t)$$

Information matrix:

$$\mathcal{I}(\beta) = -\frac{\partial^2 \log L(\beta)}{\partial \beta^2} = \sum_{i=1}^{n} \Delta_i V(\beta, X_i)$$

MPLE $\widehat{\beta}$: $\{U(\beta) = 0\}$

- $\mathcal{I}(\beta)$ is positive semi-definite
- obtain $\widehat{\beta}$ by Newton-Raphson algorithm

Breslow estimator:

$$\Lambda_0(t) = \int_0^t \lambda_0(u)du$$

$$\widehat{\Lambda}_0(t) = \sum_{i:X_i \le t} \frac{\Delta_i}{\sum_{j=1}^n Y_j(X_i) e^{\widehat{\beta}' Z_j(X_i)}}$$

$$= \sum_{i=1}^n \int_0^t \frac{dN_i(u)}{S^{(0)}(\widehat{\beta}, u)}$$

$$S_0(t) \equiv \Pr(T > t | Z = 0) = e^{-\Lambda_0(t)}$$

$$\widehat{S}_0(t) = e^{-\widehat{\Lambda}_0(t)}$$

- $Z_i - z_0 \Rightarrow \lambda_0$ and S_0 pertain to z_0

Asymptotic Properties:

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) e^{\beta_0' Z_i(u)} \lambda_0(u) du$$

$$U(\beta_0) = \sum_{i=1}^{n} \int_0^{\infty} \{Z_i(t) - E(\beta_0, t)\} \, dM_i(t)$$

$$U(\beta_0) \sim N\left(0, \mathcal{I}(\beta_0)\right)$$

$$\widehat{\beta} \sim N\left(\beta_0, \mathcal{I}^{-1}(\widehat{\beta})\right)$$

$$U'(\beta_0)\mathcal{I}^{-1}(\beta_0)U(\beta_0) \sim \chi^2_p$$

- Likelihood-type methods are used to make inferences about β_0.
- If Z is a dichotomous scalar, then $U(0)$ reduces to the two-sample log-rank statistic.

$$\widehat{\Lambda}_0(t) \sim N\left(\Lambda_0(t), V(t)\right)$$

$$\widehat{S}_0(t) \sim N\left(S_0(t), \widehat{S}_0^2(t)V(t)\right)$$

- When no covariates are involved, $\widehat{\Lambda}_0(t)$ reduces to Nelson-Aalen estimator and $\widehat{S}_0(t)$ is asymptotically equivalent to Kaplan-Meier estimator. When covariates are involved, $V(t)$ involves extra variation due to estimation of β_0.
- Construction of confidence intervals for $S_0(t)$ is similar to the one-sample case.
- It is possible to obtain simultaneous confidence bands (Lin, Fleming & Wei, 1994).

5. PBC Study

PBC: fatal chronic liver disease; no effective drugs; liver transplantation for advanced stage

Mayo Clinic database:

424	patients, $1/74 \sim 5/84$
312	randomized to DPCA vs placebo
106	followed off study
6	loss to followup early

Date of data listings: 7/86

Numbers of deaths: 125/312, 36/106

Covariates: 14 demographic, clinical, biochemical/histologic parameters.

Source: Appendix D.1 of Fleming & Harrington (1991)

Effect of DPCA on survival:

Cox model: $\lambda(t|Z) = \lambda_0(t)e^{\beta Z}$

$$Z = \begin{cases} 1 & \text{DPCA} \\ 0 & \text{placebo} \end{cases}$$

$U(0) = 1.7811,\ \mathcal{I}(0) = 31.198,\ l(0) = -639.9799$

$\widehat{\beta} = 0.0571,\ \mathcal{I}(\widehat{\beta}) = 31.153,\ l(\widehat{\beta}) = -639.9290$

Testing $H_0 : \beta = 0$

Wald: $\widehat{\beta}^2\mathcal{I}(\widehat{\beta}) = 0.10166$

Score (log-rank): $U^2(0)/\mathcal{I}(0) = 0.10168$

Likelihood ratio (LR): $2\{l(\widehat{\beta}) - l(\beta_0)\} = 0.10193$

Estimation of the hazard ratio: $\theta \equiv \lambda_1(t)/\lambda_0(t) = e^{\beta}$

Point estimate: $\widehat{\theta} = e^{\widehat{\beta}} = 0.944$

95% confidence interval: $e^{\widehat{\beta} \pm 1.96\mathcal{I}^{-1/2}(\widehat{\beta})} = (0.665, 1.342)$

Natural history model:

Usefulness:

- counseling patients
- understanding the course of PBC in untreated patients
- providing historical control information to evaluate new intervention

Model building:

1. Inexpensive, non-invasive and readily available measurements

$\Rightarrow$ elimination of stage, urine copper and SGOT

2. Step-down elimination ($p > 0.01$ using Wald statistics)
   - LR for 5 eliminated variables $= 2(554.237 - 550.603) = 7.268$ (χ^2_5)
3. Clinical and empirical evidence for transformations

- adding variables log(age), log(albumin), log(bilirubin) and log(protime) $\Rightarrow$ significantly better fit: LR = 31.926
- log(bilirubin) $\Rightarrow$ substantial improvement, hepatomegaly nonsig
- stepwise considerations of logarithmic and squared transformations for albumin, age and protime $\Rightarrow$ model with age, log(albumin), log(bilirubin), edema and log(protime)

Parameter	Est.	S.E.	Est./S.E.
Age	0.039	0.008	5.15
log (Albumin)	–2.533	0.648	–3.91
log (Bilirubin)	0.871	0.083	10.54
Edema	0.859	0.271	3.17
log (Protime)	2.380	0.767	3.10

Risk score (Mayo R-score)

$$\widehat{R} = \widehat{\beta}_1 Z_1 + ... + \widehat{\beta}_5 Z_5$$

$$\widehat{S}(t|Z) = \{e^{-\widehat{\Lambda}_0(t)}\}^{e^{\widehat{R}}}$$

Median risk score: $\widehat{R} = 5.24 \Rightarrow \widehat{S}(1) = .982, \widehat{S}(5) = .845$

Low-risk patient: bilirubin=0.5, albumin=4.5, age=52, protime=10.1, edema=0

$$\Rightarrow \widehat{R} = 3.49 \Rightarrow \widehat{S}(5) = 0.97 \Rightarrow \text{low risk of death}$$

High-risk patient: bilirubin=13.9, albumin=2.8, age=52, protime=13.8, edema=.5

$$\Rightarrow \widehat{R} = 9.19 \Rightarrow \widehat{S}(1) = 0.39 \Rightarrow \text{candidate for transplant}$$

6. Software:

Data input:

time status covariates

S-Plus:

coxph (Surv(time, status) ~ covariates, method='breslow')

SAS:

```
proc phreg;
        model time*status(0)=covariates;
```

B. Stratified Analysis

Setup: K strata, n_k $(k = 1, \ldots, K)$ subjects in kth stratum

$T_{ki} = i$th failure time of kth stratum

$C_{ki} = i$th censoring time of kth stratum

Observation time: $X_{ki} = \min(T_{ki}, C_{ki})$

Failure indicator: $\Delta_{ki} = I(T_{ki} \leq C_{ki})$

Covariates: $Z_{ki}(t) = \{Z_{1ki}(t), \ldots, Z_{pki}(t)\}'$

Model(s):

$$\lambda_k(t|Z_{ki}) = \lambda_{k0}(t)e^{\beta' Z_{ki}(t)}, \quad i = 1, \ldots, n_k; \; k = 1, \ldots, K$$

- strata-specific baseline hazard functions
- common vs. strata-specific parameters

Notation:

At-risk indicator: $Y_{ki}(t) = I(X_{ki} \geq t)$

$S_k^{(0)}(\beta, t) = \sum_{i=1}^{n_k} Y_{ki}(t)e^{\beta' Z_{ki}(t)}$

$S_k^{(1)}(\beta, t) = \sum_{i=1}^{n_k} Y_{ki}(t)e^{\beta' Z_{ki}(t)} Z_{ki}(t)$

$S_k^{(2)}(\beta, t) = \sum_{i=1}^{n_k} Y_{ki}(t)e^{\beta' Z_{ki}(t)} Z_{ki}(t)^{\otimes 2}$

Inferences:

$$L(\beta) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left\{ \frac{e^{\beta' Z_{ki}(X_{ki})}}{S_k^{(0)}(\beta, X_{ki})} \right\}^{\Delta_{ki}}$$

$$U(\beta) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \Delta_{ki} \left\{ Z_{ki}(X_{ki}) - \frac{S_k^{(1)}(\beta, X_{ki})}{S_k^{(0)}(\beta, X_{ki})} \right\}$$

$$\mathcal{I}(\beta) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \Delta_{ki} \left\{ \frac{S_k^{(2)}(\beta, X_{ki})}{S_k^{(0)}(\beta, X_{ki})} - \frac{S_k^{(1)}(\beta, X_{ki})^{\otimes 2}}{S_k^{(0)}(\beta, X_{ki})^2} \right\}$$

- Asymptotic properties for $U(\beta)$ and $\widehat{\beta}$ are the same as the unstratified case.
- Inferences can be implemented by including the *strata* option in *coxph* or the *strata* statement in *phreg*.

C. Model Misspecification

1. Model Assumptions:

- proportional hazards
- functional forms of covariates
- exponential regression function

2. Properties of $\widehat{\beta}$ for Misspecified Models

$$S^{(0)}(t) = \sum_{i=1}^{n} Y_i(t)\lambda_i(t)$$

$$S^{(1)}(t) = \sum_{i=1}^{n} Y_i(t)\lambda_i(t)Z_i(t)$$

$$s^{(r)}(t) = E\left\{n^{-1}S^{(r)}(t)\right\}, \quad s^{(r)}(\beta, t) = E\left\{n^{-1}S^{(r)}(\beta, t)\right\}$$

$$n^{-1}U(\beta) \xrightarrow{P} \int_0^\infty \left\{ \frac{s^{(1)}(t)}{s^{(0)}(t)} - \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)} \right\} s^{(0)}(t)dt \equiv h(\beta)$$

$$\beta^* : \{h(\beta) = 0\}$$

$$\widehat{\beta} \sim N\left(\beta^*, D\right)$$

$$D = \mathcal{I}^{-1}\widehat{B}\mathcal{I}^{-1}$$

(Sandwich robust variance estimator)

$$\widehat{B} = \sum_{i=1}^{n} W_i^{\otimes 2}$$

$$W_i = \int_0^\infty \left\{ Z_i(t) - E(\widehat{\beta}, t) \right\} \left\{ dN_i(t) - Y_i(t) e^{\widehat{\beta}' Z_i(t)} d\widehat{\Lambda}_0(t) \right\}$$

- D is always valid whereas $\mathcal{I}^{-1}$ may not be (Lin & Wei, 1989)
- D may be obtained by including the option "robust=T" in *coxph* or "covs" in *phreg*

Data input:

time status covariates

SAS procedure:

proc phreg covs;

model time*status(0)=covariates/ties=breslow;

strata vstrata;

- covs or covsandwich requests robust sandwich variance-covariance estimate
- strata requests stratification on variable vstrata

S-Plus function:

coxph(Surv(time, status) ~ covariates+strata(vstata),robust='T',

method='breslow')

- robust='T' requests the robust variance-covariance estimate
- strata requests stratification on variable vstrata

3. Consequences of Model Misspecification in Treatment Comparisons

- $\lambda_1(t)$ and $\lambda_0(t)$ nonproportional, but no overlapping $\Rightarrow$ loss of power, β^* =average of $\log\{\lambda_1(t)/\lambda_0(t)\}$ over t
- $\lambda_1(t)$ and $\lambda_0(t)$ cross $\Rightarrow$ poor power, β^* meaningless
- $\lambda(t|Z_1, Z_2) = \lambda_0(t)e^{\beta_1 Z_1 + \beta_2 Z_2}$, $Z_1||Z_2$:
  Omitting $Z_2 \Rightarrow \lambda_1(t)$ and $\lambda_0(t)$ are nonproportional,
  $|\beta_1^*| < |\beta_1|$, and $\beta^* - \beta_1$ is small unless $|\beta_2|$ is large
- Omitting or mismodelling baseline prognostic factors has little effect on size of test for no treatment difference but may reduce power

D. Model Checking

1. Martingales

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) e^{\beta_0' Z_i(u)} \lambda_0(u) du$$

- $E\{M_i(t)\} = 0$
- $\text{Cov}\{M_i(t), M_j(t)\} = 0 \quad (i \neq j)$
- $M_i(t) =$ observed minus model-predicted numbers of events by time t on subject i

2. Martingale Residuals

$$\widehat{M}_i(t) = N_i(t) - \int_0^t Y_i(u) e^{\widehat{\beta}' Z_i(u)} d\widehat{\Lambda}_0(u)$$

- $\widehat{M}_i(t)$ are similar to residuals in linear and Poisson regression
- $\widehat{M}_i(t)$ = observed minus model-estimated numbers of events by time t on subject i
- $\sum_{i=1}^n \widehat{M}_i(t) = 0$
- $E\{\widehat{M}_i(t)\} \approx 0$
- $\text{Cov}\{\widehat{M}_i(t), \widehat{M}_j(t)\} \approx 0 \quad (i \neq j)$
- $\widehat{M}_i = \widehat{M}_i(\infty) = \widehat{M}_i(X_i)$

3. Model Checking Techniques

(1) Functional Forms of Covariates

Individual residuals:

$\widehat{M}_{i(j-)} = \widehat{M}_i$ when Z_j is excluded from the model

Scatterplot of $\widehat{M}_i$ vs. Z_{ji} is centered around 0 if the functional form for Z_j is correct.

A smoothed plot of $\widehat{M}_{i(j-)}$ vs. Z_{ji} suggests approximate functional form for Z_j.

- good approximation if Z_j has a weak effect on T and is independent of other covariates in the model.
- highly subjective

PBC data: Fig. 9

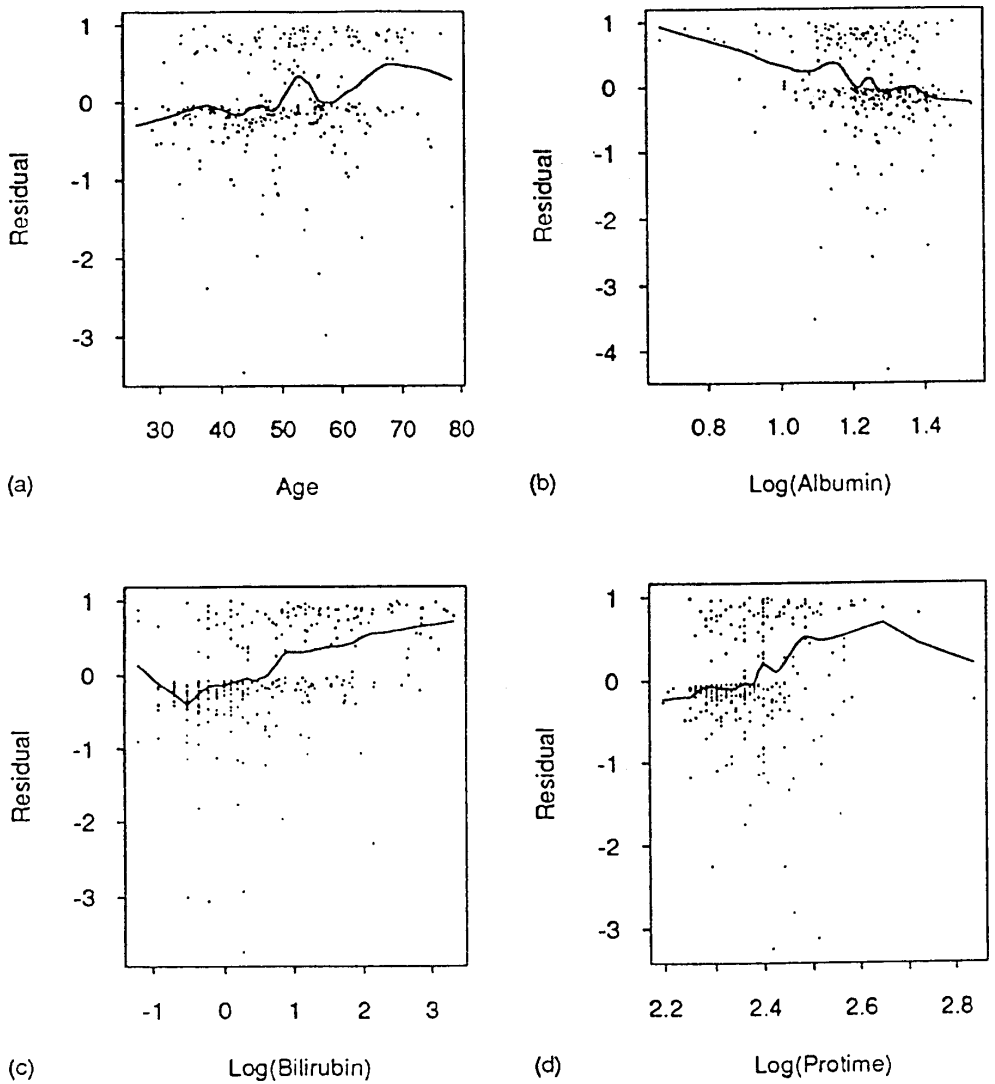


Figure 9 Martingale residuals in the PBC data. Residuals from a model with the covariate edema and three of the four continuous variables (age, log(albumin), log(bilirubin), and log(protime)) are plotted against the omitted variable. LOWESS smooths use a span of .2.

Cumulative sum of residuals:

- Take cumulative sum of $\widehat{M}_i$ over Z_j
- Approximate null distribution of cumulative-sum process by a zero-mean Gaussian process
- Compare observed process with simulated realizations from the approximate distribution

PBC data: Figs. 10-14

(2) Exponential Regression Function:

Plot cumulative sum of $\widehat{M}_i$ vs. $\widehat{\beta}' Z_i$

PBC data: Fig. 15

Fig. 10 Mayo PBC Model

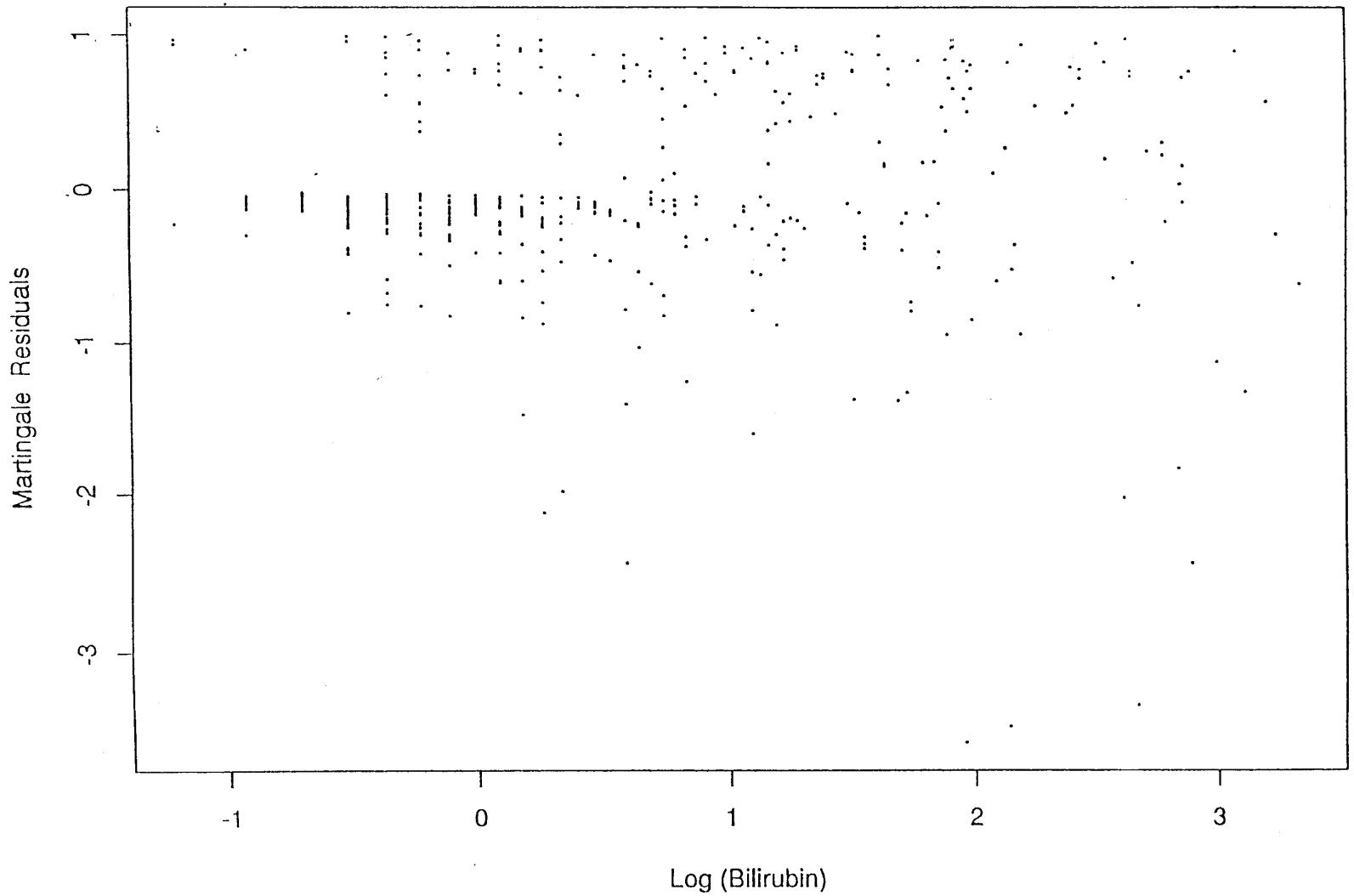


Fig. 11 Mayo PBC Model

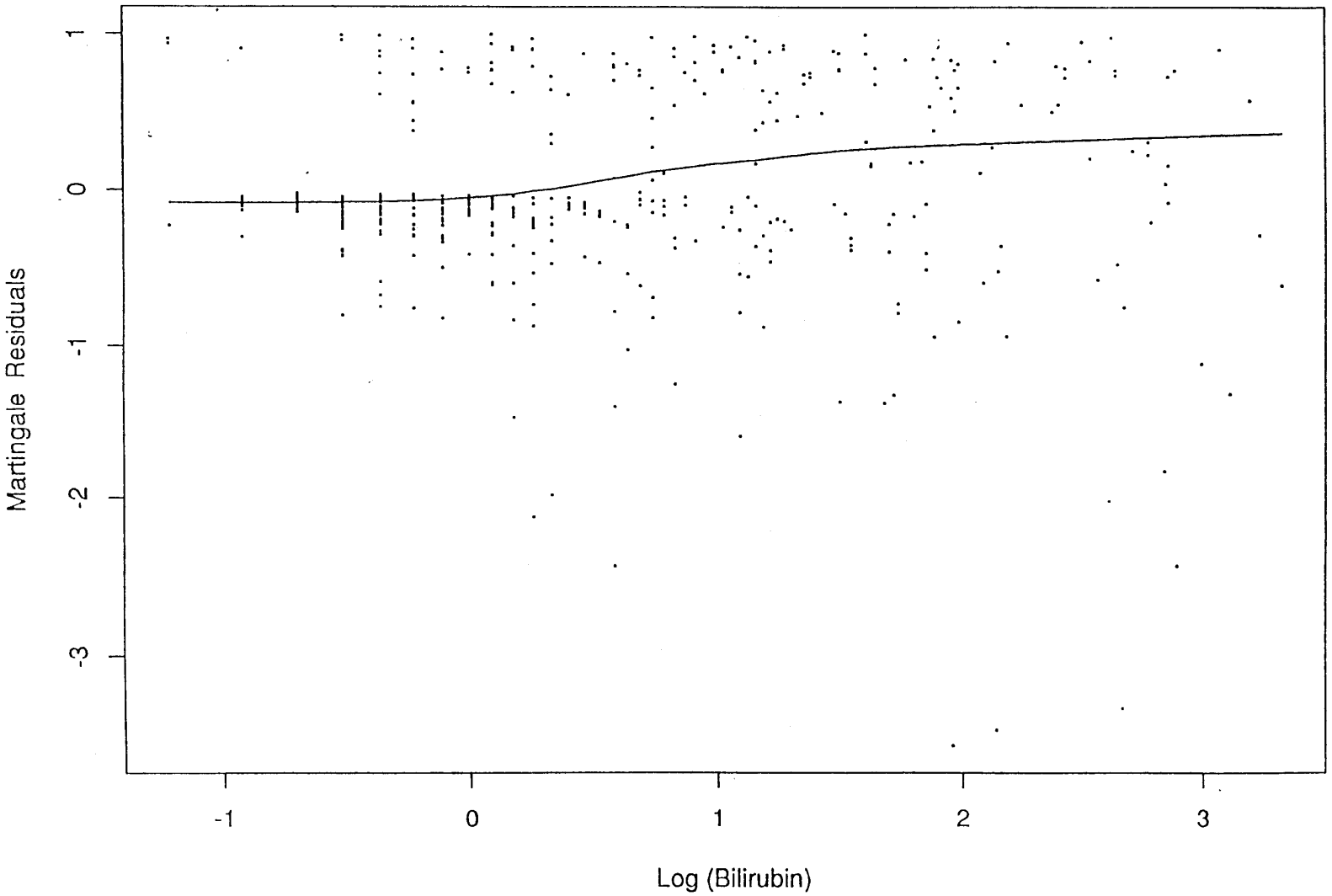


Fig.12 Mayo PBC Model

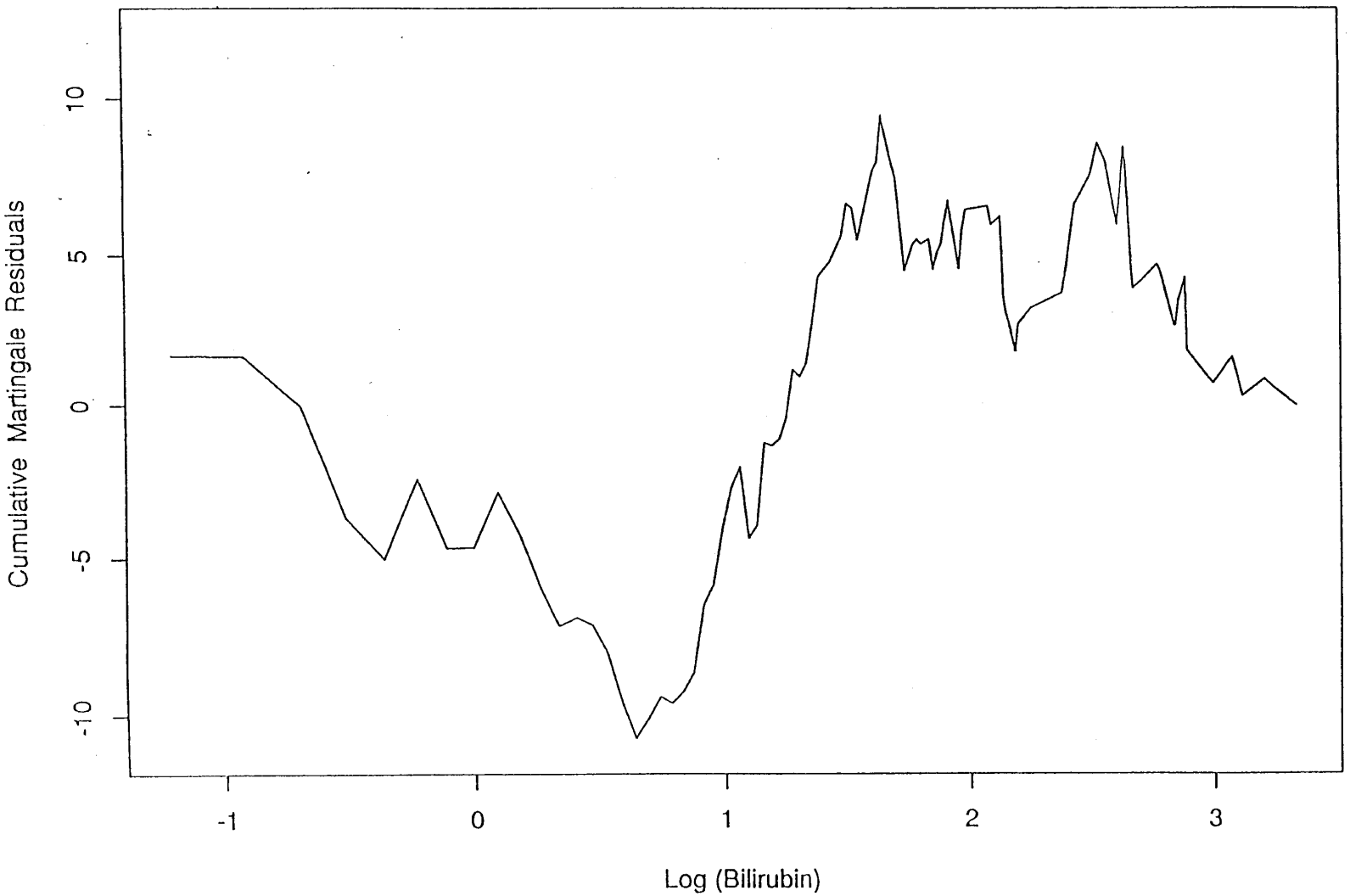


Fig. 13. Functional Form of Bilirubin

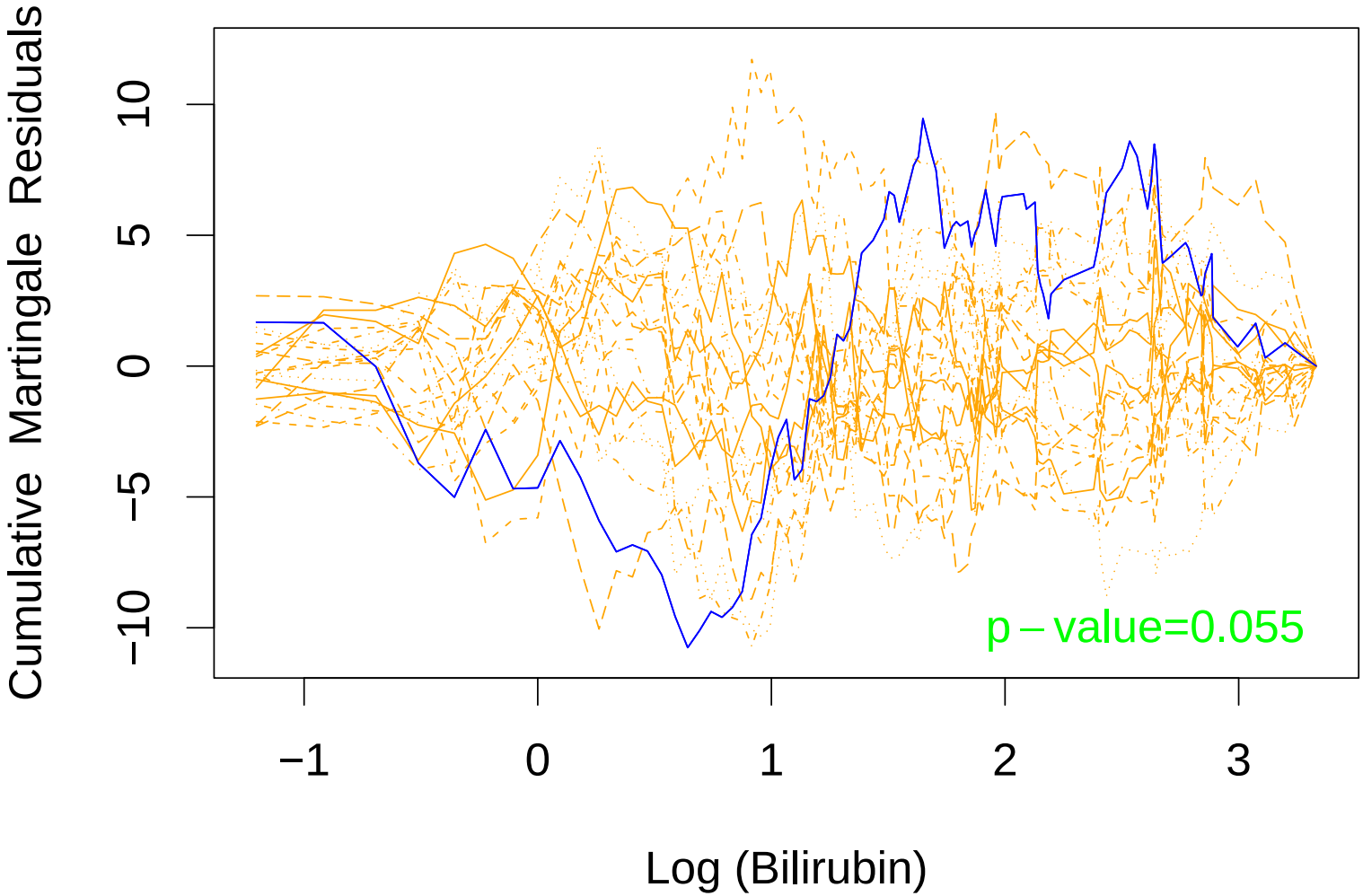


Fig. 14. Functional Form of Bilirubin

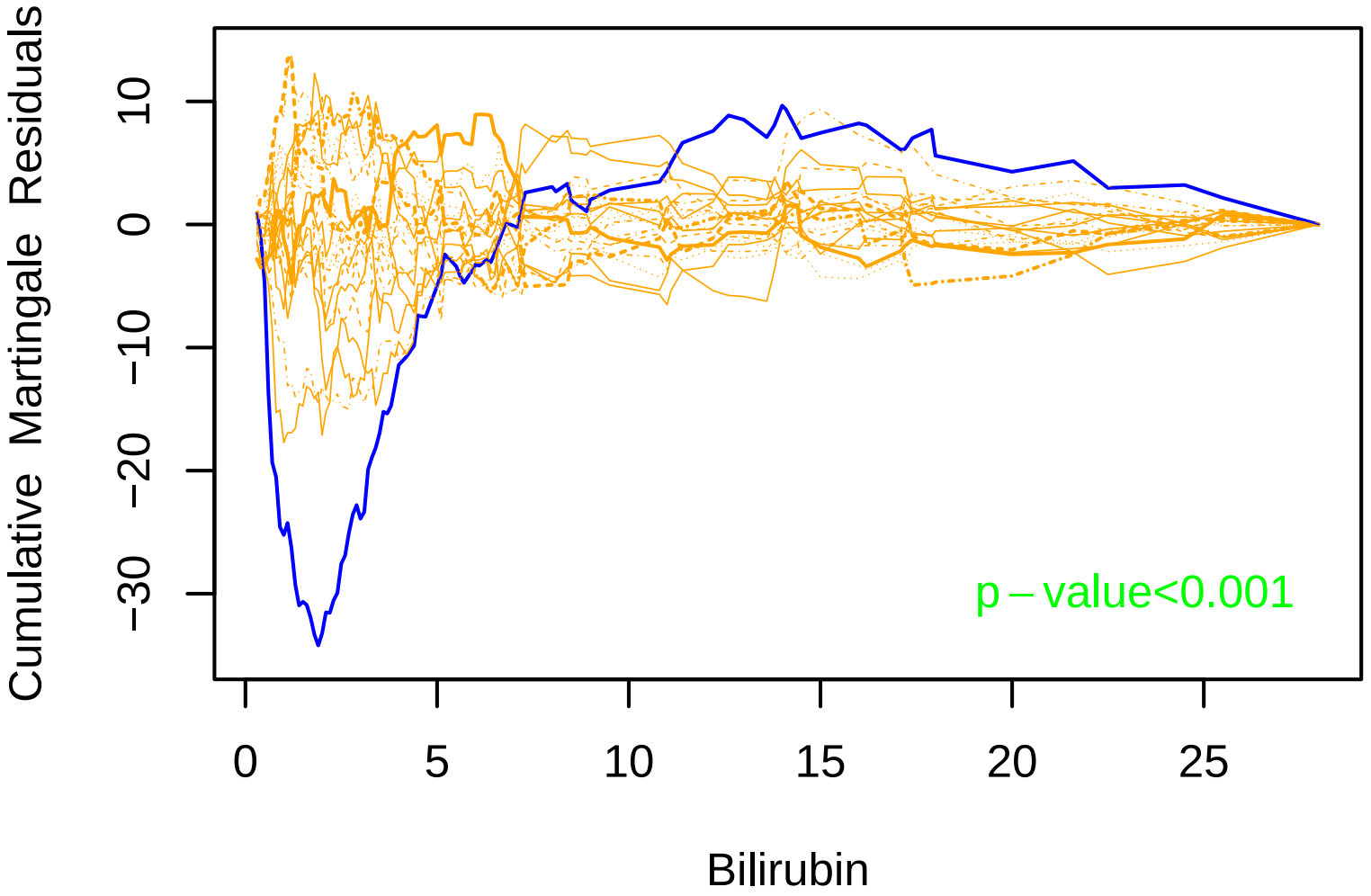
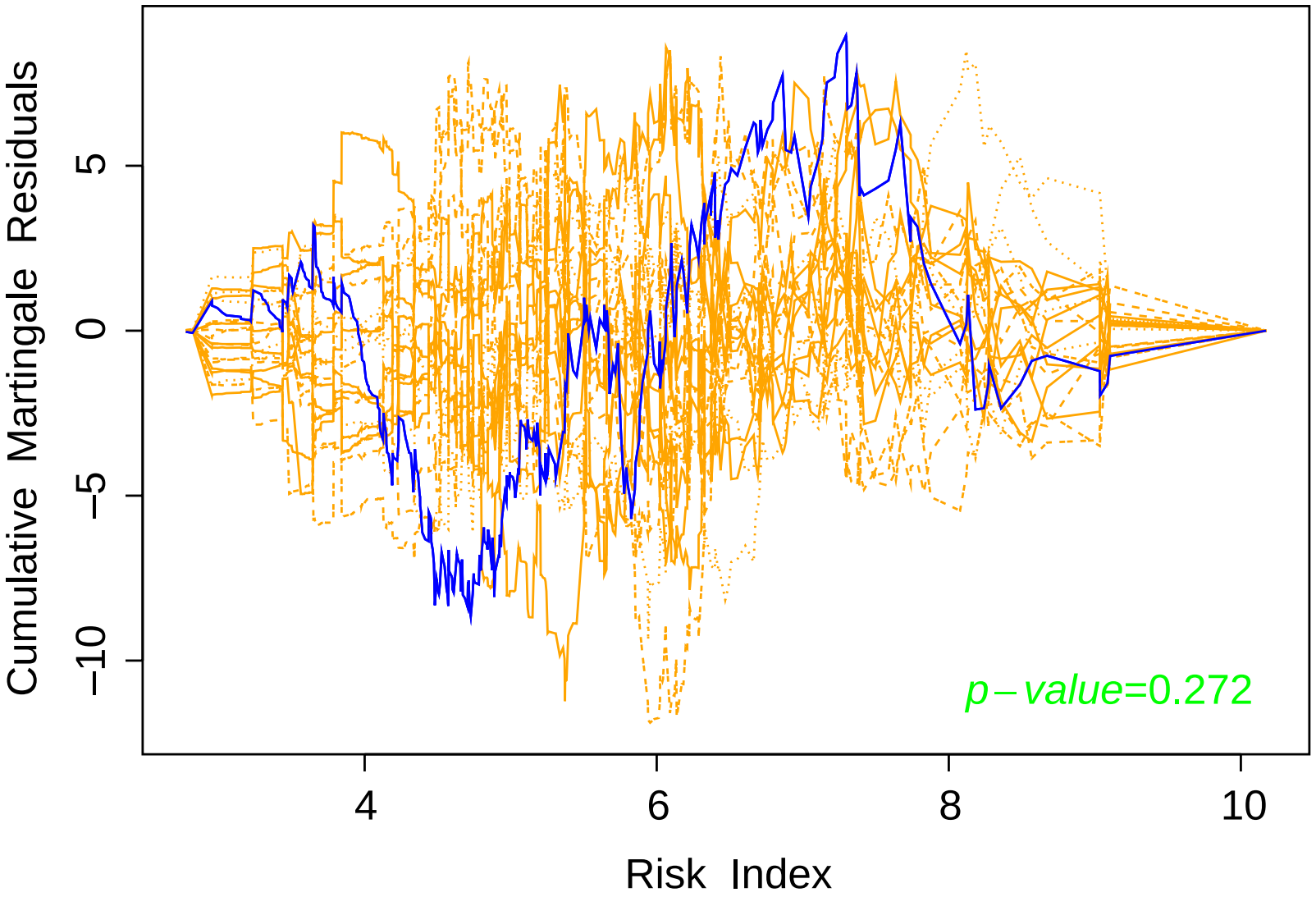


Fig 15. Exponential Regression Function



(3) Proportional Hazards:

$$U(\beta, t) = \sum_{i=1}^{n} \int_0^t \{Z_i(u) - E(\beta, u)\}\, dN_i(u)$$

$$U(\beta_0, t) = \sum_{i=1}^{n} \int_0^t \{Z_i(u) - E(\beta_0, u)\}\, dM_i(u)$$

$$U(\widehat{\beta}, t) = \sum_{i=1}^{n} \int_0^t Z_i(u) d\widehat{M}_i(u)$$

$p = 1 : \sup_{0 \le t < \infty} \mathcal{I}^{-1/2}(\widehat{\beta}) |U(\widehat{\beta}, t)| \sim \sup_{0 \le s \le 1} |B^0(s)|$

- B^0 = Brownian bridge
- critical value = 1.358 for $\alpha = 0.05$
- consistency

$p > 1$: simulate dist of $\sup_{0 \le t < \infty} \mathcal{I}^{-1/2}(\widehat{\beta})_{jj} |U_j(\widehat{\beta}, t)|$

PBC data: Figs. 16-17

Fig. 16. Proportional Hazards for log(Bilirubin)

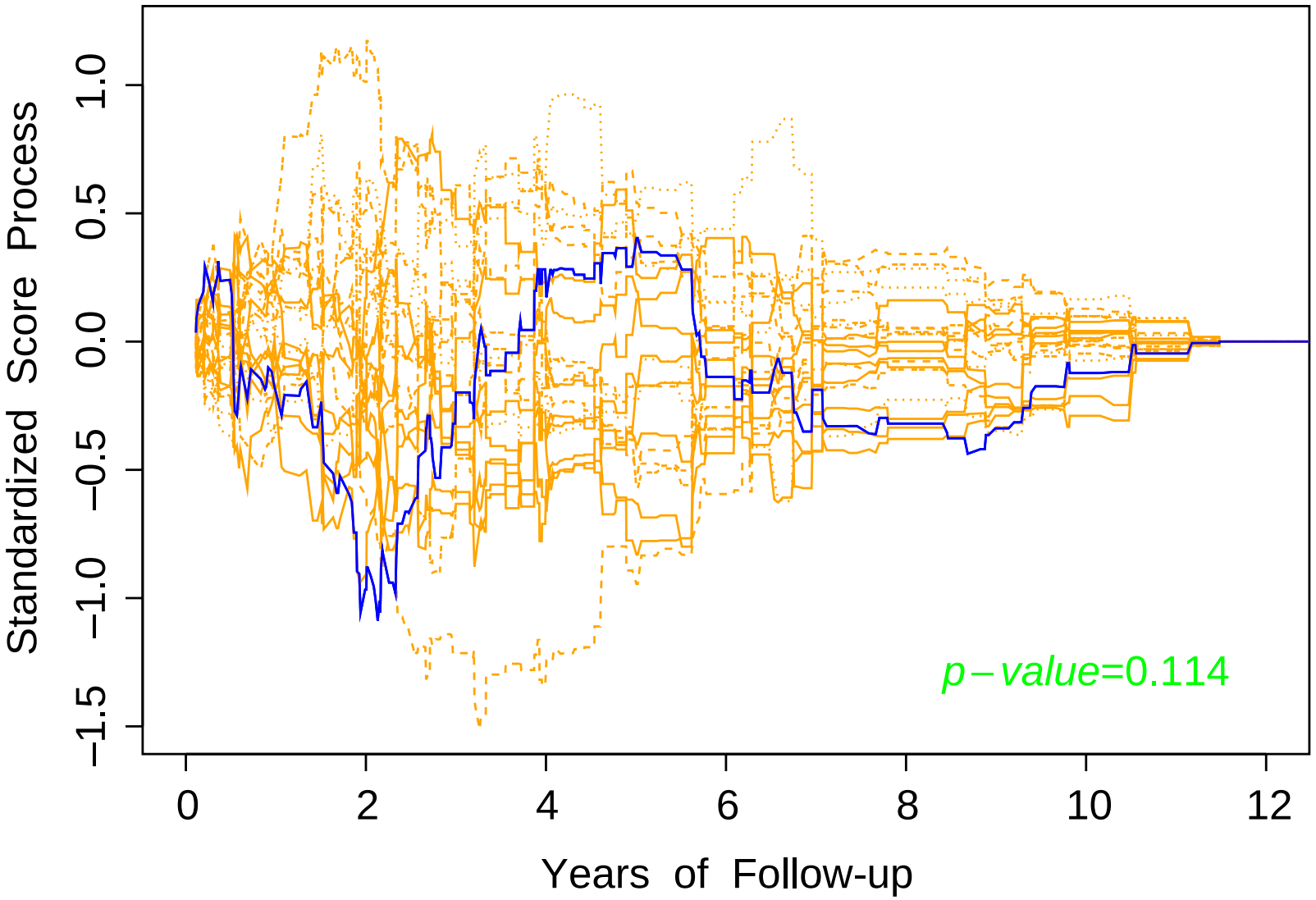
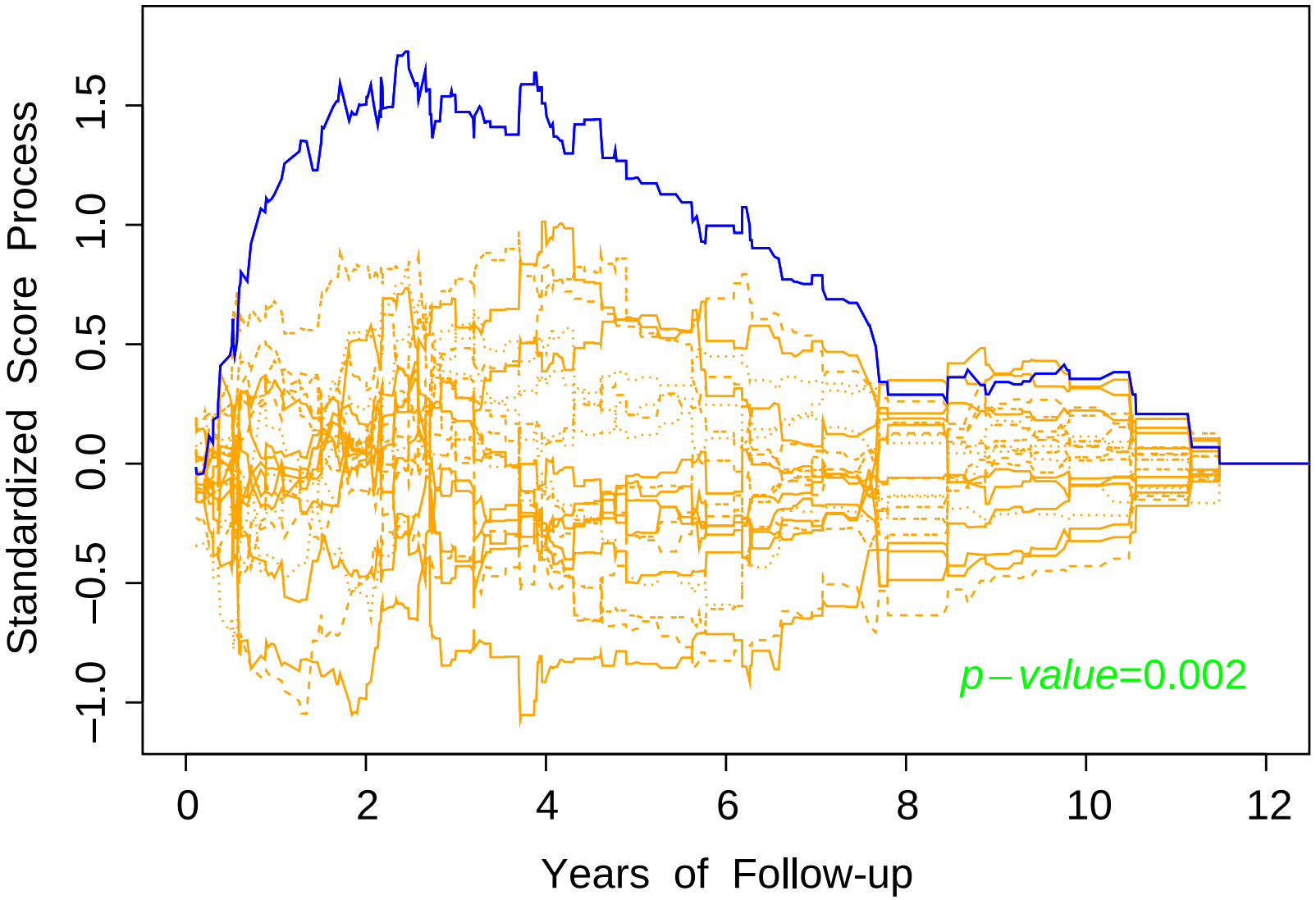


Fig 17. Proportional Hazards for Log(Protime)



ASSESS Statement

ASSESS < **VAR=***(list)* > < **PH** > < */options* > ;

The ASSESS statement performs the graphical and numerical methods of Lin, Wei, and Ying (1993) for checking the adequacy of the Cox regression model. The methods are derived from cumulative sums of martingale residuals over follow-up times or covariate values. You can assess the functional form of a covariate or you can check the proportional hazards assumption for each covariate in the Cox model. PROC PHREG uses ODS Graphics for the graphical displays. You must specify at least one of the following options to create an analysis.

VAR=(*variable-list***)**
specifies the *list* of explanatory *variables* for which their functional forms are assessed. For each variable on the list, the observed cumulative martingale residuals are plotted against the values of the explanatory variable along with 20 (or *n* if NPATHS=*n* is specified) simulated residual patterns.

PROPORTIONALHAZARDS
PH
requests the checking of the proportional hazards assumption. For each explanatory variable in the model, the observed score process component is plotted against the follow-up time along with 20 (or *n* if NPATHS=*n* is specified) simulated patterns.

The following options can be specified after a slash (/):

NPATHS=*n*
specifies the number of simulated residual patterns to be displayed in a cumulative martingale residual plot or a score process plot. The default is *n*=20.

CRPANEL
requests that a plot with four panels, each containing the observed cumulative martingale residuals and two simulated residual patterns, be created.

RESAMPLE < **=***n* >
requests that the Kolmogorov-type supremum test be computed on 1,000 simulated patterns or on *n* simulated patterns if *n* is specified.

SEED=*n*
specifies an integer seed for the random number generator used in creating simulated realizations for plots and for the Kolmogorov-type supremum tests. Specifying a seed enables you to reproduce identical graphs and *p*-values for the model assessments from the same PHREG specification. If the SEED= option is not specified, or if you specify a nonpositive seed, a random seed is derived from the time of day.

Output 64.12.3 Typical Cumulative Residual Plot Patterns

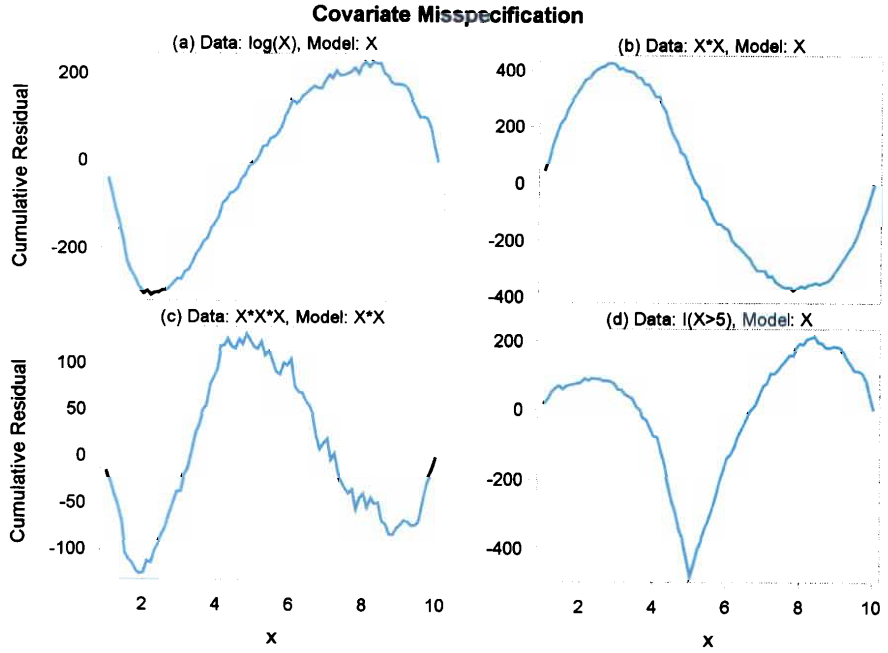


Table 64.9 Model Misspecifications

Plot	Data	Fitted Model
(a)	$\log(X)$	X
(b)	$\{X, X^2\}$	X
(c)	$\{X, X^2, X^3\}$	$\{X, X^2\}$
(d)	$I(X > 5)$	X

The curve of observed cumulative martingale residuals in Output 64.12.2 most resembles the behavior of the curve in plot (a) of Output 64.12.3, indicating that log(Bilirubin) might be a more appropriate term in the model than Bilirubin.

Log-log survival plot

$S(t|Z) = S_0(t)^{e^{\beta_0' Z}}$

$\Rightarrow \log\{-\log S(t|Z)\} = \beta_0' Z + \log\{-\log S_0(t)\}$

$\Rightarrow \log \Lambda(t|Z) = \beta_0' Z + \log \Lambda_0(t)$

$\Rightarrow \log \widehat{\Lambda}(t|Z_{(1)})$ and $\log \widehat{\Lambda}(t|Z_{(2)})$ are parallel

- need to discretize covariates
- $p = 1$: $\widehat{\Lambda}$ are stratum-specific Nelson-Aalen estimates
- $p > 1$: $\widehat{\Lambda}$ are stratum-specific Breslow estimates

PBC data: Fig. 18

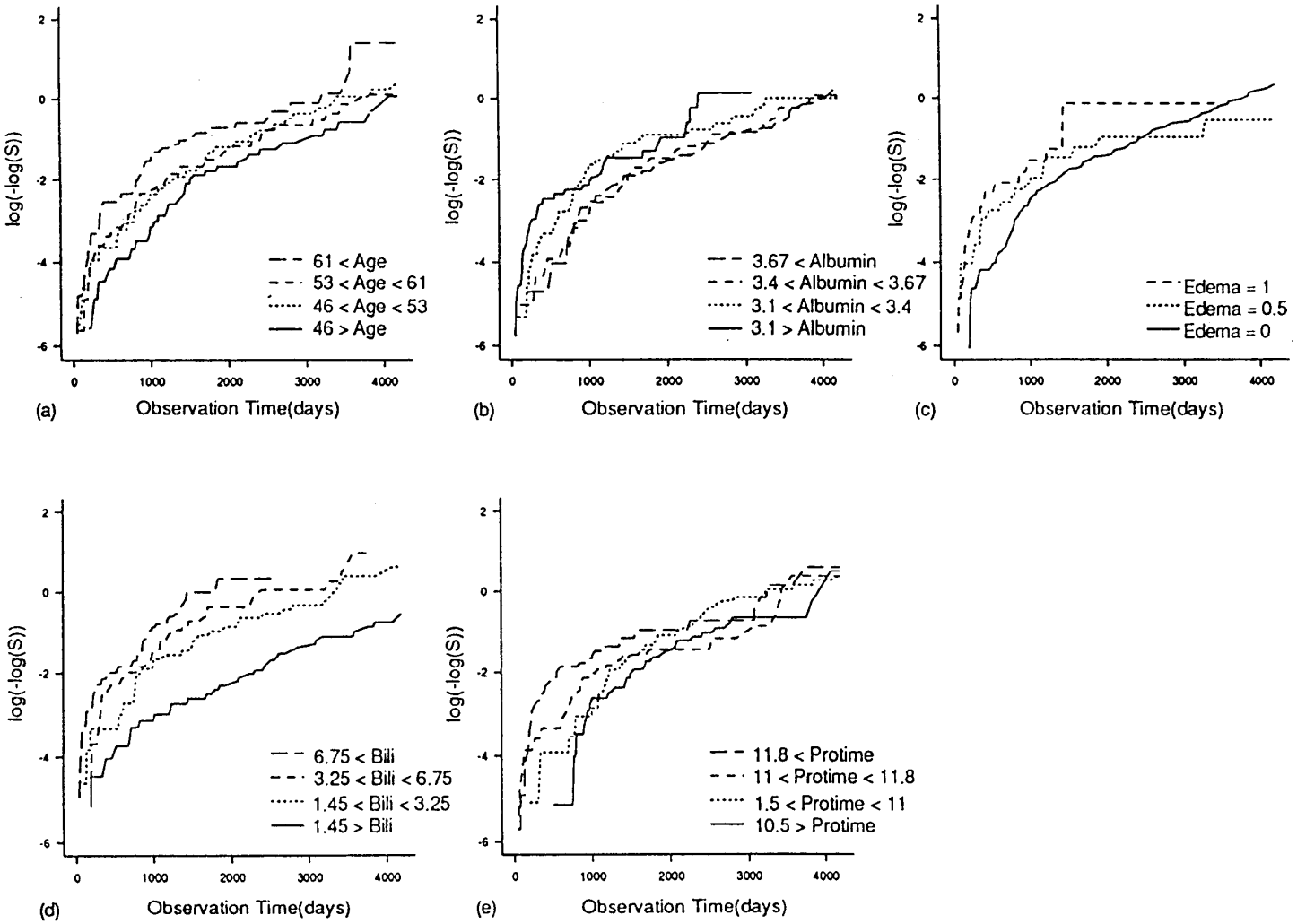


Figure 18 Logarithm of estimated baseline cumulative hazard functions in stratified models. Each model uses one of the five variables age, log(albumin), log(bilirubin), edema, and log(protime) as a stratification variable, and the other four variables as covariates. The strata are defined the lower right of each plot.

Dummy time-dependent covariates

$$\lambda(t|Z) = \lambda_0(t)e^{\beta_0' Z + \gamma_0 Z_j Q(t)}$$

- $Q(t) = t$ or $\log t$
- testing $H_0 : \gamma_0 = 0$
- sensitive to departures in the form of Q

Adjusting for non-proportional hazards

- dummy time-dependent covariates
- stratification
- robust variance
- transformation models (Zeng & Lin, 2007)

E. Covariate Adjustment in Randomized Trials

- Randomization permits valid testing of the null hypothesis of no treatment effect without recourse to any probability model and without regard to any (baseline) covariates that may or may not be measured.
- Weighted log-rank tests provide valid testing of no treatment effect regardless of whether or not there are chance imbalances in covariates.
- Randomization ensures that treatment assignment is independent of covariates. Consequently, the use of significance testing for covariate imbalance in determining which covariates to adjust is illogical.

- Covariate adjustment may yield more powerful test of treatment effect.
- The treatment effect conditional on covariates is not equal to the unconditional effect.
- Omitting or mismodelled covariates may yield biased estimate of treatment effect.
- Covariate adjustment may be used to relax the assumption of independent censoring.
- For testing no treatment effect, both unadjusted and adjusted tests may be used. However, the decision on which one is the primary test should be made beforehand and must not be based on minimization of the p-value. Covariate adjustment in estimating treatment effect with censored data requires care.

VI. SEQUENTIAL ANALYSIS

1. Introduction

Rationale: Most clinical trials recruit patients over a long period of time, and the data accumulate gradually during the course of the study. For ethical, scientific and economic reasons, it is desirable to monitor the accumulating data periodically so that the study can be terminated as soon as the superiority of one treatment is compelling or when it becomes evident that there is little or no difference between treatments.

Group sequential tests: Repeated significance tests based on data accumulated at interim analyses for possible early termination of trial.

Multiple looks effects: Repeated use of significance tests increases the overall Type I error probability.

Overall percentage of rejecting H_0 at the 5% nominal significance level after K repeated tests when there is no treatment difference

K	1	2	5	10	25	50	200
Error rate	5	8.3	14.2	19.3	26.6	32.0	42.4

2. Brownian Motion Statistics

Preliminaries:

$H_0 : \theta = 0$ vs. $H_1 : \theta = \theta_1$

$U(t) =$ "score" statistic calculated at analysis time t

$V(t) =$ amount of "information" at t

$U(t) \sim \mathcal{B}\left(0, V(t)\right)$ under H_0

$U(t) \sim \mathcal{B}\left(\theta_1 V(t), V(t)\right)$ under H_1

$\mathcal{B} =$ Brownian motion

$Q(t) = U(t)/V^{1/2}(t) \sim N(0,1)$ under H_0

Number of interim analyses: K

Times of interim analyses: $t_1 < t_2 < \ldots < t_K$

Test statistics: $Q_k = Q(t_k)$ $(k = 1, \ldots, K)$

Overall (two-sided) type I error: α

Boundary values: $c_1, \ldots, c_K$

$$\Pr_{H_0}\{|Q_k| \geq c_k \text{ for some } 1 \leq k \leq K\} = \alpha$$

Pocock boundary: $c_1 = c_2 = \ldots = c_K$

O'Brien-Fleming boundary: $c_k = (K/k)^{1/2} c_K$

- equal group size (increment of information)
- tabulated boundary values

Slud-Wei method:

Exit probabilities: $\pi_1 + \ldots + \pi_K = \alpha$

$$\Pr_{H_0}\{|Q_1| \geq c_1\} = \pi_1$$

$$\Pr_{H_0}\{|Q_1| < c_1, |Q_2| \geq c_2\} = \pi_2$$

$$\ldots$$

$$\Pr_{H_0}\{|Q_1| < c_1, \ldots, |Q_{K-1}| < c_{K-1}, |Q_K| \geq c_K\} = \pi_K$$

Lan-DeMets method:

Error spending function: $f(\tilde{t})$, $\tilde{t} = V(t)/V(t_K)$

- $f(\tilde{t})$ nondecreasing, $f(0) = 0$, $f(1) = \alpha$
- $\pi_k = f(\tilde{t}_k) - f(\tilde{t}_{k-1})$
- $f(\tilde{t}) = 2\Phi\left(z_{\alpha/2}/\tilde{t}^{1/2}\right)$ (O'Brien-Fleming)

3. Weighted Log-Rank Tests

Problem:

$$\lambda_1(x) = \lambda_0(x)e^{\beta w_0(x)}$$

$$H_0 : \beta = 0 \text{ vs. } H_1 : \beta = \beta_1$$

Notation:

- $n =$ total number of patients in the study
- $R_i =$ entry time for the ith patient
- $T_i =$ failure time measured from R_i
- $C_i =$ censoring time measured from R_i
- $Z_i =$ treatment indicator for the ith patient

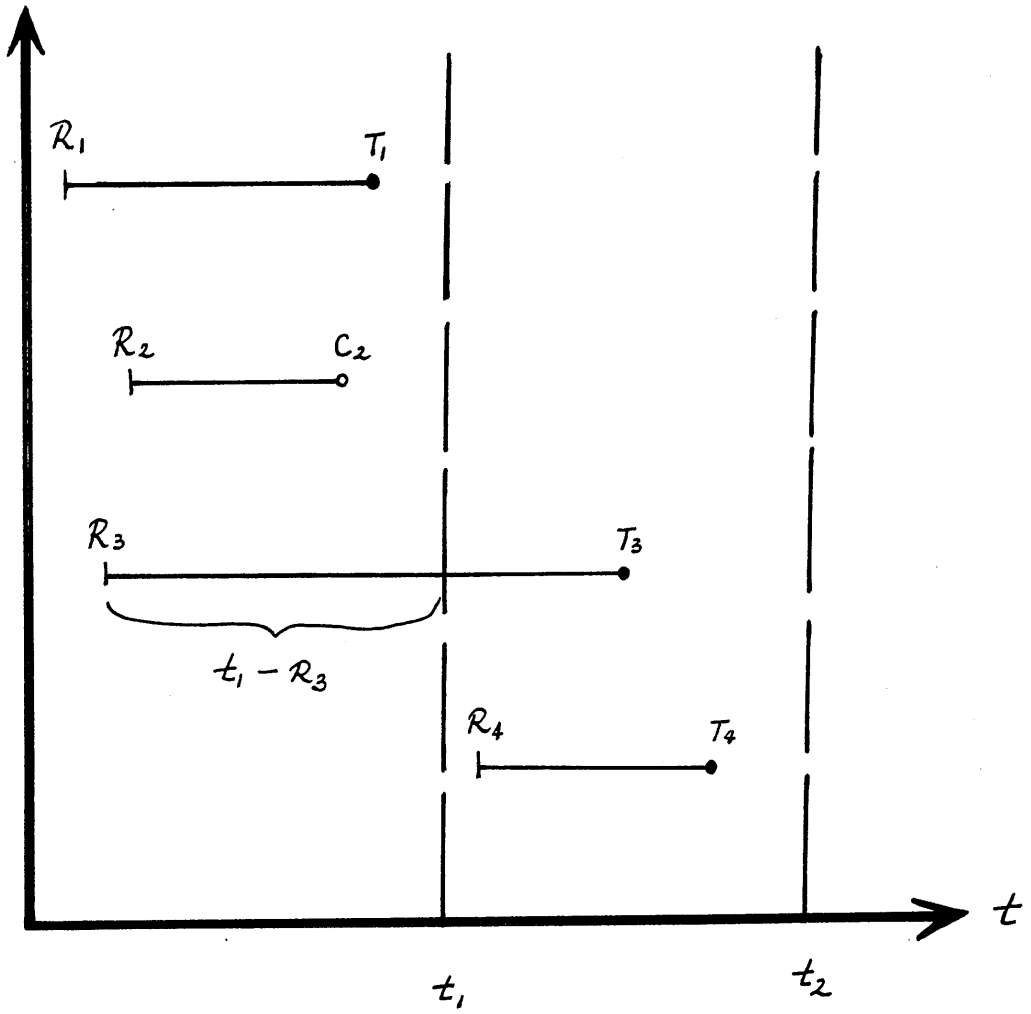
Data available at t:

$$\{X_i(t), \Delta_i(t), Z_i\} \quad (i = 1, \ldots, n)$$

- $X_i(t) = \max\{0, \min(T_i, t - R_i, C_i)\}$
- $\Delta_i(t) = I\{T_i \leq \min(C_i, t - R_i)\}$

Illustration: Fig. 19

Fig. 19. Survival Data in Sequential Trials



Sequential statistics

$$U(t) = \sum_{i=1}^{n} W\left(t, X_i(t)\right) \Delta_i(t) \left\{ Z_i - \overline{Z}\left(t, X_i(t)\right) \right\}$$

- $W(t,x) =$ weight for failure time x calculated at analysis time t
- $\overline{Z}(t,x) = \frac{\sum_{j=1}^{n} I\{X_j(t) \geq x\} Z_j}{\sum_{j=1}^{n} I\{X_j(t) \geq x\}}$

$$V(t) = \sum_{i=1}^{n} W^2\left(t, X_i(t)\right) \Delta_i(t) \overline{Z}\left(t, X_i(t)\right) \left\{ 1 - \overline{Z}\left(t, X_i(t)\right) \right\}$$

special case: $W = 1$

$V(t) = \sum_{i=1}^{n} \Delta_i(t) \mathrm{Var}(Z)$ (equal censoring)

$= D(t)/4$ (equal allocation)

Sequential properties of $U(t)$:

$U(t) \sim B(t) \equiv \text{Gaussian}\,(\mu(t), V(t))$

$W(t,x) \xrightarrow{P} w(x)$: $B(t) = \text{Brownian motion}$

- true for most weight functions, but not for Gehan and Tarone-Ware statistics

$w(x) = w_0(x)$: $B(t) = \mathcal{B}\,(\beta V(t), V(t))$

- $B(t) = \mathcal{B}\,(0, V(t))$ under H_0
- $B(t) = \mathcal{B}\,(\beta_1 V(t), V(t))$ under H_1
- Brownian motion statistics
- optimality, simplicity

4. Colon Cancer Study

Enrollment: March 1984 – October 1987

Stopping rule: O'Brien-Fleming boundary with $K = 4$ and $\alpha = 0.05$

Planned analyses: 125, 250, 375, 500 deaths

Actual analyses:

- times: December 1987, September 1989
- deaths: 125, 301

Boundary values: $c_1 = 4.006$, $c_2 = 2.582$

Lev vs. Obs: $Q_1 = 0.71$, $Q_2 = 0.004$

Lev+5-FU vs. Obs: $Q_1 = 1.163$, $Q_2 = 2.726$

REFERENCES

Fleming, T. R. & Lin, D. Y. (2000). Survival analysis in clinical trials: past developments and future directions. *Biometrics* **56**, 971-83 (Editors' Invited Paper).

Fleming, T. R. & Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* New York: Wiley

Lin, D. Y., Fleming, T. R. & Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73-81.

Lin, D. Y. & Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Am. Stat. Ass.* **84**, 1074-78.

Lin, D. Y., Wei, L. J. & Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557-72.

Zeng, D. & Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *J. R. Stat. Soc. B* **69**, 507-64.

Lin, D. Y., Yao, Q. & Ying, Z. (1999). A general theory on stochastic curtailment for censored survival data. *J. Am. Stat. Ass.* **94**, 510-21.

PART II

ANALYSIS OF MULTIVARIATE FAILURE TIME DATA

I. INTRODUCTION

1. Multivariate Failure Time Data: Each study subject may experience several events/failures, or there exists natural/artificial clustering of subjects which induces dependence among failure times of the same cluster.

2. Multiple Events Data: each study subject can potentially experience several events

(a) Ordered: natural ordering of failure times

(i) *Recurrent events*: repetitions of a phenomenon (e.g., illness)

- tumor recurrences
- infection episodes
- repeated breakdowns of machinery
- employment/unemployment cycles

(ii) *Distinct events*: successive events of different natures

- HIV-infection→AIDS→death
- randomization→cancer recurrence→death
- birth→marriage→child birth

(b) Unordered: several concurrent failure processes

- physical symptoms or diseases in several organ systems (cardiovascular, cancer, etc.)
- purchases of various consumer products

3. Clustered Data: natural/artificial clustering of study subjects

- family (twin) studies
- multi-center (group) studies
- litter matched carcinogeneicity experiments

4. Real Examples:

AIDS Study: A randmized clinical trial was conducted to assess the antiretroviral capability of ribavirin over time in AIDS patients. Blood samples for each patient were collected at weeks 4, 8 and 12. The "viral load" in each sample was measured by the number of days when virus positivity was dectected. Censored observations occurred when the culture required a longer period of time to register as virus positive than was achievable in the lab or when the sample was contaminated before positivity was dectected.

		Detected positivity		
Group	n	Week 4	Week 8	Week 12
Placbo	12	12	8	8
Low-dose ribavirin	11	8	7	6
High-dose ribavirin	13	11	11	8

Unordered multiple events

Colon Cancer Study: A randomized clinical trial was conducted in the 1980's to study the drugs Lev and 5-FU for adjuvant therapy of resected colon carcinoma. Patients with Stage C disease were randomly assigned to observation, Lev alone, or Lev+5-FU. The time to cancer recurrence and the survival time were both considered important outcome measures.

Group	Patients	Recurrences	Deaths
observation	315	155	114
Lev	310	144	109
Lev+5-FU	304	103	78

Multiple events

Bladder Tumor Study: A randomized clinical trial was conducted to assess the efficacy of thiotepa in reducing cancer recurrences in patients with superficial bladder tumors.

		Cancer recurrences						
Group	n	0	1	2	3	4	> 4	total
Thiotepa	38	20	8	3	2	2	3	45
Placebo	48	19	10	4	6	2	7	87

Recurrent events

CGD Study. Chronic granulomatous disease (CGD) is a rare immune disorder characterized by recurrent pyogenic infections. To study the ability of gamma interferon to reduce the rate of infections, a placebo controlled randomized trial was conducted by the International CGD Cooperative Study Group in the late 1980's.

		Infections			
	Patients	≥ 1	2	3	≥ 4
Placebo	65	30	5	4	3
Treatment	63	14	4	1	–

Recurrent events

Diabetic Retinopathy Study. The Diabetic Retinopathy Study was conducted by the National Eye Institute to assess the effectiveness of laser photocoagulation in delaying visual loss in patients with diabetic retinopathy. One eye of each patient was randomly selected for photocoagulation and the other eye was observed without treatment. The patients were followed over several years for the occurrence of visual loss in the left and right eyes.

Treatment	Patients (eyes)	Visual loss
Yes	1727	242
No	1727	535

Clustered data

Schizophrenia Study. In a genetic epidemiologic study of schizophrenia, 487 first degree relatives (273 males, 214 females) of 93 female schizophrenic probands were enrolled. The number of relatives of a single proband ranges from 1 to 12. An important question is whether the risk of affective illness (depression or mania or both) in the relatives is associated with the age at onset of schizophrenia of the proband.

Clustered data

5. Scientific/Statistical Questions:

- Distributions of multivariate failure times (joint, marginal and conditional distributions)
- Effects of covariates (e.g., treatment) on multivariate failure times

6. Statistical Challenges:

- Dependence of failure times within the same subject/cluster
- Censoring due to patient withdrawal/study termination
- Multiplicity of outcome measures

7. Organization of the Presentation

Marginal Cox Models for Multiple Events Data

Marginal Cox Models for Clustered Data

Intensity/Rate Models for Recurrent Events

Frailty (Random-Effects) Models

Joint Models

II. MARGINAL COX MODELS FOR MULTIPLE EVENTS DATA

1. Methods

Setup: n subjects, K potential events

$T_{ki} = k$th failure time of ith subject

$C_{ki} = k$th censoring time of ith subject

Observation time: $X_{ki} = \min(T_{ki}, C_{ki})$

Failure indicator: $\Delta_{ki} = I(T_{ki} \leq C_{ki})$

Covariates: $Z_{ki} = (Z_{1ki}, \ldots, Z_{pki})'$

Data: $(X_{ki}, \Delta_{ki}, Z_{ki})$ $(k = 1, \ldots, K; i = 1, \ldots, n)$

Marginal Cox models:

$$\lambda_k(t|Z_{ki}) = \lambda_{k0}(t)e^{\beta_k' Z_{ki}(t)},\ k = 1, \ldots, K; i = 1, \ldots, n$$

- event-specific baseline hazard functions
- event-specific regression parameters
- modeling the marginal distributions without specifying the dependence structures

(Wei, Lin & Weissfeld, WLW, 1989)

Censoring assumption: $T_{ki} \coprod C_{ki}$ given Z_{ki}

Notation:

At-risk indicator: $Y_{ki}(t) = I(X_{ki} \geq t)$

$S_k^{(0)}(\beta, t) = \sum_{i=1}^n Y_{ki}(t) e^{\beta' Z_{ki}(t)}$

$S_k^{(1)}(\beta, t) = \sum_{i=1}^n Y_{ki}(t) e^{\beta' Z_{ki}(t)} Z_{ki}(t)$

$S_k^{(2)}(\beta, t) = \sum_{i=1}^n Y_{ki}(t) e^{\beta' Z_{ki}(t)} Z_{ki}(t)^{\otimes 2}$

Partial likelihood functions:

$$L_k(\beta) = \prod_{i=1}^{n} \left\{ \frac{e^{\beta' Z_{ki}(X_{ki})}}{S_k^{(0)}(\beta, X_{ki})} \right\}^{\Delta_{ki}}$$

Score functions:

$$U_k(\beta) = \frac{\partial \log L_k(\beta)}{\partial \beta} = \sum_{i=1}^{n} \Delta_{ki} \left\{ Z_{ki}(X_{ki}) - \frac{S_k^{(1)}(\beta, X_{ki})}{S_k^{(0)}(\beta, X_{ki})} \right\}$$

Information matrices:

$$\mathcal{I}_k(\beta) = -\frac{\partial^2 \log L_k(\beta)}{\partial \beta^2} = \sum_{i=1}^{n} \Delta_{ki} \left\{ \frac{S_k^{(2)}(\beta, X_{ki})}{S_k^{(0)}(\beta, X_{ki})} - \frac{S_k^{(1)}(\beta, X_{ki})^{\otimes 2}}{S_k^{(0)}(\beta, X_{ki})^2} \right\}$$

Parameter estimators $\widehat{\beta}_k$: $\{U_k(\beta) = 0\}$

Asymptotic Properties:

$$\begin{bmatrix} \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_K \end{bmatrix} \sim N\left(\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}, \begin{bmatrix} D_{11} & \dots & D_{1K} \\ \vdots & \vdots & \vdots \\ D_{K1} & \dots & D_{KK} \end{bmatrix} \right)$$

$$D_{kl} = \mathcal{I}_k^{-1}(\widehat{\beta}_k) B_{kl} \mathcal{I}_l^{-1}(\widehat{\beta}_l)$$

$$B_{kl} = \sum_{i=1}^{n} W_{ki} W_{li}'$$

$$W_{ki} = \Delta_{ki} \left\{ Z_{ki}(X_{ki}) - \frac{S_k^{(1)}(\widehat{\beta}_k, X_{ki})}{S_k^{(0)}(\widehat{\beta}_k, X_{ki})} \right\}$$

$$- \sum_{j=1}^{n} \frac{\Delta_{kj} Y_{ki}(X_{kj}) e^{\widehat{\beta}_k' Z_{ki}(X_{kj})}}{S_k^{(0)}(\widehat{\beta}_k, X_{kj})} \left\{ Z_{ki}(X_{kj}) - \frac{S_k^{(1)}(\widehat{\beta}_k, X_{kj})}{S_k^{(0)}(\widehat{\beta}_k, X_{kj})} \right\}$$

- Robust sandwich covariance matrix estimators D_{kl} $(k \neq l)$ account for the dependence of the multiple failure times
- D_{kk} is the robust covariance matrix estimator whereas $\mathcal{I}_k^{-1}(\widehat{\beta}_k)$ is the model-based estimator.

Simultaneous Inference:

Parameters of interest: $\eta_k = \beta_{1k},\ k = 1, \ldots, K.$

Estimators: $\widehat{\eta}_k = \widehat{\beta}_{1k},\ k = 1, \ldots, K.$

Covariance matrix estimator: $\widehat{\Psi} = \widehat{\mathrm{cov}}(\widehat{\eta}_1, \ldots, \widehat{\eta}_K)$

Global Test: $H_0 : \eta_k = 0, \quad k = 1, \ldots, K$

$$Q \equiv (\widehat{\eta}_1, \ldots, \widehat{\eta}_K)\widehat{\Psi}^{-1}(\widehat{\eta}_1, \ldots, \widehat{\eta}_K)' \sim \chi^2_K$$

Estimation of Common Parameter: $\eta_1 = \ldots = \eta_K = \eta$

$$\widehat{\eta} = \sum_{k=1}^{K} c_k \widehat{\eta}_k$$

$$(c_1, \ldots, c_K)' = (e'\widehat{\Psi}^{-1}e)^{-1}\widehat{\Psi}^{-1}e$$

- $e = (1, \ldots, 1)'$
- $\widehat{\eta}/\mathrm{se}(\widehat{\eta}) \sim N(0, 1)$

Estimation of Λ_{k0}'s

$$\widehat{\Lambda}_{k0}(t) = \sum_{i:X_{ki} \leq t} \frac{\Delta_{ki}}{S_k^{(0)}(\widehat{\beta}_k, X_{ki})}$$

$$\left\{\widehat{\Lambda}_{10}(t), \ldots, \widehat{\Lambda}_{K0}(t)\right\} \sim N_K\left(\{\Lambda_{10}(t), \ldots, \Lambda_{K0}(t)\}, V(t)\right)$$

(Spiekerman & Lin, 1998)

Features:

- arbitrary dependence structures
- overall treatment effect
- ordered (recurrent or distinct) and unordered events
- total times for ordered events
- compatibility of treatment groups
- software packages

2. Software

Data input:

pid enum time status covariates

SAS procedure:

```
proc phreg covs(aggregate);
    model time*status(0)=covariates/cov ties=breslow;
    strata enum;
    id pid;
```

S-Plus function:

```
coxph(Surv(time, status) ~ covariates+cluster(pid)+strata(enum),
  method='breslow')
```

3. Examples

Bladder Tumor Study

$n = 86$, $K = 4$

T_{ki} = time to the kth tumor recurrence of the ith patient

$Z_{ki} = (Z_{1ki}, Z_{2ki}, Z_{3ki})'$

$$Z_{1ki} = \begin{cases} 0 & \text{if the } i\text{th patient was on thiotepa,} \\ 1 & \text{if the } i\text{th patient was on placebo.} \end{cases}$$

Z_{2ki} = number of initial tumors

Z_{3ki} = size of the largest initial tumor

$\eta_k = \beta_{1k}$, $k = 1, 2, 3, 4$

Parameter	Est	SE	Est/SE	p-value
η_1	.518	.308	1.681	.092
η_2	.619	.364	1.701	.089
η_3	.700	.415	1.687	.092
η_4	.651	.490	1.316	.184
$\eta = \eta_1 = \ldots = \eta_4$	.549	.285	1.924	.054

$$\widehat{\Psi} = \begin{bmatrix} .095 & .060 & .057 & .044 \\ & .132 & .130 & .116 \\ & & .172 & .159 \\ & & & .240 \end{bmatrix}$$

$Q = 3.967$, $p-$value=0.41

$c = (0.677, 0.257, -0.075, 0.141)'$

AIDS Study

$n = 36$, $K = 3$

T_{ki} = number of days to virus positivity in the kth sample of the ith patient

$Z_{ki} = (Z_{1ki}, Z_{2ki})'$

$$Z_{1ki} = \begin{cases} 1 & \text{if patient } i \text{ was on low dose ribavirin,} \\ 0 & \text{otherwise.} \end{cases}$$

$$Z_{2ki} = \begin{cases} 1 & \text{if patient } i \text{ was on high dose ribavirin,} \\ 0 & \text{otherwise.} \end{cases}$$

$\eta_k = \beta_{1k}$, $k = 1, 2, 3$

Parameter	Est	SE	Est/SE	p-value
η_1	-1.394	.525	-2.655	.008
η_2	-0.655	.523	-1.253	.210
η_3	-0.615	.554	-1.110	.267
$\eta = \eta_1 = \eta_2 = \eta_3$	-0.972	.386	-2.519	.012

$$\widehat{\Psi} = \begin{bmatrix} .245 & .051 & .107 \\ & .287 & .133 \\ & & .257 \end{bmatrix}$$

$Q = 8.53$, $p-$value=0.036

$c = (0.441, 0.340, 0.219)'$

III. MARGINAL COX MODELS FOR CLUSTERED DATA

1. Methods

Setup: n clusters, K_i members in ith cluster

T_{ki} = failure time for kth member of ith cluster

C_{ki} = censoring time for T_{ki}

Observation time: $X_{ki} = \min(T_{ki}, C_{ki})$

Failure indicator: $\Delta_{ki} = I(T_{ki} \leq C_{ki})$

Covariates: $Z_{ki} = (Z_{1ki}, \ldots, Z_{pki})'$

Data: $(X_{ki}, \Delta_{ki}, Z_{ki})$ $(k = 1, \ldots, K_i;\ i = 1, \ldots, n)$

Marginal Cox model:

$$\lambda(t|Z_{ki}) = \lambda_0(t)e^{\beta' Z_{ki}(t)}, \quad k = 1, \ldots, K_i; \; i = 1, \ldots, n$$

- $\lambda_0(\cdot)$ = arbitrary baseline hazard function
- $\beta = (\beta_1, \ldots, \beta_p)'$
- stratification
- modeling the marginal distributions without specifying the dependence structures

(Lee, Wei & Amato, 1992)

Censoring assumption: $T_{ki} \coprod C_{ki}$ given Z_{ki}

Independence working assumption: failure times within the same cluster are independent

- analogous to GEE for longitudinal data

Notation:

At-risk indicator: $Y_{ki}(t) = I(X_{ki} \geq t)$

$S^{(0)}(\beta, t) = \sum_{i=1}^{n} \sum_{k=1}^{K_i} Y_{ki}(t) e^{\beta' Z_{ki}(t)}$

$S^{(1)}(\beta, t) = \sum_{i=1}^{n} \sum_{k=1}^{K_i} Y_{ki}(t) e^{\beta' Z_{ki}(t)} Z_{ki}(t)$

$S^{(2)}(\beta, t) = \sum_{i=1}^{n} \sum_{k=1}^{K_i} Y_{ki}(t) e^{\beta' Z_{ki}(t)} Z_{ki}(t)^{\otimes 2}$

"Partial likelihood function":

$$L(\beta) = \prod_{i=1}^{n} \prod_{k=1}^{K_i} \left\{ \frac{e^{\beta' Z_{ki}(X_{ki})}}{S^{(0)}(\beta, X_{ki})} \right\}^{\Delta_{ki}}$$

"Score function":

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \sum_{k=1}^{K_i} \Delta_{ki} \left\{ Z_{ki}(X_{ki}) - \frac{S^{(1)}(\beta, X_{ki})}{S^{(0)}(\beta, X_{ki})} \right\}$$

"Information matrix":

$$\mathcal{I}(\beta) = -\frac{\partial^2 \log L(\beta)}{\partial \beta^2} = \sum_{i=1}^{n} \sum_{k=1}^{K_i} \Delta_{ki} \left\{ \frac{S^{(2)}(\beta, X_{ki})}{S^{(0)}(\beta, X_{ki})} - \frac{S^{(1)}(\beta, X_{ki})^{\otimes 2}}{S^{(0)}(\beta, X_{ki})^2} \right\}$$

Parameter estimator $\widehat{\beta}$: $\{U(\beta) = 0\}$

Asymptotic Properties:

$$U(\beta) \sim N\big(0, B(\beta)\big)$$

$$B(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{l=1}^{K_i} W_{ki}(\beta) W_{li}(\beta)'$$

$$W_{ki}(\beta) = \Delta_{ki} \left\{ Z_{ki}(X_{ki}) - \frac{S^{(1)}(\beta, X_{ki})}{S^{(0)}(\beta, X_{ki})} \right\}$$

$$- \sum_{j=1}^{n} \sum_{l=1}^{K_j} \frac{\Delta_{lj} Y_{ki}(X_{lj}) e^{\beta' Z_{ki}(X_{lj})}}{S^{(0)}(\beta, X_{lj})} \left\{ Z_{ki}(X_{lj}) - \frac{S^{(1)}(\beta, X_{lj})}{S^{(0)}(\beta, X_{lj})} \right\}$$

$$\widehat{\beta} \sim N(\beta, D)$$

$$D = \mathcal{I}^{-1}(\widehat{\beta}) B(\widehat{\beta}) \mathcal{I}^{-1}(\widehat{\beta})$$

- Robust variance estimators $B(\beta)$ and D account for intra-class dependence; naive variance estimators $\mathcal{I}(\beta)$ and $\mathcal{I}^{-1}(\widehat{\beta})$ do not.
- If $K = 1$, then D becomes the robust variance-covariance estimator for misspecified univariate Cox models.

Test statistics:

$$H_0 : \beta = 0 : \ U'(0)B^{-1}(0)U(0) \to \chi^2_p$$

$$H_0 : L\beta = 0 : \ (L\widehat{\beta})'(LDL')^{-1}(L\widehat{\beta}) \to \chi^2_r$$

- L is a $r \times p$ contrast matrix

Estimation of Λ_0

$$\widehat{\Lambda}_0(t) = \sum_{i,k:X_{ki} \leq t} \frac{\Delta_{ki}}{S^{(0)}(\widehat{\beta}, X_{ki})}$$

$$\widehat{\Lambda}_0(t) \sim N\big(\Lambda_0(t), V(t)\big)$$

(Spiekerman & Lin, 1998)

2. Software

Data input:

pid time status covariates

SAS procedure:

```
proc phreg covs(aggregate);
    model time*status(0)=covariates/ties=breslow;
    id pid;
```

S-Plus function:

coxph(Surv(time, status) ~ covariates +cluster(pid),method='breslow')

3. Examples

Diabetic Retinopathy Study

$n = 197$, $K = 2$

T_{ki} = time to visual loss of the kth eye for the ith patient

$Z_{ki} = (Z_{1ki}, Z_{2ki}, Z_{3ki})'$

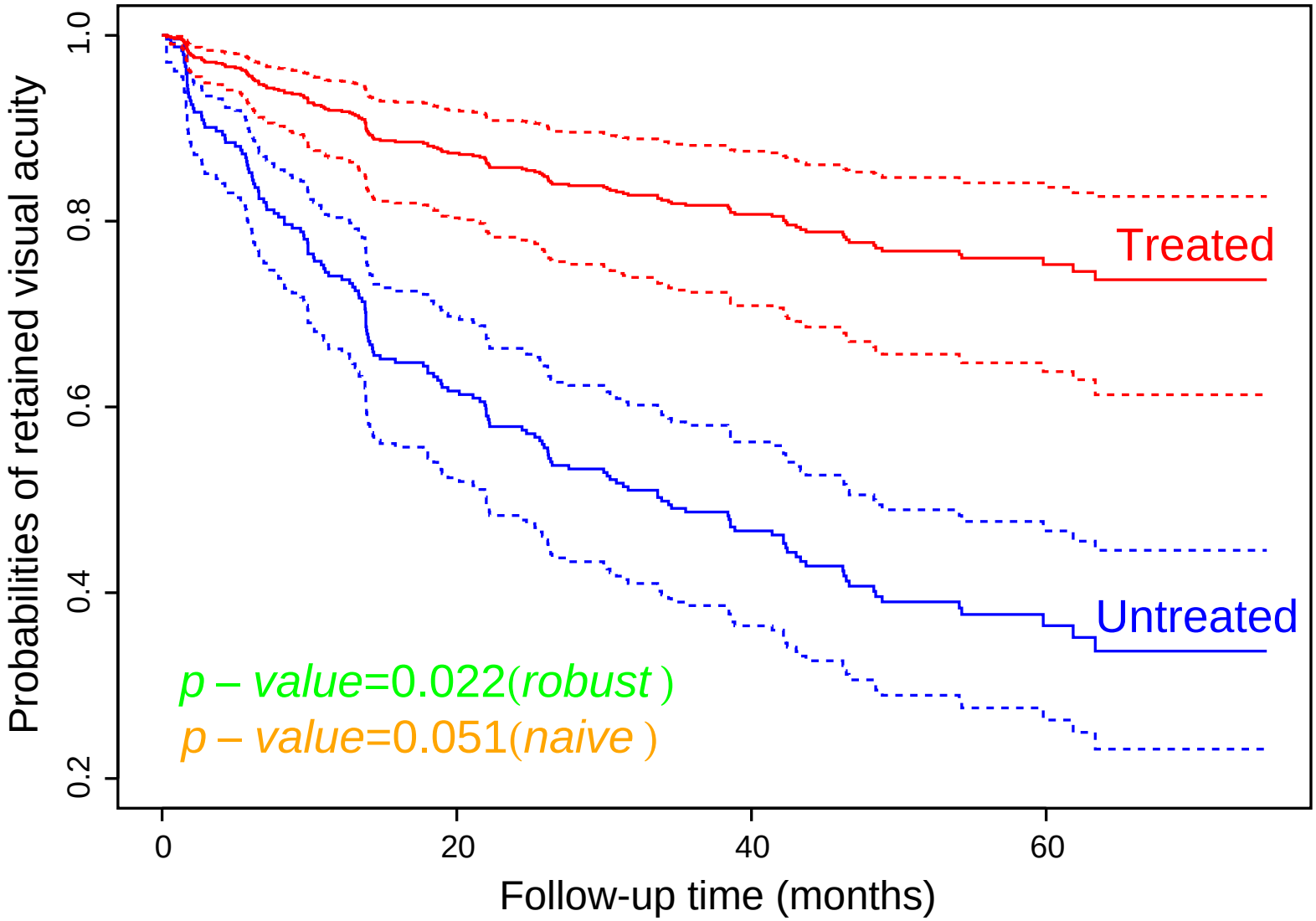
$$Z_{1ki} = \begin{cases} 1 & \text{if } k\text{th eye of patient } i \text{ was treated,} \\ 0 & \text{otherwise;} \end{cases}$$

$$Z_{2ki} = \begin{cases} 1 & \text{if patient } i \text{ had adult onset diabetes,} \\ 0 & \text{if patient } i \text{ had juvenile onset diabetes;} \end{cases}$$

$Z_{3ki} = Z_{1ki} * Z_{2ki}$

		S.E. (p-value)	
Covariate	Est.	Naive	Robust
Treatment (Z_1)	–0.43	0.22 (0.051)	0.19 (0.022)
Diabetic type (Z_2)	0.34	0.20 (0.088)	0.20 (0.082)
Interaction (Z_3)	–0.85	0.35 (0.016)	0.30 (0.005)

Fig 1. Probabilities of Retained Visual Acuity for Adult Onset Diabetes



Schizophrenia Study

$n = 93, \quad K_i = 1 \sim 12$

T_{ki} = time to affective illness of kth relative of ith proband

$Z_{ki} = (Z_{1ki}, Z_{2ki})'$

$$Z_{1ki} = \begin{cases} 1 & \text{if } i\text{th proband's age at onset} \leq 16, \\ 0 & \text{otherwise;} \end{cases}$$

$$Z_{2ki} = \begin{cases} 1 & \text{if } k\text{th relative of proband } i \text{ is male,} \\ 0 & \text{if } k\text{th relative of proband } i \text{ is female.} \end{cases}$$

	Age	Gender
Est.	–0.238	–1.244
S.E.		
naive	0.489	0.411
robust	0.517	0.408

IV. INTENSITY/RATE FUNCTIONS FOR RECURRENT EVENTS

1. Methods

Notation:

$N^*(t)$ = number of events by time t

$Z(t)$ = covariates

$\mathcal{F}_t$ = history $\{N^*(s), Z(s); 0 \leq s \leq t\}$

$dN^*(t)$ = increment of N^*over $[t, t + dt)$

Intensity function: $E\{dN^*(t)|\mathcal{F}_{t-}\} = \lambda_Z(t)dt$

Rate function: $E\{dN^*(t)|Z(t)\} = d\mu_Z(t)$

Mean function: $\mu_Z(t) = E\{N^*(t)|Z(s) : s \geq 0\}$

- external time-dependent covariates
- $d\mu_Z(t) = E\{dN^*(t)|Z(s) : s \geq 0\}$

Intensity models:

$$\lambda_Z(t) = e^{\beta' Z(t)} \lambda_0(t)$$

- $\lambda_0(t) =$ arbitrary baseline intensity function
- $\beta =$ unknown regression parameters
- Poisson process
- $E\{dN^*(t)|\mathcal{F}_{t-}\} = E\{dN^*(t)|Z(t)\} = e^{\beta' Z(t)} \lambda_0(t)dt$

(Andersen & Gill, 1982)

Rate/mean models:

$$d\mu_Z(t) = e^{\beta' Z(t)} d\mu_0(t)$$

$$\mu_Z(t) = e^{\beta' Z} \mu_0(t)$$

- $\mu_0(t) =$ arbitrary baseline mean function
- $\beta =$ unknown regression parameters
- general counting process
- $\mu_0(t) = \int_0^t \lambda_0(s) ds$
- random-effect model: $\lambda_Z(t|\eta) = \eta e^{\beta' Z(t)} \lambda_0(t)$

(Pepe & Cai, 1993; Lawless & Nadeau, 1995; Lin et al, 2000)

Notation:

$T_{ki} = k$th event time of the ith subject

$N_i^*(t) = \sum_{k=1}^{\infty} I(T_{ki} \leq t)$

$C_i =$ censoring time of the ith subject

Failure indicator: $\Delta_{ki} = I(T_{ki} \leq C_i)$

At-risk indicator: $Y_i(t) = I(C_i \geq t)$

$S^{(0)}(\beta, t) = \sum_{i=1}^{n} Y_i(t) e^{\beta' Z_i(t)}$

$S^{(1)}(\beta, t) = \sum_{i=1}^{n} Y_i(t) e^{\beta' Z_i(t)} Z_i(t)$

Censoring assumption: $C_i \coprod N_i^*(\cdot)$ given $Z_i(\cdot)$

"Partial likelihood score function":

$$U(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{\infty} \Delta_{ki} \left\{ Z_i(T_{ki}) - \frac{S^{(1)}(\beta, T_{ki})}{S^{(0)}(\beta, T_{ki})} \right\}$$

"Information matrix":

$$\mathcal{I}(\beta) = -\frac{\partial U(\beta)}{\partial \beta}$$

Estimator of β: $\{U(\widehat{\beta}) = 0\}$

Estimator of μ_0:

$$\widehat{\mu}_0(t) = \sum_{i,k:T_{ki} \leq t} \frac{\Delta_{ki}}{S^{(0)}(\widehat{\beta}, T_{ki})}$$

Asymptotic Properties:

$$U(\beta) \sim N\big(0, B(\beta)\big)$$

$$B(\beta) = \sum_{i=1}^{n} W_i(\beta) W_i(\beta)'$$

$$W_i(\beta) = \sum_{k=1}^{\infty} \Delta_{ki} \left\{ Z_i(T_{ki}) - \frac{S^{(1)}(\beta, T_{ki})}{S^{(0)}(\beta, T_{ki})} \right\}$$

$$- \sum_{j=1}^{n} \sum_{l=1}^{\infty} \frac{\Delta_{lj} Y_i(T_{lj}) e^{\beta' Z_i(T_{lj})}}{S^{(0)}(\beta, T_{lj})} \left\{ Z_i(T_{lj}) - \frac{S^{(1)}(\beta, T_{lj})}{S^{(0)}(\beta, T_{lj})} \right\}$$

$$\widehat{\beta} \sim N(\beta, D)$$

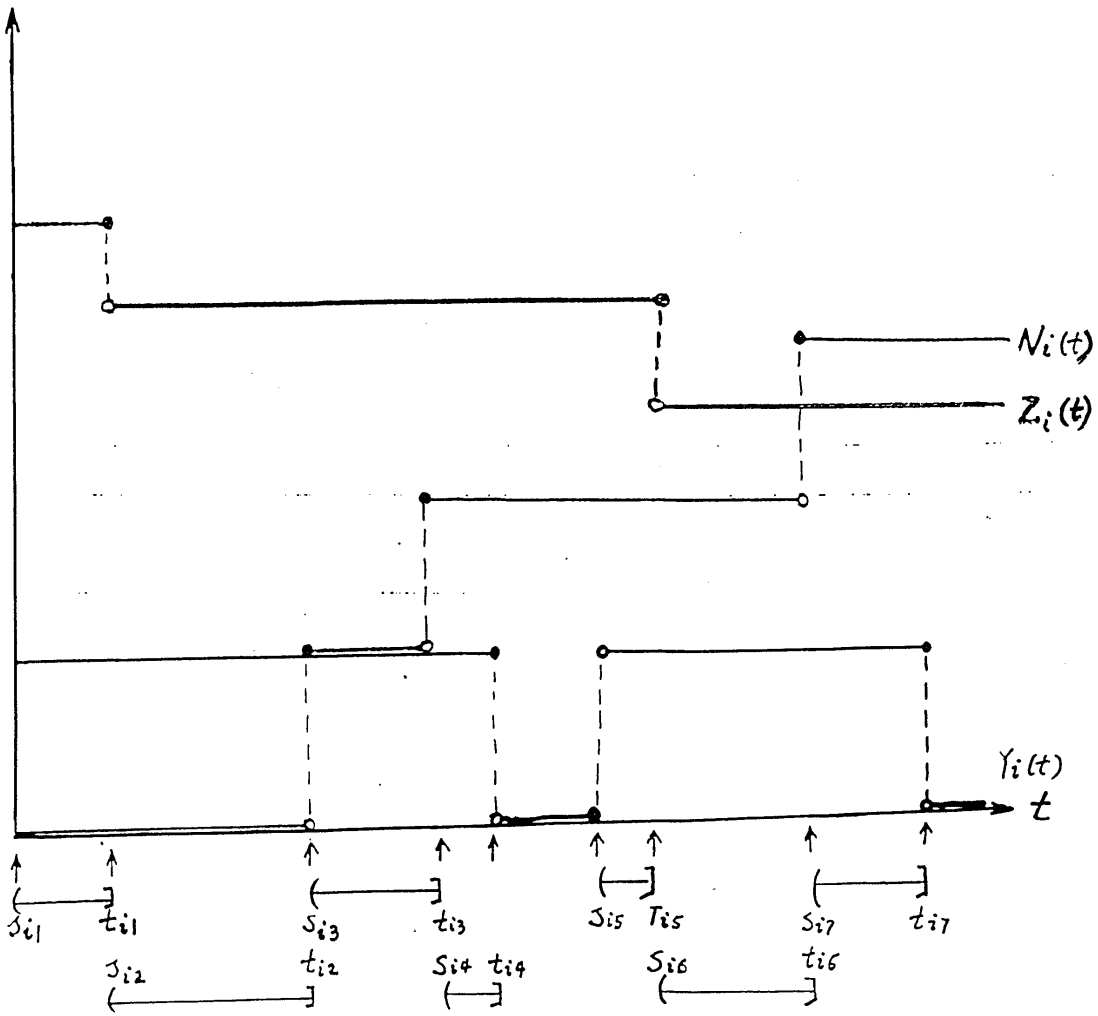
$$D = \mathcal{I}^{-1}(\widehat{\beta}) B(\widehat{\beta}) \mathcal{I}^{-1}(\widehat{\beta})$$

$$\widehat{\mu}_0(t) \sim N\big(\mu_0(t), V(t)\big)$$

- For intensity models, $\lim B(\beta) = \lim \mathcal{I}(\beta)$.
- Robust variance-covariance estimators $B(\beta)$ and D account for the dependence of multiple events on the same subject, whereas naive variance-covariance estimators $\mathcal{I}(\beta)$ and $\mathcal{I}^{-1}(\widehat{\beta})$ do not.
- Nonparametric statistic $U'(0)B^{-1}(0)U(0)$ can used to test $H_0 : \beta = 0$.
- Wald statistics can be used for testing individual parameters.

Features:

- efficient and parsimonious summarization of recurrence experience and covariate (e.g., treatment) effects
- arbitrary dependence structures
- easy/intuitive interpretations
- software packages
- recurrent events only


$N_i(t)$
$Z_i(t)$
$Y_i(t)$
t
s_{i1}
t_{i1}
s_{i2}
t_{i2}
s_{i3}
t_{i3}
s_{i4}
t_{i4}
s_{i5}
T_{i5}
s_{i6}
t_{i6}
s_{i7}
t_{i7}

2. Computational Issues

Subject i has K_i events at times $T_{i1}, ..., T_{iK_i}$

$$U(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K_i} \left\{ Z_i(T_{ik}) - \frac{\sum_{l=1}^{n} Y_l(T_{ik}) e^{\beta' Z_l(T_{ik})} Z_l(T_{ik})}{\sum_{l=1}^{n} Y_l(T_{ik}) e^{\beta' Z_l(T_{ik})}} \right\}$$

$$\mathcal{I}(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K_i} \left\{ \frac{\sum_{l=1}^{n} Y_l(T_{ik}) e^{\beta' Z_l(T_{ik})} Z_l^{\otimes 2}(T_{ik})}{\sum_{l=1}^{n} Y_l(T_{ik}) e^{\beta' Z_l(T_{ik})}} - \left[\frac{\sum_{l=1}^{n} Y_l(T_{ik}) e^{\beta' Z_l(T_{ik})} Z_l(T_{ik})}{\sum_{l=1}^{n} Y_l(T_{ik}) e^{\beta' Z_l(T_{ik})}} \right]^{\otimes 2} \right\}$$

In the data set, subject i is represented by a series of records

$$\overset{start}{s_{ij}}, \quad \overset{stop}{t_{ij}}, \quad \overset{status}{\delta_{ij}}, \quad \overset{(trt,age,...)}{Z_{ij}}; \qquad j = 1, ..., n_i;$$

where $(s_{ij}, t_{ij}]$ $(j = 1, ..., n_i)$ are disjoint time intervals, open on the left and closed on the right, which are formed by dividing the time period(s) in which subject i is at risk into finer intervals such that an event may occur only at the end of an interval and the covariates take some constant values over each interval. Specifically, subject i

is at risk throughout $(s_{ij}, t_{ij}]$, $\delta_{ij} = 1$ if the subject has an event at t_{ij}, and Z_{ij} is the value of the covariate vector Z_i over $(s_{ij}, t_{ij}]$.

$$U(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \delta_{ij} \left\{ Z_{ij} - \frac{\sum_{l=1}^{n} \sum_{m=1}^{n_l} I(s_{lm} < t_{ij} \leq t_{lm}) e^{\beta' Z_{lm}} Z_{lm}}{\sum_{l=1}^{n} \sum_{m=1}^{n_l} I(s_{lm} < t_{ij} \leq t_{lm}) e^{\beta' Z_{lm}}} \right\}$$

$$\mathcal{I}(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \delta_{ij} \left\{ \frac{\sum_{l=1}^{n} \sum_{m=1}^{n_l} I(s_{lm} < t_{ij} \leq t_{lm}) e^{\beta' Z_{lm}} Z_{lm}^{\otimes 2}}{\sum_{l=1}^{n} \sum_{m=1}^{n_l} I(s_{lm} < t_{ij} \leq t_{lm}) e^{\beta' Z_{lm}}} \right.$$

$$\left. - \left[\frac{\sum_{l=1}^{n} \sum_{m=1}^{n_l} I(s_{lm} < t_{ij} \leq t_{lm}) e^{\beta' Z_{lm}} Z_{lm}}{\sum_{l=1}^{n} \sum_{m=1}^{n_l} I(s_{lm} < t_{ij} \leq t_{lm}) e^{\beta' Z_{lm}}} \right]^{\otimes 2} \right\}$$

3. Software:

Data input: pid start stop status covariates

The ith subject contributes $(K_i + 1)$ records, where K_i is the number of observed events. For the kth record of the ith subject, start is the time of the $(k-1)$th event (or 0 if $k = 1$), stop is the time of the kth event (or censoring time if $k = K_i + 1$), and status indidates whether there is an event at the stop time.

start	stop	status
0	T_{1i}	1
T_{1i}	T_{2i}	1
$\vdots$	$\vdots$	$\vdots$
$T_{K_i-1,i}$	$T_{K_i,i}$	1
$T_{K_i,i}$	C_i	0

SAS procedures:

Intensity model:

```
proc phreg;
    model (start, stop) * status(0)=covariates/ ties=breslow;
```

Rate/mean model:

```
proc phreg covs(aggregate);
    model (start, stop) * status(0)=covariates/ ties=breslow;
    id pid;
```

S-Plus functions:

Intensity model:

```
coxph(Surv(start, stop, status) ~ covariates, method='breslow')
```

Rate/mean model:

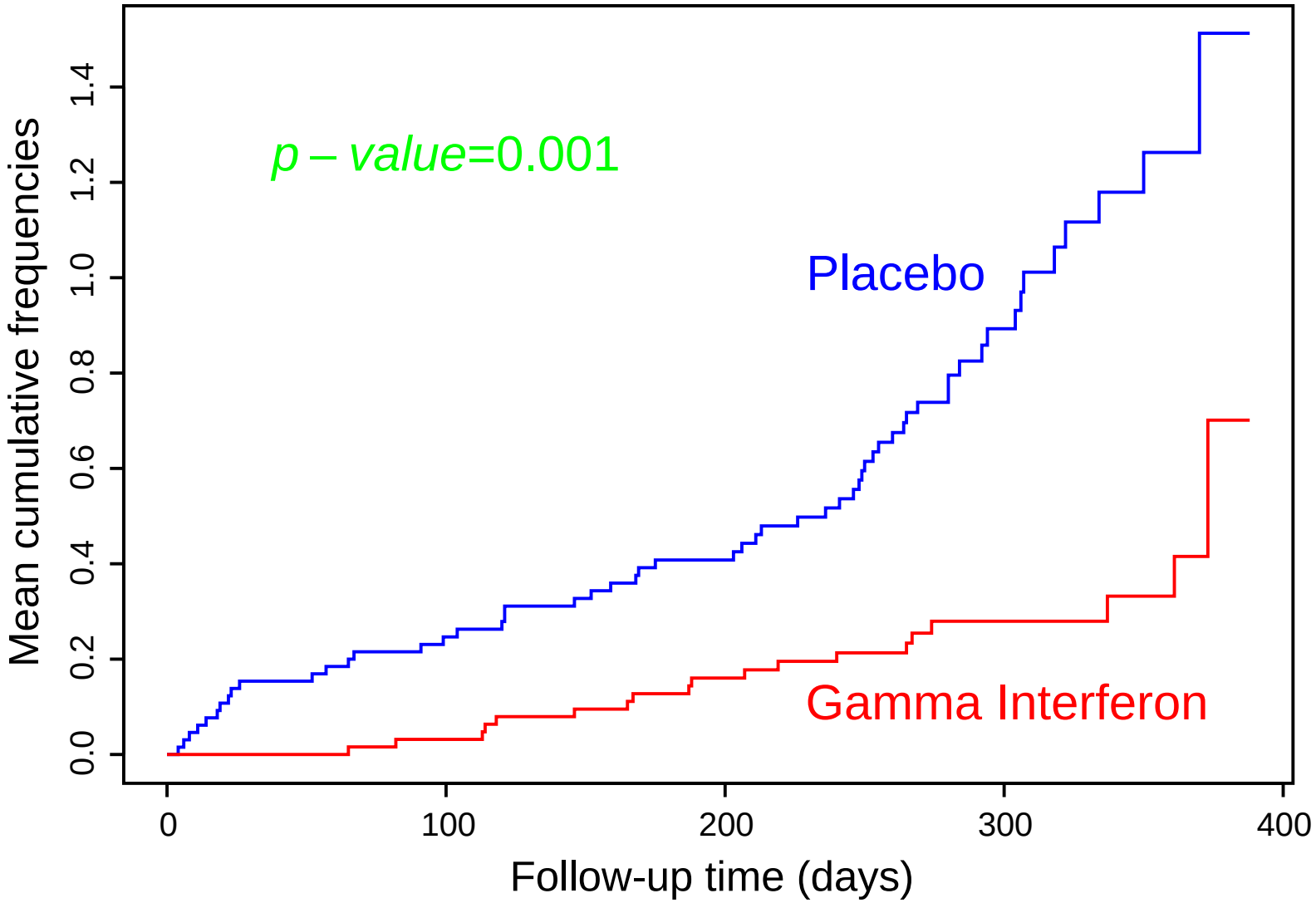
```
coxph(Surv(start, stop, status) ~ covariates +cluster(pid),
  method='breslow')
```

4. CGD Study

Two-Sample Regression

		Intensity Model		Rate Model	
	Est	SE	Est/SE	SE	Est/SE
First Event	1.09	0.34	3.27	0.34	3.27
All Events					
"Markov"	1.10	0.26	4.20	0.31	3.53
"Semi-Markov"	0.99	0.27	3.72	0.29	3.36

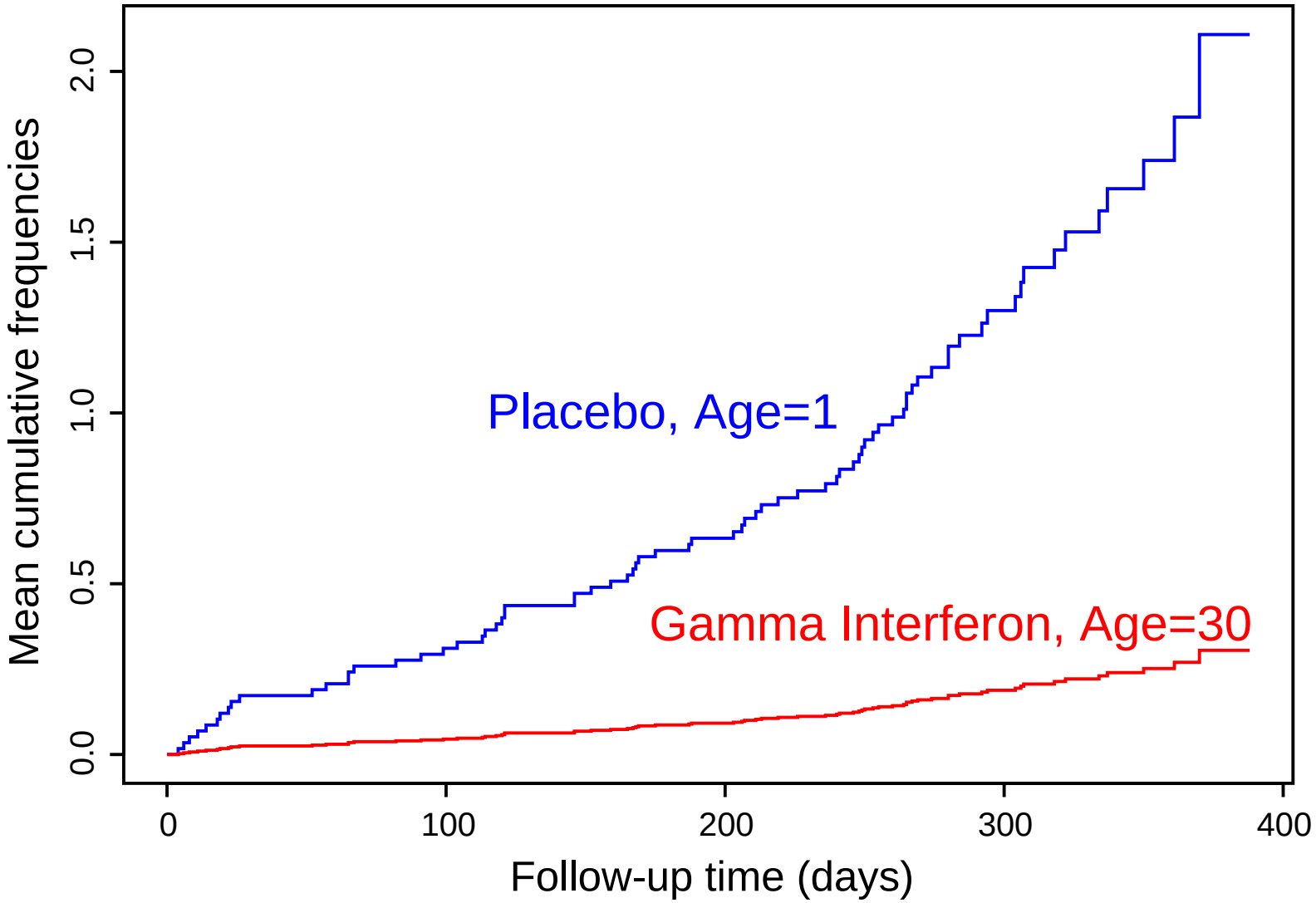
Fig 2. Estimates of Mean Cumulative Frequencies of Infections



General Regression

		Intensity Model		Rate Model	
Covariate	Est	SE	Est/SE	SE	Est/SE
Treatment	1.12	0.26	4.29	0.31	3.62
Age	–0.03	0.013	–2.31	0.014	–2.14

Fig 3. Estimates of Mean Cumulative Frequencies of Infections



V. FRAILTY (RANDOM-EFFECTS) MODELS

1. Methods

Models:

Multiple events:

$$\lambda_k(t|Z_{ki}, b_i) = \lambda_{k0}(t)e^{\beta' Z_{ki}(t)+b_i'\widetilde{Z}_{ki}(t)},\ k = 1, \ldots, K; i = 1, \ldots, n$$

Clustered data:

$$\lambda(t|Z_{ki}, b_i) = \lambda_0(t)e^{\beta' Z_{ki}(t)+b_i'\widetilde{Z}_{ki}(t)},\quad k = 1, \ldots, K_i;\ i = 1, \ldots, n$$

Recurrent events:

$$\lambda(t|Z_i, b_i) = \lambda_0(t)e^{\beta' Z_i(t)+b_i'\widetilde{Z}_i(t)},\ i = 1, \ldots, n$$

- $b_i \sim f(b; \gamma)$
- $\widetilde{Z} \subset Z$

Features:

- Population-average vs subject-specific interpretations
- Efficient estimation of covariate effects
- Accurate prediction of related events
- Characterization of intra-class dependence
- Adjustment of dependent censoring
- Joint modelling for several types of outcome measures

Likelihood (clustered data):

$$\prod_{i=1}^{n} \int_b \prod_{k=1}^{K_i} \left\{ e^{\beta' Z_{ki}(X_{ki}) + b' \widetilde{Z}_{ki}(X_{ki})} \lambda_0(X_{ki}) \right\}^{\Delta_{ki}}$$

$$\times \exp\left\{ - \int_0^{X_{ki}} e^{\beta' Z_{ki}(t) + b' \widetilde{Z}_{ki}(t)} d\Lambda_0(t) \right\} f(b;\gamma) db$$

- allow discrete Λ_0 and replace $\lambda_0(t)$ by jump size of Λ_0 at t
- maximize over β, γ and jump sizes of Λ_0 at observed event times
- (nonparametric) MLEs: $\widehat{\beta}$, $\widehat{\gamma}$, $\widehat{\Lambda}_0$
- Likelihoods for multiple events and recurrent events are similar

Asymptotic Properties:

Consistency

Asymptotic normality

Asymptotic efficiency

Variance Estimation:

Fisher information matrix for β, γ and jump sizes of Λ_0 at observed event times

Profile likelihood

Computation:

EM algorithm

- E-step: numerical integration for conditional expectations
- M-step: similar to partial likelihood and Breslow estimators

SOFTWARE

S-Plus functions:

Multiple events:

coxph(Surv(time, status) ∼ covariates + frailty (pid, dist='Gauss') + strata(enum), method='breslow')

Clustered data:

coxph(Surv(time, status) ∼ covariates +frailty (pid, dist='Gauss'), method='breslow')

Recurrent events:

coxph(Surv(start, stop, status) ∼ covariates + frailty(pid,dist='Gauss'), method='breslow')

Note: The implementation is based on penalized partial likelihood

Stata: available soon

Matlab code: http://www.bios.unc.edu/∼dzeng/Transform.html

EXAMPLES

Colon Cancer Study:

$$Z_1 = \begin{cases} 0 & \text{observation,} \\ 1 & \text{Lev+5-FU;} \end{cases}$$

$$Z_2 = \begin{cases} 0 & \text{surgery} \leq 20 \text{ days before randomization,} \\ 1 & \text{surgery} > 20 \text{ days before randomization;} \end{cases}$$

$$Z_3 = \begin{cases} 0 & \text{submucosa or muscular layer invasion,} \\ 1 & \text{serosa invasion;} \end{cases}$$

$$Z_4 = \begin{cases} 0 & \text{nodes} = 1 - 4, \\ 1 & \text{nodes} > 4. \end{cases}$$

	Cancer recurrence		Death	
Parameter	Estimate	St. error	Estimate	St. error
Treatment	–1.48	0.24	–0.72	0.28
Surgery	–0.69	0.22	–0.64	0.26
Depth of invasion	2.24	0.41	1.94	0.43
Node	2.89	0.24	3.10	0.27
Variance component	11.6	1.2		

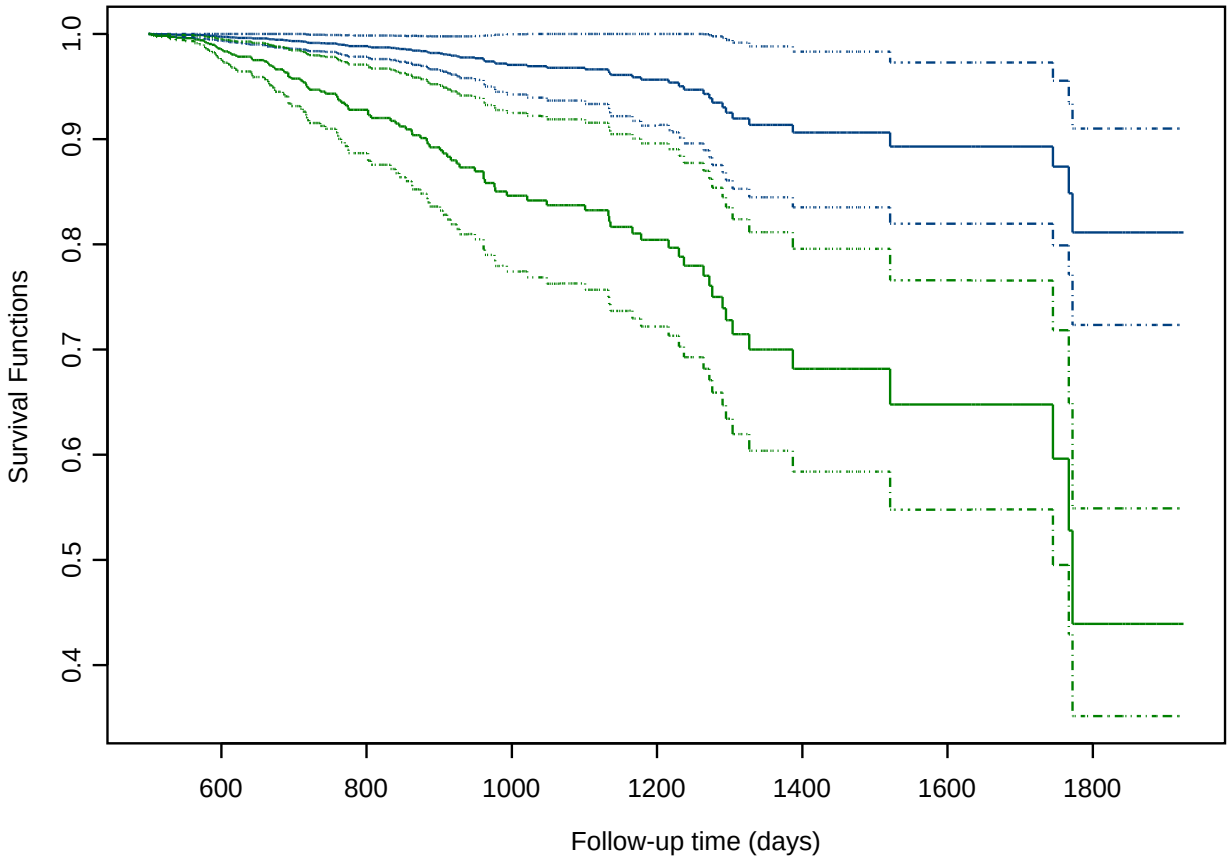


Figure 1: Fig. 4

Diabetic Retinopathy Study

	Marginal Model		Frailty Model	
Parameter	Estimate	St. error	Estimate	St. error
Treatment	–0.43	0.19	–0.52	0.23
Diabetic type	0.34	0.20	0.42	0.26
Interaction	–0.85	0.30	–1.00	0.37
Variance component			1.04	0.19

CGD Study:

		Intensity Model		Rate Model	
	Est	SE	Est/SE	SE	Est/SE
First Event	1.09	0.34	3.27	0.34	3.27
All Events					
"Markov"	1.10	0.26	4.20	0.31	3.53
"Semi-Markov"	0.99	0.27	3.72	0.29	3.36

Intensity Model With Normal Random Effect

Parameter	Est	SE	Est/SE
Treatment	1.05	0.31	3.40
Variance component	0.60	0.07	8.28

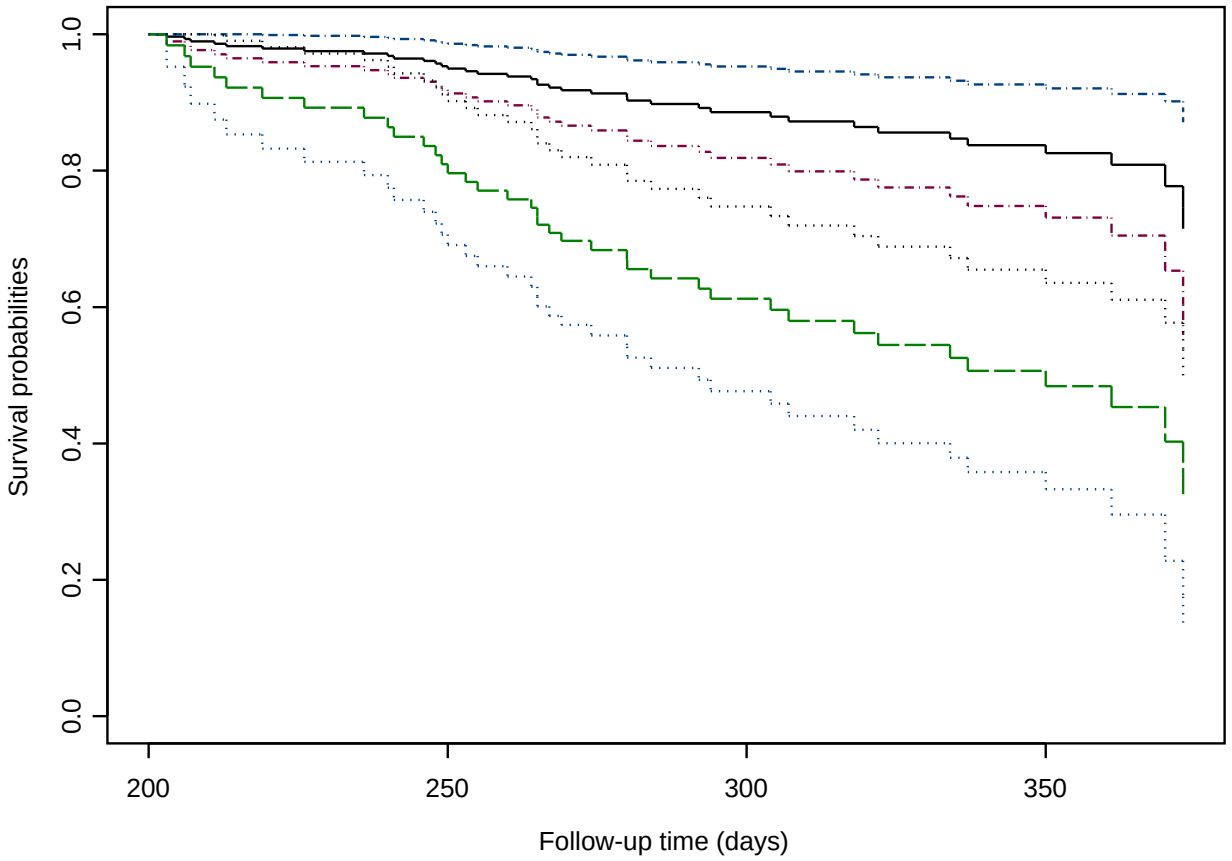


Figure 2: Fig. 5

VI. JOINT MODELS

Generalized Linear Mixed Model:

$$g\{E(Y_{ij}|X_{ij}, b_i\} = \alpha^T X_{ij} + b_i^T \widetilde{X}_{ij}$$

- Y_{ij} = response of ith subject at jth occasion
- X_{ij} = covariates of ith subject at jth occasion
- g = link function
- α = regression parameters
- $\widetilde{X} \subset X$
- $b_i \sim f(b; \gamma)$

Random-Effects Proportional Hazards Model:

$$\lambda(t|Z_i, b_i) = \lambda_0(t) e^{\beta' Z_i(t) + (\psi \circ b_i)' \widetilde{Z}_i(t)}$$

- $\widetilde{Z}_i$ = subset of Z_i
- ψ = unknown constants
- $v_1 \circ v_2$ = component-wise product of v_1 and v_2

Likelihood:

$$\prod_{i=1}^{n} \int_b \lambda(X_i|Z_i,b)^{\Delta_i} \exp\left\{ -\int_0^{X_i} \lambda(t|Z_i,b)dt \right\} \prod_{j=1}^{n_i} f_y(Y_{ij}|X_{ij};b) f(b;\gamma) db$$

- $f_y(\cdot|X;b)$ = conditional density of Y given X and b
- non-informative censoring and measurement times

Inference: Similar to frailty models (Zeng & Lin, 2007)

Software:

Stata: available soon

Matlab code: http://www.bios.unc.edu/∼dzeng/Transform.html

HIV Study

A clinical trial was conducted to evaluate the benefit of switching from zidovudine (AZT) to didanosine (ddI) for HIV patients who have tolerated AZT for at least 16 weeks (Lin and Ying, 2003). The investigators were interested in comparing the CD4 cell counts between the two groups at weeks 8, 16 and 24.

Group	Patients	Dropouts
AZT	304	174
ddI	298	147

	Transformation function			
	Logarithmic		Square-root	
	Est	SE	Est	SE
CD4 Counts				
ddI	0.506	0.215	0.613	0.261
Time	-0.041	0.005	-0.041	0.004
Dropout Time				
ddI	-0.316	0.116	-0.328	0.118
σ^2	7.421	0.575	8.994	0.772
ψ	-0.132	0.021	-0.154	0.023

APPENDIX. NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION

1. Standard Cox Model

Model: $\lambda(t|Z) = e^{\beta' Z}\lambda_0(t)$

Data: (X_i, Δ_i, Z_i) $(i = 1, \cdots, n)$

Likelihood:

$$L(\beta, \Lambda_0) = \prod_{i=1}^{n} \left\{ e^{\beta' Z_i} \lambda_0(X_i) \right\}^{\Delta_i} \exp\left\{ -e^{\beta' Z_i} \Lambda_0(X_i) \right\}$$

$$\widetilde{L}(\beta, \Lambda_0) = \prod_{i=1}^{n} \left(e^{\beta' Z_i} \lambda_i \right)^{\Delta_i} \exp\left(-e^{\beta' Z_i} \sum_{j: X_j \le X_i} \lambda_j \right)$$

Maximum likelihood estimation:

For fixed β, $\widetilde{L}(\beta, \Lambda_0)$ is maximized at

$$\lambda_i = \frac{\Delta_i}{\sum_{j \in \mathcal{R}_i} e^{\beta' Z_j}}, \quad i = 1, \cdots, n$$

Profile likelihood for β:

$$PL(\beta) = \sup_{\Lambda_0} \widetilde{L}(\beta, \Lambda_0) \propto \prod_{i=1}^{n} \left\{ \frac{e^{\beta' Z_i}}{\sum_{j \in \mathcal{R}_i} e^{\beta' Z_j}} \right\}^{\Delta_i}$$

NPMLEs:

$$\widehat{\beta} = \text{argmax} PL(\beta)$$

$$\widehat{\Lambda}(t; \widehat{\beta}) = \sum_{i: X_i \leq t} \frac{\Delta_i}{\sum_{j \in \mathcal{R}_i} e^{\widehat{\beta}' Z_j}}$$

2. Frailty Model for Clustered Failure Time Data

Data: $(X_{ki}, \Delta_{ki}, Z_{ki})$ $(k = 1, \cdots, K_i; i = 1, \cdots, n)$

Model: $\lambda(t|Z_{ki}, \xi_i) = \xi_i e^{\beta' Z_{ki}} \lambda_0(t)$, $k = 1, \cdots, K_i; i = 1, \cdots, n$

- $\xi_i \sim f(\xi; \gamma)$

Likelihood:

$$L(\beta, \gamma, \Lambda_0) = \prod_{i=1}^{n} \int_{\xi_i} \prod_{k=1}^{K_i} \left\{ \xi_i e^{\beta' Z_{ki}} \lambda_0(X_{ki}) \right\}^{\Delta_{ki}} \exp\left\{ -\xi_i e^{\beta' Z_{ki}} \Lambda_0(X_{ki}) \right\} f(\xi_i; \gamma) d\xi_i$$

NPMLEs: maximization of $L(\beta, \gamma, \Lambda_0)$ with right-continuous Λ_0

EM algorithm:

$$L_F(\beta,\gamma,\Lambda_0)=\prod_{i=1}^{n}\prod_{k=1}^{K_i}\left\{\xi_i e^{\beta' Z_{ki}}\lambda_0(X_{ki})\right\}^{\Delta_{ki}}\exp\left\{-\xi_i e^{\beta' Z_{ki}}\Lambda_0(X_{ki})\right\}$$

$$\ell_F(\beta,\gamma,\Lambda_0)=\sum_{i=1}^{n}\sum_{k=1}^{K_i}\left[\Delta_{ki}\left\{\log\xi_i+\beta' Z_{ki}+\log\lambda_0(X_{ki})\right\}-\xi_i e^{\beta' Z_{ki}}\Lambda_0(X_{ki})\right]$$

$$\beta:\ \sum_{i=1}^{n}\sum_{k=1}^{K_i}\Delta_{ki}\left[Z_{ki}-\frac{\sum_{j=1}^{n}\sum_{m=1}^{K_j}I(X_{mj}\geq X_{ki})\widehat{\xi}_j e^{\beta' Z_{mj}}Z_{mj}}{\sum_{j=1}^{n}\sum_{m=1}^{K_j}I(X_{mj}\geq X_{ki})\widehat{\xi}_j e^{\beta' Z_{mj}}}\right]=0$$

$$\Lambda_0(t):\ \sum_{i=1}^{n}\sum_{k=1}^{K_i}\frac{I(X_{ki}\leq t)\Delta_{ki}}{\sum_{j=1}^{n}\sum_{m=1}^{K_j}I(X_{mj}\geq X_{ki})\widehat{\xi}_j e^{\beta' Z_{mj}}}$$

$$\gamma:\ \sum_{i=1}^{n}E\{\partial\log f(\xi_i;\gamma)/\partial\gamma|\text{Data}\}=0$$

- $\widehat{\xi}_i=E(\xi_i|\text{Data})$

Figure Legend

Fig. 1. Estimates and pointwise 95% confidence intervals for the survival functions, namely, the probabilities of retained visual acuity, for adult onset diabetes, separated by treatment groups.

Fig. 2. Cumulative frequencies of infections for CGD patients.

Fig. 3. Cumulative frequencies of infections for 14-year old CGD patients: (a) receiving gamma interferon; (b) receiving no gamma interferon. The point estimates are shown by the solid curves, the pointwise 95% confidence limits by the dashed curves, and the 95% simultaneous confidence bands by the dotted curves.

Fig. 4. Survival probabilities of the colon cancer patients with cancer recurrences at days 500. The blue and green curves pertain to $z = (1, 1, 0, 0)$ and $z = (0, 0, 1, 1)$, respectively.

Fig. 5. Recurrence-free probabilities given first infection at 200 days: treated patient aged 20 vs placebo patient aged 5.

REFERENCES

Wei, L. J., Lin, D. Y. & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Stat. Ass.*, **84**, 1065–1073.

Lee, E. W., Wei, L. J. & Amato, D. (1992). Cox-type regression analysis for large number of small groups of correlated failure time observations. In *Survival Analysis, State of the Art*, Eds. J. P. Klein & P. K. Goel, pp. 237–247. Netherlands: Kluwer Academic Publishers.

Spiekerman, C. F. & Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Am. Stat. Ass.*, **93**, 1164–1175.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Stat. Med.*, **13**, 2233–2247.

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Stat.*, **10**, 1100–1120.

Lawless, J. F. & Nadeau, C. (1995). Some simple and robust methods

for the analysis of recurrent events. *Technometrics*, **37**, 158–168.

Pepe, M. S. & Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Am. Stat. Ass.*, **88**, 811–820.

Lin, D. Y., Wei, L. J., Yang, I. & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Stat. Soc.*, B, **62**, 711–730.

Zeng, D. & Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *J. R. Stat. Soc.*, B, **69**, 507–564.

Lin, D. Y. & Ying, Z. (2003). Semiparametric regression analysis of longitudinal data with informative drop-outs. *Biostatistics*, **4**, 385–398.