Case Study

Since this chapter uses only the simplest of those results—writing a finite sum as an integral with respect to a counting measure—a brief reading of Appendix A is more than sufficient preparation for this chapter.

0.2 A DATA SET AND SOME EXAMPLES

Between January, 1974 and May, 1984, the Mayo Clinic conducted a double-blinded randomized trial in primary biliary cirrhosis of the liver (PBC), comparing the drug D-penicillamine (DPCA) with a placebo. There were 424 patients who met the eligibility criteria seen at the Clinic while the trial was open for patient registration. Both the treating physician and the patient agreed to participate in the randomized trial in 312 of the 424 cases. The date of randomization and a large number of clinical, biochemical, serologic, and histologic parameters were recorded for each of the 312 clinical trial patients. The data from the trial were analyzed in 1986 for presentation in the clinical literature. For that analysis, disease and survival status as of July, 1986, were recorded for as many patients as possible. By that date, 125 of the 312 patients had died, with only 11 deaths not attributable to PBC. Eight patients had been lost to follow up, and 19 had undergone liver transplantation. Appendix D contains a subset of the data from the 1986 analysis.

PBC is a rare but fatal chronic liver disease of unknown cause, with a prevalence of about 50-cases-per-million population. The primary pathologic event appears to be the destruction of interlobular bile ducts, which may be mediated by immunologic mechanisms. The data discussed here, and in greater detail in Chapter 4, are important in two respects. First, controlled clinical trials are difficult to complete in rare diseases, and this case series of patients uniformly diagnosed, treated, and followed is the largest existing for PBC. The treatment comparison in this trial is more precise than in similar trials having fewer participants and avoids the bias that may arise in comparing a case series to historical controls. Second, the data present an opportunity to study the natural history of the disease. We will see that, despite the immunosuppressive properties of DPCA, there are no detectable differences between the distributions of survival times for the DPCA and placebo treatment groups. This suggests that these groups can be combined in studying the association between survival time from randomization and clinical and other measurements. In the early to mid 1980s, the rate of successful liver transplant increased substantially, and transplant has become an effective therapy for PBC. The Mayo Clinic data set is therefore one of the last allowing a study of the natural history of PBC in patients who were treated with only supportive care or its equivalent. The PBC data are used in three examples: estimating a survival distribution; testing for differences between two groups; and estimation based on a regression model.

In the notation that follows, I_A is the indicator of the event A, taking value one if A occurs, and zero otherwise. All vectors are considered column vectors, and are denoted in boldface, as in Z; Z′ denotes the transpose of the column vector Z.

: SOME PRELIMINARIES

$_{k_n^{(\omega)}})\},$

ding for each $W_l^{(n)}$ and

ch ... Condition (2.23)
di ... 2.23) of Chapter

$) - U_{l'}^{(n)}(t_{j,k_n})\}|\mathcal{F}_{t_{j,k_n}}^{(n)}\Big]$

(7.3)

', by Lemma C.7.1 and
) to show one sequence
$t \in [0, \tau]$. If we choose

$\Big\}$,

.22), (2.23) of Chapter 5 □

APPENDIX D

http://lib.stat.cmu.edu/datasets/

link 'pbc'

Data

D.1 PBC DATA

The following pages contain the data from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A description of the clinical background for the trial and the covariates recorded here is in Chapter 0, especially Section 0.2. A more extended discussion can be found in Dickson, et al. (1989) and Markus, et al. (1989).

A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo controlled trial of the drug D-penicillamine. The first 312 cases in the data set participated in the randomized trial, and contain largely complete data. The additional 112 cases did not participate in the clinical trial, but consented to have basic measurements recorded and to be followed for survival. Six of those cases were lost to follow-up shortly after diagnosis, so there are data here on an additional 106 cases as well as the 312 randomized participants. Missing data items are denoted by ".".

The variables contained here are:

- N : Case number.
- X : The number of days between registration and the earlier of death, liver transplantation, or study analysis time in July, 1986.
- δ : 1 if X is time to death, 0 if time to censoring, with an asterisk denoting that censoring was due to liver transplantation.
- Z_1 : Treatment Code, 1 = D-penicillamine, 2 = placebo.
- Z_2 : Age in years. For the first 312 cases, age was calculated by dividing the number of days between birth and study registration by 365.
- Z_3 : Sex, 0 = male, 1 = female.
- Z_4 : Presence of ascites, 0 = no, 1 = yes.
- Z_5 : Presence of hepatomegaly, 0 = no, 1 = yes.
- Z_6 : Presence of spiders, 0 = no, 1 = yes.

Z_7 : Presence of edema, 0 = no edema and no diuretic therapy for edema; 0.5 = edema present for which no diuretic therapy was given, or edema resolved with diuretic therapy; 1 = edema despite diuretic therapy.

Z_8 : Serum bilirubin, in mg/dl.

Z_9 : Serum cholesterol, in mg/dl.

Z_{10} : Albumin, in gm/dl.

Z_{11} : Urine copper, in μg/day.

Z_{12} : Alkaline phosphatase, in U/liter.

Z_{13} : SGOT, in U/ml.

Z_{14} : Triglycerides, in mg/dl.

Z_{15} : Platelet count; coded value is number of platelets per-cubic-milliliter of blood divided by 1000.

Z_{16} : Prothrombin time, in seconds.

Z_{17} : Histologic stage of disease, graded 1, 2, 3, or 4.

N	X	δ	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	Z_{13}	Z_{14}	Z_{15}	Z_{16}	Z_{17}
1	400	1	1	58.7652	1	1	1	1	1.0	14.5	261	2.60	156	1718.0	137.95	172	190	12.2	4
2	4500	0	1	56.4463	1	0	1	1	0.0	1.1	302	4.14	54	7394.8	113.52	88	221	10.6	3
3	1012	1	1	70.0726	0	0	0	0	0.5	1.4	176	3.48	210	516.0	96.10	55	151	12.0	4
4	1925	1	1	54.7406	1	0	1	1	0.5		244	2.54	64	6121.8	60.63	92	183	10.3	4
5	1504	0*	2	38.1054	1	0	1	1	0.0		279	3.53	143	671.0	113.15	72	136	10.9	3
6	2503	1	2	66.2587	1	0	1	0	0.0		248	3.98	50	824.0	96.21	63	204	11.0	3

N	X	δ	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	Z_{13}	Z_{14}	Z_{15}	Z_{16}	Z_{17}
1	400	1	1	58.7652	1	1	1	1	1.0	14.5	261	2.60	156	1718.0	137.95	172	190	12.2	4
2	4500	0	1	56.4463	1	0	1	1	0.0	1.1	302	4.14	54	7394.8	113.52	88	221	10.6	3
3	1012	1	1	70.0726	0	0	0	0	0.5	1.4	176	3.48	210	516.0	96.10	55	151	12.0	4
4	1925	1	1	54.7406	1	0	1	1	0.5	1.8	244	2.54	64	6121.8	60.63	92	183	10.3	4
5	1504	0*	2	38.1054	1	0	1	1	0.0	3.4	279	3.53	143	671.0	113.15	72	136	10.9	3
6	2503	1	2	66.2587	1	0	1	0	0.0	0.8	248	3.98	50	944.0	93.00	63		11.0	3
7	1832	0	2	55.5346	1	0	1	0	0.0	1.0	322	4.09	52	824.0	60.45	213	204	9.7	3
8	2466	1	2	53.0568	1	0	0	0	0.0	0.3	280	4.00	52	4651.2	28.38	189	373	11.0	3
9	2400	1	1	42.5079	1	0	0	1	0.0	3.2	562	3.08	79	2276.0	144.15	88	251	11.0	2
10	51	1	2	70.5599	1	1	0	1	1.0	12.6	200	2.74	140	918.0	147.25	143	302	11.5	4
11	3762	1	2	53.7139	1	0	1	1	0.0	1.4	259	4.16	46	1104.0	79.05	79	258	12.0	4
12	304	1	2	59.1376	1	0	0	1	0.0	3.6	236	3.52	94	591.0	82.15	95	71	13.6	4
13	3577	0	2	45.6893	1	0	0	0	0.0	0.7	281	3.85	40	1181.0	88.35	130	244	10.6	3
14	1217	1	2	56.2218	0	1	1	0	1.0	0.8		2.27	43	728.0	71.00		156	11.0	4
15	3584	1	1	64.6461	1	0	0	0	0.0	0.8	231	3.87	173	9009.8	127.71	96	295	11.0	3
16	3672	0	2	40.4435	1	0	0	0	0.0	0.7	204	3.66	28	685.0	72.85	58	198	10.8	3
17	769	1	2	52.1834	1	0	1	0	0.0	2.7	274	3.15	159	1533.0	117.80	128	224	10.5	4
18	131	1	1	53.9302	1	0	1	1	1.0	11.4	178	2.80	588	961.0	280.55	200	283	12.4	4
19	4232	0	1	49.5606	1	0	1	0	0.5	0.7	235	3.56	39	1881.0	93.00	123	209	11.0	3
20	1356	1	2	59.9535	1	0	1	0	0.0	5.1	374	3.51	140	1919.0	122.45	135	322	13.0	4
21	3445	0	2	64.1889	0	0	1	1	0.0	0.6	252	3.83	41	843.0	65.10	83	336	11.4	4
22	673	1	1	56.2765	1	0	0	1	0.0	3.4	271	3.63	464	1376.0	120.90	55	173	11.6	4
23	264	1	2	55.9671	1	1	1	1	1.0	17.4	395	2.94	558	6064.8	227.04	191	214	11.7	4
24	4079	1	1	44.5202	0	0	1	0	0.0	2.1	456	4.00	124	5719.0	221.88	230	70	9.9	2
25	4127	0	2	45.0732	1	0	0	0	0.0	0.7	298	4.10	40	661.0	106.95	66	324	11.3	2
26	1444	1	2	52.0246	1	0	1	1	0.0	5.2	1128	3.68	53	3228.0	165.85	166	421	9.9	3
27	77	1	2	54.4394	1	1	1	1	0.5	21.6	175	3.31	221	3697.4	101.91	168	80	12.0	4
28	549	1	2	44.9473	1	1	1	1	1.0	17.2	222	3.23	209	1975.0	189.10	195	144	13.0	4
29	4509	0	2	63.8768	1	0	0	0	0.0	0.7	370	3.78	24	5833.0	73.53	86	390	10.6	2

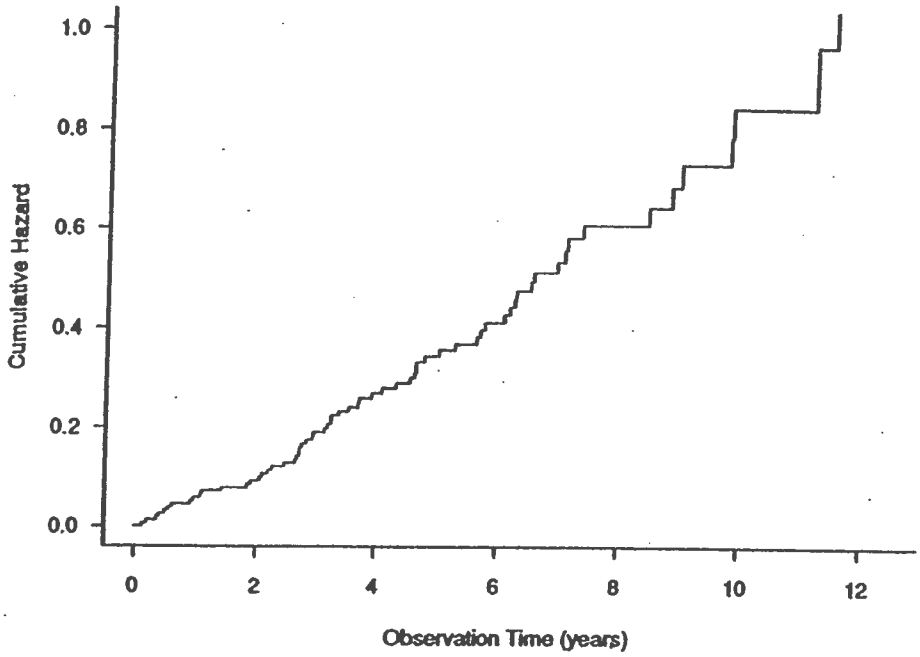


Figure 0.2.1 Nelson cumulative hazard estimate for DPCA group, PBC data.

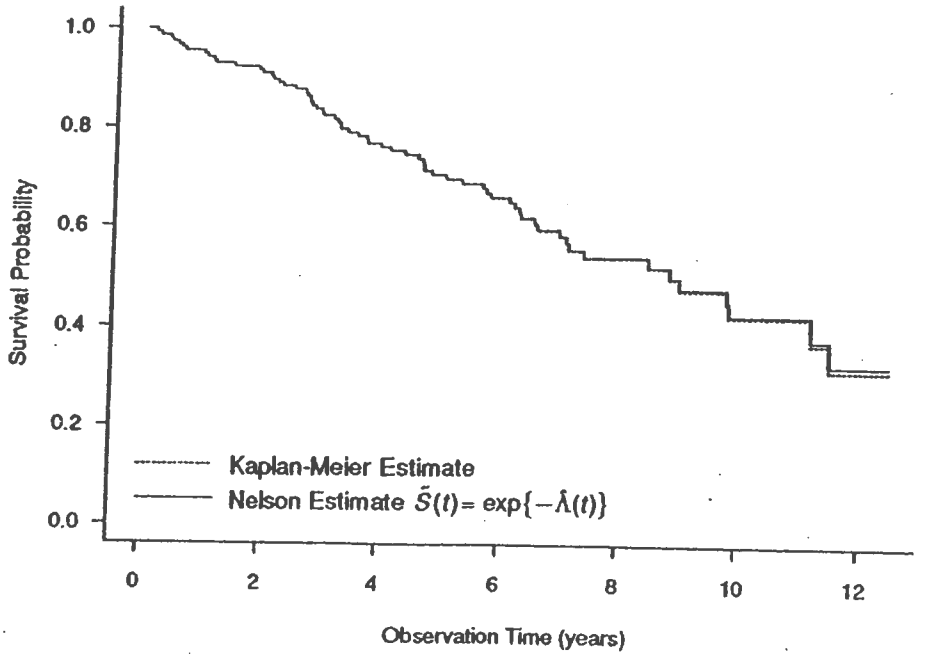
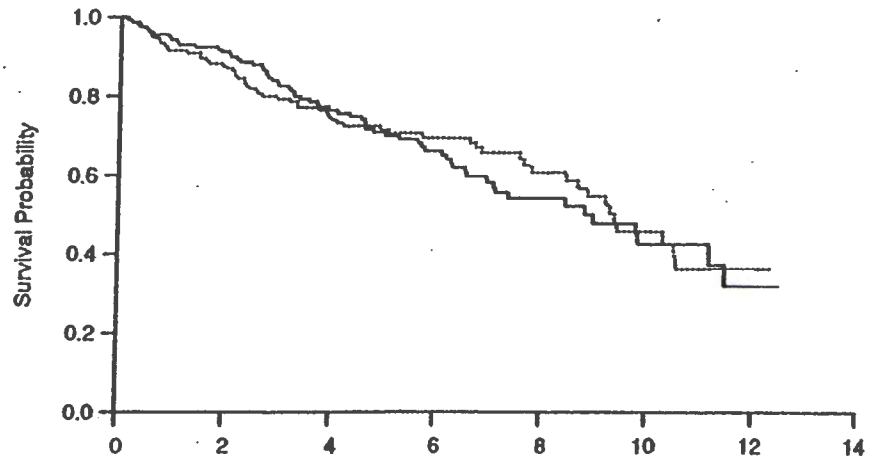


Figure 0.2.2 Estimated survival curves for DPCA group, PBC data.



	Time Interval						
Group	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 0.2.3 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.

Table 0.2.2 shows the estimated regression coefficients and their standard errors obtained by fitting this model with some of the variables or their transformations in the PBC data set. The variables in this model are age, the presence of edema, the treatment variable, and the natural logarithms of albumin, bilirubin, and prothrombin time. Chapter 4 contains a detailed description of how these variables or their transformations were chosen, so we defer that issue for now. Methods for estimating the standard errors (s.e.) of the regression coefficients and for establishing the asymptotic normality of standardized components of $\hat{\beta}$ are discussed in Chapters 4 and 8.

Table 0.2.2 Regression coefficients and their standard errors from an estimated proportional hazards model in the PBC data.

	$\hat{\beta}$	Standard Error (s.e.)	$\hat{\beta}/\text{s.e.}\hat{\beta}$
Age	0.0347	0.00891	3.89
log (Albumin)	−3.0771	0.71899	−4.28
log (Bilirubin)	0.8840	0.09871	8.96
Edema	0.7859	0.29647	2.65
log (Prothrombin Time)	2.9707	1.01588	2.92
Treatment	0.1360	0.18543	0.73

$$\hat{\Lambda}_{0q}(t) = \int_0^t \left[\sum_{i=1}^{n_q} Y_{qi}(s)\exp\{\hat{\beta}'Z_{qi}(s)\}\right]^{-1} \left\{\sum_{i=1}^{n_q} dN_{qi}(s)\right\}. \quad (3.30)$$

Section 4 illustrates the use of partial likelihood based inference in censored data regression models with a complex data set.

4.4 APPLICATIONS OF PARTIAL LIKELIHOOD METHODS

The next three sections illustrate the usefulness of proportional hazards and multiplicative intensity regression model tools in the analysis of censored survival data. This section provides background for the liver disease natural history data set and illustrates the use of partial likelihood based inference procedures. An investigation of gamma interferon in chronic granulomatous disease also is presented to show an application of the multiplicative intensity model to a data set having repeated outcome events. In Section 4.5, graphical methods and methods for analysis of residuals are discussed, using the structure of case specific martingales and martingale transforms. These methods allow a graphical approach to model building, variable transformation, checking the proportional hazards assumption, and finding leverage points and outliers. Section 4.6 then displays the application of these graphical methods to data sets, including the liver disease natural history data. Special routines were added to the UNIX statistical language S for these analyses.

Background: Liver Data

The Mayo Clinic has established a database of 424 patients having primary biliary cirrhosis (PBC). These 424 form the complete collection of all PBC patients, referred to Mayo between January 1974 and May 1984, who met standard eligibility criteria for a randomized, double-blinded, placebo-controlled, clinical trial of the drug D-penicillamine (DPCA). The patient and treating physician agreed to randomization in 312 of the 424 cases. For each of the 312 clinical trial patients, clinical, biochemical, serologic, and histologic parameters were collected. For this analysis, complete follow up to July, 1986, was attempted on all patients. By this date, 125 of the 312 had died, with only 11 deaths not attributable to PBC. Only eight were lost to follow up, and 19 had undergone liver transplantation. Appendix D contains the survival data and the entry values of the important covariates.

Because PBC is a rare disease (the prevalence of the disease has been estimated to be 50 cases-per-million population), this database is a valuable resource to liver specialists. PBC is a fatal chronic liver disease of unknown cause. The primary pathologic event appears to be destruction of interlobular bile ducts, which may be mediated by immunological mechanisms. Results of the clinical trial of 312 patients established that DPCA is not effective in PBC, in spite of the drug's immunosuppressive properties. Until recently, effective treatments for PBC did not exist, and the approach to patients with the disease was limited to supportive care.

In the early to mid 1980s, the rate of liver transplantation in patients with advanced stage PBC increased substantially, largely due to the improvement in transplantation results through the use of immunosuppressive agents such as cyclosporine and OKT3.

We first show that DPCA has a negligible effect on prognosis, then use the data from the 312 randomized cases to develop a natural history model. Such a model will be useful not only in counseling patients and in understanding the course of PBC in untreated patients, but also in providing historical control information to evaluate the efficacy of new therapeutic interventions such as liver transplantation. These evaluations of liver transplantation will be important since randomized trials comparing transplantation with non-surgical management will not be performed and since PBC is one of the most common indicators for liver transplantation in adults. In this Chapter, we present the analyses to evaluate DPCA, to develop a natural history model, and to illustrate the model's use in evaluating liver transplantation. The data set of 112 nonrandomized patients is used in model validation of the natural history model and to illustrate its use in survival prediction. Appendix D contains the data for 106 of the 112 patients, since six were lost to followup soon after their initial visit to the Mayo Clinic. Of the 106, 36 had died by July 1986, with six others having undergone transplantation.

In the database of 418 patients, the 25 transplanted patients were considered censored at the date of transplantation. This induces a small bias in a natural history model. Transplantation occurred after a median followup of 66 months for the 19 transplanted clinical trial patients and 50 months for the 6 transplanted non-trial patients.

(1)

(2)

Effect of DPCA on Patient Survival

Figure 4.4.1 presents the Kaplan-Meier estimates of survival of PBC patients following randomization to either DPCA or placebo. The curves show little separation. The median survival time of the pooled group is just under 10 years.

Under the proportional hazards assumption, the Cox regression model can be used to measure treatment effect. If treatment is coded by $Z = 0$: DPCA, $Z = 1$: Placebo, then in the model

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z), \tag{4.1}$$

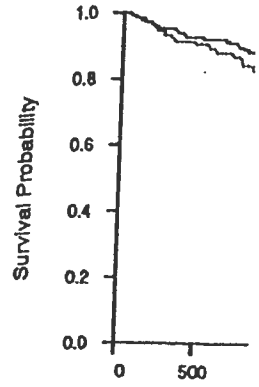
$\lambda_0(\cdot)$ represents the hazard function for death while being treated with DPCA, and β is the log of the hazard ratio; i.e., if $\lambda_1(t) \equiv \lambda(t|Z = 1)$ then, for all t,

$$\lambda_1(t)/\lambda_0(t) = e^{\beta}.$$

If $L(\beta)$ is the Cox partial likelihood for β based on the censored survival data and $\mathcal{L} \equiv \ln L$, then the score statistic is given by

$$U(\beta) \equiv \frac{d}{d\beta}\mathcal{L}(\beta),$$

PARTIAL LIKELIHOOD AI



Group 0-10
DPCA 23/1
Placebo 31/1

Figure 4.4.1 Estin

and Fisher's observed ir

For these data $U(0)$:
Hence the standardized

Later it will be establish
square with one degree
covariates in this model
no treatment effect.

The maximum parti
$\mathcal{L}(-0.0571242) = -63$
pothesis $H : \beta = 0$ is

Since the standard error
the Wald statistic for H

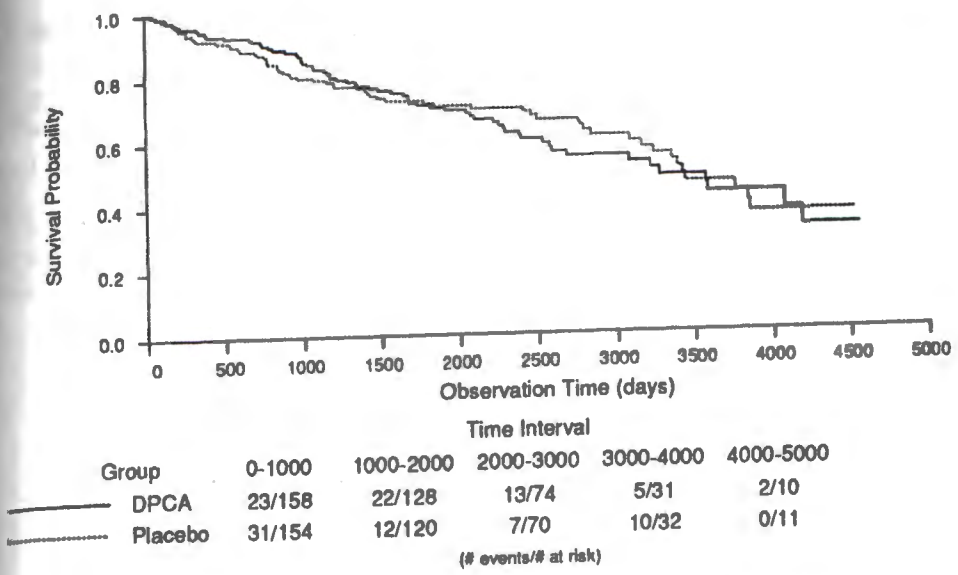


Figure 4.4.1 Estimated survival curves in DPCA and placebo groups, PBC data.

and Fisher's observed information is

$$\mathcal{I}(\beta) = -\frac{d^2}{d\beta^2}\mathcal{L}(\beta).$$

For these data $U(0) = -1.781115$, $\mathcal{I}(0) = 31.19845$, and $\mathcal{L}(0) = -639.9799$. Hence the standardized score statistic or Rao test statistic is

$$\{U(0)\}^2/\mathcal{I}(0) = 0.10168.$$

Later it will be established that this statistic is distributed asymptotically as a chi-square with one degree of freedom when $H : \beta = 0$. Since there are no nuisance covariates in this model, this score statistic is identical to the logrank statistic for no treatment effect.

The maximum partial likelihood estimate for β is $\hat{\beta} = -0.0571242$ and $\mathcal{L}(-0.0571242) = -639.9290$. Hence, the likelihood ratio statistic for the hypothesis $H : \beta = 0$ is

$$-2\{\mathcal{L}(0) - \mathcal{L}(\hat{\beta})\} = 0.10193.$$

Since the standard error for $\hat{\beta}$ is estimated by $\{\mathcal{I}(\hat{\beta})\}^{-1/2}$, where $\mathcal{I}(\hat{\beta}) = 31.1525$, the Wald statistic for $H : \beta = 0$ is

$$\hat{\beta}^2\mathcal{I}(\hat{\beta}) = 0.10166.$$

As expected, the Rao, Wald, and likelihood ratio statistics yield nearly identical results.

Under the proportional hazards assumption, the hazard ratio

$$r \equiv \lambda_1(t)/\lambda_0(t) = e^{\beta}$$

is independent of t. In large samples,

$$\hat{\beta} \sim N(\beta, \{\mathcal{I}(\hat{\beta})\}^{-1});$$

thus $\hat{r} = e^{\hat{\beta}} = 0.94448$, and a 95% confidence interval for r is

$$\exp\{\hat{\beta} \pm 1.96\{\mathcal{I}(\hat{\beta})\}^{-1/2}\} = (0.66479, 1.34184).$$

We estimate the failure rate on placebo to be 94.4% that on DPCA, and there is evidence against it being more than 134% that on DPCA. If DPCA must improve patient survival by more than a factor of 1/2 to offset the drug's expense, toxicity and inconvenience of administration, then this trial supports not using the drug in this disease.

An analysis of subsets defined by clinical, biochemical and histological variables failed to yield evidence of important survival differences between the drug and the placebo in patient subgroups.

Natural History Model for PBC

The data in Appendix D on the 312 PBC randomized patients can be used to build a statistical model for the influence of covariates on disease outcome. Table 4.4.1 provides the distributions of 14 clinical, biochemical and histological variables. With the exception of 4 missing platelet counts and two missing urine copper values, the data are complete.

For the remainder of this section, we use the model

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)\exp(\beta'\mathbf{Z}), \tag{4.2}$$

where $\mathbf{Z}' \equiv (Z_1, Z_2, \ldots, Z_K)$ is a vector of K predictors and $\beta' \equiv (\beta_1, \beta_2, \ldots, \beta_K)$ are the regression coefficients. Each predictor Z_i could be defined in a variety of ways, such as using the variables in Table 4.4.1, transformations or crossproducts of these variables, etc. In model (4.2), each individual patient is given a risk score $R \equiv \beta_1 Z_1 + \ldots + \beta_K Z_K$. Let $S(t|\mathbf{Z})$ denote the probability that patient with risk factors given by $\mathbf{Z}$ (and with risk score R) is still alive t years after time 0, and let $S_0(t)$ denote the survival function for individuals having risk score $R = 0$. Then

$$\begin{aligned} S(t|\mathbf{Z}) &= \{S_0(t)\}^{\exp(R)} \\ &= \{e^{-\Lambda_0(t)}\}^{\exp(R)}, \end{aligned}$$

where time $t = 0$ usually denotes the time the measurements in the covariate vector $\mathbf{Z}$ are obtained. In the PBC data set in Appendix D, time $t = 0$ is the date of treatment randomization. One can estimate $S(t|\mathbf{Z})$ by

$$\hat{S}(t|\mathbf{Z}) = \{e^{-\hat{\Lambda}_0(t)}\}^{\exp(\hat{R})}, \tag{4.3}$$

where the maximum l
$\hat{\beta}_1 Z_1 + \cdots + \hat{\beta}_K Z_K$,
the proportional hazar
interpretation that ever
function by the multipl
Initially model (4.2
be the 14 variables in
and Fisher's observed i
Rao or logrank statistic
right-hand column of T
of $U_j(0)$ and $U_k(0)$, ar
(3.23), $\{\mathcal{I}(0)\}_{jk}$ is the
among those at risk at

$\{c_{jk} \equiv$

Table 4.4.1 Prognosti
(312 Patients in the PB

Demographic	min
Age (years)	26.3
Sex	
Clinical	
Ascites	
Hepatomegaly	
Spiders	
Edema[1]	0: 263
Biochemical	**min**
Bilirubin	0.3
Albumin	1.96
Urine Copper	4
Pro Time	9.0
Platelet Count	62
Alkaline Phos	289
SGOT	26
Histologic	**1**
Stage	16

[1]Edema 0: No edema and n
$\frac{1}{2}$: Edema but no di
1: Edema despite d
Rao statistics computed afte
values (i.e., 4 missing Platel

where the maximum partial likelihood estimate vector $\hat{\beta}$ is used to obtain $\hat{R} = \hat{\beta}_1 Z_1 + \cdots + \hat{\beta}_K Z_K$, and where $\hat{\Lambda}_0(\cdot)$ is the Breslow estimator in Eq. (3.29). In the proportional hazards model in Eq. (4.2), each coefficient β_k has the simple interpretation that every unit increase in the kth covariate, Z_k, changes the hazard function by the multiplicative factor $\exp(\beta_k)$.

Initially model (4.2) was fit to the data with $Z' = (Z_1, \ldots, Z_{14})$ chosen to be the 14 variables in Table 4.4.1. If $U(\beta)$ and $\mathcal{I}(\beta)$ denote the score vector and Fisher's observed information matrix, respectively, the collection of univariate Rao or logrank statistics, $\{[\{U(0)\}_k]^2/\{\mathcal{I}(0)\}_{kk} : k = 1, \ldots, 14\}$, are listed in the right-hand column of Table 4.4.1. The term $\{\mathcal{I}(0)\}_{jk}$ is the estimated covariance of $U_j(0)$ and $U_k(0)$, and since $\mathcal{I}(0) = \int_0^\infty \sum_{i=1}^n V(0,t) dN_i(t)$ for V defined by (3.23), $\{\mathcal{I}(0)\}_{jk}$ is the sum (over death times) of the covariances of Z_j and Z_k among those at risk at each death time. Thus inspection of

$$\{c_{jk} \equiv \{\mathcal{I}(0)\}_{jk}/[\{(\mathcal{I}(0)\}_{jj}\{\mathcal{I}(0)\}_{kk}]^{1/2} : j \neq k\}$$

Table 4.4.1 Prognostic Factors: Summary of Univariate Statistics (312 Patients in the PBC Clinical Trial of DPCA)

Demographic	min	1st Q	med	3rd Q	max	Missing	Rao χ^2(1 d.f.)
Age (years)	26.3	42.1	49.8	56.7	78.4	0	20.86
Sex	male: 36		female: 276			0	4.27

Clinical	Absent	Present	Missing	Rao χ^2(1 d.f.)
Ascites	288	24	0	104.02
Hepatomegaly	152	160	0	40.18
Spiders	222	90	0	30.31
Edema[1]	0: 263	1/2: 29 1: 20	0	97.89

Biochemical	min	1st Q	med	3rd Q	max	Missing	Rao χ^2(1 d.f.)
Bilirubin	0.3	0.8	1.35	3.45	28.0	0	190.62
Albumin	1.96	3.31	3.55	3.80	4.64	0	70.83
Urine Copper	4	41	73	123	588	2	84.35
Pro Time	9.0	10.0	10.6	11.1	17.1	0	51.76
Platelet Count	62	200	257	323	563	4	12.15
Alkaline Phos	289	867	1259	1985	13862	0	2.58
SGOT	26	81	115	152	457	0	29.59

Histologic	1	2	3	4	Missing	Rao χ^2(1 d.f.)
Stage	16	67	120	109	0	46.49

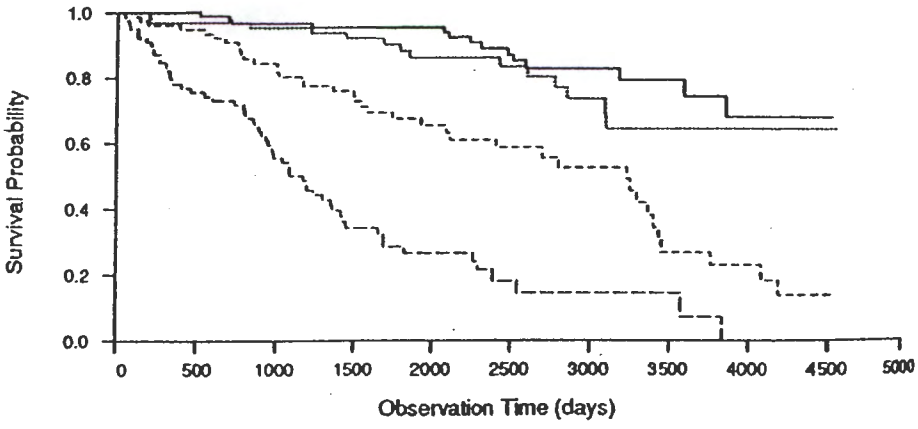
[1] Edema 0: No edema and no diuretic therapy for edema
$\frac{1}{2}$: Edema but no diuretics, or edema resolved by diuretics
1: Edema despite diuretic therapy

Rao statistics computed after six missing values were replaced by median values (i.e., 4 missing Platelet Counts, 257; 2 missing Urine Copper, 73)

is a method for finding co-linearities among the K components of Z. The largest values of c_{jk} are 0.47 between hepatomegaly and stage, 0.37 between bilirubin and SGOT, and 0.37 between bilirubin and urine copper. Bilirubin is the strongest univariate predictor of survival. One would expect and can verify that the predictive strength of the variables SGOT and urine copper are reduced in models which adjust for bilirubin. In building a parsimonious natural history model based on easily accessible variables, there is hope that readily available measurements on hepatomegaly and bilirubin will contain much of the predictive information from the invasive variable histologic stage (which requires a liver biopsy), and in the frequently unmeasured variables, urine copper and SGOT.

The score statistics in Table 4.4.1 show that nearly all 14 variables are highly significantly associated with patient survival. The Kaplan-Meier plot in Figure 4.4.2 indicates that bilirubin levels distinguish patients with good and poor prognosis.

Parsimonious but accurate models based on inexpensive, non-invasive and readily available measurements are useful in clinical science, and so the variables stage, urine copper, and SGOT were eliminated temporarily from the variable selection process. The untransformed versions of the remaining 11 variables were inserted into Eq. (4.2), and a step-down procedure was employed to eliminate variables, using the Wald statistic as a criterion for deletion of the least predictive variable. Table 4.4.2 displays the first step of the procedure, which led to the elimination of



Group	Time Interval 0-1000	1000-2000	2000-3000	3000-4000	4000-5000
0 < bili < 0.8	3/90	1/86	7/64	3/28	0/10
0.8 < bili < 1.3	3/66	5/62	4/38	2/18	0/6
1.3 < bili < 3.4	14/78	10/59	5/30	8/15	2/5
3.4 < bili	34/78	18/41	4/12	2/2	0/0

Bilirubin measured in mg/dl (# events/# at risk)

Figure 4.4.2 Estimated survival curves for four groups determined by serum bilirubin levels, PBC data.

the variable alkaline phosphatase, and the sixth step, at which each of the remaining variables has a Wald statistic exceeding 6.0. The likelihood ratio test for the five eliminated variables has the value

$$-2(-554.237 + 550.603) = 7.268,$$

and has an approximate chi-square distribution with 5 degrees of freedom. There is little evidence to retain the variables alkaline phosphatase, ascites, platelet count, sex, or presence of spiders.

In the model in Table 4.4.2(b), all variables were entered untransformed. In such a model, an increase in the value of the ith covariate, Z_i, from x to $(x + d)$ will lead to a multiplicative increase in the hazard by a factor $\exp(d\beta_i)$, independent of the value x. However, the clinical literature suggests that changes from x to $(x + d)$ in the values of variables such as bilirubin should have a greater impact on prognosis when x is small. To evaluate the need for transformations of the four continuous variables in the six variable model in Table 4.4.2(b), the variables log(age), log(albumin), log(protime), and log(bilirubin) were added. The resulting

Table 4.4.2 Results of variable selection procedure in 312 randomized cases with PBC.

(a) First Step, log likelihood −550.603

	Coef.	Std. Err.	Z stat.
Age	2.819 e-2	9.538 e-3	2.96
Albumin	−9.713 e-1	2.681 e-1	−3.62
Alk. Phos	1.445 e-5	3.544 e-5	0.41
Ascites	2.813 e-1	3.093 e-1	0.91
Bilirubin	1.057 e-1	1.667 e-2	6.34
Edema	6.915 e-1	3.226 e-1	2.14
Hepatomegaly	4.853 e-1	2.913 e-1	2.21
Platelets	−6.063 e-4	1.025 e-3	−0.59
Prothrombin Time	2.428 e-1	8.420 e-2	2.88
Sex	−4.769 e-1	2.643 e-1	−1.80
Spiders	2.889 e-1	2.093 e-1	1.38

(b) Last Step, log likelihood −554.237

	Coef.	Std. Err.	Z stat.
Age	0.0338	0.00925	3.65
Albumin	−1.0752	0.24103	−4.46
Bilirubin	0.1070	0.01528	7.00
Edema	0.8072	0.30775	2.62
Hepatomegaly	0.5903	0.21179	2.79
Prothrombin Time	0.2603	0.07786	3.34

Results computed after the four patients with missing values for platelets were assigned the median count, 257, from Table 4.4.1.

ten variable model in Table 4.4.3(a) provides a significantly better fit than the model in Table 4.4.2(b); i.e., the likelihood ratio statistic having 4 d.f. is $-2(-554.237+538.274) = 31.926$. It is apparent that the logarithmic transformation of bilirubin provides a substantial improvement and, interestingly, that the dichotomous variable hepatomegaly is no longer independently predictive. Table 4.4.3(b) presents the log likelihood and regression coefficients for the five variable model containing age, albumin, log(bilirubin), edema, and protime. The score statistic for hepatomegaly in that model is only 1.38.

The square and logarithmic transformations of albumin, age, and protime were considered by proceeding "stepwise" in the order of the Z statistics in Table 4.4.3(b) for those untransformed variables. In the model containing age, log(bilirubin), edema and protime, the score statistics for albumin, log(albumin) and $(\text{albumin})^2$

Table 4.4.3 Regression models with log transformations of continuous variables, 312 randomized cases with PBC.

(a) Log likelihood −538.274

	Coef.	Std. Err.	Z stat.
Age	−0.0289	0.07141	−0.41
log(age)	3.2248	3.71828	0.87
Albumin	1.0068	1.73450	0.58
log(Albumin)	−5.8629	5.42315	−1.08
Bilirubin	−0.0461	0.03547	−1.30
log(Bilirubin)	1.0774	0.21127	5.10
Edema	0.8238	0.30386	2.71
Prothrombin Time	−0.6175	1.14523	−0.54
log(Pro Time)	10.1928	13.36131	0.76
Hepatomegaly	0.1964	0.22628	0.87

(b) Log likelihood −541.064

	Coef.	Std. Err.	Z stat.
Age	0.0337	0.00864	3.89
Albumin	−0.9473	0.23713	−3.99
log(Bilirubin)	0.8845	0.09854	8.98
Edema	0.8006	0.29914	2.68
Prothrombin Time	0.2463	0.08426	2.92

(c) Log likelihood −540.412

	Coef.	Std. Err.	Z stat.
Age	0.0333	0.00866	3.84
log(Albumin)	−3.0553	0.72408	−4.22
log(Bilirubin)	0.8792	0.09873	8.90
Edema	0.7847	0.29913	2.62
log(Prothrombin Time)	3.0157	1.02380	2.95

were 15.94, 17.78 and 1
transformation of album
edema and protime, the
14.57 and 14.73. In the
the score statistics for p
8.01 respectively.

The log likelihood, c
transformed variables a
are in Table 4.4.3(c). Th
variables to this five vari
variables or interaction t
possible benefits of add
exercise.

The final model in T
efficient for albumin is
leads to increased hepatc
diminished. The increasi
crete the normal amount
serum bilirubin, a bile pi
which leads to an increa
in tissue, referred to as e

The Breslow estimat
estimates. For an
the 312 trial patien
$\hat{S}(1) = 0.982$ and $\hat{S}(5) =$

Consider a low-risk pa
4.5 g/dl, age, 52 years, p
of diuretic therapy (i.e., c

$\hat{R} = 0.879 * \log(0.5) - 3.($

so $\hat{R} = 3.49$. Her estima

$\hat{S}$

indicating very low risk o

In a high-risk patient
2.8 g/dl, age, 52 years, p
therapy (i.e., edema 0.5),
is

$\hat{S}$

Under most circumstance
for liver transplantation.

were 15.94, 17.78 and 14.10 respectively. This led to the choice of a logarithmic transformation of albumin. In the model containing log(albumin), log(bilirubin), edema and protime, the score statistics for age, log(age) and $(age)^2$ were 15.00, 14.57 and 14.73. In the model with age, log(albumin), log(bilirubin) and edema, the score statistics for protime, log(protime) and $(protime)^2$ were 8.34, 8.51 and 8.01 respectively.

The log likelihood, coefficients, and standard errors for the final model with transformed variables age, log(albumin), log(bilirubin), edema, and log(protime) are in Table 4.4.3(c). The additional model refinement steps which involved adding variables to this five variable model by considering either transformations of these variables or interaction terms failed to yield significantly improved prediction. The possible benefits of adding stage, SGOT or urine copper can be explored as an exercise.

The final model in Table 4.4.3(c) is biologically reasonable. The negative coefficient for albumin is consistent with the fact that, as the progression of PBC leads to increased hepatocellular damage, the liver's ability to produce albumin is diminished. The increasing damage to bile ducts reduces the liver's ability to excrete the normal amount of bilirubin from the blood, which leads to an increase in serum bilirubin, a bile pigment. Prothrombin, a protein in the plasma, is decreased, which leads to an increase in blood coagulation time. The accumulation of fluids in tissue, referred to as edema, often is associated with later stages of the disease.

The Breslow estimate of Λ_0 and Eq. (4.3) provide patient specific survival estimates. For an individual with risk score $\hat{R} = 5.24$, the median risk score in the 312 trial patients, the corresponding one- and five-year survival estimates are $\hat{S}(1) = 0.982$ and $\hat{S}(5) = 0.845$.

Consider a low-risk patient with serum total bilirubin, 0.5 mg/dl, serum albumin, 4.5 g/dl, age, 52 years, prothrombin time, 10.1 seconds, no edema, and no history of diuretic therapy (i.e., edema = 0). Her risk score is

$$\hat{R} = 0.879 * \log(0.5) - 3.053 * \log(4.5) + 0.033 * 52 + 3.016 * \log(10.1) + 0.785 * 0.0,$$

so $\hat{R} = 3.49$. Her estimated 5-year survival probability is

$$\hat{S}(5) = (0.845)^{\exp(3.49-5.24)} = 0.97,$$

indicating very low risk of death in the next five years, even without liver transplant.

In a high-risk patient with serum total bilirubin, 13.9 mg/dl, serum albumin, 2.8 g/dl, age, 52 years, prothrombin time, 13.8 sec., edema responding to diuretic therapy (i.e., edema 0.5), $\hat{R} = 9.19$ and her estimated one-year survival probability is

$$\hat{S}(1) = (0.982)^{\exp(9.19-5.24)} = 0.39.$$

Under most circumstances, such a high-risk patient would be considered a candidate for liver transplantation.

Table 4.4.4 Adjusted estimation of treatment effect, 312 randomized cases with PBC.

(a) Log likelihood −540.144

	Coef.	Std. Err.	Z stat.
Age	0.0347	0.00891	3.89
log(Albumin)	−3.0771	0.71899	−4.28
log(Bilirubin)	0.8840	0.09871	8.96
Edema	0.7859	0.29647	2.65
log(Prothrombin Time)	2.9707	1.01588	2.92
Treatment	0.1360	0.18543	0.73

A score test of the hypothesis that treatment has no effect on survival, when adjusting for the variables in Table 4.4.3(c), yields the chi-square 0.54. By Table 4.4.4, the adjusted 95% confidence interval for the ratio of placebo to DPCA hazard functions is $\exp\{0.136 \pm (0.185)(1.96)\} = (0.797, 1.646)$, which is shifted to the right of the unadjusted confidence interval obtained earlier.

Study of Gamma Interferon in Chronic Granulomatous Disease

Chronic Granulomatous Disease (CGD) is a group of inherited rare disorders of the immune function characterized by recurrent pyogenic infections which usually present early in life and may lead to death in childhood. Phagocytes from CGD patients ingest microorganisms normally but fail to kill them, primarily due to the inability to generate a respiratory burst dependent on the production of superoxide and other toxic oxygen metabolites. Thus, it is the failure to generate microbicidal oxygen metabolites within the phagocytes of CGD patients which confers the greatly increased susceptibility to these severe or even life threatening infections.

There is evidence establishing a role for gamma interferon as an important macrophage activating factor which could restore superoxide anion production and bacterial killing by phagocytes in CGD patients. In order to study the ability of gamma interferon to reduce the rate of serious infections, that is, the rate of infections requiring hospitalization for parenteral antibiotics, a double-blinded clinical trial was conducted in which patients were randomized to placebo vs. gamma interferon. Between October 1988 and March 1989, 128 eligible patients with CGD were accrued by the International CGD Cooperative Study Group. Since the study required delivering placebo injections three times weekly for a twelve month period to one-half of the patients, most being children, there was particular interest in achieving early termination of the trial if early results were extreme. A single interim analysis was to be performed as soon as patient followup was available through July 1989, six-months after the date on which one-half of the patients had been accrued.

At the time of interim analysis, twenty of 65 placebo patients and seven of 63 patients on gamma interferon each had experienced at least one serious infec-

using LOWESS in the language S. Data suggest that patients with G2% exceeding approximately 13% have a much higher rate of progression and death. Entering a dummy indicator $Z_i = 0(1)$ for G2% $< 13\%$ ($\geq 13\%$) into the Cox model for time to progression yielded strong evidence for the association (likelihood ratio statistic = 30.48), and a maximum partial likelihood estimate $\hat{\beta} = 2.27$ with standard error 0.5. Thus a 95% confidence interval for the ratio of progression rates in the $\geq 13\%$ group to the $<13\%$ group is (5.87,15.96). A plot of martingale residuals against G2% from a Cox model containing the indicator variable Z_i reveals (in Figure 4.6.2 on page 181) that the dichotomy at 13% appears to explain much of the association.

To illustrate the assessment of influence, a Cox model with G2% as a linear covariate was fit to the progression data of the 79 patients. Figure 4.6.3, on page 181, shows the standardized score residuals, the Storer-Crowley one-step influence measure, and the actual jackknife. In these data, there are a few subjects for which the Storer-Crowley (SC) influence has the wrong sign. In all these cases the SC is accurately tracking the behavior of the first iteration step after removal of subject i; i.e., a one-step jackknife also has the incorrect sign. Both the score residual and the SC measure accurately approximate the true jackknife.

(4)

Checking Functional Form in the Liver PBC Model

In Section 4.4, Cox partial likelihood methods were used to develop a survival model based on the variables age, log(albumin), log(bilirubin), edema, and log(protime). A sixth variable, hepatomegaly, had been independently predictive until the logarithmic transformation of bilirubin was introduced, (see Tables 4.4.2(b) and 4.4.3(a)). The plots of martingale residuals in Figure 4.6.4, on page 182, provide some insight into what is happening in the data. The set of 312 martingale residuals used to produce the 6 plots were obtained from a single Cox model fitting the variables age, log(albumin), edema, and log(protime). It is obvious that a linear fit to the martingale residuals is much more appropriate when the residuals are plotted against log(bilirubin) in Figure 4.6.4(a) than when they are plotted against bilirubin in Figure 4.6.4(b). In each of these two plots, a running smooth using LOWESS is displayed, along with a least squares simple linear regression line. The rise at the left-hand edge of the LOWESS smooth to the residual plot against log(bilirubin) is due entirely to three patients (#8, #108, and #163) with low bilirubin levels who died early. There is no biological support for such a pattern.

To study the role of hepatomegaly, the subset of 152 martingale residuals for patients *without* enlarged livers are plotted against log(bilirubin) and bilirubin in Figures 4.6.4(a0) and 4.6.4(b0), respectively, while the subset of 160 martingale residuals for patients *with* enlarged livers are plotted against log(bilirubin) and bilirubin in Figures 4.6.4(a1) and 4.6.4(b1), respectively. In Figures 4.6.4(a0) and 4.6.4(a1), LOWESS scatterplot smoothers and least squares lines are compared to the least square line from Figure 4.6.4(a). When the residuals are plotted against log(bilirubin), it is evident that subsetting by hepatomegaly has negligible effect on the regression lines over the range of the data. In contrast, consider when the

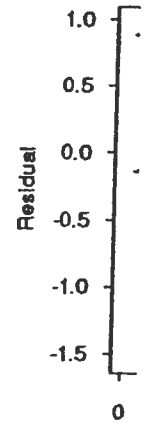


Figure 4.6.2 Martingale resi
single binary covariate indicat
LOWESS smooths use a span

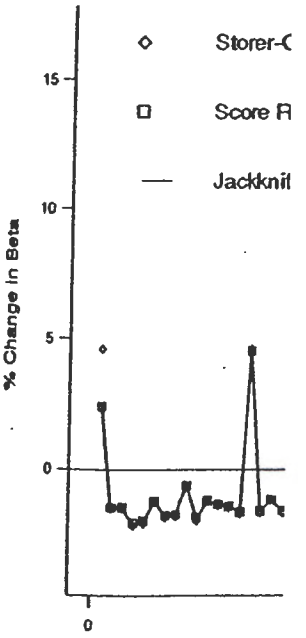


Figure 4.6.3 Influence measu
is time to disease progression

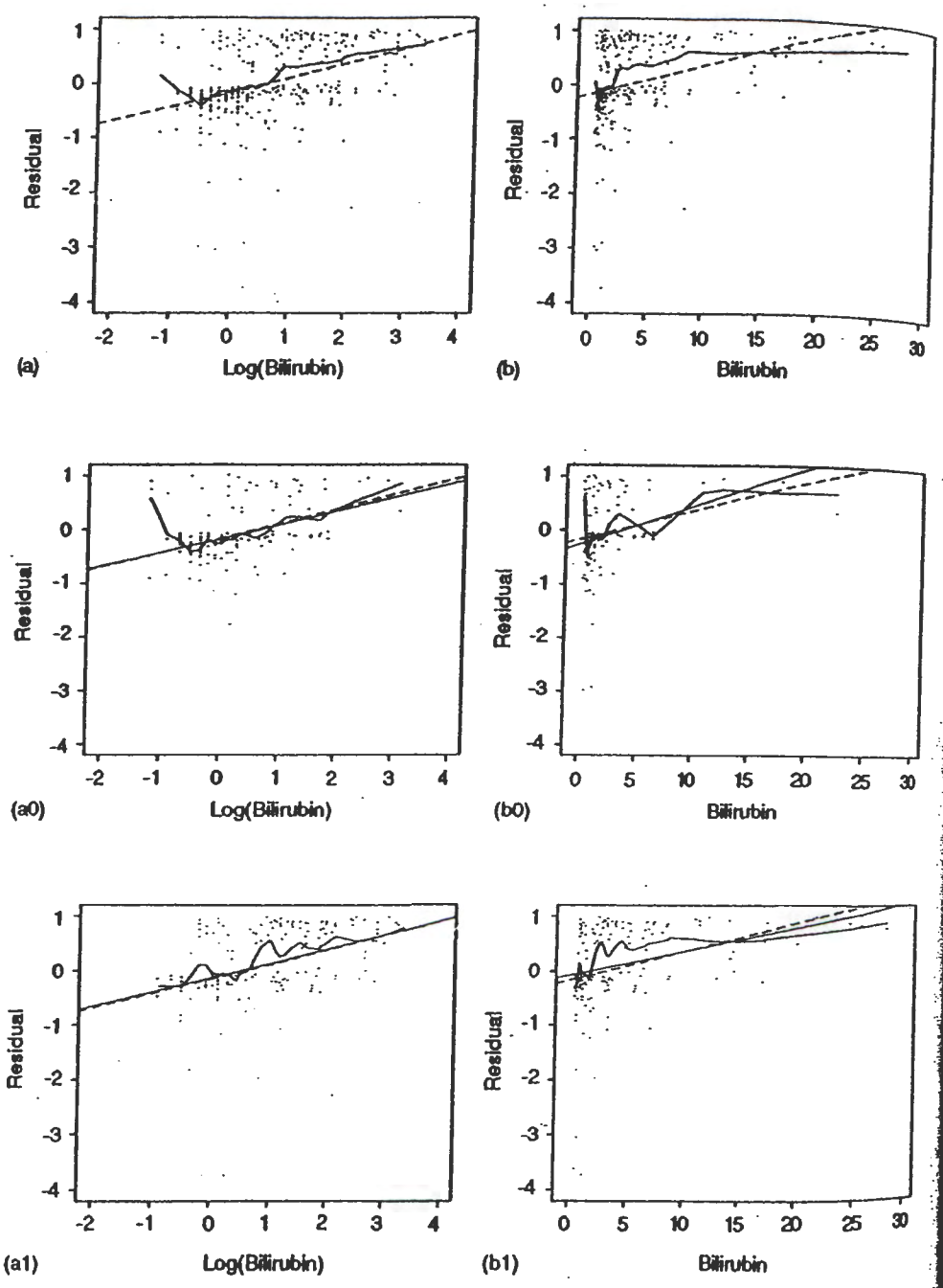


Figure 4.6.4 Martingale residuals in PBC data from a model with the covariates age, log(albumin), log(protime) and edema. Residuals are plotted against log(bilirubin) in these plots on left, and against bilirubin on the right. The top two plots use all 312 randomized cases, the middle two use the 152 cases without enlarged livers, and the bottom two use the 160 cases with enlarged livers. LOWESS smooths use a span of .2.

martingale residuals a
Figure 4.6.4(b1) revea
especially over the reg
patients without hepat
below the line from 4
visually apparent that
bilirubin rather than lo

The appropriateness
ables in the five variat
time)] is shown in Figu

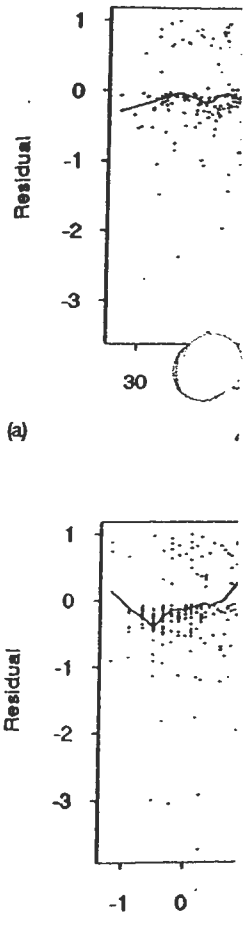


Figure 4.6.5 Martingale res
and three of the four contin
plotted against the omitted va

martingale residuals are plotted against bilirubin. For patients with hepatomegaly, Figure 4.6.4(b1) reveals the least squares line is pulled above the line from 4.6.4(b), especially over the region with bilirubin < 7, where the bulk of the data occur. For patients without hepatomegaly, the least squares line in Figure 4.6.4(b0) is pulled below the line from 4.6.4(b) over the region containing the bulk of the data. It is visually apparent that hepatomegaly is independently predictive when the variable bilirubin rather than log(bilirubin) is added to the model.

The appropriateness of the selected transformations for the four continuous variables in the five variable model [age, log(albumin), log(bilirubin), edema, log(protime)] is shown in Figure 4.6.5. For each continuous variable, martingale residuals

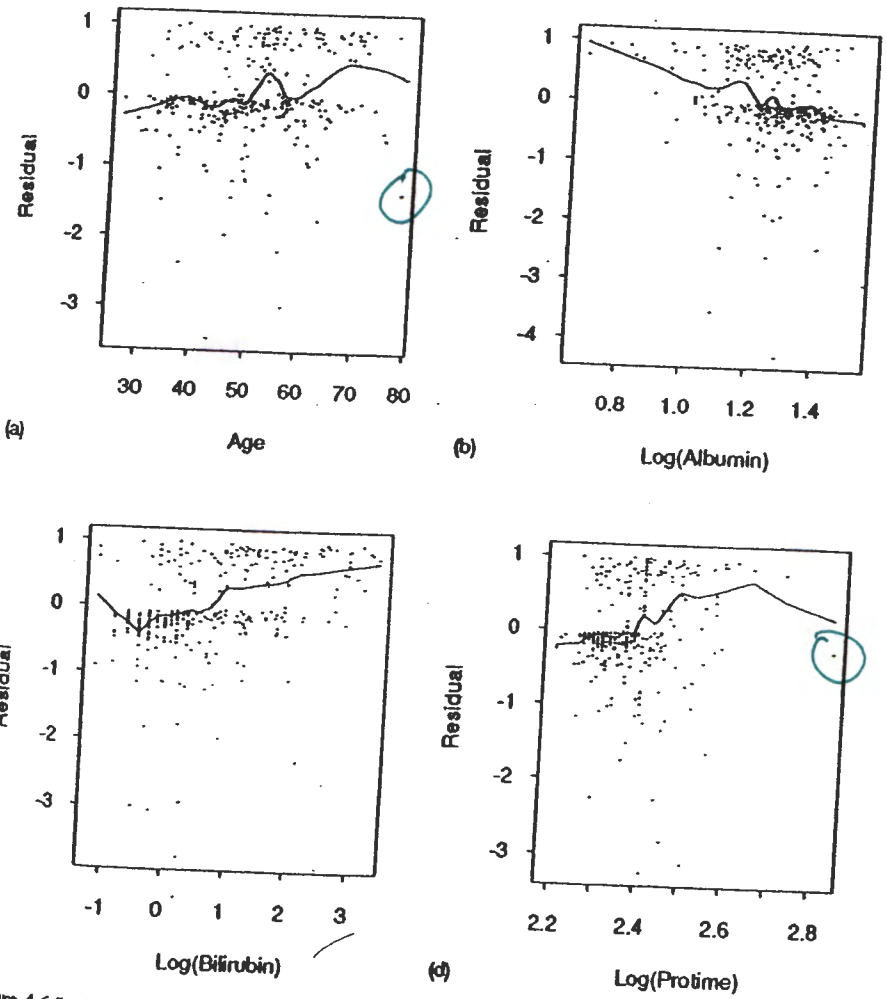


Figure 4.6.5 Martingale residuals in the PBC data. Residuals from a model with the covariate edema and three of the four continuous variables (age, log(albumin), log(bilirubin), and log(protime)) are plotted against the omitted variable. LOWESS smooths use a span of .2.

from the Cox model containing edema and the other three continuous variables are plotted against the transformed variable. Approximate linearity of each of the four LOWESS smooths provides additional support for the transformations selected in Section 4.4. The LOWESS smooth of the martingale residual plot against log(protime) departs from a linear fit in the right-hand tail, with a similar but less noticeable effect seen in the age plot as well. It appears that each of these departures is entirely due to the negative residual (i.e., unexpectedly long survival) of the two cases with extremely long protime, and high age, respectively. Later, we will inspect the influence of these two cases to see if they are largely responsible for the logarithmic transformation of the variable protime and the lack of a square transformation of the variable age.

Plots of martingale residuals in Figure 4.6.5 did not provide strong evidence for changing the transformations selected in Section 4.4 using partial likelihood based model fitting techniques. Certainly this is not always the case. For illustration, a less mature version of the PBC data was analyzed a number of years earlier in an attempt to develop a parsimonious model which would provide survival estimates $\{\hat{S}(t|Z) : 0 \leq t \leq 18 \text{ months}\}$ for an 18 month period. Variable selection procedures led to the choice of variables log(albumin), edema, and log(protime). Table 4.6.1(a) provides the score statistics, adjusting for the three variables in the model, for a number of additional variables. Log(platelets) appeared to be independently predictive and its addition made the variable platelets strongly independently predictive (see Table 4.6.1(b)). The opposite signs of their coefficients in the model in Table 4.6.1(b) make it clear that each variable is correcting for the effect of the other. To assess visually the relationship of platelets with survival, martingale residuals from the three variable model are plotted in Figure 4.6.6(a). A threshold effect is apparent; patients with serum platelet counts below 130,000 (the lower limit of the Mayo Clinic published normal range) continue to be at higher risk for death even after adjustment for albumin, edema, and protime levels. When the variable platelets is entered into the model as a dichotomous variable, Table 4.6.2 reveals the continuous variables platelets and log(platelets) provide negligible additional predictive information. Furthermore, a plot of martingale residuals against platelets from the Cox model containing the dichotomous platelet variable (in Figure 4.6.6(b)) reveals the dichotomy appears to adequately explain the association.

Looking for Influence Points in the Liver PBC Model

Returning to the five variable PBC natural history model developed in Section 4.4 and displayed in Table 4.4.3(c), one can apply the standardized score residuals in Eq. (5.16) to search for cases which are influential on parameter estimates and choice of transformation. The score residual plots for each of the five variables appear in Figure 4.6.7. The bilirubin and edema plots each have an interesting case. For bilirubin, individual #81 was a 63 year old woman with no edema, good albumin (3.65), and rather high protime (11.7). In spite of very high bilirubin (14.4), she lived seven years before dying. For the variable edema, individual #293 was a 57-year-old woman with quite poor prognostic status. In spite of low albumin (2.98), high bilirubin (8.5) and protime (12.3), and edema resistent to

approx. $\Delta_i\beta = \hat{\beta} - \hat{\beta}_{-i}$

Table 4.6.1 Regr
in an earlier versi

Variable
log(Albumin)
log(Prothrombin Tir
Edema

Rao sc

(b) Model incl

Variable
log(Albumin)
log(Prothrombi
Edema
log(Platelets)
Platelets

Table 4.6.2 A four-var
survival in an earlier ver

(Dich. Plat.: 0 =

Variable
log(Albumin)
log(Prothrombin Time)
Edema
Dich. Platelet

Rao sco

Table 4.6.1 Regression models for predicting 18 month survival in an earlier version of the PBC data.

(a) Three variable model

Variable	Coef.	Std. Err.	Chi-Square
log(Albumin)	−4.755	1.3385	12.62
log(Prothrombin Time)	9.430	1.7365	29.49
Edema	2.023	0.4546	19.80

Rao score statistics for additional variables

Variable	Chi-Square
log(Platelets)	6.47
Alk.Phos.	0.07
Platelets	2.73
Ascites	0.13
Age	0.74
log(Bilirubin)	1.83

(b) Model including linear and logarithmic platelet terms

Variable	Coef.	Std. Err.	Chi-Square
log(Albumin)	−4.764	1.4495	10.80
log(Prothrombin Time)	8.456	1.9950	17.97
Edema	2.335	0.4706	24.64
log(Platelets)	−5.423	1.5845	11.71
Platelets	0.021	0.0071	8.46

Table 4.6.2 A four-variable regression model for predicting 18 month survival in an earlier version of the PBC data.

(Dich. Plat.: 0 =Platelets< 130.000; 1 =Platelets≥ 130, 000)

Variable	Coef.	Std. Err.	Chi-Square
log(Albumin)	−4.318	1.4029	9.47
log(Prothrombin Time)	8.721	1.9670	19.65
Edema	2.615	0.4993	27.43
Dich. Platelet	−1.791	0.4928	13.20

Rao score statistics for additional variables

Variable	Chi-Square
log(Platelets)	0.01
Alk.Phos.	0.44
Platelets	0.16
Ascites	0.12
Age	0.51
log(Bilirubin)	1.01

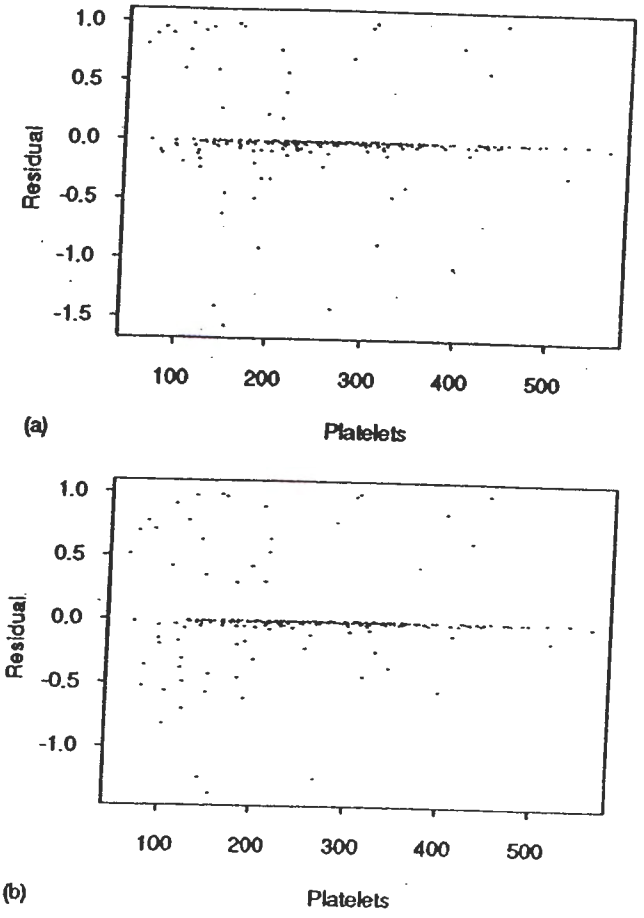


Figure 4.6.6 Martingale residuals plotted against platelets (divided by 10^3) in an earlier version of the PBC data. Residuals in (a) are from a model with the covariates log(albumin), edema, and log(protime). Residuals in (b) are from a model with these three covariates and a binary covariate indicating whether platelets $\geq 130 \times 10^3$.

diuretics, she remains alive after more than 3.5 years. After rechecking, no data errors were found for these two cases.

The most interesting score residual plots in Figure 4.6.7 are those corresponding to the variables log(protime) and age. The fanning out in the left- and right-hand tails of these two plots is a reflection of the earlier observation that influence on model fit depends on both the residual from the fit and on the extremity of its covariate value, roughly $(Z_i - \bar{Z})\times$ residual. The conjecture, supported by the martingale residual plots in Figure 4.6.5, that the two individuals with the highest protime and highest age might be influential points is confirmed in Figure 4.6.7. Individual #253, listed as the oldest patient in the study at 78.4 years of age, remained alive after nearly five years follow up despite high bilirubin (7.1), low

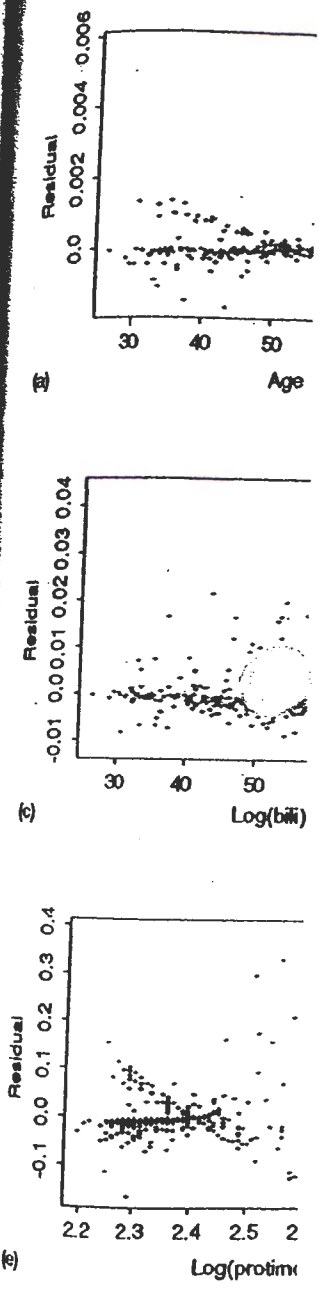
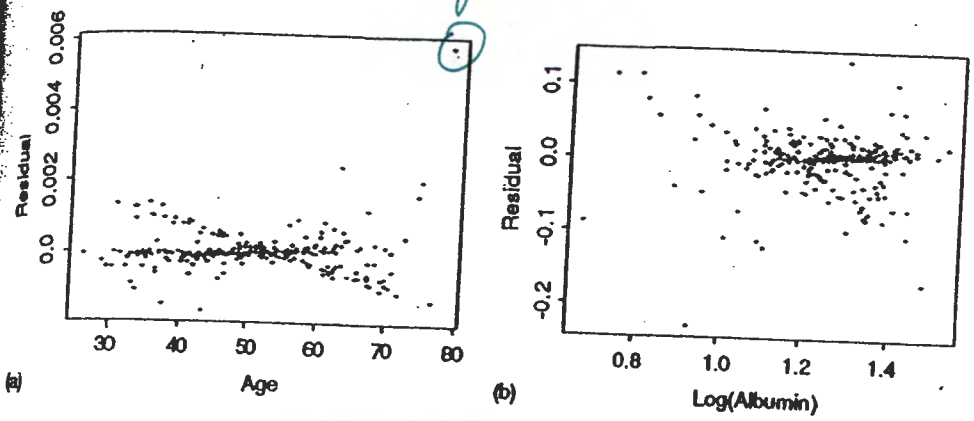


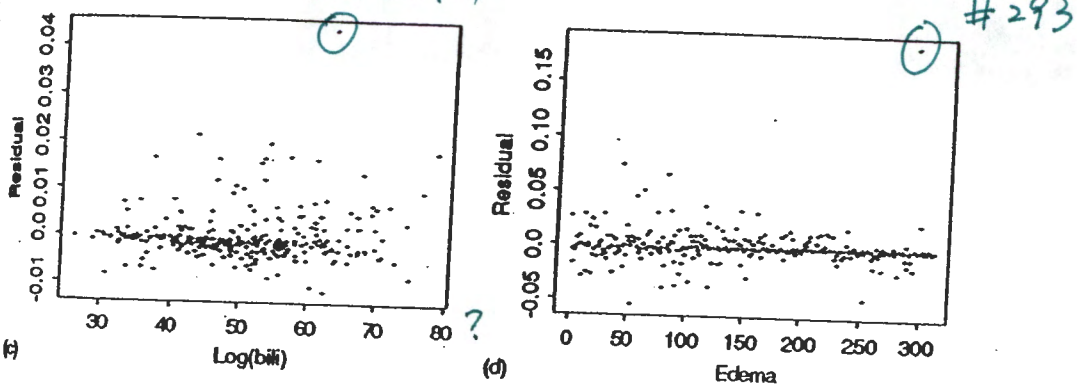
Figure 4.6.7 Standardized sc
log(albumin), log(bilirubin), ed
number.

age = 78.4



bili = 14.4

#293



protime = 17.1

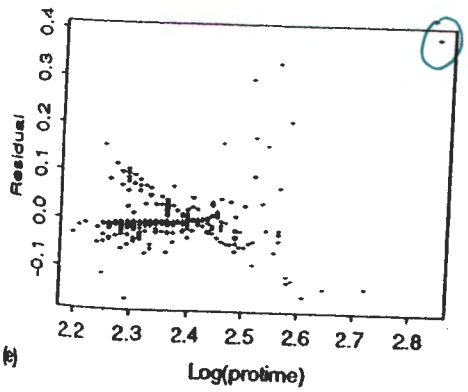


Figure 4.6.7 Standardized score residuals for PBC data from a model with the covariates age, log(albumin), log(bilirubin), edema, and log(protime). Residuals for edema are plotted against case number.

albumin (3.03), and advanced age. Individual #107, with the highest prothrombin time, has an unusual covariate profile. . This 62-year-old female had excellent albumin (4.03) and bilirubin (0.6) levels and no edema. In contrast, her protime value of 17.1 was well above the range of other prothrombin time values. The fact that she was alive after more than 9 years follow-up gave her considerable influence on the variable protime and raised concerns about the accuracy of the value 17.1

To assure the highest data quality, the entire PBC data set had been subject to considerable scrutiny throughout the lengthy period of its collection. In this sense, it was quite surprising when a rechecking of original medical records revealed database errors in the age of individual #253 (78.4 years should have been 54.4 years) and in the protime of individual #107 (17.1 seconds should have been 10.7 seconds). Since considerable additional data rechecking led to no further identification of errors, the martingale and score residual plots were effective in identifying data inaccuracies.

After correcting the protime value for #107, martingale residuals from the model containing the variables age, log(albumin), log(bilirubin), and edema were plotted against log(protime) and protime in Figure 4.6.8(a) and (b), respectively. As anticipated, the drop in the right-hand side of the martingale residual plot for log(protime) in Figure 4.6.5 is eliminated in Figure 4.6.8(a). Comparison of the smooths in Figure 4.6.8(a) and (b) provides slightly more evidence of a linear fit to the residuals when they are plotted against protime rather than against its logarithmic transformation. This graphical observation was confirmed analytically by the score statistics 10.085 and 10.428 for the variables log(protime) and protime, respectively, obtained from the model containing the variables age, log(albumin), log(bilirubin), and edema.

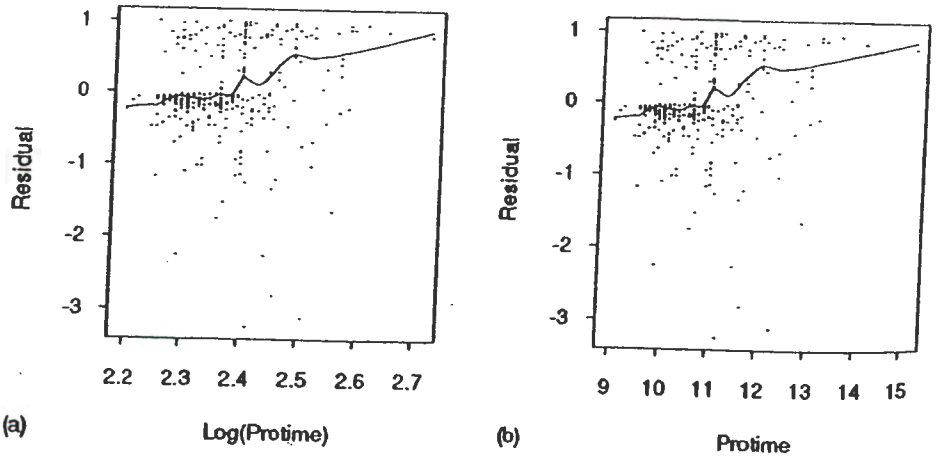


Figure 4.6.8 Martingale residuals in PBC data after the covariate protime in case 107 was corrected from 17.1 seconds to 10.7 seconds. Residuals are from a model with the covariates age, log(albumin), log(bilirubin), and edema.

APPLICATIONS OF RES

When the age of p
in the five variable r
0.0405 ± 0.0094. Th
by a multiplicative co
vide martingale resid
edema, and log(protin
#253 has been correc
this four variable mo
analytical information
formation.

The data correction
residual plots could alt
time. Since such chang
corrections) in one dat
of the new transformat
term for protime and
model would have mea
whose covariate values
for the 312 Mayo Clin
"extrapolating," e.g., u
very advanced stage of

Assessing Model A

Plots of residuals
poorly predicted by the

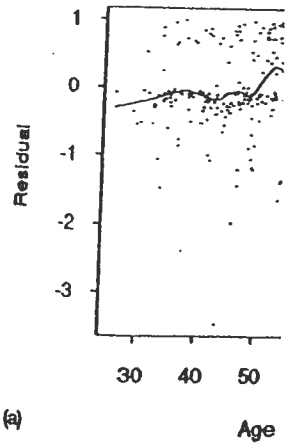


Figure 4.6.9 Martingale resid
78.4 years to 54.4 years. Resic
edema, and log(protime). LOW

When the age of patient #253 is changed from 78 to 54, the coefficient for age in the five variable model in Table 4.4.3(c) increases from 0.0333 ± 0.0087 to 0.0405 ± 0.0094. Thus each decade increase in age increases the hazard function by a multiplicative constant 1.500 rather than 1.395. Figures 4.6.9(a) and (b) provide martingale residuals from the model containing log(albumin, log(bilirubin), edema, and log(protime), plotted against age and $(age)^2$, where the age of patient #253 has been corrected. The score statistic for the variables age and $(age)^2$ in this four variable model are 19.09 and 20.44, respectively. This graphical and analytical information provides slightly more evidence in favor of the $(age)^2$ transformation.

The data corrections identified through the inspection of martingale and score residual plots could alter the transformations selected for the variables age and protime. Since such changes would be motivated by corrections (albeit very significant corrections) in one data element in each of only two cases, the evidence in favor of the new transformations is not convincing. The choice of linear vs. logarithmic term for protime and square vs. linear term for age in the PBC natural history model would have meaningful impact on survival predictions only for individuals whose covariate values are outliers relative to the distribution of covariate values for the 312 Mayo Clinic PBC patients. Thus, one should be very cautions about "extrapolating," e.g., using the model to predict survival of PBC patients in the very advanced stage of the disease.

Assessing Model Accuracy for Individual Subjects

Plots of residuals provide a graphical assessment of the cases whose outcome is poorly predicted by the model. Those with large positive values for $\widetilde{M}_i$ have more

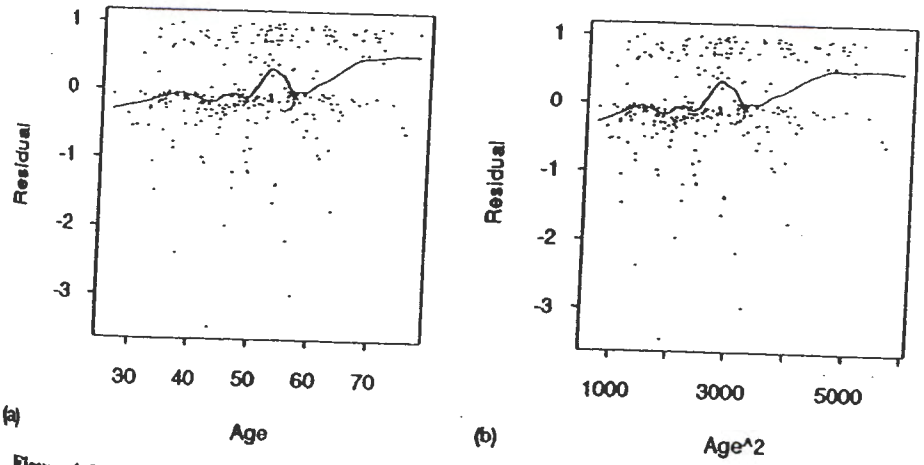


Figure 4.6.9 Martingale residuals in PBC data after the covariate age in Case 253 was corrected from 78.4 years to 54.4 years. Residuals are from a model with the covariates log(albumin), log(bilirubin), edema, and log(protime). LOWESS smooth based on a span of .2.

events than predicted, while those with large negative values have fewer. In one event models such as the Cox model, the martingale residuals are skewed, taking on values from $+1$ to $-\infty$. This skewness makes it difficult to detect outliers who "died too early," and may artificially create the appearance of some outliers who "died too late." The deviance transform symmetrizes the martingale residuals and helps alleviate this problem. In fact, the deviance residuals seem to be very nearly normally distributed when censoring is light.

Figure 4.6.10(a) provides the martingale residuals for the 312 patients in the five variable PBC natural history model displayed in Table 4.4.3(c). Each residual is plotted against the risk score $\hat{\beta}'Z_i$. In contrast, Figure 4.6.10(b) provides the deviance residuals for the same model. The deviance transform suggests that three apparent outliers in the martingale plot are, in fact, not outliers. The heavy

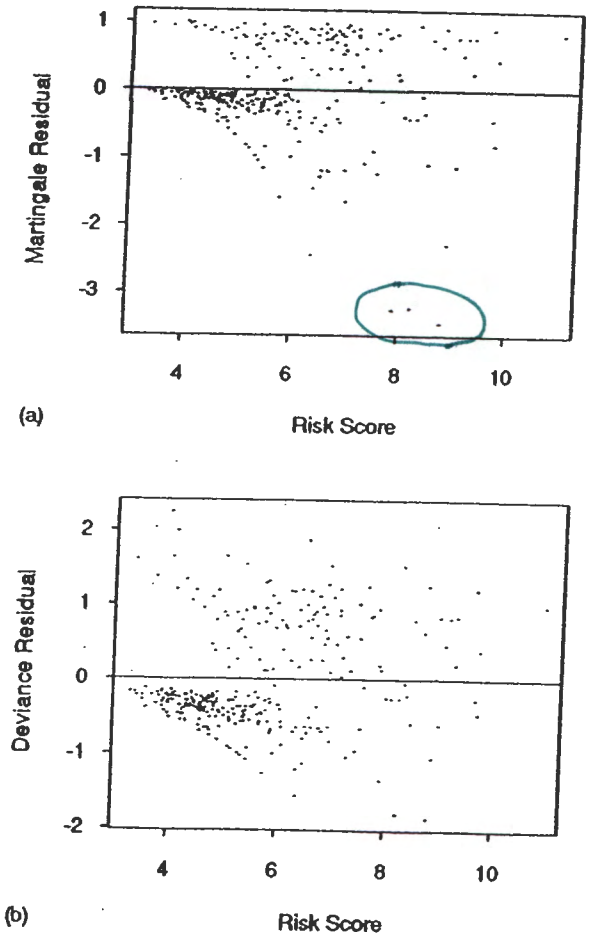


Figure 4.6.10 Residuals plotted against risk score in the PBC data. Residuals from a model with the covariates age, log(albumin), log(bilirubin), edema, and log(protime).

62% censoring leads to some non-normality in the deviance residuals, apparent by a large cluster of residuals near zero. The results in the deviance residual plot suggest examining the patients with the two largest and two smallest residuals.

Assessing Proportional Hazards in the Liver PBC Model

Throughout the model development in Section 4.4, all of the variables were assumed to satisfy the proportional hazards property. The graphical methods from Section 4.5 can be used to assess the validity of this assumption with respect to each of the five variables in Table 4.4.3(c). Figure 4.6.11 presents $\log(-\log S)$ plots (5.19) for each variable in a model containing the other four variables. For each plot, the variable being assessed (if continuous) was split into four levels so that each group had nearly 1/4 of the total number of deaths. The plots provide some evidence that log(protime) and in particular edema do not satisfy the proportional hazards assumption. For both variables, substantial early differences in the plots disappear later in time. This lack of parallelism contradicts the proportional hazards assumption.

Evidence that the variable edema violates the proportional hazards assumption is less convincing than Figure 4.6.11 might imply. Only a few patients in the two edema groups are followed beyond three years, due to the small numbers in each of these groups initially, and due to the very poor prognosis of these patients. For further insight, the standardized score residual plots, defined by the left-hand side of Eq. (5.22), appear in Figure 4.6.12. The steep rise in the plots for edema and protime suggests that the ratio of hazard functions of edema vs. non-edema patients and of high vs. low protime patients is decreasing over time, as was also implied by Figure 4.6.11. The edema plot nearly reached one of the lines $y = \pm 1.36$, while the protime plot exceeded one of the lines $y = \pm 1.63$. Thus the formal test for proportional hazards based on these plots is nearly significant at the $\alpha = 0.05$ level for edema, and is significant at less than the $\alpha = 0.01$ level for protime. In summary, re-modeling allowing for a non-proportional hazards effect for protime might be considered, while the data are inconclusive as to whether or not the variable edema violates the proportional hazards assumption. Larger numbers of patients with edema would need to be followed for a longer period of time for any graphical or analytic method to yield a definitive conclusion.

Validation of the PBC Model using Independent Data

The PBC natural history data set, described in Section 4.4 and provided in Appendix D, contains an independent set of 106 non-randomized PBC patients. These were used by Dickson et al. (1989) to assess the accuracy of the model developed using the 312 randomized patients. Two of the 106 patients had missing values for protime. Patients #359 and #368 were given imputed values 10.55 and 11.12, obtained through a regression on other covariates. The risk scores $R = Z_1\hat{\beta}_1 + \ldots + Z_K\hat{\beta}_K$ then were computed for each of the 106 cross-validation patients using $\hat{\beta}$ from Table 4.4.3(c). Using these scores, patients were divided into low, medium, and high risk subgroups. Within each subgroup, the average of the predicted

Figure 4.6.12a Standardized
points are plotted against rank

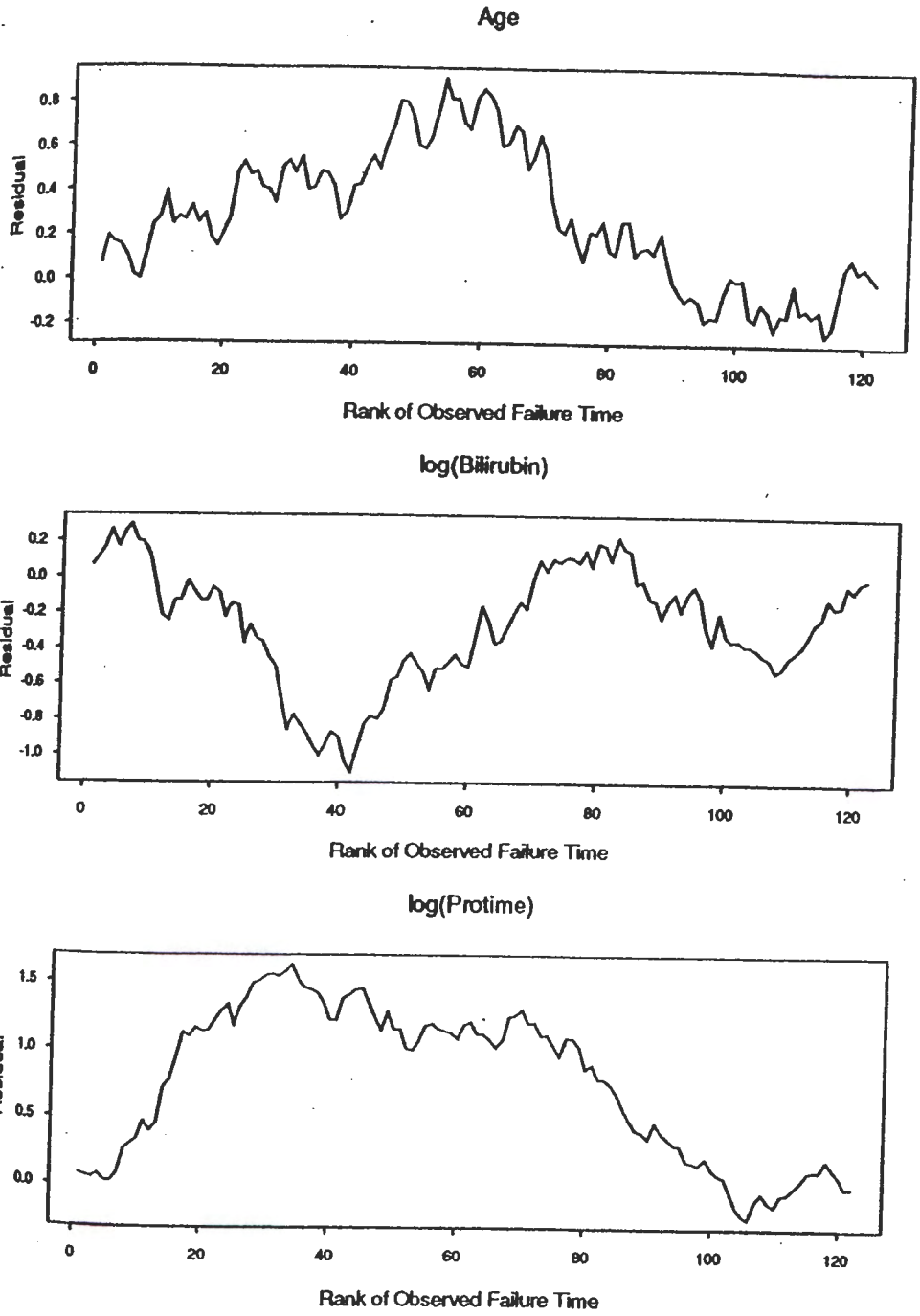


Figure 4.6.12a Standardized score residuals in PBC data. Residuals for each of the covariate components are plotted against rank of the distinct observed failure time.

cumulative score (Schoenfeld) res.
standardized

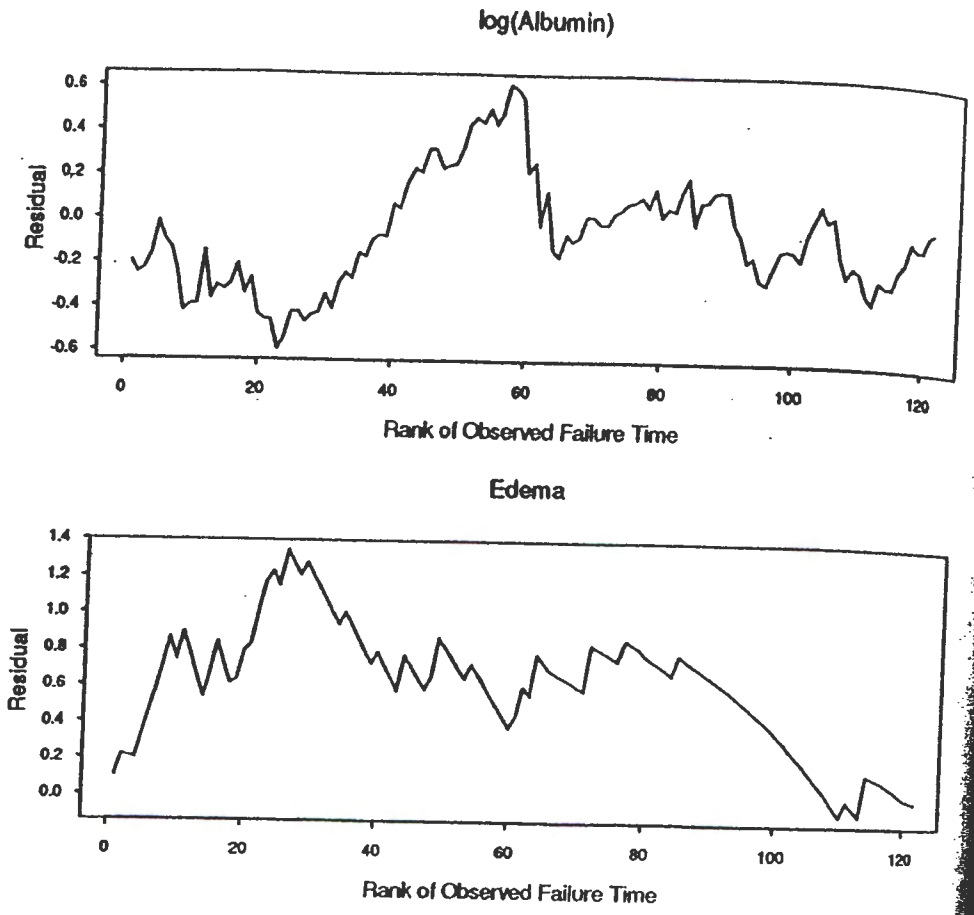


Figure 4.6.12b (*continued*)

survival curves was compared to the actual survival experience represented by a Kaplan-Meier curve. Figure 4.6.13 reveals good prediction by the model. Accuracy in the right-hand tail is difficult to assess due to the variability of the Kaplan-Meier plot. Within each subgroup, each individual's survival was compared to her own model predicted curve, with these individual comparisons then pooled over each subgroup using the one-sample logrank statistic. These statistics yielded chi-squares of approximately 0.5 in each of the three risk groups. Table 4.6.3 compares parameter estimates for the model containing age, log(albumin), log(bilirubin), edema, and log(protime) when the original 312 and the total 418 patients are used in estimation.

Additional model validation studies have revealed good prediction when the model in Table 4.6.3 using 418 patients is applied to untreated PBC patients who are in more advanced stages of disease or to patients who were managed by other

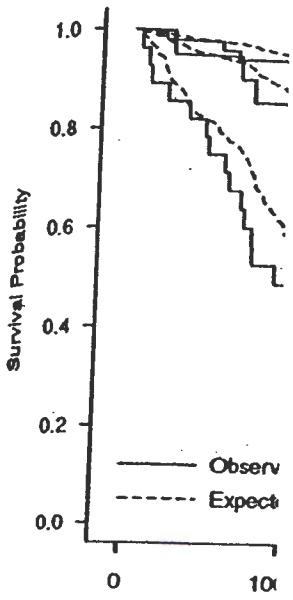


Figure 4.6.13 Predicted anc

clinical centers. Mo... v
centers other than the Ma

Evaluating Efficacy of L

To illustrate the applicatio
data were obtained for all
the University of Pittsbur
April 1980 and June 198
into three categories by t
yield equal numbers of de

Table 4.6.3 R
Survival Models,

Variable	I
log(Bilirubin)	
log(Albumin)	–
Age	
log(Protime)	
Edema	

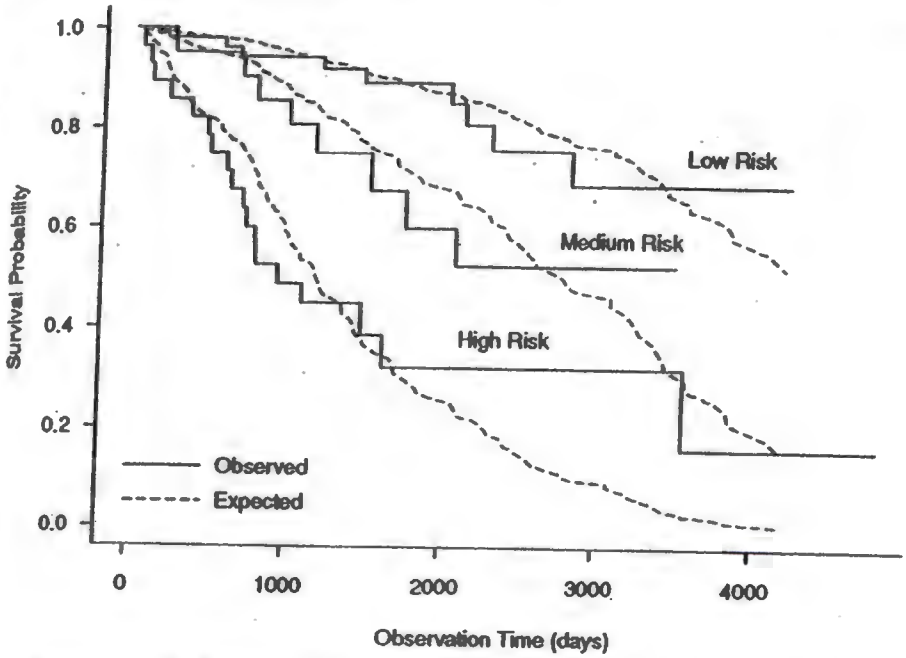


Figure 4.6.13 Predicted and observed survival curves by risk group in PBC validation data set.

clinical centers. Model validation using data from studies conducted by research centers other than the Mayo Clinic is described in Grambsch et al. (1989).

Evaluating Efficacy of Liver Transplantation

To illustrate the application of the natural history model, the covariate and outcome data were obtained for all 161 PBC patients who received liver transplantation at the University of Pittsburgh and its Baylor University satellite in Dallas between April 1980 and June 1987 (Markus et al., 1989). The 161 patients were divided into three categories by their risk scores, with category break points defined to yield equal numbers of deaths per category. The descriptive statistics for these cat-

Table 4.6.3 Regression Coefficients for Cox Regression Survival Models, PBC Data Set

	Initial Model ($n = 312$)		Final Model ($n = 418$)	
Variable	Coef.	Std. Err.	Coef.	Std.Err.
log(Bilirubin)	0.8792	0.0987	0.8707	0.0826
log(Albumin)	−3.053	0.724	−2.533	0.648
Age	0.0333	0.0087	0.0394	0.0077
log(Protime)	3.016	1.024	2.380	0.767
Edema	0.7847	0.2991	0.8592	0.2711

egories appear in Table 4.6.4. For each category, Figure 4.6.14 compares the post transplantation survival curve (using a Kaplan-Meier estimate) with the category's average predicted curve for survival had patients not been transplanted (estimated using the natural history model). The results reveal evidence that transplantation lengthens survival, although one should be cautious about issues such as generalizability of the model prediction and extrapolation required to use the model in the higher one or two risk groups in the transplantation data set.

Table 4.6.4 Descriptive Statistics of PBC Patients and Subgroups According to Risk Category

Group[1]	Variable[2]	Mean	Standard Deviation	Minimum	Maximum
1 ($N = 98$)	Age	46.5	8.5	24.7	65.1
	Bilirubin	12.1	8.2	1.1	41.6
	Albumin	3.1	0.5	1.9	4.5
	Protime	13.4	1.4	10.3	19.3
	Edema	0.36	0.42	0	1.0
	Risk score	7.36	1.04	4.37	8.65
2 ($N = 41$)	Age	47.8	6.8	34.9	64.8
	Bilirubin	24.1	12.2	6.3	52.7
	Albumin	2.7	0.6	1.8	4.5
	Protime	15.1	2.5	11.8	23.0
	Edema	0.77	0.37	0	1.0
	Risk score	9.17	0.26	8.67	9.11
3 ($N = 22$)	Age	53.8	9.0	39.9	76.6
	Bilirubin	27.8	11.7	4.5	58.5
	Albumin	2.5	0.4	1.6	3.0
	Protime	19.5	5.7	13.4	35.7
	Edema	0.93	0.23	0	1.0
	Risk score	10.44	0.47	9.95	11.58
All ($N = 161$)	Age	47.8	8.5	24.7	76.6
	Bilirubin	17.3	11.8	1.1	58.5
	Albumin	2.9	0.6	1.6	4.5
	Protime	14.6	3.4	10.3	35.7
	Edema	0.54	0.45	0	1.0
	Risk score	8.24	1.44	4.37	11.58

[1]Stratification into risk groups was performed according to Mayo Clinic model risk scores, with cutoffs chosen to produce roughly equal numbers of deaths in the first three months. Group 1: 98 patients with 10 deaths in the first three months; group 2: 41 patients with 10 deaths; group 3: 22 patients with 9 deaths.

[2]Variables are: age (years); bilirubin (mg/dl); albumin (g/dl); prothrombin time (protime, seconds); severity of edema; and risk score according to the Mayo model.

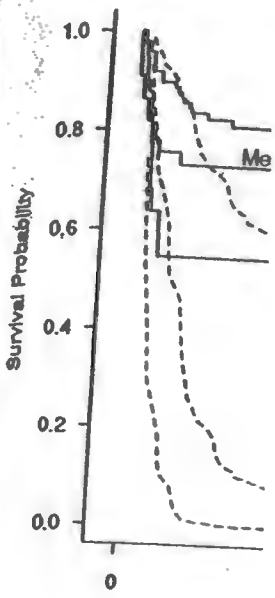


Figure 4.6.14 Predicted and Baylor University t observed curves.

4.7 BIBLIOGRAPHIC

The literature on proportion years, and a complete discus than is appropriate. This br topic.

Section 4.2

The roots of proportional ha can be found in parametric m (an exponential model) or is p Feigl and Zelen (1965) were regression; the leukemia data illustrate new methods. Kalbfl (1984) all discuss a variety o ratios to change over time. In transformation of the time axi accelerated failure time model inference for accelerated failu for the ranks of observed fail problems for computing estima

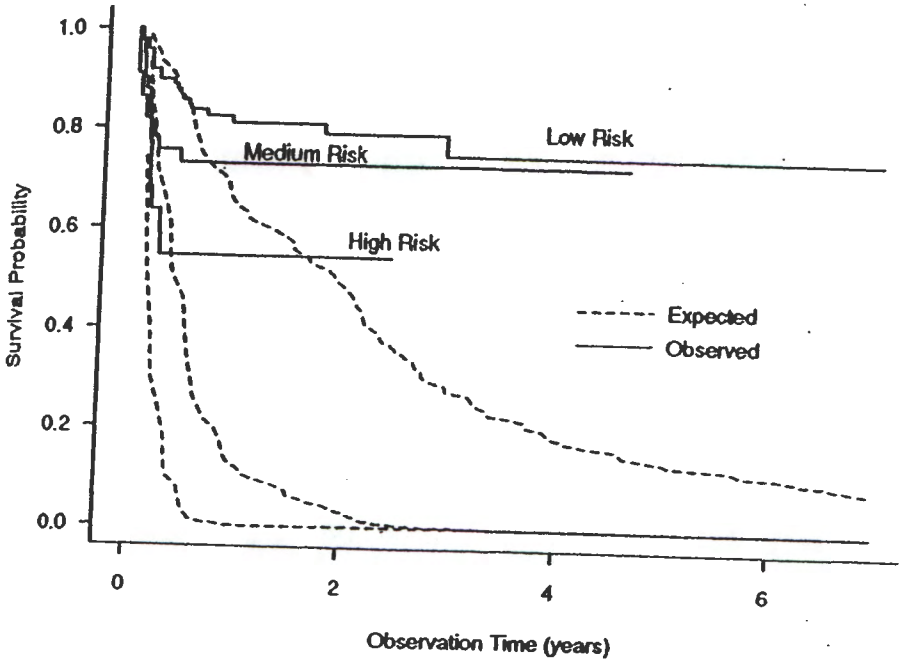


Figure 4.6.14 Predicted and observed survival curves by risk group in the University of Pittsburgh and Baylor University transplant series. Expected curves have the same ordering by risk group as the observed curves.

4.7 BIBLIOGRAPHIC NOTES

The literature on proportional hazards regression has grown quickly over the last twenty years, and a complete discussion of the contributions in this area would require more space than is appropriate. This brief summary mentions some of the influential papers on this topic.

Section 4.2

The roots of proportional hazards regression with an unspecified baseline hazard function can be found in parametric models, which often assume that the hazard function is constant (an exponential model) or is proportional to a power of the time variable (a Weibull model). Feigl and Zelen (1965) were among the first to discuss clinical applications of exponential regression; the leukemia data cited in their paper has frequently been used in the literature to illustrate new methods. Kalbfleisch and Prentice (1980), Lawless (1982), and Cox and Oakes (1984) all discuss a variety of parametric regression models, some of which allow hazard ratios to change over time. In accelerated failure time models a covariate produces a scale transformation of the time axis; we have not discussed those models here at all. Parametric accelerated failure time models are discussed in the three references above. Nonparametric inference for accelerated failure time models is based on a marginal likelihood function for the ranks of observed failure times. This approach presents some technically difficult problems for computing estimators of regression coefficients and calculating their asymptotic