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

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.

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

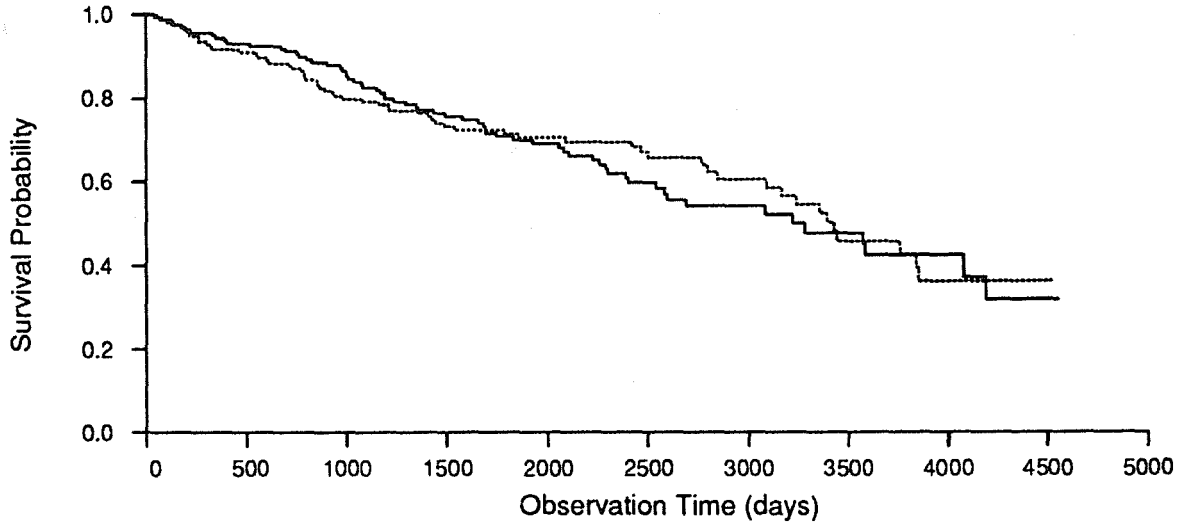
$$\lambda(t|Z) = \lambda_0(t)\exp(\beta Z), \tag{4.1}$$

$\lambda_0(\cdot)$ 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

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



		Time Interval				
Group		0-1000	1000-2000	2000-3000	3000-4000	4000-5000
———	DPCA	23/158	22/128	13/74	5/31	2/10
·····	Placebo	31/154	12/120	7/70	10/32	0/11

(# events/# at risk)

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

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

where $\mathbf{Z}' \equiv (Z_1, Z_2, \ldots, Z_K)$ is a vector of K predictors and $\boldsymbol{\beta}' \equiv (\beta_1, \beta_2, \ldots, \beta_K)$ 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 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 $\mathbf{Z}' = (Z_1, \ldots, Z_{14})$ chosen to be the 14 variables in Table 4.4.1. If $\mathbf{U}(\beta)$ and $\mathcal{I}(\beta)$ denote the score vector and Fisher's observed information matrix, respectively, the collection of univariate Rao or logrank statistics, $\{[\{\mathbf{U}(0)\}_k]^2/\{\mathcal{I}(0)\}_{kk} : k = 1, \ldots, 14\}$, 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 $\mathcal{I}(0) = \int_0^\infty \sum_{i=1}^n \mathbf{V}(0,t)dN_i(t)$ for $\mathbf{V}$ 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

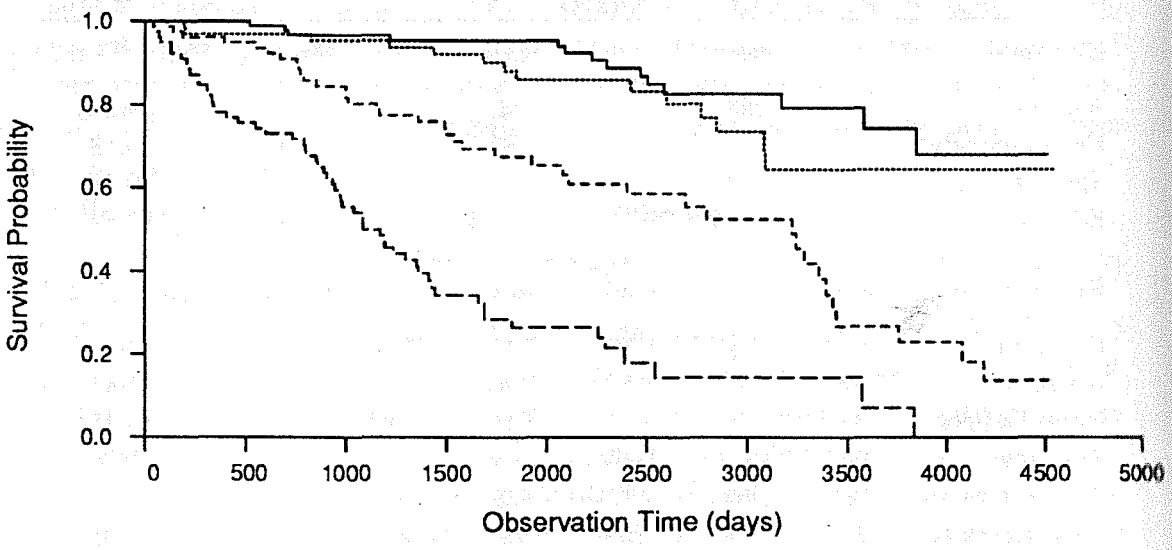
[1]Edema 0: No edema and no diuretic therapy for edema
$\frac{1}{2}$: Edema but no diuretics, or edema resolved by diuretics
1: Edema despite diuretic therapy

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 $\mathbf{Z}$. 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



	Time Interval				
Group	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 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-