0270-9139/89/1001-0001$02.00/0
HEPATOLOGY
Copyright © 1989 by the American Association for the Study of Liver Diseases
Vol. 10, No. 1, pp. 1–7, 1989
Printed in U.S.A.

Original Articles

Prognosis in Primary Biliary Cirrhosis: Model for Decision Making

E. ROLLAND DICKSON, PATRICIA M. GRAMBSCH, THOMAS R. FLEMING,† LLOYD D. FISHER† AND ALICE LANGWORTHY

Division of Gastroenterology and Internal Medicine and Section of Biostatistics, Mayo Clinic and Mayo Foundation, Rochester, Minnesota 55905

The ideal mathematical model for predicting survival for individual patients with primary biliary cirrhosis should be based on a small number of inexpensive, noninvasive measurements that are universally available. Such a model would be useful in medical management by aiding in the selection of patients for and timing of orthotopic liver transplantation. This paper describes the development, testing and use of a mathematical model for predicting survival. The Cox regression method and comprehensive data from 312 Mayo Clinic patients with primary biliary cirrhosis were used to derive a model based on patient's age, total serum bilirubin and serum albumin concentrations, prothrombin time and severity of edema. When cross-validated on an independent set of 106 Mayo Clinic primary biliary cirrhosis patients, the model predicted survival accurately. Our model was found to be comparable in quality to two other primary biliary cirrhosis survival models reported in the literature and to have the advantage of not requiring liver biopsy.

Orthotopic liver transplantation is considered to be potentially life-saving for selected patients with advanced or end-stage primary biliary cirrhosis. The availability of a model to predict survival probability for an individual patient would improve selection of patients for transplantation and the timing of that transplantation. Also, such a model could be used to help to decide which patients are appropriate, medically and ethically, for clinical trials of other treatment modalities. In addition, the model could be used for education and counseling of the patient and the family.

Using the Cox proportional hazards regression procedure (1), Roll et al. at Yale (2) and Christensen et al. in Europe (3) independently developed multivariate survival models. The Yale model used patient's age, serum bilirubin concentration, hepatomegaly and presence of portal fibrosis or cirrhosis to predict survival. The European model used age, bilirubin and albumin concentrations, presence of cirrhosis, presence of cholestasis and whether or not azathioprine was prescribed. However, neither model was developed as a medical management tool, and both models required liver biopsy.

This paper describes a pragmatic model based on inexpensive, noninvasive measurements that are universally and readily available.

PATIENTS AND METHODS

Patient Population

To develop the model, we used natural history data on the 312 primary biliary cirrhosis patients enrolled in either of two double-blind, placebo-controlled, randomized clinical trials at the Mayo Clinic evaluating the use of D-penicillamine for treating primary biliary cirrhosis. To be eligible for these trials, patients had to meet well-established clinical, biochemical, serologic and histologic criteria for primary biliary cirrhosis (4). Patient accrual took place from January, 1974, through May, 1984. One clinical trial (unpublished data) involved patients with histologic Stage 1 or 2 primary biliary cirrhosis; the other involved Stage 3 and 4 patients (4). Both trials found no therapeutic differences between control and D-penicillamine-treated patients. The study protocols required that no patient be taking any antiinflammatory or immunosuppressive medication (other than the study capsule). Therefore, it was deemed appropriate to combine all study participants to determine the natural history of primary biliary cirrhosis.

In addition, we had available 112 patients who were eligible for the trials but declined to participate. None of these patients was taking an immunosuppressive or antiinflammatory medication at the time of trial eligibility. These patients were used for model validation. It is possible that some of the cross-validation patients were exposed to antiinflammatory or immunosuppressive medication during the follow-up period. However, there has been no report of a totally effective regimen for biliary cirrhosis (5). Therefore, it is unlikely that the natural course of their disease was altered by any medication.

Data Collection

A comprehensive clinical and laboratory data base was established on each patient. The data were collected prospectively in the trial patients, by using standardized forms, definitions, and study protocols, at entry and at yearly intervals (see Table 1 for the variables measured). For the nontrial patients, the baseline data were collected from patients' records.

At entry, a liver biopsy specimen was obtained, and the

Received June 28, 1988; accepted December 12, 1988.
Supported by Research Grant AM-34238 from the National Institutes of Health.
† Present address: Department of Biostatistics, University of Washington, Seattle, Washington.
Address reprint requests to: E. Rolland Dickson, M.D., Mayo Clinic, 200 First St. SW, Rochester, Minnesota 55905.

TABLE 1. Summary of statistics for univariate prognostic factors

Demographic		RR[a]
Age (years; median)	49.8	2.1
Sex, % male	12	1.6
Race, % nonwhite	2	

Clinical	%	RR
Ascites	8	7.8
Jaundice	41	4.6
Edema	12	5.0
Varices	14	4.0
Hepatomegaly	51	3.3
Dark urine	44	3.1
Hirsutism	9	3.5
Gastrointestinal bleeding	9	3.4
Spiders	29	2.6
Splenomegaly	28	2.6
Hyperpigmentation	53	2.8
Light-colored stool	34	2.2
Xanthelasma	18	2.3
Variceal bleeding	4	3.6
Clubbing	4	3.3
Weight loss	27	1.7
Excoriations	27	1.7
Pruritus	70	1.6
Fetor	1	3.0
Fatigue	73	1.3
Abdominal pain	21	0.8
Bone pain	15	1.3
Bruising	37	1.2
Fever	3	0.7

Histologic	%	RR
Stage		
1	5	
2	22	5.0
3	38	8.6
4	35	21.4

Biochemical	Median	RR
Total serum bilirubin	1.4	6.0
Urine copper	73	3.7
Serum albumin	3.55	0.3
Prothrombin time	10.6	2.8
Sedimentation rate	55	3.3
Serum calcium	9.6	0.5
Hepatic copper	168	2.8
Aspartate aminotransferase	115	2.8
Serum copper	168	1.9
Platelet count	257	0.6
Total serum cholesterol	310	1.7
Total serum triglyceride	108	1.8
Serum alkaline phosphatase	1,259	1.7
Serum phosphorus	3.7	0.6
Serum creatinine	0.9	0.7
γ-Globulin	1.76	1.4
Immunoglobulin M	4.9	0.9
Ceruloplasmin	64.5	1.6

[a] RR = relative risk.

histologic stage was determined according to the method of Ludwig et al. (6).

At updating of follow-up in July, 1986, 125 of the original 312 patients had died; their median time on trial was 39 months. Of the remainder, 160 were still alive and being followed, with a median time on trial of 76 months. The rest either were lost to follow-up (eight patients) or had undergone liver transplantation (19 patients). For those lost to follow-up, we used the most recent available follow-up; this group had a median follow-up of 66 months. For the transplantation patients, the median time from entry into the trial until transplantation was 47 months.

To validate the statistical model, we used 106 of the 112 patients qualifying for but declining participation in the randomized studies. These 112 included 36 patients who had died with median follow-up of 26 months, six who had undergone transplantation with a median follow-up of 50 months and 64 who still were living with a median follow-up of 54 months. The remaining six patients had been lost to follow-up within a few months and therefore provided no information for validation.

Development of the Model

Definitions: Death from any cause was treated as a failure for purposes of survival analysis. Of the 125 deaths in the primary biliary cirrhosis study group, only 20 were not attributable to primary biliary cirrhosis. Transplant patients were censored at the date of transplantation. The initial time point for survival modeling was the date of determination of eligibility for the trials, and all of the clinical, biochemical and demographic risk factors were assessed on that date.

Univariate Survival Modeling: The 45 potential prognostic variables (described in Table 1) were examined individually. Some variables, such as presence or absence of a sign or symptom, are naturally dichotomous. The continuous variables were dichotomized by splitting at the median. The relationship between each variable and survival was determined by computing the relative risk for failure, comparing patients with the sign or symptom to those without it and comparing patients above the median to those below it.

For histologic stage, which does not dichotomize easily, relative risk was computed by comparing each of Stages 2, 3 and 4 with Stage 1.

Multivariate Survival Modeling: The Cox proportional hazards regression model (1, 7) was used. In this model, each individual patient is given a risk score

$$R = X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_k,$$

in which $X_1, \ldots, X_k$ are the levels of k prognostic variables and $\beta_1, \ldots, \beta_k$ are called regression coefficients. Larger values of R mean greater risk (poorer prognosis); smaller values (including negative ones) mean better prognosis. For example, suppose two patients differ in their R score by an amount d. Then, by the Cox model's proportional hazards assumption, at every point in time, the patient with the higher R has exp(d) times as much risk of dying as the patient with the lower R.

Let S(t,X) give the probability that a patient with risk factors given by $X = \{X_1, X_2, \ldots, X_k\}$ and with risk score R will still be alive t years later. Suppose we know the survival function,

$S_0(t)$, for individuals having risk score R_0. It follows from the proportional hazards assumption that we obtain a very simple formula for S(t,X), given by

$$S(t,X) = \{S_0(t)\}^{\exp(R-R_0)}$$

Standard techniques presented by Kalbfleisch and Prentice (8) allow us to estimate $S_0(t)$ from the data, and the regression coefficients $\{\beta_1, \ldots, \beta_k\}$ can be estimated by the method of maximum likelihood estimation applied to the Cox partial likelihood. Each coefficient β_i has the simple interpretation that every unit increase in the ith covariate, X_i, increases the risk of dying by the multiplicative factor $\exp(\beta_i)$.

Model Estimation and Assessment: These were done by computer with the SAS procedure PHGLM (9). In selecting variables for the model, we used both the forward and the backward stepwise variable selection procedures separately. For the continuous variables, we considered as the candidates the variable itself and also standard transformations of it: square, logarithm and square root. Variables had to have $p \leq 0.01$ in both the forward and the backward procedure to be retained in the model. The appropriateness of the proportional hazards assumption was examined by preparing log(−log) plots of the survival function and by the Z:PH statistic (10), which tests each variable in a Cox model for proportional hazards and is implemented in PHGLM.

To assess how well the model fit the data from the 312 study patients, we compared the actual survival experience with the survival predicted by the model for five different groups with nonoverlapping levels of risk. R was computed for each patient. These Rs were ranked and divided into five groups such that there were equal numbers of deaths (25 deaths) in each group. The survival curve predicted by the model for each group was found by computing the estimated survival function for each patient and then averaging these functions within the groups. The actual group survival experience was computed by the Kaplan-Meier method (11). The predicted and actual survival curves were compared graphically.

Cross-Validation of the Model

To validate the model, we assessed how well it predicted survival in the independent set of 106 Mayo primary biliary cirrhosis patients, and we compared its performance with that of the European model (3).

The European estimate of $S_0(t)$ was kindly provided to us by Douglas Altman, one of the developers of the European model.

Some data for the cross-validation patients were missing. Sixteen patients lacked values for one or two of the following: cirrhosis, cholestasis and prothrombin time. To keep all patients in the cross-validation analysis, we estimated the missing values from regression equations (ordinary multiple regression for prothrombin time and logistic regression for the binary variables cholestasis and cirrhosis) developed by using the variables from the 90 patients with complete information. Backward elimination was used to select the variables for each regression. As a check, we also carried out all cross-validation analyses with only the data from the 90 patients with no missing information. The conclusions were very similar, so those analyses are not presented here.

Rs were computed by each model for each of the 106 patients in the cross-validation data set. These two sets of Rs were centered by subtracting the mean of each set from all values in the set. A scatterplot of these Rs was drawn. The Rs were related by Pearson product-moment correlation and ordinary least-squares regression. The patients were divided into three groups—low, medium and high risk—and the survival predictions of the two models were compared with the actual survival in each group. The groups were formed in a manner designed to be neutral between the two models. The estimated Rs for each model were ranked separately and the mean of the two ranks was computed for each patient. Then the patients were ranked on the basis of these means and divided into three groups having roughly equal numbers of deaths.

The mean survival function for each group for each model was computed as described above and compared with the Kaplan-Meier survival curve for that group, both graphically and by one-sample log-rank tests (12). There is a technical difficulty in the use of one-sample tests in this context because the mean survival function is random, not fixed as assumed by the test. Therefore, the one-sample log-rank tests are more likely to reject the null hypothesis of fit between data and model than the nominal p values would indicate.

RESULTS

Initial Mayo Model

At trial entry, the 312 patients who were used for model development can be characterized as follows: 35% with cirrhosis, 9% with albumin less than 3.0%, 18% with total bilirubin more than 5 mg per dl and 0.6% with prothrombin time at least 3 sec above normal. The 45 predictor variables are presented in Table 1 with descriptive statistics. Many variables are highly predictive of survival as indicated by relative risks much different from unity.

When all 45 variables were used in multivariate Cox proportional hazards modeling, the stepwise selection procedure produced a model based on 10 variables: serum bilirubin concentration, serum albumin concentration, age, urinary copper excretion, prothrombin time, serum aspartate aminotransferase, histologic stage, excoriations, hirsutism and xanthelasma. This model was unsatisfactory for two reasons. First, the variable selection procedure involved a large number of candidate variables relative to the number of deaths and, therefore, could have produced some spuriously significant factors. Second, the variables did not all meet our criteria of practicality. The measurement of urinary copper is not readily available and the determination of histologic stage requires liver biopsy and consequently is invasive and expensive.

Therefore, a second model was developed by using stepwise modeling on a subset of 12 noninvasive, easily collected variables that require only clinical evaluation and a blood sample: age, serum albumin, bilirubin, alkaline phosphatase, cholesterol and aspartate aminotransferase values, prothrombin time, platelet count and the presence or absence of spiders, hepatomegaly, ascites and edema. To avoid selection bias, these variables were chosen without regard to their presence or significance level in the first model.

From these 12 variables, the forward and backward stepwise selection procedures chose the same five variables: $\log_e$(bilirubin), $\log_e$(albumin), $\log_e$(prothrombin time), age and presence or absence of edema. All p values were less than 0.01. To control for the fact that the presence of edema can be modified by use of diuretics as well as by the progression of the disease, we modified the

edema variable to take diuretic use into account by coding the variable as follows:

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

Clinically, this scale should parallel the severity of edema.

This five-variable model was examined further. The 10 possible cross-product terms formed to assess two-factor interactions were added to the model one at a time, but no interactions were found to be statistically significant at the 0.05 level. Assessment of the validity of the proportional hazards assumption revealed some concern about proportionality for edema. Presence of edema induces a striking increase in failure rate over the first 18 months of follow-up but appears to have only a small effect beyond that time. However, the effect of edema on long-term failure rates cannot be assessed because only 19 patients with edema survived beyond 18 months. Therefore, we could not reliably stratify by edema, so we retained edema as one of the predictors. The column of Table 2 labeled "Initial Mayo model" gives the β coefficients, standard errors and p values for the five risk factors.

The model fits the data reasonably well (Fig. 1). There is a broad range of experience from the lowest risk patients (quintile 1), whose chance of surviving 5 years is >96%, to the highest risk patients (quintile 5), whose chance of surviving 1 year is <52%. The likelihood ratio

TABLE 2. Regression coefficients for Mayo Cox regression survival models[a]

	Initial Mayo model (n = 312)	Final Mayo model (n = 418)
Bilirubin ($\log_e$)		
β	0.8792	0.8707
S.E.	0.0987	0.0826
χ^2	79.30	111.03
p	<0.0001	<0.0001
Albumin ($\log_e$)		
β	−3.053	−2.533
S.E.	0.724	0.648
χ^2	17.78	15.27
p	<0.0001	0.0001
Age		
β	0.0333	0.0394
S.E.	0.0087	0.0077
χ^2	14.76	26.53
p	0.0001	<0.0001
Prothrombin time ($\log_e$)		
β	3.016	2.380
S.E.	1.024	0.767
χ^2	8.68	9.64
p	0.0032	0.0019
Edema (and therapy)		
β	0.7847	0.8592
S.E.	0.2991	0.2711
χ^2	6.88	10.04
p	0.0087	0.0015

[a] β = point estimate of the regression coefficient; S.E. = standard error of estimate; χ^2 = chi-square test statistic for assessing significance of coefficient; p = p value for the test.

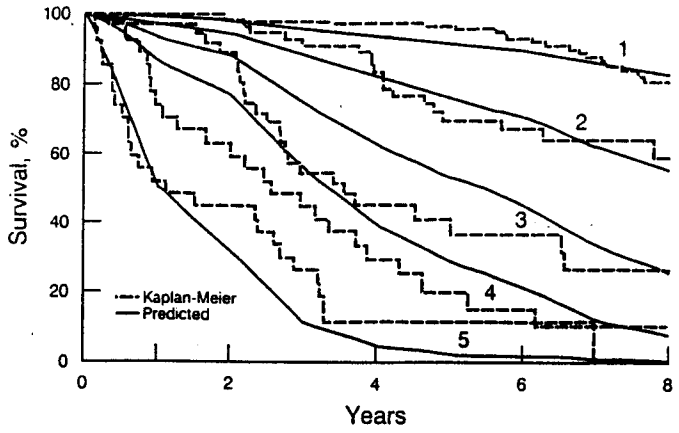

FIG. 1. Actual (Kaplan-Meier) and predicted survival curves for risk quintiles for the 312 Mayo study patients. Time zero is time of entry into trial.

χ^2 statistic for this model is nearly as large as for the 10-variable model—199.2 and 218.7, respectively—and the separation among risk group quintiles is comparable (data not presented for 10-variable model). Therefore, this five-variable model, involving only noninvasive, easily collected variables, provided nearly the same amount of prognostic information for patient survival as did the 10-variable model developed from all 45 variables.

Cross-Validation

Table 3 compares the prognostic variables and R values among three sets of patients: patients used for development of the European model, patients used for development of the Mayo model and independent set of Mayo patients used for cross-validation. The data on the European patients were obtained by averaging the data from the two European treatment groups from Table 2 of their paper (3). The three groups of patients are very similar in terms of risk and individual prognostic variables.

In a scatterplot of Rs (centered at 0.0 by subtracting the means) for the 106 Mayo cross-validation patients calculated by each of the two models, the points lie scattered closely around the line of identity and have a high correlation (Pearson's $r = 0.92$; $p < 0.001$) (Fig. 2). Therefore, the two models assess relative risk very similarly but not identically. In fact, regression of the European scores on the Mayo scores has an estimated slope (±S.E.) of 1.17 ± 0.04, which is significantly greater than 1.0 ($p < 0.05$). This means that, on average, the difference in Rs between any two patients taken at random would be about 17% greater by the European model than by the Mayo model.

The predictions of the two models for the 106 cross-validation patients were compared with the actual survival experience (Fig. 3). The two models are similar, and both order the risk groups correctly. However, the difference between the low- and the high-risk survival predictions is greater for the European model than for the Mayo model, reflecting the greater spread in European model risk scores described above. The European model tends to predict a worse survival than the actual experience, particularly for the higher-risk groups. None

TABLE 3. Characteristics of patients in European, Mayo study and Mayo cross-validation studies

	European[a] (n = 248)	Mayo study (n = 312)	Mayo cross-validation (n = 106)
Age (years; mean)[b]	54.8	50.0	52.9
Bilirubin (mg/dl; mean)[c]	1.99	1.78	1.75
Albumin (gm/dl; mean)	3.46	3.52	3.43
Prothrombin time (sec; mean)		10.73	10.75
On diuretic treatment (%)	13.4	15.7	19.8
With edema (%)	7.5[d]	12	12
Stage 1 (%)	13	5	5
Stage 2 (%)	43.5	22	25
Stage 3 (%)	15	38	35
Stage 4 (%)	28.5	35	35
Central cholastasis (%)	17		17
On azathioprine (%)	51	0	0
Risk score (European model)			
10th percentile	1.25	(1.26)[e]	1.43
50th percentile	3.00	(2.93)[e]	2.80
90th percentile	5.10	(5.68)[e]	5.66
Risk score (Mayo model)			
10th percentile		4.03	4.24
50th percentile		5.24	5.41
90th percentile		7.67	7.45

[a] From Table 2, p. 1087, of Ref. (3).
[b] Ranges were: European, 25 to 78; Mayo study, 26 to 78; Mayo cross-validation patients, 33 to 75.
[c] $\log_e$ used; geometric mean reported.
[d] Ascites used; edema data not presented in Ref. (3).
[e] Cholestasis estimated from equation developed on 106 cross-validation patients.

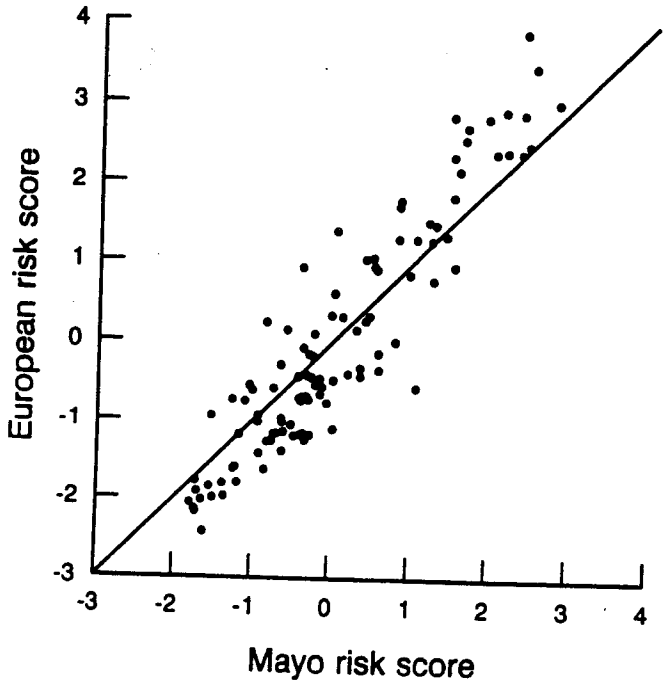

FIG. 2. Scatterplot of Rs from Mayo and European models for the 106 Mayo cross-validation patients. r = 0.92; p < 0.001

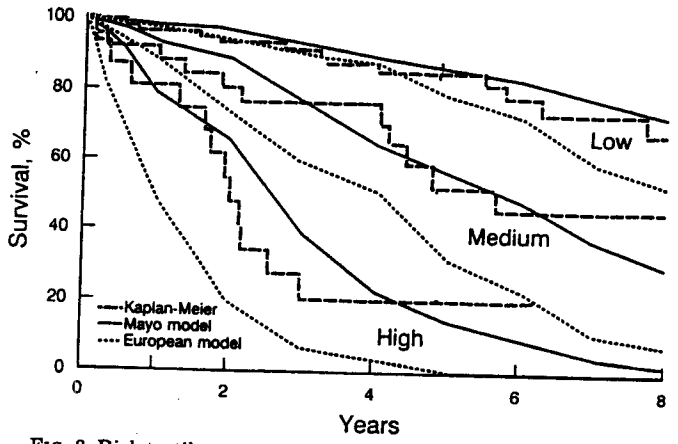

FIG. 3. Risk tertiles comparing Mayo and European models to actual (Kaplan-Meier) survival curves for the 106 Mayo cross-validation patients. Time zero is time of entry into trial.

of the log-rank tests showed a significant difference between the Mayo model and the actual experience (p = 0.53, 0.69 and 0.57 for low-, medium- and high-risk groups, respectively), whereas two of the three log-rank tests for the European model showed a significant discrepancy (p = 0.92, 0.03 and <0.001). As discussed in the "Patients and Methods" section, these p values must be interpreted with caution. It seems reasonable to conclude that the Mayo and European models behave similarly when applied to the cross-validation data set. The superiority of the Mayo model might have been expected because this validation was performed on a Mayo patient data base. Both give reasonably accurate predictions, but the Mayo model appears to be somewhat more accurate for the Mayo cross-validation patients.

Final Mayo Model

Because of the good fit of the Mayo model to the independent data set and the similarity of the original study patients and the cross-validation patients, the Mayo data sets were combined to increase the accuracy of the Mayo model. The last column of Table 2 gives the

results. Compared with the original model, with the combined set the β coefficients changed very little and the standard errors decreased, as would be expected because of the increased sample size.

Use of the Mayo Model

The Mayo model is intended to be used for medical management. It can be utilized with a sophisticated hand calculator or is easily programmed into a computer.

The first step is to compute R. With the β coefficients from Table 2 and the values of the patients' risk variables, one obtains:

$$R = 0.871 \log_e(\text{bilirubin in mg/dl}) + -2.53 \log_e(\text{albumin in gm/dl})\dagger + 0.039 \text{ age in years} + 2.38 \log_e(\text{prothrombin time in sec}) + 0.859 \text{ edema}$$

Then, to obtain the probability of survival for at least t more years, one reads $S_0(t)$ from Table 4 and computes $S(t) = \{S_0(t)\}^{\exp(R-5.07)}$. The value 5.07 is chosen so that $S_0(t)$ gives the survival probability for an individual with R = 5.07, which is the mean R from the combined data set of 418 patients. Thus, $S_0(t)$ is meaningful clinically as well as theoretically.

Example 1: Consider a hypothetical low-risk patient with the following variables: serum total bilirubin = 0.5 mg per dl; serum albumin = 4.5 gm per dl; age = 52 years; prothrombin time = 10.1 sec; no edema; not on diuretic therapy.

Insertion of these values into the above formula gives:

$$R = 0.871 \times \log_e(0.5) - 2.53 \times \log_e(4.5) + 0.039 \times 52 + 2.38 \log_e(10.1) + 0.859 \times 0.0 = 3.12$$

We gave this patient a score of 0.0 for edema in keeping with the definition of the edema variable. This R and the $S_0(t)$ function are used to compute this patient's survival probability for any period up to 7 years. For example, the estimated 5-year survival is $0.774^{\exp(3.12-5.07)}$ because $S_0(t)$ at 5 years is 0.774. The computed value is 0.96; this patient has a 96% chance to survive for at least 5 more years.

If the patient has a reasonable quality of life and no complications, it would be premature to consider liver transplantation at this time, and caution should be used in considering a medical treatment associated with potentially severe side effects.

† Albumin should be measured by serum protein electrophoresis.

TABLE 4. Underlying survival function for the final Mayo model

t (years)	1	2	3	4	5	6	7
$S_0(t)$	0.970	0.941	0.883	0.833	0.774	0.721	0.651

$S_0(t)$ gives the survival probabilities for a patient with risk score 5.07, the mean of the combined Mayo data set.

Example 2: Consider a high-risk patient with: serum total bilirubin = 13.9 mg per dl; serum albumin = 2.8 gm per dl; age = 52 years; prothrombin time = 13.8 sec; no edema; on diuretic therapy. For this patient, the calculation is:

$$R = 0.871 \times \log_e(13.9) - 2.53 \times \log_e(2.8) + 0.039 \times 52 + 2.38 \log_e(13.8) + 0.859 \times 0.5 = 8.39$$

This R is much higher than R in the previous example, indicating a poorer prognosis. The estimated 1-year survival is $0.970^{\exp(8.39-5.07)} = 0.43$. Under most circumstances, such a high-risk patient would be considered a candidate for liver transplantation.

DISCUSSION

Comparison with Other Models. Table 5 compares the Mayo model with the two other multivariate survival models for primary biliary cirrhosis. All three have similar numbers of variables and involve age and bilirubin concentration. The Rs for the three models are strongly and significantly correlated, but the Mayo and European models show higher correlation (Pearson's r = 0.92) than the Yale and Mayo models (r = 0.76). One would expect the correlation between Yale and Mayo models to be smaller because, with the exception of age, the Yale model uses only discrete variables whereas the Mayo and European models use mostly continuous ones. The Yale model used the estimated date of onset of primary biliary cirrhosis as the initial time point, although the variables in the model were measured at the date of diagnosis. The Mayo and European models used the date of entry into or eligibility for a clinical trial. This difference in initial time points could attenuate the correlation.

The most important difference among the models is the fact that the Mayo model does not require liver biopsy whereas both the European and the Yale model use histologic stage and the European model also utilizes the presence or absence of central cholestasis. The Mayo model does not need histologic stage because there is a strong association [Kendall's τ correlation coefficient for ordered data, 0.38 ($p < 0.0001$)] between stage and Mayo R and therefore the information provided by histologic stage does not make a significant additional contribution.

TABLE 5. Comparison of description of three Cox regression models for survival with primary biliary cirrhosis

Variables used		
Yale (2) (n = 238)	European (3) (n = 216)	Mayo (n = 418)
Variables not requiring liver biopsy		
Bilirubin	Bilirubin	Bilirubin
Age	Age	Age
Hepatomegaly	Albumin	Albumin
	Use of azathioprine	Prothrombin time
		Edema
Variables requiring liver biopsy		
Portal fibrosis vs. bridging fibrosis or cirrhosis	Cirrhosis	None
	Central cholestasis	

Concluding Comments. The Mayo model is a promising tool for survival modeling in PBC. It should enhance the clinician's decision-making process with respect to the management of primary biliary cirrhosis patients. We have seen examples illustrating its use in decisions concerning the timing of liver transplantation. It can also be useful in evaluating the efficacy of such a therapeutic option. Although liver transplantation has been accepted clinically as life-saving in various liver diseases, including primary biliary cirrhosis, no controlled trials have been performed to assess the efficacy of liver transplantation. Indeed, because of a marked improvement in survival after transplantation during the past 5 years, randomization of patients with advanced liver disease to a nontransplant control group has been considered clinically inappropriate. The Mayo model can be used as a mathematical control group. Of course, a mathematical control group is an imperfect substitute for a randomized control group, but it may be the only practical alternative in some situations. By applying the Mayo model to a primary biliary cirrhosis transplantation population, one can compare the actual posttransplantation survival curve to the estimated survival in the absence of transplantation, as predicted by the Mayo model. Additional validation of the model with other independent data sets is desirable and could result in further refinement and improvement.

REFERENCES

1. Cox DR. Regression models and life-tables (with discussion). J R Stat Soc [B] 1972; 34:187–202.
2. Roll J, Boyer JL, Barry D, et al. The prognostic importance of clinical and histologic features in asymptomatic and symptomatic primary biliary cirrhosis. N Engl J Med 1983; 308:1–7.
3. Christensen E, Neuberger J, Crowe J, et al. Beneficial effect of azathioprine and prediction of prognosis in primary biliary cirrhosis. Final results of an international trial. Gastroenterology 1985; 89:1084–1091.
4. Dickson ER, Fleming TR, Wiesner RH, et al. Trial of penicillamine in advanced primary biliary cirrhosis. N Engl J Med 1985; 312:1011–1015.
5. Wiesner RH, Grambsch PM, Lindor KD, et al. Clinical and statistical analyses of new and evolving therapies for primary biliary cirrhosis. Hepatology 1988; 8:668–676.
6. Ludwig J, Dickson ER, McDonald GS. Staging of chronic nonsuppurative destructive cholangitis (syndrome of primary biliary cirrhosis). Virchows Arch A 1978; 379:103–112.
7. Christensen E. Multivariate survival analysis using Cox's regression model. Hepatology 1987; 7:1346–1358.
8. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: John Wiley & Sons, 1980.
9. SAS Institute, Inc. SUGI supplemental library user's guide, version 5 ed. Cary, North Carolina: SAS Institute, Inc., 1986.
10. Harrell FE Jr, Lee KL. Verifying assumptions of the Cox proportional hazards model. In: Proceedings of the eleventh international conference of the SAS user's group. Atlanta, Georgia, February 9–12, 1986: 823–828.
11. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958; 53:457–481.
12. Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. Biometrika 1982; 69:553–566.

PBC data on 16 demographic, clinical, biochemical, and histologic measurements made at the time of randomization. Patient age and sex are the demographic variables. Clinical measurements recorded the presence or absence of ascites, hepatomegaly, spiders, and edema. Ascites is an accumulation of fluid in the abdominal cavity; hepatomegaly is a swelling or enlargement of the liver; spiders are vascular lesions formed by the dilation of a small group of blood vessels, generally occurring on the upper chest and arms; and edema is a swelling caused by excess fluid in subcutaneous tissue. Biochemical measurements included the levels of bilirubin, albumin, urine copper, alkaline phosphatase, SGOT, cholesterol and triglycerides. Bilirubin is a liver bile pigment; albumin is a protein found in the blood; urine copper is self-explanatory; alkaline phosphatase and SGOT are enzymes. Serum cholesterol and triglycerides are blood lipoproteins. In addition, the patient's platelet count and prothrombin time were recorded. Prothrombin is a blood coagulation agent, and prothrombin time is the time until a blood sample begins coagulation in a certain laboratory test. Finally, Appendix D contains the histologic stage of the disease, graded 1, 2, 3, or 4, with higher stage disease denoting worse prognosis. Specific coding and units of measurement for all of the variables can be found in that Appendix.

The proportional hazards model (Cox, 1972) is frequently used to estimate the effect of one or more covariates on a failure time distribution. Let $\mathbf{Z}' = (Z_1, \ldots, Z_p)$ denote p measured covariates on a given individual with censored failure time observation $(X = \min(T, U), \delta)$. In the proportional hazards model,

$$\begin{aligned} \lambda(t|\mathbf{Z}) &\equiv \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\{t \le T < t + \Delta t \mid T \ge t, \mathbf{Z}\} \\ &= \lambda_0(t) \exp(\beta' \mathbf{Z}), \end{aligned} \tag{2.3}$$

where $\lambda_0(t)$ is a baseline hazard corresponding to $\mathbf{Z}' = (0, \ldots, 0)$, $\beta' = (\beta_1, \ldots, \beta_p)$ is a vector of regression coefficients, and $\beta'\mathbf{Z}$ is an inner product. Standard likelihood methods cannot be used to estimate β when no parametric model is used for λ_0. The components of β can be estimated, however, by maximizing a "partial likelihood" function with respect to β, even when λ_0 is left completely unspecified. This approach is discussed in detail in Chapter 4, and we provide only an outline here.

The partial likelihood function in the proportional hazards model is based on a conditional probability argument. Suppose $(A_1, B_1), (A_2, B_2), \ldots, (A_K, B_K)$ is a collection of pairs of events. Then the likelihood of all $2K$ events is:

$$\begin{aligned} &P\{A_K B_K A_{K-1} B_{K-1} \ldots A_1 B_1\} \\ &\quad = \left[\prod_{k=2}^{K} P\{A_k B_k \mid A_{k-1} B_{k-1} \ldots A_1 B_1\}\right] P\{A_1 B_1\} \\ &\quad = \left[\prod_{k=2}^{K} P\{A_k \mid B_k A_{k-1} B_{k-1} \ldots A_1 B_1\}\right] P\{A_1 \mid B_1\} \end{aligned}$$