GASTROENTEROLOGY 1985;89:1084–91

Beneficial Effect of Azathioprine and Prediction of Prognosis in Primary Biliary Cirrhosis

Final Results of an International Trial

ERIK CHRISTENSEN, JAMES NEUBERGER, JOHN CROWE, DOUGLAS G. ALTMAN, HANS POPPER, BERNARD PORTMANN, DEBORAH DONIACH, LEO RANEK, NIELS TYGSTRUP, and ROGER WILLIAMS

Division of Hepatology, Medical Department A, Rigshospitalet; Department of Hepatology, Hvidovre Hospital, Copenhagen, Denmark; Liver Unit, King's College Hospital and Medical School, London, England; Stratton Laboratory for the Study of Liver Diseases, Mount Sinai School of Medicine of the City University of New York, New York, New York; Department of Immunology, Middlesex Hospital, London, England; and Statistical Research Unit, Danish Medical and Social Science Research Councils

The effect of azathioprine on survival of patients with primary biliary cirrhosis was studied prospectively in a multinational, double-blind, randomized clinical trial including 248 patients of whom 127 received azathioprine and 121 placebo. There were 57 deaths in the azathioprine group and 62 in the placebo group. The actual survival was slightly longer during azathioprine than during placebo treatment. Using Cox multiple regression analysis and adjusting for slight imbalance between the two treatment groups, the therapeutic effect of azathioprine was statistically significant ($p = 0.01$), with azathioprine reducing the risk of dying to 59% of that observed during placebo treatment (95% confidence interval 40%–90%) or improving survival time by 20 mo in the average patient. Furthermore, azathioprine slowed down progressing incapacitation. Side effects of azathioprine were relatively few. The analysis revealed that the following five variables independently implied poor prognosis: high serum bilirubin, old age, cirrhosis, low serum albumin, and central cholestasis. These factors were combined to a "prognostic index" for prediction of outcome in new patients. The index was validated on independent patient data. On the basis of these results we recommend azathioprine as a routine treatment of primary biliary cirrhosis.

The etiology of primary biliary cirrhosis (PBC) is unknown, but because immune reactions seem to be involved in the pathogenesis of the disease, an immunosuppressive drug like azathioprine may be expected to be effective. A controlled trial of azathioprine versus no treatment performed at the Royal Free Hospital in London found that the development of cirrhosis was not prevented and survival was not improved by the treatment (1). However, because only 45 patients were included in the study, the risk that a therapeutic effect might have been

Received June 27, 1984. Accepted April 22, 1985.

Address requests for reprints to: Erik Christensen, M.D., Department of Hepatology 233, Hvidovre Hospital, University of Copenhagen, DK-2650 Copenhagen, Denmark.

This work was supported by grant No. 512-10526 and 12-3500 from the Danish Medical Research Council and by grants from the Professor Brøchner-Mortensen Foundation and the British Society of Gastroenterology. This work was also supported by the Wellcome Foundation. J.N. was supported by Ciba-Geigy Ltd.

The authors thank the Wellcome Foundation for the supply and distribution of the tablets. The authors also thank Jean-Pierre Benhamou, Clichy, France; Jan de Groote, Leuven, Belgium; Kurt Iversen, Copenhagen, Denmark; Erik Juhl, Hvidovre, Denmark; Steven Mistilis, Sidney, Australia; Helmer Ring-Larsen, Hvidovre, Denmark; Juan Rodés, Barcelona, Spain; Fenton Schaffner, New York, New York; and Geoffrey Watkinson, Glasgow, Scotland, for their cooperation; Per Kragh Andersen for suggesting the method for calculating confidence intervals for the estimated survival function; and Tove Thomsen for assistance. The authors also thank the Department of Data Processing at Rigshospitalet for generous access to their computer.

0016-5085/85/$3.30

Abbreviations used in this paper: PBC, primary biliary cirrhosis; PI, prognostic index.

missed (a type II error) was considerable. In 1980 a preliminary report of an international trial of azathioprine versus placebo was published (2). The results at that time indicated weak trends toward a beneficial effect of azathioprine, but the duration of follow-up was too short to provide a definite conclusion. In spite of the higher number of patients studied there still was a high risk of not detecting a substantial therapeutic effect.

This report presents the final and more precise results of the international trial on the therapeutic effect of azathioprine in PBC based on a follow-up of longer duration. The detailed investigation of prognostically important variables using proportional hazards regression is described and a prognostic index derived from these variables is evaluated.

Patients and Methods

Patients fulfilling the following criteria were included in the trial: clinical picture and histologic features compatible with PBC, alkaline phosphatase activity greater than twice the upper limit of normal in the absence of evidence of extrahepatic biliary obstruction, and no antimetabolites (e.g., azathioprine, 6-MP) administered within 6 mo. No patient received such drugs before the study. Patients were entered regardless of age, duration of symptoms, or histologic stage at the time of diagnosis. Informed consent was obtained before entry into the study and approval was also given by the ethical committee in each hospital. Patients were randomized to azathioprine or placebo separately for each center and for each sex by the sealed envelope technique. The identically looking tablets contained 50 mg of azathioprine or lactose. Patients weighing 40 kg or less received six tablets per week. For each 10-kg increment in body weight the dosage was increased by two tablets per week to a maximum dosage of 100 mg/day. For the first 2 wk, half of the indicated dose was given. Leukocytes and platelet counts were made every 2 wk for 2 mo, and monthly thereafter. If the leukocyte count dropped below 2000/μl or the platelet count below 20,000/μl, treatment was temporarily discontinued.

Clinical assessment was carried out at entry into the trial and at 6-mo intervals thereafter. This included an estimation of the degree of incapacitation based on the number of days during the previous 100 days spent in each of the following categories: (a) normal health; (b) reduced well-being; (c) without capacity for work, but out of bed; (d) at home in bed; (e) in bed in the hospital. The degree of incapacitation was calculated as follows:

$$\text{Incapacitation index} = (0a + 1b + 2c + 3d + 4e)/4.$$

Standard liver function tests and measurements of serum immunoglobulins and autoantibodies were also performed approximately every 6 mo. Before entry into the trial and at yearly intervals thereafter, a liver biopsy was performed. The biopsy specimen was assessed by a histopathologist (H. P. or B. P.) without knowledge of the clinical condition of the patient or the treatment given.

The actual survival curves (Kaplan–Meier plots) for the treatment and control groups were compared using the logrank test (3), taking the survival time as the time from entry into the trial to the last observation (either death or censoring). Thus the follow-up on patients withdrawn from treatment was continued and the total period of observation was used in the survival analyses according to the "intention to treat" principle. The effect of therapy on the progression of incapacitation was analyzed in a similar way, using the time from entry into the trial to the first occurrence of an incapacitation index >25 in patients who had lower values at entry into the trial, as described previously (2,4). To refine the survival analysis, the simple (marginal) relation between all clinical, biochemical, and histologic variables and survival was analyzed in each of the two treatment groups and for both groups together. Variables showing a relation with survival either significantly different ($p < 0.05$) in the two treatment groups ("therapeutic variables") or significant in either or both of the treatment groups ("prognostic variables") were retained so that further multivariate analysis could be performed to determine which variables had independent influences on survival and to adjust for differences in such factors between the two treatment groups at the time of entry into the trial. We used the Cox multiple regression model for censored survival data (5). According to this model the hazard or risk of death $\lambda(t)$ at time t after randomization for a patient with variables $z_1 \cdots z_p$ is

$$\lambda(t) = \lambda_0(t) \exp(b_1 z_1 + \cdots + b_p z_p), \qquad t > 0,$$

where $\lambda_0(t)$ is the so-called underlying hazard and $b_1 \cdots b_p$ are regression coefficients. If a regression coefficient b_i is positive, higher values of the corresponding variable z_i indicate higher hazard or worse prognosis, and vice versa if b_i is negative. If b_i is zero the corresponding variable z_i has no influence on survival.

Before inclusion in the Cox analysis, the distributions of the continuous variables were checked for normality (6) and, if necessary, a logarithmic transformation was made in order to prevent undue influence on the regression analysis of a few extreme observations.

The assumption of proportional hazards was checked by observing constant vertical differences (independent of time) between plots of estimates of the logarithm of the integrated (cumulative) hazard function against time for various levels of each variable (7). For variables with more than two levels, including the continuous variables, the lines of the plots for equally spaced strata were checked for equal spacing in order to fulfill the assumption of linearity. Only for age was a transformation necessary; of several possible scorings, the exponential gave the best fit. For each variable in the final model the assumption of proportional hazards was also tested by the goodness of fit test proposed by Andersen (8), and in no case was the assumption rejected.

Because 25 variables were significant by marginal analysis they had to be entered in the Cox analysis stepwise in three groups. Only those variables found to be significant at the 5% level were retained in the model at each step. Thereafter it was checked that none of the eliminated variables that had at some stage been significant in the model could not be reintroduced. This was performed by

backward elimination in a model that included such variables together with those that had maintained their significance. The final regression analysis and subsequent calculations were based on 216 patients with complete data.

The analysis revealed a number of prognostic variables but no therapeutic variables except the treatment itself (azathioprine or placebo), so the final Cox regression model had the following form:

$$\lambda(t) = \lambda_0(t) \exp(b_1 z_1 + \cdots + b_p z_p + b_{tr} z_{tr}),$$

where b_{tr} is the overall treatment effect coefficient and z_{tr} is the indicator for one of the treatment groups (i.e., $z_{tr} = 0$ for azathioprine; $z_{tr} = 1$ for placebo). This can be rewritten to define a prognostic index (PI):

$$\text{PI} = \log_e[\lambda(t)/\lambda_0(t)] = b_1 z_1 + \cdots + b_p z_p + b_{tr} z_{tr}.$$

Higher values of the PI mean higher risk, i.e., worse prognosis (shorter survival), and lower (including negative) values mean better prognosis. To facilitate interpretation, the PI and the estimated integrated underlying hazard function $\hat{\Lambda}_0(t)$ can be combined into an estimate of the survivorship function $S(t, z)$ for patients with covariates $z = (z_1 \cdots z_p)$, namely, the estimate

$$\hat{S}(t, z) = \exp\{[-\exp(\hat{PI})]\hat{\Lambda}_0(t)\}.$$

The Appendix describes the method for deriving 95% confidence limits for the survival function.

After estimation of $\hat{\Lambda}_0(t)$ (see Appendix), a graph of the estimated probability of surviving a given time, e.g., the 5-yr survival probability as a function of the PI, can be constructed. The median survival time (the time span the patient will survive with 50% probability), another measure of prognosis, can be estimated for a patient with given variables as the shortest time for which the estimated probability of surviving that long is <0.5. It is not possible to estimate confidence limits for these functions as they depend on the values of the individual variables z_i.

The PI was validated on another group of patients, i.e., the control group of the controlled trial of penicillamine versus placebo, which was started when inclusion of patients into the present trial ended, and which was conducted in the same way as this trial (4). The individual survivorship functions for the new group of patients were estimated as described above. Patients were divided into three groups according to their PI values. The average estimated survivorship functions in the three groups (see Appendix) were compared with Kaplan–Meier plots of the observed data, and in each group the difference was tested using the one sample logrank test (9).

Results

Between October 1971 and December 1977, 248 patients were entered into the trial from seven national centers. The distribution of patients according to treatment and center is shown in Table 1. Table 2 shows that the two groups were broadly similar at entry into the trial. Formal significance testing was not carried out as this does not indicate the clinical importance of any differences (10). It is of interest in view of the subsequent results that the difference in bilirubin level between the azathioprine and the placebo groups was not statistically significant.

Patients were followed up to March 1983. Of the 248 patients, 63 were lost to follow-up (29 in the azathioprine group and 34 in the placebo group). Twenty patients were withdrawn in the azathioprine group for the following reasons: rash (6), nausea and vomiting (6), marrow depression (2), liver transplant (1), pregnancy (1), hair loss (1), immune complex disease (1), cancer of the sigmoid (1), and myocardial infarction (1). Ten patients in the placebo group were withdrawn for the following reasons: rash (2), nausea and vomiting (1), marrow depression (1), headache (2), rash and arthralgia (1), steatorrhea (1), and liver transplant (2). The number of dropouts and withdrawals was not significantly different in the two groups.

Clinically, azathioprine slowed down progressing incapacitation as evidenced by a reduction in the risk of developing an incapacitation index >25 [relative risk (azathioprine/placebo) = 0.63, p = 0.05].

As shown in Table 3, the causes of death were not significantly different in the two groups. The actual survival curves for the two groups are shown in Figure 1. There was a tendency (p = 0.10) toward a longer survival in the azathioprine group compared with the placebo group.

The Cox multiple regression analysis revealed that six variables, including the treatment (azathioprine or placebo), had a significant independent prognostic influence (Table 4). Thus high serum bilirubin, older age, presence of cirrhosis, presence of central cholestasis, and placebo therapy (all having positive *b* coefficients) were associated with poor prognosis, whereas high albumin (having a negative coefficient) had a beneficial effect on prognosis. Of the prognostic variables, the serum bilirubin was by far the most important. The significant therapeutic effect in the multivariate analysis is corrected for differences in prognostic factors between the two treatment groups.

Table 1. Patients Included by Treatment and Center

Center	Azathioprine	Placebo	Total[a]
London	70	71	141
Copenhagen	29	34	63
New York	7	6	13
Barcelona	8	3	11
Clichy	7	2	9
Leuven	6	2	8
Sidney	0	3	3
Total[b]	127	121	248

[a] Total patients from the same center. [b] Total patients receiving the same treatment.

Table 2. Comparison of Treatment Groups at Start of Trial

Variable	Azathioprine	Placebo
General		
Mean age (yr) (range 25–78 yr)	54.7	54.9
Males	9%	12%
Duration of history <12 mo	74%	70%
Clinical		
Pruritus	74%	73%
Jaundice	60%	56%
Pigmentation	53%	54%
Xanthomata	26%	26%
GI hemorrhage	18%	16%
Ascites	6%	9%
Incapacitation index		
>10	35%	42%
>25	16%	18%
>50	2%	7%
Cholestyramine treatment	35%	29%
Diuretic treatment	10%	17%
Laboratory		
Mean bilirubin (μmol/L) (3–20)[a]	37.2	30.9
Mean alkaline phosphatase (IU/L) (30–85)[a]	661	490
Mean alanine aminotransferase (IU/L) (7–40)[a]	107	105
Mean cholesterol (mmol/L) (3.0–8.3)[a]	7.94	7.94
Mean albumin (g/L) (35–50)	34.8	34.4
Mean IgG (g/L) (6–16)[a]	15.5	15.1
Mean IgA (g/L) (1.25–4.25)[a]	2.98	2.55
Mean IgM (g/L) (0.5–2.0)[a]	3.01	2.88
Mitochondrial antibodies	93%	86%
Histologic		
Stage I	14%	12%
Stage II	44%	43%
Stage III	15%	15%
Stage IV	27%	30%
Intralobular hepatitis	27%	34%
Intralobular inflammation	49%	37%
Granulomas	23%	22%
Bile duct destruction	30%	32%
Proliferation and destruction of ductules	80%	80%
Piecemeal necrosis	66%	78%
Lymphoid follicles	36%	44%
Fibrosis without cirrhosis	62%	57%
Cirrhosis	27%	30%
Central cholestasis	18%	16%
Peripheral cholestasis	37%	40%

[a] Antilog of mean of logarithmic values. GI, gastrointestinal; Ig, immunoglobulin.

Thus the slight imbalance, especially with respect to bilirubin (the azathioprine group having a higher bilirubin than the placebo group), explains why the uncorrected treatment effect was not statistically significant (Figure 1). After adjusting for imbalance in prognostic variables using the Cox regression model, a statistically significant beneficial effect of azathioprine on survival was revealed (p = 0.01) as seen in Figure 2. It should be noted that adjusting only for the imbalance in serum bilirubin made the treatment effect highly significant. Adjusting also for the other prognostic variables had little additional effect as they were well balanced between the treat-

Table 3. Main Cause of Death

Cause	Azathioprine	Placebo
Liver failure (with or without renal failure)	21	22
Liver failure and GI bleeding	16	19
GI bleeding	5	7
Malignancy	3	6
Hepatocellular carcinoma	1	1
Bronchogenic carcinoma	0	3
Cancer of the pancreas	0	1
Cancer of the stomach	1	0
Carcinoma of the gallbladder	0	1
Peritoneal carcinosis	1	0
Infection	6	3
Septicemia	2	0
Bronchopneumonia	3	1
Pneumococcal meningitis	0	1
Miliary tuberculosis	0	1
Unspecified	1	0
Cardiovascular disease	4	2
Myocardial infarction	3	0
Cerebrovascular hemorrhage	0	1
Pulmonary embolus	1	0
Unspecified	0	1
Miscellaneous	1	2
Liver transplant	0	1
Esophageal perforation	0	1
Automobile accident	1	0
Unspecified	1	1
Total	57	62

GI, gastrointestinal.

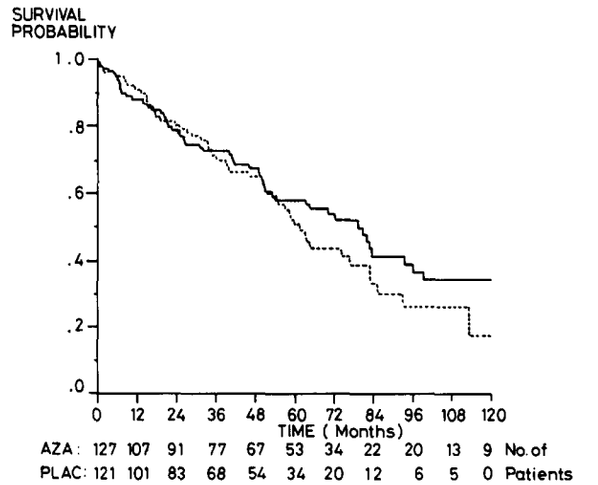

Figure 1. Actual Kaplan–Meier survival curves for patients treated with azathioprine (*AZA*) (——) and placebo (*PLAC*) (-----) (p = 0.10).

Table 4. Significant Prognostic Variables and Their Regression Coefficients in the Final Cox Regression Model

Variable	Scoring	Regression coefficient (*b*)	Standard error [*SE*(*b*)]	Normal deviate [*b*/*SE*(*b*)]	p
Serum bilirubin	Log_{10} (value in μmol/L)	2.51	0.32	7.94	<0.0001
Age	Exp [(age in years −20)/10]	0.0069	0.0016	4.27	<0.0001
Cirrhosis	Absent: 0 Present: 1	0.88	0.22	4.07	<0.0001
Serum albumin	Value in g/L	−0.050	0.018	2.77	0.006
Central cholestasis	Absent: 0 Present: 1	0.68	0.27	2.47	0.01
Therapy	Azathioprine: 0 Placebo: 1	0.52	0.21	2.50	0.01

ment groups. The curves in Figure 2 are higher than in Figure 1 because cirrhosis and central cholestasis were set to zero.

Splenomegaly, ascites, pigmentation, and jaundice tended to have a harmful prognostic influence but not significantly so ($p > 0.2$) and, therefore, these variables were not included in the final model presented in Table 4. The center where the patients had been followed had no significant influence on the model.

No variable had a statistically significant influence on the treatment effect, that is, no therapeutic variable was identified. Weak trends toward an increased beneficial effect of azathioprine were observed with low alanine aminotransferase, low immunoglobulin M, high cholesterol, hepatomegaly, and proliferation and destruction of ductules in the liver biopsy specimen. It should be emphasized, however, that these trends were far from statistical significance ($p > 0.2$). Therefore none of these variables was included in the final model.

The prognostic variables presented in Table 4 can be used to estimate prognosis from any given patient's data at the time of diagnosis by calculating the PI. If, for example, a patient has the following variables: serum bilirubin 32 μmol/L ($z_1 = \log_{10}32 = 1.51$), age 68 yr ($z_2 = \exp[(68 - 20)/10] = 121.5$), no cirrhosis ($z_3 = 0$), serum albumin 32 g/L ($z_4 = 32$), no central cholestasis ($z_5 = 0$), and is treated with azathioprine ($z_{tr} = 0$), then

$$PI = 2.51 \times 1.51 + 0.0069 \times 121.5 + 0.88 \times 0 - 0.05 \times 32 + 0.68 \times 0 + 0.52 \times 0 = 3.0.$$

The distribution of the PI in our patients is shown in Figure 3.

Using Figure 4 it is possible to derive from the value of the PI for a given patient the estimated probability of surviving 2, 5, or 8 yr. Figure 5 shows the estimated median survival time for any value of the PI. No estimates can be made beyond the time of the last death (99 mo). For the example presented above, the estimated probabilities of surviving 2, 5, and 8 yr are 91%, 65%, and 28%, respectively (Figure 4). The estimated median survival time is 80 mo (Figure 5). If such a patient had been allocated to placebo, the PI would have been increased by 0.52 ($b_{tr}z_{tr} = 0.52 \times 1$) to 3.52. Using Figures 4 and 5 the estimated 5-yr survival probability would have been 50% and the estimated median survival time ~60 mo. Thus the estimated gain in survival time of giving this patient azathioprine would be ~20 mo.

The estimated survival curve for this patient, who in fact had the median PI in this study, is shown in the middle of Figure 6 together with 95% confidence limits. The estimated median survival time and its 95% confidence limits can be read as the times for which the confidence interval for the survival function covers 0.5. In the same figure the estimated

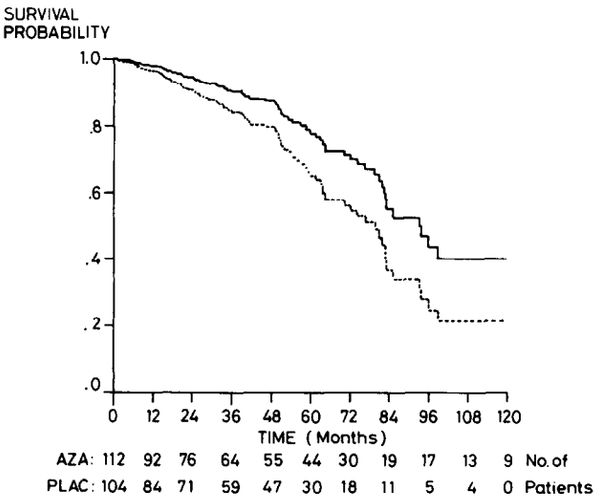

Figure 2. Estimated survival functions for patients treated with azathioprine (*AZA*) (——) and placebo (*PLAC*) (-----) based on the final Cox regression model with cirrhosis and central cholestasis absent and mean values of the other variables ($p = 0.01$).

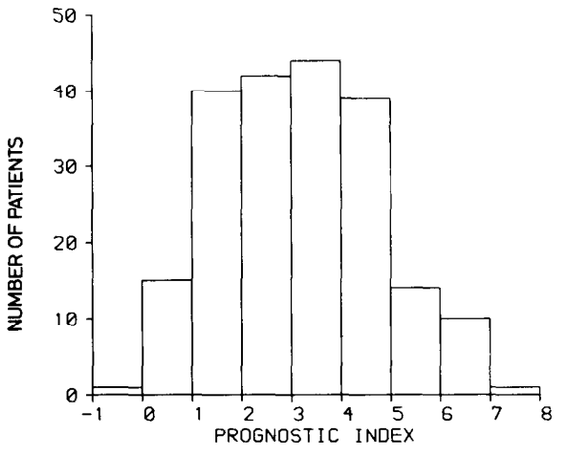

Figure 3. Distribution of the prognostic index for 216 patients with complete data.

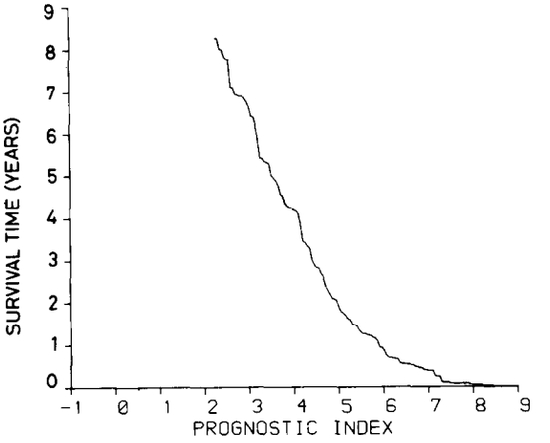

Figure 5. Estimated median survival time by the prognostic index.

survival curves for patients at the 10th and 90th percentile of the distribution of the PI are shown.

The final model was validated as described in Patients and Methods using data from a further group of 85 placebo-treated patients (4). Figure 7 shows the observed and expected survival functions for the new group. The difference was not statistically significant ($p = 0.4$).

Discussion

This trial has demonstrated a statistically significant beneficial effect of azathioprine on survival in patients with PBC. This is in contrast with the controlled trial at the Royal Free Hospital (1) and the early report of this trial (2), which did not find any statistically significant beneficial effect of the drug. However, the chance of not finding a substantial beneficial effect (the type II error) was considerable (a) because of the small sample size (45 patients) in the first case (1) and (b) because of a follow-up of limited duration in the other (2).

Even with longer follow-up the actual survival curves of the present study revealed only a tendency toward a beneficial effect of azathioprine (Figure 1).

The purpose of randomization is to obtain comparable treatment groups, but random allocation does not guarantee complete balance. Random imbalance may occur. For this reason several writers have suggested that imbalance in known or suspected prognostic variables, implying a difference in "spontaneous" prognosis between the treatment groups, should be detected and adjusted for to ensure a fair comparison between the treatments (11–14). For this purpose we performed the multivariate Cox regression analysis, which revealed that five variables had significant independent prognostic influence. Of these, serum bilirubin has long been recognized as a

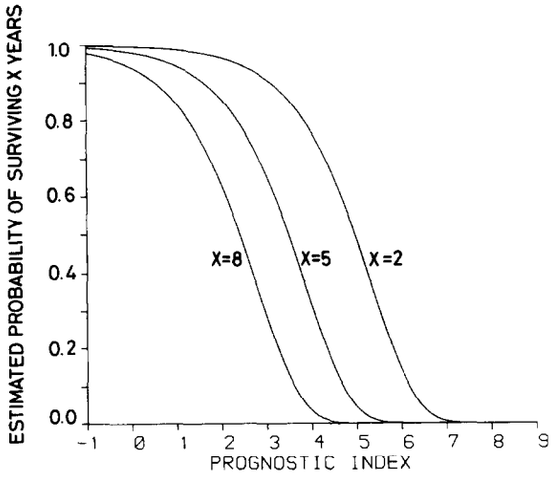

Figure 4. Estimated probability of surviving 2, 5, and 8 yr by the prognostic index.

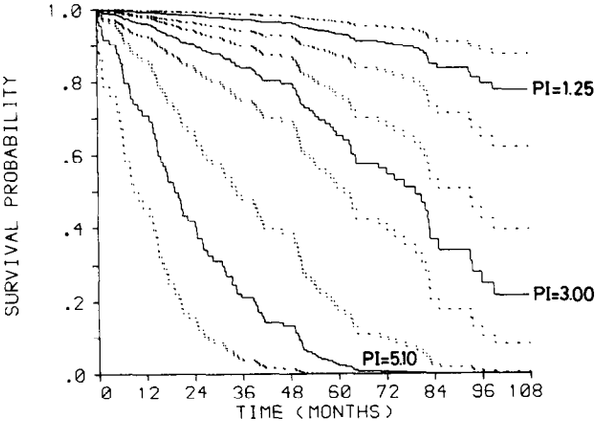

Figure 6. Estimated survival functions with 95% confidence limits for 3 patients at the 10th, 50th, and 90th percentiles of the distribution of the prognostic index (PI) (PI = 1.25, 3.00, and 5.10, respectively).

very important prognostic variable in PBC (15–19). Bilirubin was not well balanced between the treatment groups, the level being slightly but not significantly higher in the azathioprine than in the placebo group, implying a slightly poorer spontaneous prognosis in the azathioprine group than in the placebo group. By adjusting for imbalance in all five prognostic variables (or for imbalance in bilirubin alone), a statistically significant beneficial effect of azathioprine was revealed (Figure 2). The imbalance obscured the beneficial effect of azathioprine in the simple survival analysis. The reason why the rather small imbalance could have such a marked effect is that the prognostic influence of bilirubin is very strong. The multivariate analysis excluded 32 patients with incomplete data. This cannot, however, explain the difference in the estimated effect of therapy with and without adjustment, as the unadjusted comparison of survival in the two therapy groups for only those 216 patients with full data was smaller and less significant. The treatment effect observed in this trial is not very big. Azathioprine, however, reduces the risk of dying (the hazard) to 59% of that observed during placebo treatment [exp(−0.52) = 0.59], with the 95% confidence interval being 40%–90%. For the average patient the gain in survival time was estimated to be 20 mo. In addition, progressing clinical incapacitation seems to be slowed down by the treatment.

We did not identify any variable as having an influence on the magnitude of the therapeutic effect, so that the relative benefit of azathioprine treatment appears to be the same for any patient. The difference in median estimated survival times and hence the gain obtained by active treatment in terms of added survival time, however, is greater in absolute numbers for patients with a relatively good prognosis than for patients with a poorer prognosis.

In addition to serum bilirubin, the Cox regression analysis revealed four other independent prognostic variables. Age has previously been found to be an important prognostic variable in PBC (18,19). This study showed that the risk increases markedly with increasing age, especially over 60 yr, and we found that the exponential scoring of age fit best in the model.

Cirrhosis also had an independent prognostic influence, in contrast to the earlier histologic stages, confirming the previous finding that patients in stage 1, 2, and 3 have a similar prognosis (18). A decreased serum albumin reflects advanced chronic liver disease of almost any etiology. We found serum albumin to be an independent prognostic factor in PBC in contrast with Roll et al. (19), who found albumin to be significant only in a marginal analysis. Central cholestasis implies a poor prognosis and probably

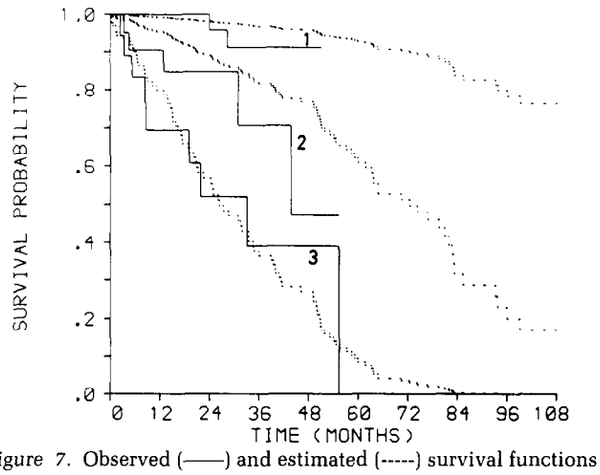

Figure 7. Observed (——) and estimated (-----) survival functions of three groups of placebo-treated patients from a controlled trial of D-penicillamine (Reference 4): group 1: PI < 2.5, n = 44, $O = 2$, $E = 0.9$; group 2: 2.5 < PI < 3.75, n = 22, $O = 5$, $E = 2.7$; group 3: PI > 3.75, n = 19, $O = 9$, $E = 9.0$. O and E are the observed and expected numbers dying in each group ($\chi^2_3 = 3.2$, $p = 0.4$) (Reference 9).

reflects an additional aspect of cholestasis that is not reflected in the serum bilirubin level.

In agreement with Lee et al. (20), we found that the presence of granulomas had a significant association with a good prognosis in the total group of patients by marginal analysis. The association was weak, however, and in the multivariate analysis granulomas were replaced by the more powerful prognostic variables included in the final model.

The multivariate analysis provided the basis for the PI by which the characteristics of a patient can be used to estimate his or her prognosis. The index is relatively simple and can easily be calculated on a pocket calculator. The interpretation of the index is facilitated by the graphs showing the estimated 2-, 5-, and 8-yr survival probabilities and the estimated median survival time as a function of the PI (Figures 4 and 5). The PI can be estimated for each of the therapeutic alternatives and thus the gain in terms of time added to the median survival time to be expected by administering azathioprine can be found.

The validation of the final model using an independent set of data is based on relatively few patients of whom only 16 died [because of shorter follow-up time (4)]. Because of the way the regression model was derived, one would expect the difference in observed survival between the high and low PI groups to be slightly less than predicted through a form of regression to the mean (21). The results suggest that the effect, if present, is small (Figure 7), and that the PI derived from the azathioprine trial does give valid information about the prognosis of new patients.

Although the demonstrated beneficial effect of

azathioprine on survival is not dramatic, the drug must be considered an advance in therapy especially when the alternatives, including penicillamine, are considered. The therapeutic effect of penicillamine varies somewhat in different trials (4) but may be less than that of azathioprine. The number of side effects of azathioprine is relatively small; the drug was well tolerated in most of our patients. This is in marked contrast to penicillamine, where a larger proportion of the patients experienced side effects necessitating withdrawal from treatment (4). Thus, until even better treatments can be found, azathioprine may be regarded as one of the best available medical treatments for PBC. Estimation of the prognosis using the PI may be of value for optimal timing of potentially hazardous treatments such as liver transplantation (22,23).

Appendix

The cumulative underlying hazard function $\Lambda_0(t)$ was estimated as a step function as suggested by Breslow (24). Tsiatis (25) derived the standard error of this estimate and also of $\hat{\Lambda}(t, z_0) = \hat{\Lambda}_0(t) \exp(\hat{b}^T z_0)$, where z_0 is any vector of covariates. He showed that the standard error of the estimated survival function may be estimated from

$$SE\,[\hat{S}(t, z_0)] = \hat{S}(t, z_0) \times SE[\hat{\Lambda}(t, z_0)].$$

This estimate, however, may lead to confidence limits falling outside the range (0,1). This can be avoided by constructing confidence limits for $\log\hat{\Lambda}_0(t)$, which is unbounded, and back-transforming these limits. Using the same principle as Tsiatis [see Rao (26)] but transforming in the other direction we have

$$SE[\log\hat{\Lambda}(t, z_0)] = SE[\hat{\Lambda}(t, z_0)]/\hat{\Lambda}(t, z_0) = s_{lh},$$

and pointwise 95% confidence limits for the log cumulative hazard function can be estimated as $\log [\hat{\Lambda}(t,z_0)] \pm 1.96 \times s_{lh}$. Thus 95% confidence limits for $\hat{S}(t,z_0)$ can be estimated as

$$\exp\{-\exp[\log\hat{\Lambda}(t, z_0) \pm 1.96 \times s_{lh}]\}.$$

These limits are asymmetric and constrained to lie within (0,1).

A similar approach was adopted for the validation of the PI using new data. The expected survival function for a new group of patients was calculated by first averaging the individual estimated log cumulative hazard functions and then back-transforming. This is equivalent, however, to estimating the survival function for the average value of the PI in that group of patients.

References

1. Heathcote J, Ross A, Sherlock S. A prospective controlled trial of azathioprine in primary biliary cirrhosis. Gastroenterology 1976;70:656–60.
2. Crowe J, Christensen E, Smith M, et al. Azathioprine in primary biliary cirrhosis: a preliminary report of an international trial. Gastroenterology 1980;78:1005–10.
3. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. Br J Cancer 1977; 35:1–39.
4. Neuberger J, Christensen E, Portmann B, et al. Double-blind controlled trial of D-penicillamine in patients with primary biliary cirrhosis. Gut 1985;26:114–9.
5. Cox DR. Regression models and life tables (with discussion). J R Statist Soc B 1972;34:187–220.
6. Royston JP. An extension of Shapiro and Wilk's W test for normality to large samples. Appl Statist 1982;31:115–24.
7. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: John Wiley & Sons, 1980.
8. Andersen PK. Testing goodness-of-fit of Cox's regression and life model. Biometrics 1982;38:67–77.
9. Breslow NE. Analysis of survival data under the proportional hazards model. Int Stat Rev 1975;1:45–58.
10. Lavori PW, Louis TA, Bailar JC III, Polansky M. Designs for experiments—parallel comparisons of treatment. N Engl J Med 1983;309:1291–8.
11. Brown BW. Designing for cancer clinical trials: selection of prognostic factors. Cancer Treat Rep 1980;64:499–502.
12. Lachin JM. Statistical analysis of the randomized clinical trial. In: Tygstrup N, Lachin JM, Juhl E, eds. The randomized clinical trial and therapeutic decisions. New York: Marcel Dekker, 1982:155–97.
13. Lewis JA. Clinical trials: statistical developments of practical benefit to the pharmaceutical industry (with discussion). J R Statist Soc A 1983;146:362–93.
14. Altman DG. A fair trial? Br Med J 1984;298:336.
15. Sherlock S. Primary biliary cirrhosis (chronic intrahepatic obstructive jaundice). Gastroenterology 1959;37:574–86.
16. Popper H, Schaffner F. Nonsuppurative chronic cholangitis and chronic hepatitis. In: Popper H, Schaffner F, eds. Progress in liver disease, Vol 3. New York: Grune & Stratton, 1970; 336–54.
17. Shapiro JM, Smith H, Schaffner F. Serum bilirubin: a prognostic factor in primary biliary cirrhosis. Gut 1979;20:137–40.
18. Christensen E, Crowe J, Doniach D, et al. Clinical pattern and course of disease in primary biliary cirrhosis based on an analysis of 236 patients. Gastroenterology 1980;78:236–46.
19. Roll J, Boyer JL, Barry D, Klatskin G. The prognostic importance of clinical and histologic features in asymptomatic and symptomatic primary biliary cirrhosis. N Engl J Med 1983; 308:1–7.
20. Lee RG, Epstein O, Sherlock S, et al. Granulomas in primary biliary cirrhosis: a prognostic feature. Gastroenterology 1981; 81:983–6.
21. Copas JB. Regression, prediction and shrinkage (with discussion). J R Statist Soc B 1983;45:311–54.
22. Jones EA. Primary biliary cirrhosis and liver transplantation. N Engl J Med 1982;306:41–3.
23. Scharschmidt BF. Human liver transplantation: analysis of data on 540 patients from four centers. Hepatology 1984; 4:95s–101s.
24. Breslow NE. Contribution to the discussion of the paper by DR Cox. J R Statist Soc B 1972;34:216–7.
25. Tsiatis AA. A large sample study of Cox's regression model. Ann Statist 1981;9:93–108.
26. Rao CR. Linear statistical inference. New York: John Wiley & Sons, 1965:386.