Methotrexate (MTX) Plus Ursodeoxycholic Acid (UDCA) in the Treatment of Primary Biliary Cirrhosis

Burton Combes,1 Scott S. Emerson,2 Nancy L. Flye,2 Santiago J. Munoz,3 Velimir A. Luketic,4 Marlyn J. Mayo,1 Timothy M. McCashland,5 Rowen K. Zetterman,5 Marion G. Peters,6 Adrian M. Di Bisceglie,7 Kent G. Benner,8 Kris V. Kowdley,9 Robert L. Carithers, Jr.,9 Leonard Rosoff, Jr.,10 Guadalupe Garcia-Tsao,11 James L. Boyer,11 Thomas D. Boyer,12 Enrique J. Martinez,12 Nathanael M. Bass,13 John R. Lake,13 David S. Barnes,14 Maurizio Bonacini,15 Karen L. Lindsay,15 A. Scott Mills,4 Rodney S. Markin,5 Raphael Rubin,16 A. Brian West,11,17 Donald E. Wheeler,1,18 Melissa J. Contos,4 and Alan F. Hofmann19

This placebo-controlled, randomized, multicenter trial compared the effects of MTX plus UDCA to UDCA alone on the course of primary biliary cirrhosis (PBC). Two hundred and sixty five AMA positive patients without ascites, variceal bleeding, or encephalopathy; a serum bilirubin less than 3 mg/dL; serum albumin 3 g/dL or greater, who had taken UDCA 15 mg/kg daily for at least 6 months, were stratified by Ludwig’s histological staging and then randomized to MTX 15 mg/m² body surface area (maximum dose 20 mg) once a week while continuing on UDCA. The median time from randomization to closure of the study was 7.6 years (range: 4.6-8.8 years). Treatment failure was defined as death without liver transplantation; transplantation; variceal bleeding; development of ascites, encephalopathy, or varices; a doubling of serum bilirubin to 2.5 mg/dL or greater; a fall in serum albumin to 2.5 g/dL or less; histological progression by at least two stages or to cirrhosis. Patients were continued on treatment despite failure of treatment, unless transplantation ensued, drug toxicity necessitated withdrawal, or the patient developed a cancer. There were no significant differences in these parameters nor to the time of development of treatment failures observed for patients taking UDCA plus MTX, or UDCA plus placebo. The trial was conducted with a stopping rule, and was stopped early by the National Institutes of Health at the advice of our Data Safety Monitoring Board for reasons of futility. In conclusion, methotrexate when added to UDCA for a median period of 7.6 years had no effect on the course of PBC treated with UDCA alone. Supplementary material for this article can be found on the HEPATOLOGY website (http://www.interscience.wiley.com/jpages/0270-9139/suppmat/index.html). (HEPATOLOGY 2005;42:1184-1193.)

See Editorial on Page 1009
When the present treatment trial was designed for patients with primary biliary cirrhosis (PBC), ursodeoxycholic acid (UDCA) was emerging as a promising therapeutic agent. The important beneficial effects of UDCA reported in the uncontrolled observations of Poupon, Leuschner and their respective associates\textsuperscript{1,2} were subsequently confirmed in many clinical trials. However, only four of them dealt with relatively large numbers of patients in a double-blind, randomized, placebo-controlled manner.\textsuperscript{3-6} All demonstrated improvement in laboratory tests that serve as markers of cholestasis (\textit{i.e.}, serum bilirubin, alkaline phosphatase, gamma glutamyl transpeptidase), and of inflammation (\textit{i.e.}, aspartate aminotransferase, alanine aminotransferase, immunoglobulin M). Improvements in histological features were reported in three of the trials.\textsuperscript{3,4,6}

The most important endpoints of beneficial medical therapy in PBC are the prevention of liver transplantation and prolongation of survival without transplantation. None of the above studies demonstrated a statistically significant improvement in these endpoints during the placebo-controlled phases of the respective trials. Moreover, liver disease continued to progress in patients on UDCA alone.

During this same period, uncontrolled data from Kaplan and associates\textsuperscript{7-9} suggested that oral pulse methotrexate (MTX) might be effective in the management of precirrhotic PBC. Additional short-term observations of MTX alone\textsuperscript{10} or in combination with UDCA\textsuperscript{11,12} provided conflicting data about the usefulness of MTX. In general, most of the investigators involved in these latter assessments indicated the need for long-term, well-controlled trials to examine the potential efficacy of MTX. The current study was designed to take advantage of the potential beneficial effects of both UDCA and MTX as a two arm, randomized, double-blind, placebo-controlled clinical trial of UDCA plus MTX versus UDCA plus placebo.

**Patients and Methods**

**Study Design.** In the current trial of UDCA plus MTX versus UDCA plus placebo, the primary measure of treatment outcome was liver transplant–free survival as measured by the distribution of time to transplant or death from all causes, whichever came first. Secondary endpoints included comparison of treatment arms with respect to overall survival, time to clinical decompensation (development of ascites, hepatic encephalopathy, variceal bleeding, transplant, or death), development of varices, changes in biochemical tests, liver histology, and symptomatology and sense of well-being. The study design was reviewed and approved by the institutional review boards at each of the clinical centers.

**Patient Selection.** Investigators at 12 geographically diverse clinical centers in the USA screened 535 patients with PBC for possible entry into our trial. Patients’ clinical records were reviewed and various clinical, laboratory, radiology, and pathology tests were performed to assure (1) that patients would satisfy our inclusion criteria for PBC; (2) that they had not already demonstrated exclusion criteria which would keep them from qualifying for the MTX/placebo phase; and (3) if eligible for the trial, that they would proceed to the next step.

The intention was to study the effect of MTX on the progression of PBC in 20- to 69-year-old patients of either sex and any race and with only moderately advanced disease at study entry. Patients were to have had a diagnosis of chronic cholestatic liver disease of at least 6 months duration, a positive antimitochondrial antibody test, levels of serum alkaline phosphatase at least 1.5 times the upper limit of normal, and a liver biopsy within the 6 months prior to randomization (and on UDCA at least 6 months) with histological findings compatible with the diagnosis of PBC. To be judged adequate for staging of disease, the liver biopsy must have been at least 2 cm long if cirrhosis was not detected. Asymptomatic patients must have had a histological stage greater than stage I using the Ludwig’s classification.\textsuperscript{13} Patients could not have markedly advanced PBC, and thus patients having a history of a serum bilirubin of 3.0 mg/dL or greater, a serum albumin less than 3.0 g/dL, or a history of ascites, hepatic encephalopathy, or variceal bleeding were not eligible for randomization. At screening 393 of the 535 patients were determined to meet the defined inclusion criteria (see Supplementary Table 1, available at the HEPATOLOGY website [http://interscience.wiley.com/jpages/0270-9139/ suppmat/index.html]).

As shown in this supplementary table, patients were excluded from the study if they had clinical, serological, or histological evidence of liver disease of other etiology, had a history of alcohol abuse within the 2 years before study enrollment, were treated with immunosuppressive agents, rifampin, or dilantin in the months preceding randomization, had a history of malignant disease, were HIV positive, had other major illnesses that could limit life span, or were pregnant or unwilling to use adequate forms of birth control to avoid pregnancy. Of the 385 patients meeting the screening inclusion and exclusion criteria, 300 patients progressed to a pre-entry evaluation phase during which they were treated with UDCA alone at a dose of approximately 15 mg/kg/d. At the end of this UDCA phase, the patients were again screened for meeting the
inclusion and exclusion criteria given above, as well as for having an acceptable hematological profile, adequate renal and pulmonary function, no radiological or ultrasound evidence of biliary obstruction, and a liver biopsy within the last 6 months consistent with a diagnosis of PBC.

Randomization. Between January 1994 and March 1998, 265 patients who signed informed consent documents were randomized with equal probability in a double-blind fashion to receive UDCA plus MTX (132 patients) or UDCA plus placebo (133 patients). Ten patients who failed to meet only one of the eligibility criteria were reviewed by the study principal investigator (B. Combes), and were judged suitable for randomization despite no liver biopsy within the last 6 months (1 patient on the MTX arm whose biopsy was 9.2 months prior to randomization, and 1 patient on the placebo arm whose biopsy was 9.7 months prior to randomization); a percent predicted lung diffusing capacity (DLCO) of 45% in one patient on the MTX arm; or a creatinine clearance between 50 and 60 mL/min/1.73 m² (5 patients on the MTX arm, and 2 patients on the placebo arm). In addition, in later, postrandomization review of medical records, two patients on the MTX arm were found to have previous bilirubin measurements of 5.3 and 7.9 mg/dL, respectively. These values had never been confirmed. In each instance, every subsequent serum bilirubin value was less than 1.4 mg/dL over a period of seven years in one, and less than 2.0 mg/dL from the time of screening to toxicitiy occurred was then attempted.

Modification of Methotrexate Dose. Because there is no current evidence that UDCA affects blood elements or induces side effects other than diarrhea in a small number of patients, the development of cytopenia, of mucositis, significant nausea or anorexia were initially considered to be related to MTX, and MTX dose was altered in accord with a predefined rating for the common side effects and bone marrow toxicity of MTX (Supplementary Table 2). Toxicity was rated as either mild (acceptable), moderate (requiring alteration of dose), or severe (requiring discontinuation of therapy).

For moderate toxicity, weekly dosage was reduced by a quarter or a third, and the toxicity was monitored weekly until resolved. The dosage of MTX was then increased by 2.5 mg per week until a dose of 2.5 mg less than the original toxic dose was reached, provided toxicity did not recur. A subsequent return to the original dose at which toxicity occurred was then attempted.

For severe toxicity, MTX was stopped completely while the toxic reaction was being managed. Gastrointestinal and hematological findings usually improved fairly rapidly. Once resolved, MTX was to be restarted at half the toxic dose, and then increased 2.5 mg per week at monthly intervals provided toxicity did not recur, until a weekly dose 2.5 mg less than the original toxic dose was reached. If recurrent toxicity was not observed, cautious increase to full dose was attempted.

Other reasons listed in the protocol for decreasing the dose or stopping methotrexate included the appearance of allergic reactions, severe skin rash, pulmonary symptoms or chest x-ray findings suggestive of pulmonary fibrosis, severe exacerbation of liver disease, and worsening of renal function. MTX was to be withdrawn if evidence of alcohol abuse arose or if the patient became pregnant or would no longer practice birth control. Study medication was stopped in patients who developed a cancer.

Dose modification could be carried out without the local investigator breaking the medication code, because in all instances dosage would be temporarily reduced or...
stopped. Nevertheless, when deemed necessary by our external safety monitors, the treatment code could be broken for their use in assisting with the management of our patients.

**Schedule of Patient Visits and Investigations.** Patients were seen and had blood drawn at weeks 2 and 4, then monthly for the first 6 months, bimonthly for the next 6 months, then at 3-month intervals for the duration of the study. Blood was drawn 1 week after the preceding dose of MTX and on the day of, but preceding the next dose of MTX. Symptoms of liver disease and of potential toxicity were assessed at each visit by history and with the aid of a diary. Complete histories, physical examinations, chest x-rays and pulmonary function studies, including measurements of DLCO were to be obtained at least annually, the latter because of concerns about possible MTX-induced lung injury. Patients were to have a liver biopsy and upper endoscopy after 24 months on MTX or its placebo, and subsequently at additional intervals of 2 years.

**Evaluation of Compliance.** Patients were given known quantities of study medication at appropriate intervals and instructed in how to keep a log of medicine intake. The log was checked, and unused medicine counted at appropriate return visits, and before a new supply of medicine was given to the patient. The log and pill counts were kept in the permanent record for each patient.

**Evaluation of Adverse Experiences.** The adverse experiences reported by patients during their study visits were grouped within broad categories defined by organ system. In addition, adverse experiences were categorized across all organ systems infections, bleeding events, neoplastic events, and cancers. Serious adverse experiences occurring at any clinical center were reported promptly to a central committee monitoring such events.

**Evaluation of Treatment Response.** The primary and several secondary measures of treatment outcome were based on the distribution of time to treatment failure as defined by a hierarchy of clinical and subclinical outcomes. Times to death, transplant, activation for transplant, and clinical deterioration as defined by development of variceal bleeding, hepatic encephalopathy, ascites, or disabling pruritus were obtained from the routine follow-up of patients. Subclinical deterioration was defined as a doubling of serum bilirubin from baseline to at least 2.5 mg/dL, a decrease in serum albumin to a level less than 2.5 g/dL, or an increase in PT/INR to 1.3 or greater. In order to be judged a subclinical deterioration, the corresponding threshold must have been exceeded on two consecutive clinic visits.

Liver histology was evaluated as the average stage and fibrosis scores on liver biopsies obtained every 2 years and scored independently by a panel of five pathologists in a central core. Development of varices was evaluated by endoscopies performed every 2 years. Because patients with varices at screening were eligible for randomization, only a subset of the trial participants were evaluable for the endpoint of development of new varices. Patients who terminated study treatment prior to liver transplant or death were encouraged to continue all regular clinic visits, and patients who agreed were considered evaluable for all measures of treatment response. Some patients declined to have further biopsies, endoscopies, and/or serum chemistries measured, but were willing to be followed for clinical events, and these patients are considered fully evaluable for all endpoints that could be observed without invasive procedures. Patients who withdrew consent for all further follow-up contribute information only up to the time of their withdrawing their consent to be studied, although as described in the statistical methods, exploratory analyses imputed missing measurements for these patients.

**Monitoring of the Clinical Trial.** See HEPATOLOGY website (http://interscience.wiley.com/jpages/0270-9139/suppmat/index.html).


**Results**

**Baseline Characteristics.** Baseline characteristics are summarized in Table 1. The patients in each treatment arm were well matched. Over 90% of the randomized patients were women. Serum bilirubin was 1.0 mg/dL or less in 85%, 1.1 to 2.0 mg/dL in 13%, and 2.1 to 2.9 mg/dL in only 4 patients (approximately 1.5%).

**Subject Disposition.** The interventional aspect of the study was stopped in November 2002, when the time from randomization was a median of 7.6 years (range 4.6 to 8.8 years). Prior to that time, liver transplants were received by 7 patients on the MTX arm and 7 patients on the placebo arm. Deaths were observed in 7 patients on the MTX arm and 11 patients on the placebo arm. Two of the deaths observed on the placebo arm occurred posttransplantation. Hence a total of 14 and 16 liver transplants or deaths were observed on the MTX and placebo arms, respectively.
Forty-one patients on the MTX arm and 47 patients on the placebo arm discontinued taking study drug prior to the end of the interventional phase and prior to experiencing the primary endpoint of liver transplant or death.

By the seventh year postrandomization, approximately one third of patients in both arms discontinued study treatment with no statistically significant differences between the treatment arms. Table 2 presents the numbers of patients discontinuing treatment early for each of several categories of reasons for early termination.

The overwhelming majority of patients who discontinued their study drug were still followed for occurrence of the study endpoints. Only 11 patients prematurely withdrew consent for follow-up of transplant-free survival status: 3 in the MTX arm and 8 in the placebo arm. The cumulative proportion withdrawing from the study in this manner was 1%, 3%, and 4.5% at 1, 2, and 6 years.
post randomization, respectively, with no statistically significant difference between treatment arms (P = .14 from proportional hazards model). Such continued follow-up of patients was crucial, because patients who discontinued study treatment early were most often symptomatic, and this was associated with an increased risk of liver transplantation and/or death in the time period immediately following that discontinuation. In an analysis of transplant-free survival in which discontinuation of study drug is modeled as a time-varying covariate, patients who discontinued their study drug had a 42-fold higher risk of the primary endpoint (95% CI, 14-fold to 124-fold higher; P < .001). The actuarial estimate of the cumulative probability of liver transplantation or death during the first year postdiscontinuation was 19.2% in the MTX arm and 18.6% in the placebo arm; the estimated two year failure rates were 25.7% and 28.3%, respectively.

Following the recommendation by the Data Safety Monitoring Board (DSMB) for early termination of the study on October 31, 2002, the clinical centers advised the 173 patients (89 in the MTX arm, 84 in the placebo arm) still taking their study drug to discontinue the experimental therapy. The median time between the decision to terminate the study and patients’ discontinuation of study drug was 13 days, with only 3 patients continuing to take study medication for more than 6 weeks after the end of the interventional phase of the trial.

During the postintervention phase of the trial (i.e., following the October 31, 2002 DSMB recommendation), 110 patients in the MTX arm and 98 patients in the placebo arm consented to continue to be followed in regular clinic visits for an average of 381 days (median 398 days, range 110 to 460 days). An additional 12 MTX patients and 19 placebo patients consented to be followed for liver transplantation and vital status during this period.

**Treatment Failures.** The importance of stratification by histological stage is summarized in Table 3, which presents the number of patients with each type of treatment failure observed during the interventional phase of the trial, as well as the numbers when the additional follow-up period is included. Also presented is the number of treatment failures for the hierarchical study endpoints. The number of patients with treatment failures was greater in those who had stage III-IV histology (stratum 2) than in those with stage I-II histology (stratum 1) at entry. In both instances, treatment failures (i.e., number of patients; number of failures per patient) were comparable in patients receiving either methotrexate or its placebo when assessed separately for stratum 1 and stratum 2 patients.

**Comparison of Treatment Groups by Time to Event.** Figure 1 displays the Kaplan-Meier estimates for the probability of transplant-free survival by randomization strata and treatment arm during the interventional phase of the trial. Evident is a trend toward worse transplant-free survival in the patients having advanced histological stage at randomization, but the treatment arms within each stratum show very comparable results. Similar patterns were observed for the other hierarchical defi-
nitions of treatment failure, as well as when the experience in the additional follow-up period is included for each of the analyses.

Table 4 provides the results of analyses of event-free survival (EFS) for the three hierarchically defined treatment failures when follow-up is limited to the interventional phase of the clinical trial, as well as when additional follow-up is included. Seven year EFS probabilities are similar across treatment groups in each stratum. The hazard ratio estimates are close to 1.0, and the 95% CIs exclude the alternatives which the study was adequately powered to detect.

**Laboratory Measurements.** Patients had blood chemistry measurements at intervals of 3 months (serum bilirubin and albumin) or 6 months (prothrombin time [PT]). Patients who died, who had received liver transplants, or who withdrew consent for regular clinic visits were missing data for these measurements. Hence, comparisons of treatment groups over time with respect to these measurements are potentially subject to bias, if the treatment groups differed in the loss to follow-up. In an attempt to at least partially account for some of this bias, we compared serum bilirubin, serum albumin, and PT/international normalized ratio (INR) measurements both when analysis was restricted to the available data, when the last measurement was carried forward, and when missing data was imputed using linear trends for observed disease progression. In no case were there significant differences between the treatment arms for either bilirubin, albumin, or PT/INR measurements.

**Endoscopic Prevalence of Varices.** Supplementary Table 3 presents the prevalence of varices at the protocol specified examinations according to treatment arm, randomization stratum, and presence of varices at randomization. Also provided is the \( P \) value from a Mantel-Haenszel statistic testing for differences in prevalence of varices across treatment groups, adjusted for randomization stratum and baseline presence of varices. While there are clear trends toward higher prevalence of varices among those patients who had varices at baseline and among patients having higher histological stage of liver disease (stratum 2), the two treatment arms had comparable prevalence of varices, with no statistically significant differences found at any time of follow-up.

**Histological Stage of Liver Disease.** There were no statistically significant differences between the treatment arms with respect to either stage and fibrosis score averaged from independent readings by four or five independent pathologists. More detailed analysis of the biopsy readings will be described in a separate manuscript.

**Incidence of Adverse Experiences by Treatment Group.** Supplementary Table 4 presents the number of
Table 4. Estimates of Treatment Effect for Event-Free Survival Using Hierarchical Definitions of Treatment Failure, 7 Year Event-Free Survival Probabilities (95% CI)

<table>
<thead>
<tr>
<th>Stratum 1</th>
<th>Stratum 2</th>
<th>Hazard Ratio (95% CI; P Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention Phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death or Transplant</td>
<td>0.93 (0.85-1.00)</td>
<td>0.94 (0.88-1.00)</td>
</tr>
<tr>
<td>Death, Transplant, or Clinical Deterioration</td>
<td>0.87 (0.79-0.96)</td>
<td>0.92 (0.85-1.00)</td>
</tr>
<tr>
<td>Death, Transplant, Clinical Deterioration, or Subclinical Progression</td>
<td>0.85 (0.76-0.95)</td>
<td>0.89 (0.80-0.98)</td>
</tr>
<tr>
<td><strong>Additional Follow-up</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death or Transplant</td>
<td>0.93 (0.87-1.00)</td>
<td>0.95 (0.89-1.00)</td>
</tr>
<tr>
<td>Death, Transplant, or Clinical Deterioration</td>
<td>0.88 (0.79-0.96)</td>
<td>0.93 (0.86-1.00)</td>
</tr>
<tr>
<td>Death, Transplant, Clinical Deterioration, or Subclinical Progression</td>
<td>0.86 (0.77-0.95)</td>
<td>0.90 (0.82-0.98)</td>
</tr>
</tbody>
</table>

subjects experiencing adverse experiences of any severity level (mild, moderate, severe) during the intervention phase of the study. All adverse events are included whether or not the patient was continuing to take study drug (i.e., an intent-to-treat analysis).

There were no striking differences in the incidence of adverse experiences across treatment groups, with the possible exception of a trend toward higher bone marrow suppression in the MTX group (blood/lymphatic disorders were overwhelmingly leukopenia, thrombocytopenia and anemia). The statistical significance of such differences was only demonstrated when adverse experiences of all levels of severity were included.

We did not observe any instances of clinical pneumonitis, acute respiratory distress syndrome or respiratory failure that we could attribute to MTX, in contrast to the experiences reported by others in PBC patients receiving MTX.12,19

**Pulmonary Measurements.** None of the analyses of annually measured forced vital capacity (FVC) and DLCO showed statistically significant differences across the treatment groups. More detailed analyses of these pulmonary measurements will be presented in a future manuscript.

**Postintervention Phase Follow-up of Laboratory Measurements.** Patients were followed for up to a year after discontinuation of the study drug in order to assess the possibility of “rebound” effects upon termination of MTX therapy. Data obtained after halting the intervention phase of the study were combined with data on patients who had taken study drug at least four years and then prematurely discontinued study drug for any reason. Supplementary Table 5 presents statistics on the change from the last measurement while on study drug to measurements 3, 6, 9, and 12 months off study drug for these patients. In no case was the difference in average laboratory measurements statistically significant.

**Biliary Bile Acid Composition.** A total of forty specimens of bile selected randomly from patients in a single treatment center on UDCA and MTX or UDCA plus placebo, in strata 1 and 2, at years 2, 4 and 6 after randomization were analyzed for bile acid composition in the laboratory of Alan F. Hofmann by a high pressure liquid chromatographic method.20 Bile acid composition showed enrichment in UDCA of a magnitude similar to that observed in other UDCA PBC studies,5,6,12 and (Table 5) was comparable in patients receiving either MTX or its placebo, and in stratum 1 and stratum 2. The inference is that MTX did not affect the circulating bile acid pool, and that a lack of a MTX effect was not due to interference with UDCA metabolism.

**Discussion**

In this randomized, double-blind clinical trial of MTX therapy added to UDCA, we did not observe evidence of a clinically important benefit of MTX therapy on trans-

Table 5. Bile Acid Composition in Bile of Patients Receiving UDCA and Either Mtx or Its Placebo

<table>
<thead>
<tr>
<th>UDCA and Either Mtx or Its Placebo</th>
<th>2 years</th>
<th>4 years</th>
<th>6 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDCA Cholic Acid &amp; Chenodeoxycholic Acid &amp; Deoxycholic Acid</td>
<td>n = 6</td>
<td>n = 8</td>
<td>n = 8</td>
</tr>
<tr>
<td><strong>2 years MTX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 6</td>
<td>35.8 ± 16.1</td>
<td>27.4 ± 18.9</td>
<td>21.7 ± 18.0</td>
</tr>
<tr>
<td>Placebo</td>
<td>40.8 ± 45.9</td>
<td>24.8 ± 69.0</td>
<td>24.0 ± 24.0</td>
</tr>
<tr>
<td><strong>4 years MTX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 8</td>
<td>35.4 ± 47.75</td>
<td>28.6 ± 50.7</td>
<td>23.8 ± 24.5</td>
</tr>
<tr>
<td>Placebo</td>
<td>40.3 ± 55.52</td>
<td>29.9 ± 76.4</td>
<td>26.1 ± 53.9</td>
</tr>
<tr>
<td><strong>6 years MTX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 5</td>
<td>38.6 ± 17.6</td>
<td>28.3 ± 18.6</td>
<td>24.6 ± 8.7</td>
</tr>
<tr>
<td>Placebo</td>
<td>41.5 ± 30.0</td>
<td>26.9 ± 39.9</td>
<td>22.3 ± 18.7</td>
</tr>
</tbody>
</table>
plant-free survival, time to clinical deterioration, or time to subclinical progression. Furthermore, there was no discernible benefit of MTX on prevalence of varices or histological liver stage. On the other hand, there was also no evidence of substantial toxic effects of MTX therapy, beyond a slight increase in low-grade bone marrow suppression.

This study was hampered by an overall low event rate. At the time of study design, it was anticipated that the 5-year transplant-free survival would be 83%, whereas the experience in this trial corresponded instead to a 94.5% 5-year, and 84.3% 8-year transplant survival probability. The lower event rate meant that patients had to be followed for a longer time period, and this in turn was likely responsible for a higher rate of patients discontinuing study drug: Kaplan-Meier estimates suggest a 22% probability of discontinuing drug within 5 years, and a 38.5% probability in 8 years. Early termination of study drug was strongly associated with liver transplantation or death within a year, and indeed many of those discontinuing treatment did so for reasons attributable to the progression of disease. It is noteworthy that we did not detect differences between the treatment arms with respect to early discontinuation of study drug.

This study was conducted with a group sequential stopping rule, and guided by that stopping rule, the DSMB recommended early termination of the study when only 30 of the planned 50 events had been observed. The statistical inference presented on the time to event endpoints adjusts for this early termination of the study, and based on that inference, we are able with high confidence to rule out the magnitude of beneficial treatment effect the study could have detected even with the full planned sample size. Had the study continued to observe the full 50 events, it would have required an estimated hazard ratio of 0.567 or less to reject the null hypothesis of no treatment effect. A nonsignificant result at that analysis would have failed to rule out a benefit of MTX therapy corresponding to a hazard ratio of 0.33. After observing 30 events, the estimated hazard ratio was 0.86, with a 95% confidence interval ruling out any benefit greater than that corresponding to a hazard ratio of 0.44 or less. Thus, the results of this trial are sufficiently precise to decide against the alternative of a markedly beneficial MTX effect. Based on the results of the trial after observing 30 events, assuming that the patients would continue to receive any benefit of MTX therapy, we estimate only a 5% chance that the final results at 50 events might have been statistically significant. Given the relatively high rate of patients discontinuing study drug, however, if continued MTX therapy is required to obtain any benefit from the drug, the chances of observing a statistically significant effect at the final analysis might be markedly less than 5%.

In this report of the trial results, we have presented some data regarding blood chemistry measurements and histological liver stage over time. The interpretation of such statistics is always complicated by the possibility that patients with certain values are at higher risk of death and thus are removed from the analyses at later time points. Hence, trends in population statistics may more represent the change in the mix of patients considered, rather than time trends in individual patients. This issue becomes more complicated when patients are also removed from the data set by liver transplantation or withdrawal from follow-up, either of which might be influenced by the true value of the unobserved measurements. While we did note some increased tendency for placebo patients to withdraw from follow-up 5 or more years after randomization, for the most part the availability of data was similar for the two treatment arms.

The results of a number of other therapeutic trials combining MTX and UDCA have been published. In general, the number of patients assessed was too few, and the duration of therapy too short to provide adequate insight into the long-term effects of such treatment.

In this study, we estimated eight year survival probability as 93%, and eight year transplant-free survival probability as 84%. Such low event rates have large bearing on our ability to explore the effects of treatment in a population of patients with only moderately advanced PBC. We can consider the number of patients needed in order to obtain a statistically significant result with high probability in a clinical trial using either overall survival or transplant-free survival as a primary endpoint. When using the log rank statistic to compare the distributions of time to an event in a randomized clinical trial, we need to observe 88 events in order to have 90% power to detect a twofold improvement in the risk of an event. In a study that followed subjects for an average of 8 years (with an expected 7% rate of death and 16% rate of transplant or death), this would then suggest that we would need to randomize a total of 1,257 subjects when using the endpoint of overall survival, or 550 subjects when using the endpoint of transplant-free survival. Longer studies could obviously randomize fewer subjects, but they would have greater problems with ensuring subject compliance with an unproven therapy. Alternatively, a more-sick population with a higher rate of events could be used, although the ability to generalize any observed results is more problematic: A failure to demonstrate an effect may occur because patients needed to be treated earlier in the course of their disease.
Acknowledgment: We are indebted to Renate Davis (UT Southwestern) for invaluable administrative support and preparation of this manuscript; to clinical coordinators Judy Otey and Marilyn (Ann) Heard-Sakhaee (UT Southwestern); Carol Campbell-Sargeant (MCV of Virginia Commonwealth Medical School); Melissa Loyet (Washington University, Saint Louis); Sharon Westerberg and Donna McGill (Albert Einstein Medical Center; Jefferson Medical College); Nancy Shura (University of Southern California); Ellen Katz (Emory); Paula Cox, Tamsin Sarich and Dorothy Rush (University of Washington); Suni Wilson (Oregon Health Sciences University); Michelle Kiwi and Marilyn McCormick (Cleveland Clinic); Lily Luu and Cherie Bremmer-Kamp (UC San Francisco); June Kupson (Yale); Tamara Bernard and Jenny McCarthy (Nebraska) for superior interactions with patients. Statistical Core - Scott S. Emerson, Robert L. Carithers, Jr., Nancy L. Flye Pathology Core - A. Scott Mills, Rodney S. Markin, Melissa J. Contos, Raphael Rubin, A. Brian West, Donald E. Wheeler Data Safety and Monitoring Board - Henry C. Bodenheimer, Nora V. Bergasa, Rolland C. Dickson, Gregory T. Everson, Keith D. Lindor, Elizabeth C. Wright.

References


