Summary

We sought to examine the effect of methotrexate (MTX) when added to ursodeoxycholic acid (UDCA) on overall survival and progression-free survival in patients with primary biliary cirrhosis (PBC). We performed a two-arm, randomized, double-blind, placebo controlled, clinical trial of MTX plus UDCA versus placebo plus UDCA in 265 patients with moderately advanced PBC. Primary endpoints were progression-free and overall survival, with progression-free survival defined as the absence of splenomegaly, esophageal varices, or hepatic encephalopathy. Kaplan-Meier estimates of percent of patients surviving without progression did not differ between treatment groups at 3, 6, or 9 years (9-year progression-free survival 78.8% in the placebo group, 71.2% in the MTX group, 95% confidence interval of the difference in percentile surviving without progression -49.6% to 64.8%, p=0.80). Likewise, estimates of overall percent of patients surviving did not differ between treatment groups at 3, 6, or 9 years (9 year survival 93.4% in the placebo group, 90.8% in the MTX group, 95% confidence interval of the difference in percentile surviving -43.0% to 48.3%, p=0.91). We conclude that MTX adds no additional progression-free or overall survival benefit to monotherapy with UDCA in patients with moderately advanced PBC.

Comment: multicenter

Comment: How many events? Statistical information is roughly proportional to the number of events in the presence of censoring.

Comment: This CI is so wide that it is to the point of absurdity. Clearly, focusing on the 9 year survival is not very precise here. Using survival at fixed point in time is OK, but PH regression based HR is more standard (that does not necessarily mean better, just standard)

Comment: This is rather a strong statement given the width of your CI. Certainly it is clinically important to have an 40% better survival at 9 years.

Background

Primary biliary cirrhosis is a slowly progressive liver-specific autoimmune disease typically occurring in middle aged adults, characterized by a female to male ratio of 10 to 1[1]. Autoimmune destruction of intrahepatic bile ducts leads to decreased bile secretion and hepatic retention of toxic substances. Hepatic fibrosis and cirrhosis ensues, resulting in ascites, hepatic encephalopathy, jaundice, coagulopathies, bleeding esophageal varices, temporal and proximal muscle wasting, and in some cases death[1]. Hepatomegaly is observed in 70% of cases. Fatigue and pruritis are the most common presenting symptoms of disease[2], although hyperlipidemia, hypothyroidism, osteopenia, and other autoimmune diseases such as Sjögren's syndrome and scleroderma can also be observed. Antimitochondrial antibodies are observed in 90-95% of cases, and their presence, along with persistent elevations in liver enzymes (> 6 months) and liver biopsy pathology demonstrating asymmetric destruction of bile ducts, together provide the diagnosis.

Comment: pruritus

Treatment typically involves UDCA, which reduces the likelihood of liver transplantation or death at 4 years[3, 4]. Although effective with few side effects, UDCA does not prevent disease progression in all patients and is not effective in advanced disease. Methotrexate is an immunomodulatory agent with potential treatment synergy with UDCA in patients with PBC[5, 6]. In order to investigate the benefits of dual therapy, we performed a randomized, double-blind, placebo controlled trial of MTX plus UDCA in patients with PBC. Our goals were to: 1) determine if dual therapy could prolong time to disease progression as defined by the presence of ascites, hepatic encephalopathy, or esophageal varices, and 2) determine if dual therapy could prolong time to death.

Questions of Interest

1) Does treatment with MTX plus UDCA in patients with PBC increase the proportion of patients *surviving without disease progression* at 3, 6, and 9 years, compared to treatment with placebo plus UDCA?
2) Does treatment with MTX plus UDCA in patients with PBC increase the proportion of patients *surviving* at 3, 6, and 9 years, compared to treatment with placebo plus UDCA?

Comment: You will have a multiple comparison problem, which will decrease your power to detect an effect. Again, HR is more standard, but using, say, the 6 year survival prob would be OK (that would have had a bit more precision).

Source of the Data

The data comes from a multi-site, double-blind, randomized, placebo controlled trial of MTX plus UDCA as a treatment for PBC. Study participants were men and women between the ages of 20 and 69 with moderately severe disease at the time of enrollment. Participants had to meet a number of inclusion and exclusion requirements, including no history of alcohol abuse in the previous 2 years, no liver disease of unknown or other etiology, and current treatment with

UDCA (among others). Patients were stratified by the severity of their liver disease (early and advanced staged disease), then randomized to UDCA and MTX (maximum dose 20 mg/week or 15 mg/1.73 m^2 body surface area) or UDCA and placebo. Patients continued on their regimen until closure of the study unless death ensued, drug toxicity necessitated withdrawal, the patient developed cancer, or voluntary withdrawal ensued. Patients were followed regularly with blood draws, physical examinations, and measures of pulmonary and hepatic function. Treatment outcomes were measured in terms of overall survival and progression-free survival as defined above.

Data collected from patients included demographic variables (age, sex, weight, and height); a marker of duration of disease (time from diagnosis of PBC to randomization); two markers of hepatocellular damage (serum alkaline phophatase level and alanine aminotransferase level); an indicator of disease stage (early—stage 1 or 2—or advanced—stage 3 or 4—disease); indicators of hepatic inflammation/cirrhosis (splenomegaly, and platelet counts); liver function indicators (serum bilirubin, cholesterol, and albumin levels, and clotting times); treatment group indicator (MTX or placebo) and a treatment adherence indicator variable. Time to disease progression or last follow-up and time to death or last follow-up were also included along with the indicator variables of progression or death.

Comment: This is rather technical for a manuscript. Scientifically what matters is that you had potentially censored measures of time to progression/death. The reader does not really care how they were represented in your data set.

Data were collected on a total of 265 patients over the course of nearly 10 years. The dataset includes a single row corresponding to each patient, including some or all of the variables above. Missing data include: one missing weight, one missing height, 2 missing alanine aminotransferase values, 4 missing prothrombin times, and 5 missing cholesterol values. For the purposes of analyzing survival time, we can consider this dataset complete, as we have survival time estimates for all patients.

Comment: Good to note

Though randomization should dispense with any confounding, we examined the data for any unlike distribution of variables across treatment groups, which could produce confounding in our analysis. These variables include those associated with liver inflammation/cirrhosis, hepatocellular damage, and liver-function indicators. Demographic variables, disease stage at randomization, and splenomegaly are potential effect modifiers. Our analysis will examine these variables across treatment groups and their potential effect on the response.

Comment: Careful with this jargon, here. This would generally mean that you were doing something like assessing the PH assumption. Better terminology would be "to assess comparability of treatment arms"

Statistical Methods

All statistical analyses were conducted using the Stata software package. To assess the appropriateness of our model, we obtained summary measures such as mean, standard deviation, median, quartiles, maximum, and minimum for all variables by treatment groups. We used these summary measures to compare differences between treatment groups. Also, we created scatterplots (not included) and survival curves (Figure 1) of variables stratified by treatment groups in order to assess for confounders and effect modifiers. We examined disease stage as a possible modifier of treatment effect. We found that stratifying by disease stage (early= stage 1 and 2 versus advanced = stage 3 and 4) did not alter the effect of treatment on progression-free or overall survival (Figure 1). This is not unexpected condsidering that patients were stratified by severity of liver disease before randomization. Therefore, we present results with these stage categories combined.

Comment: Tell us about the statistical methods used for the results presented here, and ignore things that are quite standard in the process of analyzing your data. List the main analysis methods first, then describe any exploration you did (Your main question did not mention any aspect of confounding or effect modification)

Comment: This is a Result

We then used the Kaplan-Meier method to estimate overall survival or progression-free survival probabilities. These probabilities estimate a time point beyond which a subject remains free of progression or death. This method also provides 95% confidence intervals for these probabilities. These confidence intervals suggest that if true survival or progression-free survival probabilities were within the corresponding confidence intervals, data obtained would not be unusual.

Comment: AVOID data driven analyses. This is acceptable as an exploratory analysis, but you should have some pre-specified analysis.

Comment: Actually, the probability is the estimate. The time point is chosen by you.

We chose 3 time points (3, 6, and 9 years) to compare estimated survival or progression-free survival probabilities by treatment groups. We chose these time intervals to represent clinically relevant information on short, intermediate, and long-term (as long as the study would reasonably allow) outcomes. We used p-values to evaluate the null hypothesis of

Comment: What was the primary endpoint?

Comment: You would need to tell us how you made inference about the difference in survival probabilities (you probably used Greenwood's formula and a normal approximation, but there are other approaches possible)

equal overall survival or progression-free survival by treatment group. Survival curves estimated survival or progression-free survival probabilities at each time point (Figure 1).

Comment: The figure is a "Result"

Our Kaplan-Meier estimates took censored data into account. That is, in our data some subjects were observed to have died or progressed while other subjects dropped from the study without this data available. Because they may have lived or remained progression-free longer than subjects with observations, we cannot ignore them or assume death or progression at the point they were lost to follow up. In order to deal with this issue, the Kaplan-Meier method assumes that subjects who dropped from the study remain and are equally likely to die or have progression at any particular time point similar to subjects still in the study.

Results

We first performed univariate and bivariate descriptive statistics to assess for errors and validity of assumptions, to examine the experiment's materials and methods, and to get preliminary estimates of scientifically relevant hypotheses about the data. We examined the demographic and baseline liver function/disease stage variables across treatment groups to assess for confounding (Table 1). We detect no concerning differences between the MTX group and the placebo group with respect to these variables at randomization. This is not unexpected considering the randomized nature of this clinical trial. This verifies equal distribution of patients between the 2 groups with respect to these potential confounders/effect modifiers. We therefore did not stratify the dataset by any of these variables when performing descriptive and inferential analysis of the outcomes. The only noticeable between-group difference is that of weight, which we do not believe would be likely to have an effect on survival or disease progression.

Comment: Don't tell us about this, just do it.

Comment: Common phrasing: "imbalance across treatment groups"

Comment: Soften your wording here: It takes an infinite sample size to prove equality.

Study subjects in both groups were overwhelmingly female (reflecting the demographics of PBC patients in the general population). Mean age was around 50 years. Measurements of age, weight, height, and duration of disease were approximately normally distributed in both groups. The distributions of alkaline phosphatase and alanine aminotransferase were observed to be somewhat right-skewed in both treatment groups (the standard deviations are greater than one half the mean). Indeed, these distributions are characterized by several large right outliers. The average disease stage was slightly larger in the MTX group. Time to first censoring when considering time to disease progression was 39 days. Time to first censoring when considering time to death was 49 days.

Comment: Good to make these comments first, then talk about imbalance across groups.

Comment: Very good to talk about length of follow-up. The minima you discuss here are of interest. So would be the median, quartiles of follow-up time as estimated by a KM estimate of the censoring distribution.

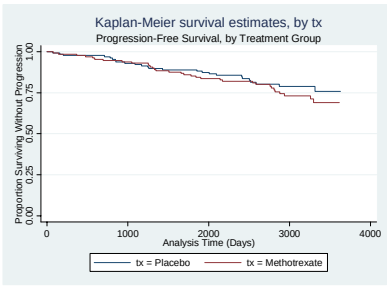
Table 1. Descriptive statistics for patient characteristics by treatment group

	Methotrexate Group Mean (SD) (n=132)	25^{th}, Median, 75^{th} Percentiles	Placebo Group Mean (SD) (n=133)	25^{th}, Median, 75^{th} Percentiles
Age (years)	50.4 (8.7)	45, 50, 56	52.2 (8.6)	47, 53, 59
Sex (% female)	92.5	-	92.5	-
Weight (kg)	70.1 (14.4)	54.9, 68.3, 77.1	73.4 (16.2)	62.7, 70.3, 84.4
Height (cm)	163.7 (7.9)	157.5, 162.2, 170.3	163.7 (8.1)	157.4, 162.7, 168.2
Duration of Disease (days)	42.1 (41.5)	341, 983, 1647	44.1 (39.3)	376, 867, 1979
Alkaline Phosphatase (U/l)	242.9 (145.7)	140, 208.5, 304.5	244.9 (187.4)	120, 200, 277
Alanine Aminotransferase (U/l)	54.3 (41.4)	27, 42, 67	50.2 (41.6)	26.5, 38, 61
Bilirubin (mg/dl)	0.66 (0.44)	0.4, 0.6, 0.8	0.72 (0.39)	0.5, 0.7, 0.9
Albumin (g/dl)	4.0 (0.35)	3.8, 3.95, 4.2	4.0 (0.34)	3.8, 4, 4.2
Cholesterol (mg/dl)	239.2 (58.2)	204, 233, 266	235.8 (58.8)	199, 224.5, 267
Platelets (1000 cells/cu mm)	243.5 (88.6)	179, 231.5, 299	234.7 (83.2)	166, 235, 291
Prothrombin Time (seconds)	11.2 (1.1)	10.6, 11.35, 12	11.4 (1.1)	10.9, 11.5, 12.1
Splenomegaly (%)	8.3	-	10.5	-
PBC Histologic Stage (1-4)	2.57	2, 3, 3	2.50	2, 3, 3

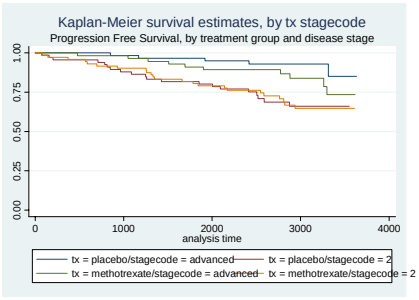
Histologic Stage 1-4 increases as disease progression worsens. Duration of Disease represents the time from diagnosis with PBC to randomization. SD= standard deviation. PBC= primary biliary cirrhosis.

Figure 1. Kaplan-Meier Curves of overall and progression-free survival. Panel (a): progression-free survival, stratified by treatment group. Panel (b): progression-free survival, stratified by treatment group and disease stage. Panel (c): overall survival, stratified by treatment group. Panel (d): overall survival, stratified by treatment group and disease stage.

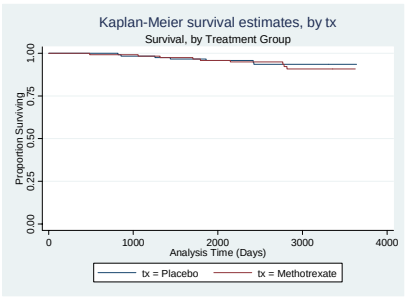
a)



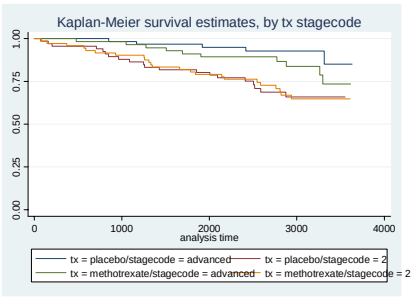
b)



c)



d)



We next performed a Kaplan-Meier analysis of the progression-free and overall survival times, stratified by treatment arm and disease stage (Figure 1). Because disease stage was not found to substantially alter the treatment effect, we do not include inference stratified by this variable. Survival percentiles (time at which 90%, 85%, and 80% of subjects survived progression-free, and times at which 99%, 95% and 90% survived) were similar for the 2 treatment groups (slightly larger in the placebo group, in general), and without evidence of large qualitative differences (Table 2). We chose to analyze these percentiles (rather than, say, 75th 50th, and 25th) because they were most relevant and representative of the data—reflecting scientifically relevant information.

Comment: This should not be a travelog. Just give the results in a logical order. This may or may not be the exact chronological order you did the analyses (though it usually is)

Comment: avoid data driven analysis methods

Comment: Data driven descriptive statistics like this are appropriate in the sense that you could not estimate the median and 25th percentile.

Table 2. Kaplan-Meier estimates of percentile with progression-free and overall survival times by treatment group.

	Time	
	Methotrexate Group	Placebo Group
Progression-Free Survival		
90th percentile	3.50 years	3.42 years
85th percentile	4.88 years	6.53 years
80th percentile	7.44 years	7.85 years
Overall Survival		
99th percentile	1.33 years	2.24 years
95th percentile	5.73 years	6.55 years
90th percentile	Not observed (>10 years)	Not observed (>10 years)

We then performed inferential analysis on the between-group differences in the Kaplan-Meier estimates of 3, 6, and 9 year survival and progression-free survival (Table 3). Methotrexate has no statistically significant effect on these survival measures. The differences between treatment groups in proportion of patients surviving overall or progression free were small (less than 8%) at all time periods. The 95% confidence intervals of the difference in overall and progression-free survival probability include zero, and have p-values much larger than 0.05. Therefore we cannot, with 95% confidence, reject the null hypothesis of no difference in either progression-free or overall survival at 3, 6, or 9 years as a function of treatment with MTX versus placebo. The confidence bounds, using the 3 year progression-free survival interval as an example, indicate that the results we observed would not have been atypical if the true difference in proportion of patients surviving was between 26.8% higher in the methotrexate group and 29.6% higher in the placebo group.

Comment: but look at the size of your CI!!!

Comment: pretty neat trick for survival probabilities in the neighborhood of 92% - 98% (This is a setting where I would likely use the log(- log S(t)) estimates to compare survival were I doing this in real life. This would just avoid the nonsensical CI when we have so little precision.

Table 3. Inferential statistics presenting point estimates of progression-free and overall survival by treatment group and difference in survival between groups.

Analysis Time (years)	Point Estimate of Percentage Surviving			95% Confidence Interval of the Difference	p-value
	Placebo Group	Methotrexate Group	Difference*		
3	98.36%	98.38%	-0.02%	-29.6% – 26.8%	0.99
6	95.7%	94.8%	0.9%	-38.0% – 39.7%	0.96
9	93.4%	90.8%	2.6%	-43.0% – 48.3%	0.91

Analysis Time (years)	Point Estimate of Percentage Surviving Progression-Free			95% Confidence Interval of the Difference	p-value
	Placebo Group	Methotrexate Group	Difference*		
3	92.2%	93.0%	-0.8%	-43.0% – 41.3%	0.97
6	85.7%	82.0%	3.7%	-46.4% – 53.7%	0.89
9	78.8%	71.2%	7.6%	-49.6% – 64.8%	0.80

* Difference = Kaplan-Meier estimate of proportion surviving in placebo group minus proportion surviving in methotrexate group.

Discussion

In patients with moderately advanced PBC, patients treated with MTX exhibited no trend toward improved progression-free or overall survival over patients treated with placebo when added to UDCA. In fact, MTX demonstrated a slight deleterious effect in the long-term, as evidenced by the steeper slope and lower final end-point of the Kaplan-Meier curve in this group. This study was conclusive insofar as survival time estimates were available on all subjects and missing data was scarce and therefore uninformative.

Comment: Excellent to note

Lack of observed benefit to MTX has many possible explanations. Methotrexate is known to have many deleterious side effects, which could magnify the pathophysiological consequences of PBC, resulting in decreased positive clinical effects of the medication. Methotrexate is immunomodulatory, and it could be that the pathophysiology of moderately severe PBC is beyond mediation by the immune system, obviating any positive effect of this medication. Lastly, it may be that MTX is simply ineffective at treating PBC, regardless of level of disease progression.

Confounding was not a major issue with our analysis as randomization ensured equal distribution of patients between the MTX and placebo treatment groups. We considered the possibility that stage of disease might modify the treatment response of MTX. However, we did not find evidence that stage substantially changed the effect of treatment (i.e. there was no effect shown on proportion surviving, irrespective of disease state at baseline). This is not unexpected, considering that patients were stratified by severity of liver disease prior to randomization. A limitation of this study, however, is the small number of patients exhibiting

Comment: Stated a little strongly for a manuscript (In class, I always make a point of saying “randomization precludes confounding in some sense”, but that “sense” is over replicated experiments.

abnormal physiologic measures related to liver function and disease at baseline (might methotrexate have an effect in patients with already poorly functioning livers?).

We performed an intent-to-treat analysis, keeping each patient in her respective treatment group regardless of side effects requiring cessation of treatment. We did not address whether MTX improves clinical endpoints when administered as monotherapy. Likewise, the effect of this treatment on subjective outcomes was not assessed. These issues warrant further study.

In conclusion, MTX appears to be ineffective at improving progression-free or overall survival in patients with moderately severe PBC when used in tandem with UDCA. For the time being, it appears that UDCA monotherapy or UDCA plus colchicine are the treatments of choice for primary biliary cirrhosis[7].

Comment: WAY too strongly stated for a study with so little precision.

References

1. Nishio A, Keeffe EB, Gershwin ME. Primary biliary cirrhosis: lessons learned from an organ-specific disease. *Clin Exp Med.* Dec 2001;1(4):165-178.

2. Bergasa NV. Pruritus and fatigue in primary biliary cirrhosis. *Clin Liver Dis.* Nov 2003;7(4):879-900.

3. Poupon RE, Lindor KD, Pares A, Chazouilleres O, Poupon R, Heathcote EJ. Combined analysis of the effect of treatment with ursodeoxycholic acid on histologic progression in primary biliary cirrhosis. *J Hepatol.* Jul 2003;39(1):12-16.

4. Degott C, Zafrani ES, Callard P, Balkau B, Poupon RE, Poupon R. Histopathological study of primary biliary cirrhosis and the effect of ursodeoxycholic acid treatment on histology progression. *Hepatology.* Apr 1999;29(4):1007-1012.

5. Kaplan MM, DeLellis RA, Wolfe HJ. Sustained biochemical and histologic remission of primary biliary cirrhosis in response to medical treatment. *Ann Intern Med.* May 1 1997;126(9):682-688.

6. Bonis PA, Kaplan M. Methotrexate improves biochemical tests in patients with primary biliary cirrhosis who respond incompletely to ursodiol. *Gastroenterology.* Aug 1999;117(2):395-399.

7. Vuoristo M, Farkkila M, Karvonen AL, et al. A placebo-controlled trial of primary biliary cirrhosis treatment with colchicine and ursodeoxycholic acid. *Gastroenterology.* May 1995;108(5):1470-1478.