ABSTRACT

The efficacy of using methotrexate as a treatment for primary biliary cirrhosis (PBC) remains uncertain, with existing studies providing mixed results. Therefore, we compared methotrexate to placebo in a clinical trial evaluating its potential for decreasing disease progression and the rate of disease-related mortality. Two hundred sixty-five patients with PBC were randomized to receive either methotrexate or placebo. Our sample was predominantly female (92.4%) and middle-aged (mean=51.3, SD=8.7), which is representative of the typical PBC cohort, as this disease is more prevalent among women and is typically diagnosed between the ages of 35 and 60 years. Since there is a low mortality rate in our sample, this outcome becomes difficult to analyze and interpret; instead our analyses focused on progression free survival. Using Kaplan-Meier survival analyses and Cox regression to examine hazard ratios, our analyses were stratified by bilirubin level, as this lab measure has been shown to be predictive of decreased mortality among PBC patients. We were unable to demonstrate any significant difference in either mortality (low bilirubin hazard ratio: 2.02 (95% CI: 0.62, 6.58), p=0.24; insufficient number of events in high bilirubin strata) or progression-free survival (low bilirubin hazard ratio: 1.61 (95% CI: 0.84, 3.09), p=0.16; high bilirubin hazard ratio: 0.98 (95% CI: 0.39, 2.46), p=0.96). Overall, the current data does not provide evidence to suggest that methotrexate is beneficial to PBC patients in terms of slowing disease progression or mortality rates.

Comment: multicenter

Comment: Quite often we would give the number randomized to each arm.

Comment: in a double-blind fashion

Comment: How long were the subjects followed? Descriptive statistics for the censoring distribution would be most helpful here.

Comment: True, but you should analyze the primary endpoint. So your justification for your endpoint should be based on what the protocol specified. (In this case, progression free survival is closer to the one specified in the protocol.)

Comment: Again, this should be done only if this were pre-specified. Randomization was stratified by stage, so adjustment for stage was appropriate. But in that case, we would give a single estimate after adjustment for the precision variable, rather than separate estimates in the strata.

INTRODUCTION

Primary biliary cirrhosis (PBC) is a serious, chronic liver disease, which slowly destroys the bile ducts in the liver. This deterioration prevents the proper excretion of bile, leading to bile buildup and cirrhosis. If left untreated, the eventual result is liver failure and the need for liver transplantation.

The cause of PBC remains unknown (Poupon, 2000). PBC is more often found in women than men and is most often diagnosed in individuals in the age range of 35 to 60 years. There is some evidence to suggest that PBC may be an auto-immune disease, meaning that an affected individual's own immune system is attacking his or her liver.

In recent years, the incidence and prevalence of PBC have increased (Parés and Rodés, 2003). This is due, in part, to the ability to diagnose the disease early and in asymptomatic patients. PBC is a progressive disease, but the rate of progression tends to vary between patients (Poupon, 2000). However, death usually occurs within five to ten years of development of symptoms.

For patients exhibiting symptoms and in the later stages of PBC, several variables have been identified as associated with survival. Among them are age, bilirubin, albumin, prothrombin time, and advanced histological stage (Parés and Rodés, 2003). However, bilirubin levels have been consistently recognized as the best predictor of survival (Parés and Rodés, 2003; Poupon, 2000). Together, these variables and their relative importance in predicting survival are used in model building, which can then be utilized to efficiently evaluate new treatments.

One treatment for PBC receiving attention is methotrexate, the focus of this study. In 1988, Kaplan et al. was first to report an improvement in two PBC patients treated with low-dose methotrexate (Poupon, 2000). Similar results were then observed in a sample size of nine women. Yet, Hendrickse et al. (1999), in a placebo-controlled trial, found no association between methotrexate and survival. The strength of these results have been questioned, however, because the trial included only 60 patients, a small sample size, and patients at entry into the trial were all in a non-advanced stage of disease (Poupon, 2000). The disparate results of these and similar studies fuel the debate over whether methotrexate as an effective treatment for PBC is hope or hype (Kiyasu, 1993).

Therefore, in light of this uncertainty, the objective of this study is to investigate the efficacy of the treatment drug, methotrexate, in treating patients exhibiting the auto-immune pattern of PBC. More specifically, we explore whether methotrexate results in decreased mortality or progression free survival relative to the placebo group.

METHODS

Source of Data

Our data is from 265 patients with PBC who participated in a double blind randomized clinical trial receiving either methotrexate or placebo. Data is available on the demographics of each patient, multiple variables that indicate the severity of the disease for each patient, the length of time each patient remained on either methotrexate or placebo, the length of time until death, and the length of time until clinical progression of the disease. Because the trial was randomized, the variables that were measured are not potential confounders in our analysis.

Comment: multicenter

Comment: You might soften your wording. Some readers will always worry about "imbalances" in the randomization. In some sense, we can definitely state there is no confounding over replicated experiments. But your critics will always worry that your results were spurious. You can get the idea across with softer wording.

Statistical Methods

Analyses were performed using STATA 9.0 software (STATA Corp., College Station, TX, 2006). Survival time was examined for each group using Kaplan-Meier methods for right-censored data. The Kaplan-Meier approach assumes non-informative censoring, meaning that there is independence between censoring and the outcomes of interest (death and clinical progression of disease). In other words, those who are censored must be equally likely to have an event than those who are not censored after the point of censoring. As this is difficult to determine based solely on the data, we assume that those who have left the study are no more and no less likely to die or experience disease progression. It is necessary to use the Kaplan-Meier approach due to the partial information gained from each censored individual; we know that a censored subject survived at least until the time of censoring, but it is inappropriate to claim that they experienced an event at that time. It is also inappropriate to claim that they did not ever experience an event. Therefore, we assume that their probability of an event is equal on average to those still in the study, and therefore each censored individual becomes represented by those still in the study; the Kaplan-Meier methodology reflects this. In this study, our first censored event occurred at two years. Because PBC death usually occurs within five to ten years of diagnosis, our interest would be in ten year survival. However, our longest observation was nine years; therefore, we looked at equal intervals leading up to this endpoint (i.e. three, six, and nine years) from the Kaplan-Meier curve.

Comment: Your discussion of the methods is good when judged at the level of a collaborator. A journal might not allow all of this discussion, so you might lead off with a very succinct description of the methods used: KM and PH regression.

Comment: Very good to note both here and in the Discussion.

Additionally, hazard ratios were examined using Cox regression with the "robust" option. Using the "robust" option allows us to not have to assume proportional hazards. The hazard is the instantaneous risk of an event, and the hazard ratio is simply the comparison of the hazard in the two groups. If we assume proportional hazards, then we are claiming that at all times in our data, the true hazard in one group is a constant multiple of the true hazard of the other group. This claim may not hold true, for example, if the drug has some cumulative effect over time.

Comment: The technical description would mention the "Huber-White sandwich estimator".

Confidence intervals and P-values were constructed from our survival estimates and Cox regression. P-values represent the probability of observing data as or more extreme to that observed, if the null hypothesis were true. While we are interested in decreased mortality, we used two-sided p-values to address the possibility of drug toxicity. The meaning of the confidence intervals is as follows: If the true mean lies within the constructed confidence interval, then the data observed are not unusual. In our analyses, we stratified by bilirubin level, dichotomized into groups of low (≤1mg/dl) and high (>1mg/dl) bilirubin (May, 2002). This is to avoid comparing groups that may have quite different disease profiles, as high bilirubin level has been shown to be the best predictor of poor prognosis (Parés and Rodés, 2003; Poupon, 2000).

Comment: But in presenting your analyses the way you did, you seem to be more concerned with effect modification than merely adjusting for important predictors of disease (a concern about precision).

You should have presented a single analysis as your primary endpoint.

(Parenthetically, I will note that we need to qualify the results you note. If one considers patients only over a short period of time and include patients with severe disease, you might find that bilirubin is not as important. However, I do believe that bilirubin is the best indicator over a longer period of time.)

RESULTS

Two hundred sixty-five patients with PBC were randomized to receive either methotrexate or placebo in a clinical trial. One hundred thirty-three patients were assigned to the placebo group and 132 were assigned to the treatment arm of the study. Demographic data and disease-related variables were summarized for the entire sample and for the treatment groups (see Table 1). No significant differences were expected across groups, as all subjects were randomized for treatment.

Comment: I agree for most of the variables here. Did you think there were any substantial differences that would suggest the treatment groups were not comparable?

I do note that the length of time the drug was taken is a post-randomization variable, and thus could differ across groups.

Mortality rates between the two treatment groups were compared using a Kaplan-Meier survival analysis, stratifying by bilirubin level. We categorically defined bilirubin level into two groups as either low (≤1 mg/dl) or high (>1 mg/dl) (May, 2002). The descriptive graph did not reveal a clear trend toward better or worse survival in terms of mortality for the two treatment groups (see Figure 1B).

To compare the patients' mortality for the two treatment groups stratified by bilirubin levels, we looked at: a) the estimates of the years at which 90%, 80%, and 75% of the subjects are estimated to still be surviving, and b) the estimated probabilities of surviving three, six, or nine years (see Table 2). Hazard ratios estimated using Cox regression did not reveal any treatment differences for the low bilirubin sample (see Table 3). Patients with low bilirubin levels who were receiving methotrexate were 2.03 times more likely to experience death than those receiving placebo, but this difference was not significant ($\chi 2 = 2.02$, $p = 0.24$, 95% CI = 0.62 to 6.58). As there were no deaths at all in the methotrexate group with high bilirubin levels, Cox regression for this stratum becomes extremely difficult to interpret.

Comment: Before this, you might describe the length of follow-up for the patients, as well as some measure of their compliance with the prescribed treatment. I note that Table 1 contains info on this, but you did not comment in your Results section.

Disease progression between the two treatment groups was compared using a Kaplan-Meier survival analysis, stratifiying by bilirubin level. While patients with high bilirubin levels tended to exhibit greater disease progression than those with low bilirubin levels, there was no clear trend toward any effect of treatment on either of these two groups (see Figure 1A).

Comment: Again, you should have analyzed the entire sample, perhaps adjusting for stage (most justifiable given the randomization) or biilrubin (if pre-specified).

Comment: What does this mean?

Hazard ratios estimated using Cox regression did not reveal any significant differences between treatment groups at either bilirubin strata (see Table 3). Patients with low bilirubin levels who were receiving methotrexate were 1.61 times more likely to experience disease progression than those on placebo, but this difference was not significant ($\chi 2$ = 2.02, p = 0.16, 95% CI = 0.84 to 3.09). Similarly, patients with high bilirubin levels who were receiving methotrexate were 0.97 times more likely to experience disease progression than those on placebo, but this difference was also not significant ($\chi 2$ = 0.00, p = 0.96, 95% CI = 0.39 to 2.46).

DISCUSSION

Consistent with previous studies, patients with higher bilirubin levels were observed to have worse survival in terms of disease progression and mortality. The extraordinary importance of bilirubin for survival has been confirmed in multiple studies as previously mentioned. For example, Shapiro et al. (1979) observed the following: for patients with bilirubin levels above 2 mg/dl, 6 mg/dl, and 10 mg/dl, the mean survival was 4, 2, and 1.4 years, respectively. While all patients in this study had baseline bilirubin levels less than 3 mg/dl, a similar relationship between survival and bilirubin was observed.

Based on the results of this clinical trial, methotrexate does not appear to be an effective treatment drug for PBC. Within the low and high bilirubin groups, progression free survival was not improved in the patients receiving methotrexate relative to the placebo group. In fact, for patients in the low bilirubin range especially, progression free survival tends to be worse for those receiving methotrexate (but this was not statistically significant).

Comment: did we have enough precision?

A similar pattern was observed in the mortality analysis, with the exception of the high bilirubin methotrexate group, where no deaths were observed. However, it must be emphasized that only 16 deaths were observed in the entire sample population. This censoring prevents a complete view of mortality, despite using the Kaplan-Meier approach to handle the censoring problem. Specifically, we could not estimate the probability of decreased mortality and 95% confidence intervals for three, six, and nine year survival times for the high bilirubin methotrexate group. Because of this limitation, the results of this clinical trial provide a more complete picture of progression free survival than mortality. Ideally, follow-up will continue for patients in this trial; the longer study period and reduced censoring is necessary to discern the efficacy of methotrexate as a PBC treatment, in terms of improving patient survival time. Likewise, to analyze the effects of methotrexate over time, it would be helpful to have multiple bilirubin measurements at various time points throughout the study.

Comment: An interesting question: Is it worthwhile to continue follow-up in this trial?

One assumption made during this study and analysis warrants further exploration. We assumed an intent to treat study design. However, analysis of the measure of days between treatment randomization and the patient stopping intake of either methotrexate or placebo raises doubt as to whether this is a valid assumption. Certainly, if patients were not taking their respective treatment or if there was significant variation in treatment duration, then there are potentially severe implications for the results.

Comment: Yes and no. Perhaps people just got lazy, and had they taken more drug it would have been effective. But perhaps they experienced toxicity and if we forced them to take more drug they would have died sooner. But I agree there are implications, and showing a bit of the analysis (how did time taking drug compare to time of follow-up?) would be helpful.

Strengths of this study are the study design and larger sample size. In some of the first studies exploring methotrexate as a treatment drug for PBC, there was no control group (Kaplan et al., 1988). Because this study was designed as a double blind, randomized, placebo-controlled

clinical trial, the results are more reliable than previous studies. Likewise, the 265 patient sample size is much larger than previous studies (Hendrickse et al., 1999; Poupon, 2000).

Comment: Statistical information is proportional to the number of events. How did this study compare to those other studies with respect to number of events.

Additional research is needed to determine whether methotrexate is an effective treatment for PBC patients generally. This study was narrowly focused on a subset of PBC patients exhibiting the auto-immune pattern of PBC. Also, more research is needed to determine if there is a dose-response relationship for methotrexate and progression free survival and decreased mortality.

In conclusion, our study did not find evidence to suggest that methotrexate is an effective treatment for patients with the auto-immune pattern of PBC, in terms of slowing disease progression and mortality. The expected relationship between bilirubin levels and progression free survival was observed. While patients with elevated bilirubin levels tended to exhibit earlier disease progression and slightly higher rate of mortality, this did not differ across treatment groups.

REFERENCES

Angulo, P. and E.R. Dickson. 1999. Methotrexate in treatment of primary biliary cirrhosis: The hype and the hope. Gastroenterol 117(2): 492-95.

Bach, N., Bodian, C., Bodenheimer, H., Croen, E., Berk, P.D., Thung, S.N., Lindor, K.D., Therneau, T., and F. Schaffner. 2003. Methotrexate therapy for primary biliary cirrhosis. Am J Gastroenterol. 98 (1):187-93.

Bonis P.A. and M. Kaplan. 1999. Methotrexate improves biochemical tests in patients with primary biliary cirrhosis who respond incompletely to ursodiol. Gastroenterol 117(2): 395-99.

Hendrickse, M. T., Rigney, E., Giaffer, M. H., Soomro, I., Triger, D. R.., Underwood, J. C. E., and D. Gleeson. 1999. Low-dose methotrexate is ineffective in primary biliary cirrhosis: Long-term results of a placebo-controlled trial. Gastroenterol 117: 400-407.

Kaplan, M.M., Knox, T.A., and S. Arora. 1988. Primary biliary cirrhosis treated with low-dose oral pulse methotrexate. Ann Int Med 109:429-31.

Kiyasu, P. 1993. Methotrexate for PBC: Hope or hype? Am J Gastroenterol 88(3): 455-56.

May, L.D., Lefkowitch, J.H., Kram, M.T., and D. E. Rubin. 2002. Mixed hepatocellular–cholestatic liver injury after pioglitazone therapy. Ann Internal Med 136(6): 449-52.

Parés, A. and J. Rodés. 2003. Natural history of primary biliary cirrhosis. Clin Liver Dis 7: 779-94.

Poupon, R. 2000. Treatment of primary biliary cirrhosis. Baillieres Best Pract Res Clin Gastroenterol 14(4): 615-28.

Shapiro, J.M., Smith, H., and F. Schaffner. 1979. Serum bilirubin: a prognostic factor in primary biliary cirrhosis. Gut 20: 137-40.

APPENDIX: FIGURES AND TABLES

Table 1. Demographic and Clinical Characteristics of Study Population

Placebo Group (N=133)					
	N	Mean	SD	Min	Max
Age (years)	133	52.19	8.64	25	67
Sex (% male)	133	7.52	0.26	0	1
Height (cm)	132	163.69	8.07	149.5	191.9
Weight (kg)	132	73.35	16.17	46	149.1
Duration (years)	133	3.67	3.27	0.51	14.43
Stage	133	2.50	--	--	--
Splenomegaly (%)	133	10.53	--	--	--
Serum Bilirubin (mg/dl)	133	0.72	0.39	0.1	2.3
Serum Albumin (g/dl)	133	4.00	0.34	3.1	4.9
Serum AP (U/l)	133	244.92	187.44	63	1127
Serum ALT (U/l)	132	50.17	41.61	9	314
Time to Clot (sec)	133	11.37	1.08	8.6	13.9
Serum Cholest (mg/dl)	130	235.77	58.81	128	560
PLT (1000 cells/cu mm)	133	234.71	83.18	77	561
Stop drug (years)	133	5.81	2.35	0.11	9.79
Methotrexate Treatment Group (N=132)					
	N	Mean	SD	Min	Max
Age (years)	132	50.38	8.67	31	70
Sex (% male)	132	7.58	--	--	--
Height (cm)	132	163.71	7.92	142.8	181.6
Weight (kg)	132	70.12	14.42	42.7	114.2
Duration (years)	132	3.51	3.46	0.55	17.93
Stage	132	2.57	--	--	--
Splenomegaly (%)	132	8.40	--	--	--
Serum Bilirubin (mg/dl)	132	0.66	0.44	0.1	2.8
Serum Albumin (g/dl)	132	4.00	0.35	3	5.9
Serum AP (U/l)	132	242.85	145.67	50	930
Serum ALT (U/l)	131	54.29	41.43	9	199
Time to Clot (sec)	128	11.24	1.13	8.5	14.6
Serum Cholest (mg/dl)	130	239.18	58.18	137	472
PLT (1000 cells/cu mm)	132	243.53	88.60	86	619
Stop drug (years)	132	5.96	2.50	0.13	8.96

Duration: duration of disease (from diagnosis to randomization); AP: alkaline phosphatase; ALT: alanine transaminase; Cholest: cholesterol; PLT: count of platelets circulating in blood

Comment: The format of this table makes comparisons across treatment groups (the primary comparison of interest) very difficult. Try to have treatment groups as columns.

Also, the Stop Drug data (which you never define or comment on in your Results) is a post-randomization variable, and thus randomization does not protect you here.

Figure 1A. Kaplan-Meier Survival Estimates for Disease Progression Stratified by Treatment and Bilirubin Level

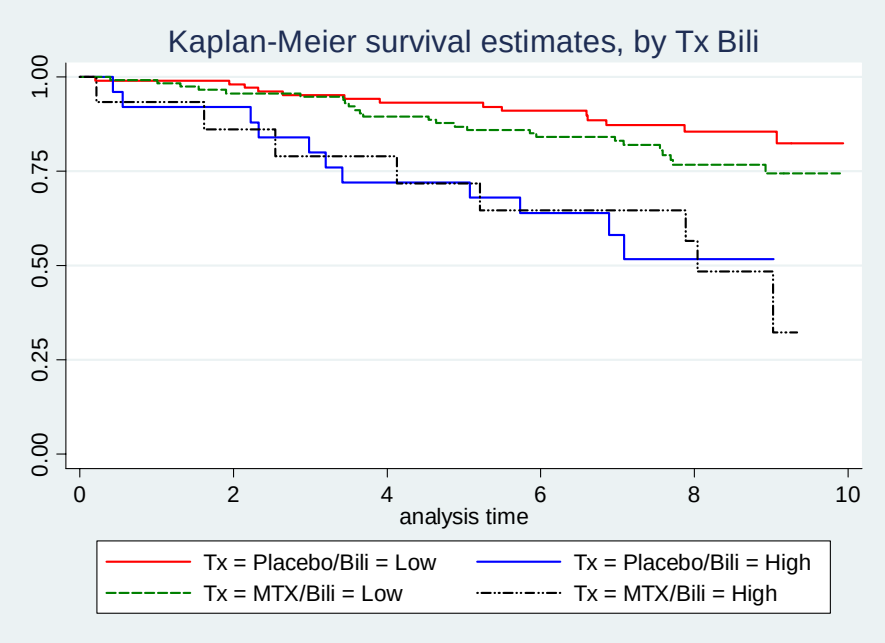

Figure 1B. Kaplan-Meier Survival Estimates for Mortality Rates Stratified by Treatment and Bilirubin Level

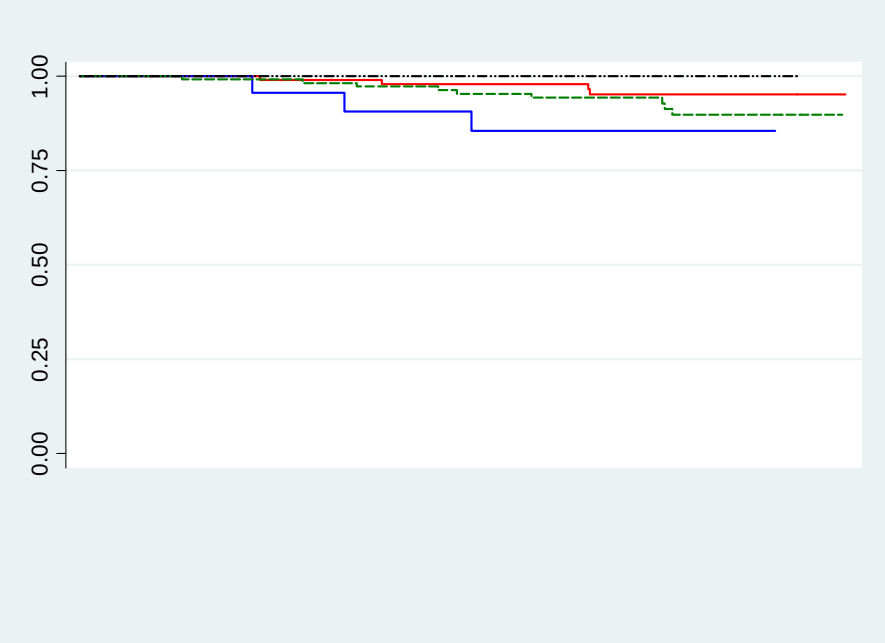

Table 2. Survival Analysis for Disease Progression by Treatment Group Stratified by Bilirubin Level

Group	Percentiles for Survival Time (years)			Probability of Surviving (95%CI) (%)		
	90%	80%	75%	3 Years	6 Years	9 Years
Disease Progression						
Placebo						
Low bili	6.60	>9.94	>9.94	95.13 (88.70, 97.95)	91.05 (83.49, 95.24)	85.50 (76.06, 91.42)
High bili	2.22	2.98	3.42	80.00 (58.44, 91.15)	64.00 (42.21, 79.38)	51.72 (28.85, 70.48)
MTX						
Low bili	3.69	7.59	8.93	94.78 (88.75, 97.62)	84.16 (76.04, 89.71)	74.44 (63.81, 82.37)
High bili	1.61	2.54	4.13	78.97 (47.91, 92.71)	64.62 (34.68, 83.52)	48.46 (20.98, 71.46)
Mortality						
Placebo						
Low bili	>9.97	>9.97	>9.97	98.99 (93.05, 99.86)	97.94 (92.00, 99.48)	95.22 (87.63, 98.20)
High bili	5.10	>9.06	>9.06	95.65 (72.93, 99.38)	85.58 (61.53, 95.13)	85.58 (61.53, 95.13)
MTX						
Low bili	7.72	>9.93	>9.93	98.21 (93.01, 99.55)	94.32 (87.78, 97.41)	89.83 (81.00, 94.69)
High bili	>9.36	>9.36	>9.36	NA	NA	NA

Note: CI = confidence interval; bili = bilirubin

Comment: Again, the primary comparison should be across treatment groups, so your table should be laid out in a way to facilitate that comparison.

Table 3. Survival Analysis for Mortality by Treatment Group Stratified by Bilirubin Level.

	Hazard Ratio	95% CI		P-value
Disease Progression				
Low bili	1.61	0.84	3.09	0.16
High bili	0.98	0.39	2.46	0.96
Mortality				
Low bili	2.02	0.62	6.58	0.24
High bili	4.12e-17	1.16e-17	1.46e-16	<0.0001

Note: CI = confidence interval; bili = bilirubin

Comment: As you noted, the lack of any events in the MTX high bilirubin group makes these results quite suspect. You have essentially estimated a zero probability of an event, and due to the mean variance relationship, this estimates zero variability. The Wald statistic is not trustworthy here. The more trustworthy score statistic (logrank statistic obtained from sts test) gives a P value of about 0.19.