Summary

The purpose of this analysis is to examine for evidence of pulmonary toxicity among patients taking methotrexate for treatment of primary biliary cirrhosis. The primary study was a randomized, double-blinded, placebo controlled trial testing whether methotrexate might prolong the overall or progression free survival of patients with primary biliary cirrhosis. Data was collected during the study period regarding lung function, as measured by DLCO. Two hundred sixty five patients were enrolled in the study. Of these, there was baseline and follow-up data of lung function available for 119 patients randomized to receive methotrexate, and 125 patients who received placebo. There were no major differences in the age, sex, weight, height, severity or duration of disease between these two groups. Decline in lung function was defined as a decrease in DLCO over the time of the study enrollment. The mean change in DLCO was defined as measurement at the time the patient discontinued the study drug or placebo minus the measurement at baseline. In the placebo group, there was a 1.51 ml/min/mmHg decrease in DLCO over the time of the study (95% confidence interval [-.78, 2.24]). The mean decrease in DLCO in the methotrexate group was a 1.59 ml/min/mmHg over the time of the study (95% CI [-.85, 2.33]). There was no statistically significant difference found in the mean change of the two groups (p = 0.5620). In this study of methotrexate for treatment of primary biliary cirrhosis, there was no worsened decline in lung function, as measured by DLCO, in the patients who received methotrexate as compared to placebo.

Comment: multicenter

Comment: how long

Comment: what is a CI for the difference between MTX and Plc?

Comment: Stated way too strongly unless you justify that the CI has ruled out all differences that were clinically meaningful

Background

Primary biliary cirrhosis is a disease in which the scarring of intrahepatic bile ducts impairs the secretion of bile into the intestinal tract. As a result, bilirubin levels increase in the tissues and there is an onset of jaundice. The condition can progress so that the liver becomes damaged and lessens its ability to detoxify chemicals, produce proteins, and metabolize fats and glucose. It has been suggested that the onset of disease is due to some autoimmune component in which the patient's own immune system damages his or her liver, although the disease does affect more women than men. The disease is often diagnosed when the patient is between the ages of 35 to 60 years old and death usually occurs 5 to 10 years after the first diagnosis.

Methotrexate, a drug used in the treatment of cancers and diseases such as rheumatoid arthritis and psoriasis, is being looked at as a potential treatment to be administered to patients with primary biliary cirrhosis. Patients with primary biliary cirrhosis may have impaired liver function, which can lead to a decreased ability to metabolize drugs. Therefore increased toxicities resulting from treatment with this hepatically-metabolized drug are of concern and need to be monitored in patients receiving it. Due to its' use as a treatment of other diseases, there is some data on toxic effects of the drug at high doses, including impairment of lung function. This is commonly measured as a patient's diffusing capacity of the lung for carbon monoxide, or DLCO. It is a measurement of the ability of gases to diffuse into the patient's blood from inhaled air in their lungs. Low values of DLCO suggest some degree of impaired lung function. Since the aim of administering methotrexate to patients with primary biliary cirrhosis is to improve the survival of patients, impaired lung function is a toxic effect that is to be avoided since it alone can affect the survival of a patient that is sufficiently ill.

Comment: no apostrophe

Question of Interest

The questions of interest for this study are 1.) whether treatment with methotrexate in conjunction with UDCA imparts improved survival or progression-free survival relative to treatment with UDCA alone (placebo) and 2.) whether treatment of primary biliary cirrhosis with the study drug methotrexate leads to impaired lung function as measured by DLCO compared to the placebo control group. Only the latter question will be addressed in this paper. There is a potential discrepancy between the two questions in that if the methotrexate does impair lung function to the point where the patient's health is put in danger, it may in fact cause a decrease in patient survival, regardless of the progress of the disease.

Comment: I agree. And it can work either way. It could be that MTX impairs lung function, but that the net effect is improved survival because it helps PBC so much.

Source of Data

A total of 265 patients with primary biliary cirrhosis and exhibiting the auto-immune pattern of the disease were used as subjects for a randomized, double blinded, placebo controlled clinical trial in order to determine if the drug methotrexate affected the survival. These patients were randomized to receive either methotrexate or placebo in conjunction with UDCA, and data were obtained on the patients' general history (age, sex, weight, height), markers for liver disease or malfunction (including bilirubin and albumin levels in the blood, cholesterol and amino alanine transferase (ALT) serum levels, and the stage of disease), as well as information on the patient outcome during the period of observation (the amount of time until death or progression of disease was observed). However, for this study, the focus will be on measures of DLCO: the baseline DLCO or the value at randomization, the DLCO at the time the patient discontinued the drug or placebo (final DLCO), and the minimum DLCO value measured while the patient was taking the drug or placebo.

There were 8 patients in the placebo group and 12 patients in the treatment group that had values for the baseline DLCO but were lacking values for the minimum and final DLCO. One patient in the treatment group was missing a value for baseline DLCO. Since we were unable to look at the change in DLCO during the duration of the study for these individuals, they provided no information on the effect methotrexate may have on lung function and they were removed from the analyses. Factors such as age, underlying lung disease, smoking, anemia and poor effort with the test can affect DLCO measurements. However, with the use of randomization, these factors should be equally distributed between the treatment and placebo groups, thereby minimizing confounding effects.

Comment: This is not necessarily true. Randomization works for toxicity as well. We do not have to have baseline values, though using them may afford us more statistical preision.

Statistical Methods

The statistical package Stata version 9.1 was used to analyze the data to address the question of impaired lung function. For those patients having all three DLCO measurements available, these values were analyzed using the t-test. Under the assumption that each of the DLCO variables (baseline, final, and minimum) in each treatment group approximates a bell-shaped curve (is distributed normally), the t-statistic is the standardized difference between the means in each treatment group. In Stata, the t-test gives the probability that there is no real difference in the mean DLCO value between the treatment groups. For each of the treatment groups the t-test also gives the mean of

Comment: We do NOT need normality of the variables' distribution. We only need that the sample mean be normally distributed. (I keep saying this over and over again—it takes a long term to unlearn things that are mistaught)

Comment: the t statistic is always the standardized difference in means, whether or not the t statistic is approximately normally distributed. (Recall that we can use Chebyshev's thm here.)

Comment: NO. This is a Bayesian interpretation. The P value gives the probability of observing such "extreme" data when the null hypothesis is true.

the DLCO variable and an interval around the mean; we can be 95% certain that the true mean of the DLCO variable in the patient population lies in this interval. The width of this interval reflects the degree of spread of the data. The t-test was performed on each of the three DLCO variables by treatment group. In addition the t-test was performed on the difference between final and baseline measurements, and the difference between final and minimum measurements by treatment group. In each case the assumption of normality of each variable is reasonable by inspection of a scatterplot. Analysis on the geometric means and the medians was not performed, as each of these measures is insensitive to outliers, which are of prime concern. Low outliers may represent patients who either have lower lung function at baseline or are especially sensitive to the toxic effects of the drug. High outliers may represent patients with very good lung function at baseline or are especially insensitive to the toxic effects of the drug.

In this analysis, observations of patients missing any DLCO measurements have been dropped: of 265 patients, one had a missing baseline DLCO measurement and 20 had missing final and minimum DLCO measurements. These missing data can be regarded as non-informative, since there was no considerable difference in patient demographics and lab values on those patients missing DLCO values and those having DLCO values

Results:

Baseline characteristics of both treatment groups after the exclusion of the patients who lacked any pertinent DLCO measurements are summarized in Table 1. The placebo arm consisted of 125 patients and the methotrexate arm consisted of 119 patients. The two treatment arms have comparable characteristics. The placebo group had 92.8% women and the methotrexate group had 92.5% women. The mean body mass index (BMI) for the placebo group was 27.2 kg/m^2 and 26.2 kg/m^2 for the methotrexate group. Moreover, the duration of treatment in both groups is comparable, and hence comparisons of the minimum DLCO measurements across the two treatment groups, as well as examination of the difference between minimum and baseline DLCO measurements, are both reasonable and justifiable. Both the placebo group and methotrexate group had a median stage of disease of 3, and death was observed in similar percentages between both treatment groups, observations consistent with a correct randomization process.

Comment: No, but the t-test can be "inverted" to obtain a CI. Stata is nice in that its ttest function also provides the CI.

Comment: Careful here. Given your misinterpretation of the P value, I am betting you also do not understand what "95% certain" means. What is your probability space (or to a layperson, your denominator)?

Comment: Yes, but it also reflects the sample size, so the width of the CI is a pretty poor measure of the spread of the data.

Comment: Same problem. First, I can't imagine how you can judge normality from a scatterplot. Second, I have spent a lot of time on the fact that you do not need normality of the data.

Comment: Good motivation for your choice of summary measure. You probably would not get this into a published manuscript, but I think it would probably be beneficial.

Comment: There is no information in your data that can tell you whether missing data is ignorable. What we care about is whether the people missing data might have had lower measurements of DLCO than the people who were not missing data. If we tended to see that the older people were missing data more often, for instance, then we might especially worry that the missing data were nonignorable. But the absence of any such patterns is not a guarantee. Hence, it is good for you to comment on whether there were any differences in baseline variables, but use softer language. "Statistics means never having to say you are certain."

Comment: Many times we would present the data for those with and without data. We certainly would want to be sure that missing data was not more common in

Comment: Could this lower sample size on the MTX arm be a sign of decreased lung function?

Table 1. Descriptive statistics for patient characteristics by treatment group

Placebo Arm	N	Mean	Sd	Min	p25	P50	P75	Max
Age (years)	125	52.28	8.63	25	47	53	59	67
Sex – proportion male	125	0.072	-	-	-	-	-	-
Weight (kg)	124	73.15	16.17	46	61.95	70.05	84.35	149.1
Height (cm)	124	163.82	8.21	149.5	157.25	162.70	168.65	191.9
BMI (kg/m^2)	124	27.19	5.48	17.10	23.33	26.76	29.70	53.39
Duration of disease (years)	125	3.64	3.22	0.5	1.02	2.37	5.37	14.42
Duration of treatment (years)	125	6.13	2.00	1.25	5.33	6.33	7.86	8.79
Proportion of clinical progression	125	0.168	-	-	-	-	-	-
Proportion of death	125	0.04	-	-	-	-	-	-
Methotrexate Arm	**N**	**Mean**	**Sd**	**min**	**p25**	**P50**	**P75**	**Max**
Age (years)	119	50.23	8.73	31	44	50	56	70
Sex – proportion male	119	0.075	-	-	-	-	-	-
Weight (kg)	119	70.41	14.42	42.7	59.4	68.3	77.1	114.2
Height (cm)	119	163.81	8.03	142.8	157.2	162.56	170.5	181.6
BMI (kg/m^2)	119	26.16	4.78	17.6	22.54	25.31	29.22	43.21
Duration of disease (years)	119	3.6	3.57	0.55	0.92	2.7	4.5	17.94
Duration of treatment (years)	119	6.45	1.99	0.93	5.47	7.08	8.00	8.96
Proportion of clinical progression	119	0.22	-	-	-	-	-	-
Proportion of death	119	0.067	-	-	-	-	-	-

Comment: Having the MTX and Plc groups side by side facilitates comparisons.

I note that many journals might not allow all these descriptive statistics. I would argue that for the duration of treatment it is extremely important that we have comparability of the entire distribution.

Comment: This is an indicator for a censored variable. Can you justify giving this proportion when the follow-up differs? (Same goes for the clinical progression)

There were 264 patients with a baseline DLCO measurement and 244 patients with a DLCO measurement at the end of the study. Table 2 and Figure 1 presents the DLCO measurements at baseline, at the end of the study and the minimum measurement while under treatment.

Table 2. Descriptive statistics for outcomes by treatment group

DLCO (ml/min/mmHg)		Obs	Mean	Std. Dev	Min	P25	P50	P75	Max
Baseline	Placebo	125	19.87	5.02	11.5	16.6	18.6	22.3	41
	Methotrexate	119	20.35	5.06	8.9	16.4	20.1	23.7	37.7
Minimum	Placebo	125	16.70	4.17	8.6	14.3	16.2	18.6	32.7
	Methorexate	119	17.01	4.35	7	14.6	17.2	19.2	34.6
Final	Placebo	125	18.36	4.96	8.5	15.7	18.2	20.8	39.4
	Methotrexate	119	18.76	4.77	8.6	15.4	18.8	21.4	35.8

Comment: I would have included the differences in this table as well. Fact is, I have greatest interest in the min and maz for the minimum and final, but it could be that the max drop is also clinically important (especially since we have not normalized for body size, age, sex).

Note that I also have less problem with MTX and Plc by rows here, because I can easily compare each measurement.

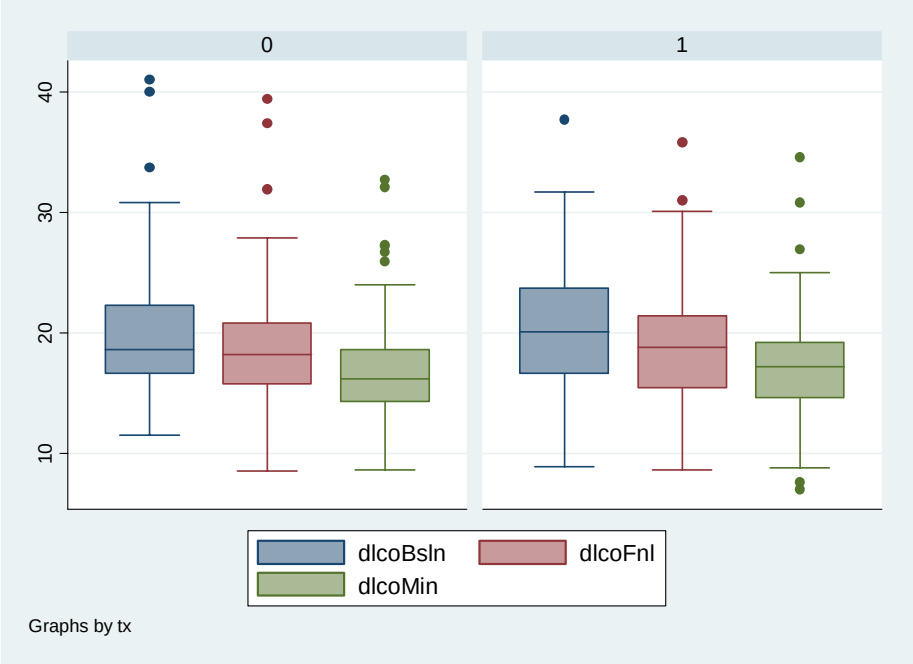

Figure 1. Box plots of DLCO data obtained during course of treatment (0=placebo, 1=treatment).

The analysis of the change in the DLCO measurements was restricted to the 125 patients in the placebo group and the 119 in the methotrexate group with DLCO measurements available at the beginning and the end of the study. At the time of randomization, the average DLCO measurement in the placebo group was 19.87 ml/min/mmHg (SD 5.02 ml/min/mmHg). At the final DLCO measurement, the average DLCO was 18.36 ml/min/mmHg (SD 4.96 ml/min/mmHg). In the methotrexate group, the average DLCO measurement at baseline was 20.35 ml/min/mmHg (SD 5.06 ml/min/mmHg) and at the end of the study duration the average DLCO was 18.76 ml/min/mmHg (SD 4.77 ml/min/mmHg). There were no statistically significant differences between the two groups' DLCO at baseline, minimum and final (Table 3).

Since the DLCO values in each treatment group are highly correlated, a higher precision can be attained in the analysis by examining the difference between the final and baseline DLCO values than by examining the final value alone. However, the minimum DLCO value for each patient is highly dependent on the number of measurements made on this patient, while the baseline and final measurements are not, so this comparison may not be valid as the number of measurements made is not known. In the present case this comparison may be valid, as the difference in the time from diagnosis with PBC to randomization between the two treatment groups is not statistically significant (p = 0.9161, 95% confidence interval for difference in years [-.8138, .9058]). Hence, analysis

Comment: Yes. This was why it was so important to see comparability of the time on drug. But this has to do with the extreme value distribution and its dependence on sample size, not on the use of the baseline measurement. (It is also affected by aging, calendar year effects, etc.)

Comment: This is irrelevant. We care about number of DLCO measurements made and time on treatment. These have to do with the time on drug, not duration of disease.
This was pointed out to you by your referees.

of the difference between minimum and baseline DLCO values across the two groups is reasonable.

Comment: I believe it is just as appropriate to take the difference MIN – BSLN as it is to take the difference FNL – BSLN.

Table 3 summarizes the differences between final and baseline DLCO measurements, and the differences between minimum and baseline measurements, across treatment groups. The placebo group averaged a decrease of 1.52 ml/min/mmHg (SD 0.36 ml/min/mmHg) over the placebo treatment period. The methotrexate group averaged a decrease of 1.59 ml/ min/mmHg (SD 0.37 ml/min/mmHg) over the treatment period.

The difference between the change in DLCO measurement from baseline to final in the two treatment arms was not statistically significantly different from 0 (lower one-sided P =0.5620). A 95% confidence interval suggests that the observed results would not be unusual if the true average change over the treatment period in a population treated with placebo were anywhere between a decrease of 0.95 ml/min/mmHg and an increase of 1.11 ml/min/mmHg as compared to the treatment group.

A similar comparison between the measurements of both groups at their minimum DLCO measurement under treatment and at baseline shows that the difference is also not significant (lower one-sided P =0.6471). A 95% confidence interval suggests that the observed results would not be unusual if the true average change over the treatment period in a population treated with placebo were anywhere between a decrease of 0.73 ml/min/mmHg and an increase of 0.1.08 ml/min/mmHg as compared to the treatment group.

Comment: In both cases, you would probably want to comment on the maximum decrease. We do not always live and die by statistical significance in these cases. Anecdotal observations are also important. Luckily you did show the minimum DLCO measurements in Table 2, and we could see that there were no "catastrophic" losses of lung function. A word or two in the text to this effect would be a good idea.

Table 3. Inferential statistics by treatment group

DLCO (ml/min/mmHg)	Treatment	Mean	Std. Err	Std. Dev	[95% Conf. Interval]	P value (one sided)
Baseline	Placebo	19.87	0.45	5.02	[18.98, 20.76]	P=0.3723
	Methotrexate	20.35	0.46	5.06	[19.43, 21.27]	
Minimum	Placebo	16.70	0.37	4.17	[15.96, 17.44]	P=0.6482
	Methotrexate	17.01	0.40	4.35	[16.22, 17.80]	
Final	Placebo	18.36	0.44	4.96	[17.48, 19.24]	P=0.5996
	Methotrexate	18.76	0.44	4.77	[17.89, 19.62]	
(Final-Baseline)	Placebo	-1.51	0.37	4.11	[-2.24, -0.78]	
	Methotrexate	-1.59	0.37	4.09	[-2.33, -0.85]	
	Difference	0.082	0.52		[-0.95 1.11]	P=0.5620
(Minimum - Baseline)	Placebo	-3.17	0.30	3.36	[-3.77, -2.58]	
	Methotrexate	-3.35	0.23	3.81	[-4.04, -2.65]	
	Difference	0.17	0.46		[-0.73, 1.08]	P =0.6471

Discussion:

The primary goal of this randomized, double-blinded, placebo-controlled trial for the treatment of primary biliary cirrhosis was to compare progression of disease and survival during treatment with UDCA in conjunction with either methotrexate or placebo. However, with concern for potential pulmonary toxicity associated with use of

methotrexate (1), a secondary function of the study was to assess for evidence of lung toxicity in the treatment groups. While both treatment groups had a decline in their lung function over the period of the study, this analysis did not show any significant differences in the manifestation of lung toxicity between the two treatment groups. Lung toxicity was assessed by measurement of DLCO, the diffusing capacity of carbon monoxide from the inhaled air to the pulmonary circulation.

Previous studies of methotrexate to treat cancers and other inflammatory diseases have shown evidence of pulmonary toxicity (2,3). However, those studies tended to use doses of methotrexate higher than the doses used in this study. Our current study did not allow us to assess a dose-response in the degree of lung toxicity, but it is possible that there is increased risk of lung toxicity as the dose of methotrexate increases.

Additionally, it should be considered that the single measure of DLCO may not fully capture all cases of pulmonary toxicity associated with methotrexate. While it is a reasonable measure for the loss of alveolar-capillary interface for diffusion of gas, there may be other manifestations of lung disease that were not examined. If desired, future studies could also consider spirometry to assess for obstructive or restrictive defects in lung function, as well as symptomatic shortness of breath scores and clinical symptoms as rated by the patient.

There are other potential limitations to this current study. While this was a relatively large trial of the use of methotrexate for patients with PBC, it is possible that with a larger sample size we would be able to detect smaller differences between the two groups. Additionally, not all study participants were followed for the entire duration of the study; there was censoring. Perhaps if followed for a longer period of time, there would be more manifestations of lung toxicity, which could lead to differences between the two groups. Finally, while there were no large differences in the demographics of the two treatment groups, it should be noted that males were quite underrepresented in this study.

Comment: This only matters if smaller differences would be clinically important. Would they? (Yeah, I don't know either. But I would at least comment on this issue.)

In conclusion, this current study does not give any evidence of an increased risk of lung toxicity with use of methotrexate at the goal dose of 13-15mg/kg/day in patients with primary biliary cirrhosis as compared to similar patients who are taking placebo.

Comment: Not knowing the clinically important difference, I would soften my wording by including the interpretation of the CI in the sentence.

References:

1. Cannon GW. Methotrexate pulmonary toxicity. Rheum Dis Clin North Am. 1997 Nov;23(4):917-37.
2. Kremer JM et al. Clinical, laboratory, radiographic, and histopathologic features of methotrexate-associated lung injury in patients with rheumatoid arthritis: a multicenter study with literature review. Arthritis Rheum. 1997 Oct;40(10):1829-37.
3. Shapiro CL et al. Drug-related pulmonary toxicity in non-Hodgkin's lymphoma. Comparative results with three different treatment regimens. Cancer 1991 Aug 15;68(4):699-705.