The effect of treatment with methotrexate for primary biliary cirhossis on lung function, as measured by change in diffuse lung carbon monoxide.

Summary: Low dose methotrexate treatment (MTX) has been proposed as a potentially promising treatment for primary biliary cirrhosis (PBC), which accounts for many of the liver transplantations in the USA. However, methotrexate treatment has been found to contribute to lower lung function in conditions other than PBC, and if it occurs concomitant with treatment of PBC, could exacerbate the reduced lung function already a characteristic of these patients. We therefore tested for impaired lung function as measured by diffused lung carbon-monoxide (DLCO) during a long term randomized placebo control trial of MTX on PBC. No significant impairment of the mean change in DLCO (baseline – minimum) were found between placebo or MTX groups for early (-0.65, (-1.94, 0.64), p=0.1605) or advanced (0.16, (-1.12, 1.44), p=0.5972) PBC liver stages. Likewise, no statistically significant effect of MTX was detected using the ratio of geometric means: (-3.74%, (-11.59%, 3.56%), p=0.3206) and (1.85%, (-5.47%, 8.65%), p=0.3991), respectively. An exploratory analysis considering duration of treatment also could not reveal a significant impairment of DLCO in patients grouped by shorter than or longer than 6 years of MTX treatment. In conclusion, this study was unable to establish evidence of lung toxicity by MTX among early or advanced stage PBC patients.

Comment: how long?

Comment: multicenter

Comment: How many patients?

Comment: Why present subgroup analyses first? And what are the units of DLCO. And how did you compute differences: MTX – Plc or Plc – MTX?

Comment: Why the multiple comparisons?

Comment: I would agree. Now can we use the inability to detect a difference as proof that we are not worried about toxicity?

Background: Methotrexate (MTX) is an anti-folate compound approved by the FDA in 1953 for oncological treatments, and has more recently been tested as an anti-inflammatory and immunosuppressant (1), where it has been shown to be beneficial in low doses for treatment of autoimmune disorders such as psoriasis, and rheumatoid arthritis. Side effects of low dose MTX include opportunistic lung infections, such as *Pneumocystis carini* among the elderly and interstitial lung injury, which occurs more often in those 60 years and older (2).

According to pharmacokinetic studies, dose individualization is needed only at therapy initiation. Acute MTX toxicity symptoms (i.e., nausea and fatigue) occur in about 30% of patients and are attributed to folate depletion. Folate therapy and MTX dose division can alleviate acute toxicity. In the absence of folate supplementation, high BMI may aggravate MTX hepatotoxicity. During the early phases of MTX, hepatotoxicity can arise, but typically resolves within the first month of therapy (2). Severe hepatotoxicity (stage 3 or 4), which may occur later, depends on the weekly and cumulative MTX dosages (2).

Despite potentially complicating liver distress, MTX was piloted as a treatment for primary biliary cirrhosis (PBC) in 1996, and a potential benefit was suggested for early disease states (3). In 1998 a study showed MTX decreased inflammation and bile duct injury after 2 years of MTX treatment (4). PBC is in part an inflammatory-mediated biliary disease of unknown cause, arising more often in middle-aged women.

PBC is one of the top five predecessors to liver transplantations in the USA (5). Studies of liver transplant recipients prior to operation indicated that PBC patients are likely to have reduced diffusing capacity for carbon monoxide (6-8). In an attempt to slow PBC progression, a long period of MTX was trialed in a randomized control trial by stage of disease. The potential for lung complications due to MTX in PBC patients was surveyed using diffused lung carbon monoxide measures. We report here our findings regarding lung function in this MTX-for-PBC trial population.

Comment: Why not start of the Background with the disease you are treating, and then talk about the treatment?

Questions of Interest: We address the question "Does treatment with the study drug lead to impaired lung function as measured by diffused lung carbon monoxide (DLCO)?" The DLCO measures used to address this question include DLCO at time of randomization (DLCO), a minimum DLCO obtained while patient was taking MTX or placebo, and a final DLCO measure once the patient discontinued MTX or placebo. We specifically address the questions "Does treatment with MTX lead to impaired lung function as measured by larger differences between baseline and minimum DLCO?" In an exploratory question, we address "If we consider time on MTX or placebo, would this affect or influence our conclusion regarding impairment of lung function?"

Source of the Data: The data provided for the analysis was selected from a multi-center, double-blind randomized, placebo-controlled trial of MTX in which the primary question was whether MTX might prolong survival, or survival free from progression of liver disease in a group of patients with cirrhosis with a possible auto-immune component. Patients were divided into groups by stage of disease as assessed from liver biopsy prior to randomization. The data consist of the demographic measures age, sex, weight (1 missing) and height (1 missing) and possible cirrhosis disease severity measures at the time of randomization for 265 enrolled patients (of 535 screened) including: time from diagnosis to randomization , levels of alkaline phosphatase, alanine amino transferase (1 missing), bilirubin, albumin, cholesterol (5 missing), indicator of enlarged spleen (splenomegaly, 1 missing), platelet count, time to blood clotting (1 missing) and stage of disease. Measures of lung function were baseline DLCO (at time of randomization, 1 missing) and the minimum DLCO (20 missing) and final DLCO (20 missing) observed while on treatment. Patients were on either MTX or placebo for variable amounts of time.

Eight and twelve patients were missing both minimum and final DLCO measures in the placebo and MTX arms, respectively, and 1 patient in the MTX arm was missing baseline, but not minimum or final DLCO. These 21 patients as a group had stopped taking drug far earlier than the other 244 patients (17/21 vs. only 1/244 patients had stopped by day 445, 20/21 vs. 12/244 by day 606). These patients did not differ substantially in any measures at time of randomization but were more likely to die or progress earlier in the study (see Statistical Methods). As these 21 patients were missing the measures necessary for our analysis, we did not include them in our inferential analyses. We found an apparent error in the data for four patients as they could not possibly have received drug while they were dead. We therefore truncated the time-on-treatment to equal time until death for these 4 patients.

Statistical Methods: The data analysis was conducted using Stata version 9.2 (Stata Corp., College Station TX). Demographics, liver function data at randomization and all DLCO measures were evaluated for anomalies in the distribution within subgroups defined by stage and treatment assignment. Distributions were characterized by their mean, standard deviation, and range. Histograms were examined for potential anomalies, for example bimodality or outliers. Scatterplots and Lowess-smooth curves compared baseline demographic characteristics and measures of disease severity against the outcome, providing guidance about the use of transformed data and helping to identify potential bivariate anomalies. Based upon these scatterplots, the mean difference and ratio of geometric means (to reduce the influence of outliers and variability in outcome across groups defined by the change in DLCO) were selected as summary measures of the change in DLCO from baseline to minimum.

Comment: How long did it last? 3 days? 100 years? Might this be germane enough that you could give us a hint?

Comment: Very good to find out and note. This would of course be fixed prior to final publication of the analysis. In fact, in real life I would sort this out before even proceeding further with the analysis. This was not real life for you, however, and you took a very appropriate tack.

Comment: And then what? If there were errors you would fix them, but otherwise you would just note them and proceed by including them in the analysis.

Comment: In a real paper, you would not mention all this checking for errors. It would be presumed.

Comment: Wrong motivation on a couple counts. First, you should choose your summary measure prior to looking at the data. Second, it is not clear that you would want to downweight outliers: They might be your cases of severe toxicity.

Participants missing a baseline, minimum, or final DLCO measurement were excluded from the inferential analyses (see Source of Data). PBC-stage-specific comparability of excluded and included participants was assessed by comparison of available data on baseline variables and DLCO. Time-on-treatment was defined as the time until earlier of either termination of treatment, last follow-up, or death. The fact that complete DLCO data on all included patients exists removes censoring as an issue for this analysis. Since the study sample was randomized within strata defined by PBC stage ("early" as Stages 1-2 and "advanced" as Stages 3-4), descriptive analyses were stratified by disease stage and treatment assignment, whereas inferential analyses compared treatment effects within stage strata.

Comment: In real life, I would hate for people to exclude a treated patient because they were missing a baseline value. Arguably the patient could not be ethically treated if you knew they were not going to be used in the analysis. A better approach would be to impute the baseline data, or to use an analysis that does not require the baseline data. (This, of course, is not real life, so this was fine here, so long as you noted it.)

Comment: You would want to "adjust for stage", not do subgroup analyses. By doing subgroups, you decrease the sample size and lessen your power to detect toxicity.

To determine if an association exists between methotrexate treatment and potential decrease in lung function, the mean changes in DLCO from baseline to minimum were compared between the treatment groups for each PBC stage. The same comparison was conducted using the ratio of geometric means. Ninety-five percent confidence intervals and one-sided p-values for t-tests of no difference from the null (i.e., ≥ 0 for the mean difference and ≥ 1 for the ratio of geometric means) were estimated for both comparisons. For analyses involving use of the geometric mean, point estimates and 95% confidence intervals are presented as the percent change in the geometric mean of methotrexate as compared to the placebo groups. Confidence intervals not including the value postulated by the null hypothesis or one-sided p-values <0.05 were considered indicative of a significant difference between stage-specific treatment groups.

Due to concerns about the effect of chronic exposure to methotrexate (see Background),the effect of time on treatment on any association, or lack thereof, between treatment and the outcome was investigated by calculating stratified mean differences of the change in DLCO with accompanying 95% confidence intervals and one-sided p-values of the probability of observing a difference of ≥ 0. A lack of a priori estimates of a relevant threshold of time on treatment, the median time on treatment (6.3 years), rounded to the nearest year, was selected as the cutoff for dichotomizing this variable.

Comment: Regression and scatterplots could also have been used.

Results: Table 1 presents descriptive statistics on randomized groupings with respect to disease stage for all available data on the 265 subjects. Within PBC-stage-groups, the treatment groups demonstrated similar means and ranges for age, height, weight and BMI. The treatment-specific prevalence of splenomegaly in the advanced PBC stage-group was double that of the early disease stage-group.

Comment: Hardly surprising. That is why we call it "advanced".

Within stage group we asked whether there was an effect of methotrexate on lung function as measured by the difference from baseline to minimum DLCO (delta-DLCO) and found no statistically significant effect (Figure 1A and Table 2). Within early stages, we found a 0.6502 ml/min/mmHg (95% CI: 0.643 lower through 1.943 higher) higher mean delta-DLCO in methotrexate versus placebo treatment groups. Within advanced stages, we found a 0.1594 ml/min/mmHg (95% CI: 1.438 lower to 1.119 higher) lower mean delta-DLCO in methotrexate versus placebo treatment groups. The null hypothesis that there is no impairment of lung function, as measured by DLCO, due to MTX compared to placebo was not rejected for both early and late stage disease patients (lower one-sided P-value = 0.1605 and 0.5972, respectively).

Comment: Bad idea. Averaging the treatment effectg across strata would have been okay, but you really lost power with this analysis within subgroups.

Comment: It is very much of interest to comment on the minimum measurements within each group. Sometimes there are a few people who experience a very severe toxicity. We need to note these. It will often be anecdotal, but we note them anyway.

Comment: Did we rule out all clinically meaningful effects.

Within stage group we also asked whether there was an effect of methotrexate versus placebo on lung function as measured by the percent difference in geometric mean ratio of baseline to minimum DLCO and found no statistically significant effect (Figure 1A and Table 2). Within early stages, we found a 3.74 % (95% CI: 3.56% lower through 11.59% higher) higher geometric mean ratio for MTX versus placebo. Within late stages, we found a 1.85% (95% CI:

Comment: There is a multiple comparison issue here. Since both analyses were not statistically significant, the conclusion "we couldn't find a problem no matter what we did" is of some interest, but I do get worried about what you might have done if one analysis had been significant.

8.65% lower to 5.47% higher) lower geometric mean ratio for MTX versus placebo. The null hypothesis (i.e., the geometric mean ratio is greater than or equal to 1.0) that there is no impairment of lung function due to methotrexate treatment versus placebo treatment was not rejected for both early and late stage disease patients (lower one-sided P-value = 0.1603 and 0.6955, respectively).

Figure 1B is a descriptive scatterplot showing that there is no impressive differences in geometric mean ratio of baseline and minimum DLCO according to possible effects of age, stage group and drug treatment on lung function. Table 2 summarizes statistical significance on all the analyses performed.

Comment: See my comments. These were not geometric means you showed!!

Lastly, an exploratory analysis was performed to compare the difference of the mean change in baseline DLCO measurements to minimum DLCO measurements with respect to disease stage and time on study drug (dichotomized at 6 years). Once again, when stratified by years on study drug, the null hypothesis that methotrexate impairs lung function, as measured by DLCO, was not rejected for both early ($\leq$ 6 years on drug and > 6 years on drug: lower one-sided P-value = 0.1712 and 0.3318, respectively) and late stage disease patients ($\leq$ 6 years on drug and > 6 years on drug: lower one-sided P-value = 0.5875 and 0.7727, respectively). For subjects taking the study drug less than or equal to 6 years, the methotrexate group had a 0.8667 (95% CI: 0.9593 lower to 2.6927 higher) higher mean change than the placebo group with early stage disease. In subjects with advanced stage disease, the methotrexate group had a 0.1810 (95% CI: 1.816 lower to 1.454 higher) lower mean change than the placebo group. For subjects taking the study drug longer than 6 years, the methotrexate group had a 0.5686 (95% CI: 1.316 lower to 2.052 higher) higher mean change than the placebo group with early stage disease. In subjects with advanced stage disease, the methotrexate group had a 0.6686 (95% CI: 2.442 lower to 1.105 higher) lower mean change than the placebo group.

Discussion: We assessed the effect of methotrexate treatment on lung function, as measured by the change in DLCO. We were unable to detect a significant effect, using both the mean (sensitive to outliers) and the geometric mean (largely invariant to outliers). We also considered whether duration of exposure would change our conclusions, which it did not. We can not conclude that the doses of methotrexate used for this clinical trial detrimentally affected lung function.

Comment: This seems to hint at a reasonable motivation for looking at both summary measures, but you would still have multiple comparison issues. So I would think you would need to provide a little bit more motivation to ensure how you would have interpreted conflicting results. Perhaps the GM analysis would have been relied upon more to find systematic effects across the dose group and the analysis of means would have been looked at for clinically meaningful results, whether or not stat significant.

Comment: Yes. But can you conclude that MTX did not detrimentally affect lung function? What was your precision?

PBC is primarily a female disease of unknown causes that have both autoimmune and non-immune mediated mechanisms (5). Low-dose methotrexate has been used to treat autoimmune disorders; however acute toxicity and chronic toxic effects can occur with this drug even in low doses. Both PBC and LDMTX have been linked to lower DLCO measurements (6-8). Age and BMI have been implicated as factors that can also contribute to an increased likelihood of lung problems. The distributions of these variables were similar for all compared groups and were likely not a factor in this analysis.

Comment: I would say “a disease in females”

Several limitations of the current analysis in addressing the question of the effect, if any, of methotrexate treatment for PBC on lung toxicity are 1) the use of the maximum decrease in DLCO vs. the change between baseline and final treatment visit, 2) the need to estimate stage-specific versus an overall effect, 3) and the wide confidence intervals for the effect estimates.

Comment: Good.

Use of the maximum decrease in DLCO from baseline as the primary measure of lung function pertained to an interest in the magnitude of maximum effect the treatment may have had on lung function. The alternative outcome, change in DLCO from baseline to final measurement, would address an interest in the chronic, or long term, effect of the treatment.

Since the first task in assessing any potential toxicity is to establish the existence of an association, we chose the outcome most likely to capture larger differences, yet still controlling for within-individual characteristics (i.e., inclusion of baseline measurements in calculating the outcome). If a significant association had been found, we would have been concerned about controlling for variation in the number of DLCO measurements taken per participant. Time on treatment may have acted as a proxy for the number of DLCO measurements, as the study design suggested that each patient was scheduled to be measured annually. We tried to address this issue in the exploratory component of our analysis, using time-on-treatment as the proxy to number of DLCO measurements per participant.

We estimated PBC-stage-group-specific effects because the trial methodology randomized participants to a treatment group according to pre-defined PBC-stage groups. Presumably due to prior knowledge of the effects of PBC on lung function, the trial administrators wished to control for stage-group-specific effects and thereby better tease out any association between the treatment and clinical outcomes. The option to derive and a PBC-stage independent estimate of association could produce more generalizable results, applicable to a broader set of situations.

Finally, since the trial administrators were adjusting the levels methotrexate based upon toxic symptoms other than lung function, it is possible that they adjusted methotrexate dose below a level at which they might have seen a possible effect on the lung function of these patients.

Comment: I believe we call this "appropriate medical treatment". You know. "First do no harm." I would argue that it is entirely appropriate to modify treatments to avoid toxicity, and that we want to find out whether we can successfully do this. Sometimes our treatment will therefore not be efficacious in some patients. Other times, we titrate everyone to the point of efficacy (insulin in diabetics comes to mind).

Bibliographic references:

1. Methotrexate. http://en.wikipedia.org/wiki/Methotrexate
2. Grim J, Chladek J, Martinkova J. Pharmacokinetics and pharmacodynamics of methotrexate in non-neoplastic disease. Clin Parmocokinet 2003; 42 (2): 139-151.
3. Hoekstra M, van Ede AE, Haagsma CJ, van de Laar MAFJ, Huizinga TWJ, Kruijsen MWM, Laan RFJM. Factors associated with toxicity, final dose and efficacy of methotrexate in patients with rheumatoid arthritis. Ann Rheum Dis 2003; 6 (9)2:423-6.
4. Bach N, Thung SN, Schaffner F. The histologic effects of low dose methotrexate therapy for primary biliary cirrhosis. Arch Pathol Lab Med 1998; 122(4): 342-5.
5. Talwalkar JA, Lindor KD. Primary biliary cirrhosis. Lancet 2003; 362(7): 53-61.
6. Chang SW, Ohara N. Pulmonary intravascular phagocytosis in liver disease. Clin Chest Med. 1996; 17(1): 137-50.
7. Fallon MB, Abrams FA. Scenario number one: hepatopulmonary syndrome. Liver Transpl Surg. 1996; 2 (4): 313-9.
8. Mohamed R, Freeman JW, Guest PJ, Davies MK, Neuberger JM. Pulmonary gas exchange abnormalities in liver transplant candidates. Liver Trans 2002; 8(9): 802-8.

Figure Legend
A: Comparison of the percent difference in geometric mean ratio (natural logarithm[baseline DLCO – minimum DLCO]) between placebo and methotrexate groups, by PBC stage.
B: Scatter plot of the Geometric Mean Ratio of baseline DLCO to minimum DLCO versus age at baseline, by treatment and stage groups.

Table 1: PBC-stage-specific descriptive data by randomized treatment groups

	Stage 1-2				Stage 3-4			
	Placebo (female = 62, male=2)		**Methotrexate** (female = 55, male=5)		**Placebo** (female = 61, male=8)		**Methotrexate** (female = 67, male=5)	
	Mean (SD)	Range (min-max)	Mean (SD)	Range (min-max)	Mean (SD)	Range (min-max)	Mean (SD)	Range (min-max)
Age (years)	50.2 (9.2)	25 - 67	50.4 (8.0)	31 - 70	54.0 (7.6)	33 - 66	50.4 (9.2)	31 - 69
Height (cm)	164.0 (8.0)	149.7 - 191.9	164.7 (7.7)	148.2 - 179.0	163.4 (8.2)	149.5 – 185.1	162.9 (8.0)	142.8 - 181.6
Weight (kg)	75.7 (17.1)	47.3 - 149.1	69.5 (14.8)	44.4 - 114.2	71.2 (15.0)	46 - 108	70.6 (14.2)	42.7 - 106.6
BMI (kg/m^2)	28.09 (5.7)	18.2 - 53.4	25.6 (5.0)	18.6 - 43.2	26.6 (5.0)	17.1 – 41.0	26.5 (4.3)	17.6 - 36.7
LIVER STATUS								
Bilirubin (mg/dl)	0.65 (0.30)	1.0 - 1.8	0.66 (0.47)	0.1 - 2.8	0.79 (0.44)	0.1 – 2.3	0.67 (0.42)	0.1 - 2.1
Albumin (g/dl)	4.04 (0.30)	3.2 - 4.6	4.01 (0.37)	3.3 - 5.9	3.97 (0.36)	3.1 – 4.9	3.99 (0.34)	3.0 - 4.7
Cholesterol* (mg/dl)	240.8 (63.9)	128 - 560	240.7 (62.1)	137 - 440	230.9 (53.5)	141 - 378	237.9 (55.1)	140 - 472
Prothrombin time (sec)	11.30 (0.98)	8.9 - 13.5	11.4 (1.20)	8.5 - 14.6	11.44 (1.17)	8.6 – 13.9	11.12 (1.07)	9.0 - 13.3
Platelets (1000/ mm^3)	269.3 (81.5)	81.4 - 561.0	264.5 (94.9)	119 - 619	202.6 (71.6)	8.6 – 13.9	226.1 (79.5)	86 - 546
AlkPhos* (U/I)	213.1 (192.5)	63 - 1127	238.6 (134.1)	73 - 996	274.4 (179.0)	73 - 996	246.4 (155.5)	50 - 930
ALT* (U/I)	44.68 (44.09)	9 - 314	52.35 (42.22)	9 - 175	55.17 (38.86)	13.0 – 252.0	55.93 (40.98)	10 - 199
Splenomegaly (%)	6.25%		6.7%		14.5%		12.5%	
Years since diagnosis of PBC	3.14 (3.04)	0.54 - 13.80	3.17 (3.17)	0.55 - 15.50	4.16 (3.42)	0.51 - 14.40	3.78 (3.67)	0.55 - 17.90
LUNG FUNCTION								
Baseline DLCO	21.00 (4.90)	12.7 - 40.0	21.08 (5.11)	10.2 - 31.7	18.57 (4.87)	11.5 – 41.0	19.63 (4.94)	8.9 - 37.7
Final DLCO	19.42 (5.06)	8.5 - 39.4	18.67 (4.40)	8.9 - 31.0	17.33 (4.68)	9.2 – 31.9	18.71 (5.15)	8.6 – 35.8
Minimum DLCO	17.52 (4.08)	8.7 - 32.7	16.92 (4.19)	7.0 - 30.8	15.91 (4.15)	8.6 – 27.3	16.99 (4.54)	7.6 - 34.6
Years on drug	6.2 (2.0)	0.68 - 8.80	6.0 (2.4)	0.13 - 8.96	5.4 (2.6)	0.11 – 8.66	5.9 (2.6)	0.30 - 8.75

* = serum measures

Abbreviations:
Stage = histologic stage of liver disease
DLCO = diffused lung carbon monoxide (ml/min/mmHg)
AlkPhos= alkaline phosphate
ALT = alanine aminotransferase
progress= disease progression

Comment: Very good (and important) to have included these ranges so we could assess whether any patients had severe toxicity. This would likely just be anecdotes, but important anecdotes. As it was, both groups had similarly low minima.

Comment: Also very important to provide.

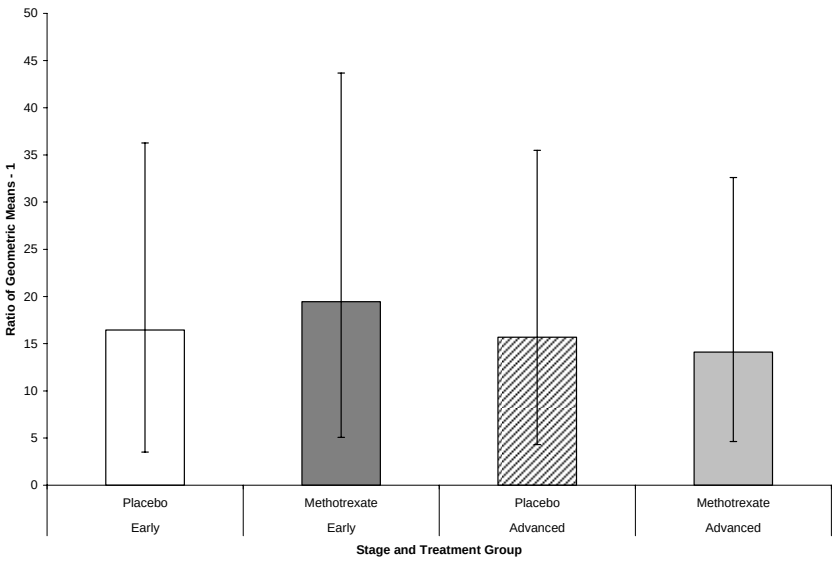
Table 2: Effect of Methotrexate vs. Placebo on Lung Toxicity as Measured by DLCO

	n (placebo, drug)	**Difference in mean change between Baseline DLCO and Minimum DLCO between placebo group and methotexrate group**	**95 % confidence interval**	**p-value**
Early Stage Disease	(62,53)	-0.6502	(-1.943, 0.643)	0.1605
Advanced Stage Disease	(63,66)	0.1594	(-1.119, 1.438)	0.5972
	n (placebo, drug)	**% Difference in Geometric Mean Ratio of (Baseline DLCO:Minimum DLCO) between placebo group and methotrexate group**	**95 % confidence interval**	**p-value**
Early Stage Disease	(62, 53)	-3.74%	(-11.59%, 3.56%)	0.1603
Advanced Stage Disease	(63, 66)	1.85%	(-5.47%, 8.65%)	0.6955

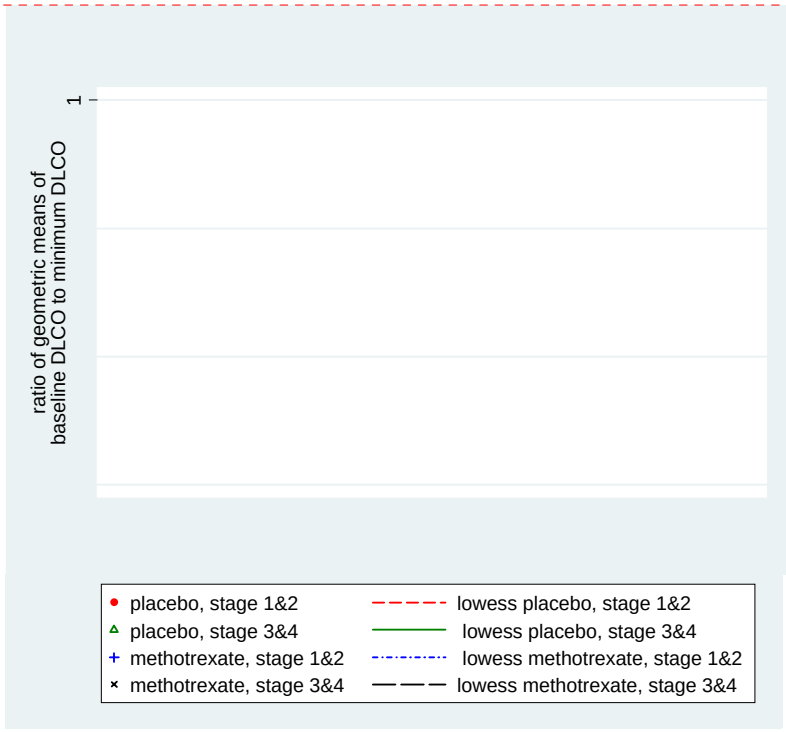
Abbreviations
DLCO = diffused lung carbon monoxide (ml/min/mmHg)

Figure 1

A.



B.



Comment: Your Y axis is labeled incorrectly. These are not geometric means, but are instead the log ratio of measurements for each individual. Ratios of positive measurements have to be positive. Geometric means have to be positive. It is not the geometric mean until you summarize for the group.

Also, a more interesting graph would probably have been by time on study, if you are constrained to two figures. But this is still of interest. I would have looked at ratio minimum to baseline.