Randomized, double blind, prospective trial investigating the efficacy of Methotrexate in induction and maintenance of steroid free remission in ulcerative colitis (MEthotrexate Response In Treatment of UC - Merit-UC)

DATA ANALYSIS PLAN

Version 1.4 5/30/2015

Anastasia Ivanova, PhD
Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Christopher F. Martin, MSPH
Center of Gastrointestinal Biology and Disease
Biostatistics Core
University of North Carolina

Principal Investigator:
Hans H. Herfarth, MD, PhD
Division of Gastroenterology and Hepatology
University of North Carolina

Co-Investigators:
Kim Isaacs, MD, PhD
Division of Gastroenterology and Hepatology
University of North Carolina

James Lewis, MD, MSCE
Center for Clinical Epidemiology and Biostatistics,
University of Pennsylvania

Mark Osterman, MD
Center for Clinical Epidemiology and Biostatistics,
University of Pennsylvania

Bruce Sands, MD, MS
Dr. Henry Janowitz Division of Gastroenterology
The Mount Sinai Medical Center, New York

The statistical analyses of the primary and secondary endpoints will be performed with the statistical support of Joseph Galanko, PhD, who is the biostatistician at the Biostatistics Core of the Center for Gastrointestinal Biology and Disease at the University of North Carolina and Christopher Martin, MSPH, who is the director of the Biostatistics Core of the Center for Gastrointestinal Biology and Disease at the University of North Carolina. SAS software (SAS Institute, Cary NC) will be used to perform all analyses.

1 STATISTICAL ANALYSIS

The primary aim is to test effectiveness of MTX in maintaining clinical remission over 32 weeks of therapy compared to placebo. Analyses of this aim will be conducted at three time points: (1) final analyses at the conclusion of the trial and (2) one interim analysis, one each at the end of years two and three, for review by the DSMB. The interim analysis is planned to allow the DSMB to determine whether treatments are so convincingly different that continuation of the trial would be unethical, and also whether side effects of treatment are too severe to warrant continued therapy given the potential risk:benefit ratio. Details and methods of these final and interim analyses are described below.

In addition, we plan to conduct a single test of open label induction phase remission rates after the first 75 patients have completed this phase of the trial. Accrual of sufficient numbers of patients into the randomized maintenance trial depends on the effectiveness of MTX during the 16-week open label induction phase. Therefore, after the first 75 patients have completed the induction phase, we will assess whether observed response rates are sufficient to meet goals for accrual to randomization in a timely manner, and also to determine whether inclusion/exclusion criteria should be modified to target subgroups of patients most likely to benefit from induction treatment.

1.1 Analysis of induction phase response rates

Each patient will be categorized as a success or failure with respect to response at 16 weeks. Response at week 16 is defined as either clinical remission or clinical response: (1). Clinical remission is a partial Mayo score of no more than 2 points with no individual subscore exceeding 1 point, *and* no use of steroids at week 16; (2) clinical response, defined by a reduction of the partial Mayo score ≥ 2 points and at least 25% with an accompanying decrease in the subscore for rectal bleeding of at least 1 point or an absolute subscore for rectal bleeding of 0 or 1, *and* no use of steroids at week 16 and a partial Mayo score ≤ 5.

Because only patients who achieve response during the induction phase are randomized into the maintenance therapy, successful accrual into the randomized maintenance trial depends on the minimal proportion achieving response during induction. Based on expected numbers of patients available for enrollment into induction phase, and the number required to meet minimum power in the randomized trial, we have determined that induction rates need to a minimum of 30% or more to ensure sufficient accrual into the trial during the time allowed.

Patients who discontinue treatment prior to 16 weeks for any reason will be categorized as failures. As soon as 75 patients have completed the 16-week induction phase, we will compute the point estimate and Clopper-Pearson exact 95% confidence interval of the proportion of successful induction therapy.

Patients who discontinue treatment prior to 16 weeks for any reason will be categorized as failures. To achieve our enrollment goals in the planned time frame, we estimate that the rate of remission/response during induction should be at least 30%. Therefore, we will assess the success rate as soon as the first 75 patients have completed the 16-week induction phase. If the proportion achieving remission does not exceed 30% (23 or more patients achieving this outcome), the sponsor and the DSMB will review the data and decide if the trial should be stopped or modified and continued.

If the proportion achieving remission/response exceeds 30%, we will continue enrollment as planned. However, we will utilize this assessment of remission/response rates to investigate possible revision of inclusion criteria to improve accrual. Specifically, we will compare remission/response rates across strata of patients defined by pre-enrollment therapy: steroid-dependent only, azathioprine/6-MP failure, azathioprine/6-MP intolerance, anti-TNF failure, or anti-TNF loss of response. If one or more of these subgroups has significantly higher remission/response rates, we will explore whether revision of inclusion criteria to preferentially recruit patients most likely to respond to induction therapy will enhance accrual rates into the randomized phase.

We recognize that there is a degree of uncertainty around the remission/response rate yet our decision to stop the trial would be based on a point estimate of response of at least 30%. With 75 evaluable subjects, the confidence interval around the 30% response rate would be 20% to 42%. Thus, with 75 subjects, we will have relatively precise estimates of response/remission rates. Even were the response/remission rate as high as 42%, the medication would be viewed as relatively modestly effective. We would reserve the right to discuss this stopping rule with the DSMB if the proportion of the responders who have achieved remission is high, understanding that greater value may be placed on remission than response by patients and treating physicians.

1.2 Primary analyses

The primary aim of the trial is to test whether primary and secondary outcomes after 32 weeks of maintenance therapy differ between patients randomized to MTX versus those randomized to placebo. As described previously, one interim analysis of primary outcomes is planned for DSMB review during the course of the trial. Both the primary and interim analyses will be based on intent-to-treat therapy. That is, patients will be classified according to the treatment arm to which they were randomized, without respect to compliance, drop out, or other drugs taken. We also plan secondary per-protocol analyses after completion of primary intent-to-treat analyses.

The total period of enrollment is 48 weeks, consisting of 16 weeks open label induction, followed by 32 weeks of randomized maintenance therapy for patients who achieve remission during the induction phase. The follow-up time for planned primary comparisons of treatment-specific treatment survivor functions is restricted to the 32 week maintenance phase. The colonoscopy exam at the end of follow-up occurs at week 48 of enrollment, which is week 32 of follow-up. Thus, we refer to outcome assessment below as occurring at week 32.

1.3 Definition of primary and secondary outcomes

The primary outcome is relapse-free survival, comprised of two components, both of which must be met to be categorized as relapse-free: total week 32 Mayo score not exceeding 2 points, with all individual subscores not exceeding 1 point and relapse free survival defined by a numerical stable Mayo score throughout 32 weeks of maintenance therapy without increase of 3 or more points in the partial Mayo clinic score (excluding sigmoidoscopy) compared to the partial Mayo score of the individual patient at

randomization at week 16 ***and*** no steroid use or other immunosuppressive medication (anti-TNF agents, thiopurines, cyclosporine, tacrolimus) to control disease activity throughout the 32 week maintenance period. For most patients, time-to-failure or censoring will occur at the final visit (32 weeks) when the Mayo score is determined, however for patients who relapse due to initiation of steroid treatment or the need for other prohibited medications to control disease activity before 32 weeks, time-to-failure will be the week in which steroid therapy was started.

Six secondary outcomes are also defined:
(1) mucosal healing, defined as an absolute subscore for endoscopy no more than 1 at week 32,
(2) relapse of disease, defined as an increase of 3 or more points in the partial Mayo clinic score (excluding sigmoidoscopy) with an absolute clinical Mayo score ≥ 4, or re-treatment with steroids during maintenance. Any patients for whom outcomes cannot be assessed for any reason will be classified as failures (relapses) with the event time assigned as the date that the patient was withdrawn from the study.
(3) steroid free clinical remission as defined as a Mayo score of ≤ 2 points with no individual subscore exceeding 1 point or steroid free clinical response defined as a reduction from baseline in the clinical Mayo score of ≥ 2 points and at least 25%, with an accompanying decrease in the rectal bleeding subscore of ≥ 1 point or an absolute rectal bleeding subscore of 0-1 point and a clinical Mayo score ≤5 and stool calprotectin levels <250 mcg/g stool at week 16 of the induction period in the subgroup of patients with calprotectin >250mcg/g stool at screening.
(4) steroid free clinical remission as a Mayo score of ≤ 2 points with no individual subscore exceeding 1 point and stool calprotectin levels of ≤ 50mcg at week 16 of the induction period in the subgroup of patients with calprotectin >250mcg/g stool at screening.
(5) steroid free clinical remission as defined as a Mayo score of ≤ 2 points with no individual subscore exceeding 1 point or steroid free clinical response defined as a reduction from baseline in the clinical Mayo score of ≥ 2 points and at least 25%, with an accompanying decrease in the rectal bleeding subscore of ≥ 1 point or an absolute rectal bleeding subscore of 0-1 point and a clinical Mayo score ≤5 and stool calprotectin levels <250 mcg/g stool at week 32 of the maintenance period in the subgroup of patients with calprotectin >250mcg/g stool at screening.
(6) steroid free clinical remission as a Mayo score of ≤ 2 points with no individual subscore exceeding 1 point and stool calprotectin levels of ≤ 50mcg at week 32 of the maintenance period in the subgroup of patients with calprotectin >250mcg/g stool at screening.

Patient characteristics of interest

For assessment of randomization success, and for assessment of possible confounding of treatment effects with respect to primary and secondary outcomes, we will use the following list of patient factors: age, sex, previous medication use (5-ASA, azathioprine/-MP, anti-TNF agents), extent of disease (proctitis, left-sided disease, or pan colitis), steroid dependent disease at the time of enrollment, entry Mayo disease activity score, calprotectin at inclusion and smoking status.

2 ANALYSIS PLAN

2.1 Descriptive Statistics

We will compute the treatment-arm-specific proportions and exact 95% confidence intervals of patients achieving each primary and secondary outcome of interest.

We will describe and summarize the bivariate distributions of patient characteristics (age, sex, duration of disease, previous therapies, smoking status, disease extent, disease severity, calprotectin) within treatment arms. We will compute medians, interquartile ranges, means and standard errors of continuous variables. Categorical variables will be summarized using proportions in each level. We will compare distributions across treatment arms as follows: (1) for categorical variables, Fisher's exact chi-square tests of association in 2-by-X tables, or (2) for continuous variables, either Student's t-tests of differences in means (for normally distributed variables), or Wilcoxon rank-sum tests for non-normally distributed variables. Because any imbalance in the two randomized groups is by definition a chance occurrence, these descriptive analyses will be used to highlight potential areas of substantial unbalance between the study arms and to inform adjusted analyses of treatment effect.

2.2 Analysis of primary and secondary outcomes

All statistical tests for the primary and secondary outcomes will be 2-sided using alpha=0.05, except as described below with regard to the pre-specified interim analyses of the primary outcome for the DSMB.

The primary aim is to test whether the relapse-free remission in patients randomized to MTX maintenance therapy following MTX-induced remission is superior to that in similar patients randomized to placebo maintenance using the intention-to treat (ITT) principle. This primary outcome will be determined for each patient, along with the time-to-failure or, for patients who do not relapse, time-to-censoring of 32 weeks. These data will be used to compute survivor functions within treatment groups. We will then compare treatment-group-specific survivor functions using the non-parametric logrank test. This tests the null hypothesis

$$H_0 : S_1(t) = S_2(t) \text{ for all } t \leq 32$$

where $S_j\ (t)$ is the survival function in group j at time t. For this test, we will use alpha=.05 to account for the two planned interim analyses (see below).

We fully expect randomization to result in balanced distribution of potential confounders across treatment groups. Further, as randomization will be by permuted blocks within site, we also do not expect confounding by study site. We will nonetheless assess possible confounding by any variables found to differ across treatment groups, and, also by site if treatment distribution differs across sites.

We will also assess heterogeneity of effect by covariates of interest, and by site, using Mantel-Haenzel methods to compute and compare crude and summary incidence rate ratios, with accompanying tests for heterogeneity. If any heterogeneity is observed, we will report stratum-specific treatment effects.

We will repeat these analyses for the secondary outcomes.

2.3 Exploratory analyses of factors associated with remission at week 32 of Maintenance Period

If a treatment effect on relapse-free survival is observed, we will also conduct exploratory analyses of patient characteristics predicting remission versus failure among MTX-treated patients. We will use bivariate methods to assess associations of each patient factor of interest (described above) with relapse-free remission at week 32, followed by multivariable logistic regression modeling having relapse-free remission as the dependent variable.

2.4 Exploratory analyses of factors associated with successful induction therapy

We will conduct exploratory analyses of patient factors predicting achievement of remission and clinical response following 16 week of open label MTX treatment (Induction Period). The methods used will parallel those just described used for exploratory analyses of factors associated with relapse-free remission at week 32 of the Maintenance Period, except the outcomes of interest will be remission and clinical response at 16 weeks induction therapy.

3 DSMB INTERIM ANALYSES AND EARLY STOPPING RULES FOR INTERIM ANALYSES

Interim analyses of treatment effect will be conducted when 50% of patients expected to enrol in the maintenance phase will have completed the study (which will be approx. at the end of year 2). Results of the interim analyses will only be provided only to the Data Safety Monitoring Board, not to the investigators. This interim analysis will use the methods described above using logrank tests to detect treatment group differences in survivor functions with respect to the primary outcome, relapse-free maintenance of remission. The absolute and relative difference in survivor functions, if any, will also be summarized. Accompanying these results will be summaries of the distributions of covariates of interest by treatment group, as explained in Descriptive Analyses. In addition, summaries of adverse events will be calculated by treatment group with accompanying chi-square tests to indicate statistically different incidence of adverse events by treatment group.

By definition, interim analyses require multiple analyses of the data. To avoid inflating the overall type I error rate for the principal analysis of efficacy, we will use the O'Brien-Fleming method of adjusting type I error rate for planned interim analyses. This design utilizes very low type I error rates early in the trial, to preserve a level at or near the traditional 0.05 for the final analysis. In addition, an inflating type I error rate is used for sequential interim analyses, increasing the probability of stopping the trial in later analyses when more data are available. As this trial accrues patients for over three years, there will be a total of two analyses (one interim analyses and one final analysis). The suggested stopping boundaries for the analysis of primary treatment effect are shown in table 1.

Table 1. Proposed stopping boundaries

Analysis	Z	P
1	2.800	.0050
2	1.977	.0480

Since the DSMB might not meet when exactly half of the patients completed the study, it might be preferable to use the error spending function $f(t) = \min\{2-2\Phi(0.025/\sqrt{t}),0.05\}$, where Φ is the cumulative distribution function of standard normal and t is the information fraction observed in the trial at the time of interim analysis.

Any decision to stop the trial early due reasons of efficacy or safety will be made soley by the Data Safety Monitoring Board; these a priori stopping boundaries are merely intended to serve as a guideline to those decisions.

4 REFERENCES

Armitage P, Berry G. Statistical Methods in Medical Research. 2nd ed Oxford: Blackwell Scientific Publications; 1987.

Chow, S., Shao, J., and Wang, H. (2003), Sample Size Calculations in Clinical Research, Boca Raton, FL: CRC Press.

Clopper, C. J., and Pearson, E. S. (1934), "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," Biometrika 26, 404–413.

Collett, D. 2003a. Modelling Binary Data. 2nd ed. London: Chapman & Hall/CRC.

Freedman, L. S. 1982. Tables of the number of patients required in clinical trials using the logrank test. Statistics in Medicine 1: 121–129.

Landis J, Sharp T, Kuritz S, Koch G. Mantel-Haenszel Methods. In: Armitage P, ed. Encyclopedia of Biostatistics. London: Wiley & Sons; 1997.

Machin, D., M. J. Campbell, S. B. Tan, and S. H. Tan. 2009. Sample Size Tables for Clinical Studies. 3rd ed. Chichester, UK: Wiley–Blackwell.

Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. American Journal of Epidemiology. 1989;129:125-37.

Schoenfeld, D. A. 1981. The asymptotic properties of nonparametric tests for comparing survival distributions. Biometrika 68: 316–319.