Supplementary Online Content 2

FAST Uveitis Trial original statistical analysis plan, final statistical analysis plan, and summary of changes

FAST Uveitis Trial

Statistical Analysis Plan

Confidential

Prepared by:
FI Proctor Foundation
Data Coordinating Center
[redacted]
Email: [redacted]

TABLE OF CONTENTS

1 Introduction

This document (Statistical Analysis Plan, SAP) describes the planned analysis and reporting for **the FAST (First-line Antimetabolites as Steroid-Sparing Treatment) Uveitis Trial**, University of California, San Francisco. It includes specifications for the statistical analyses and tables to be prepared for the final Clinical Study Report.

The proposed FAST Uveitis Trial is a block randomized, observer-masked, comparative effectiveness, Phase III clinical trial to compare the efficacy of mycophenolate mofetil (CAS 128794-94-5) to methotrexate (CAS 59-05-2) for the treatment of non-infectious uveitis requiring steroid-sparing therapy.

The content of this Statistical Analysis Plan meets the requirements stated by the US Food and Drug Administration and conforms to the American Statistical Association's Ethical Guidelines.[1, 2]

The following documents were reviewed in preparation of this Statistical Analysis Plan:

- FAST Uveitis Trial Manual of Operations
- ICH Guidance on Statistical Principles for Clinical Trials[2]
- Statistical Analysis Plan (prepared by C. McCulloch), Steroids for Corneal Ulcers Trial (T. Lietman and N. Acharya, principal investigators)
- Statistical Analysis Plan (prepared by T. Porco), Mycotic Ulcer Treatment Trial (T. Lietman and N. Acharya, principal investigators)

The planned analyses described in this Statistical Analysis Plan will be included in future manuscripts. Note, however, that exploratory analyses not necessarily identified in this Statistical Analysis Plan may be performed to support the analysis. All post-hoc or unplanned analyses which have not been delineated in this Statistical Analysis Plan will be clearly documented as such in the final Clinical Study Report, manuscripts, or any other document or submission.

The final SAP is subject to the approval of an appointed Data and Safety Monitoring Committee.

The following individuals contributed to this document: N. Acharya, T. Lietman, N. Nardone, and T. Porco.

2 Investigational Plan

2.1 Study Design

The proposed study is an international, multicenter, block randomized, observer-masked, comparative effectiveness clinical trial to determine which treatment, methotrexate or mycophenolate mofetil, is more effective as first-line corticosteroid-sparing treatment for patients with non-infectious intermediate, posterior and panuveitis requiring corticosteroid-sparing therapy.

Full details which specify the definition of treatment success are given in the FAST Manual of Operations.

2.2 Study Population

Eligible volunteers diagnosed with non-infectious uveitis who have given informed consent will be enrolled in this trial. Specific eligibility and exclusion criteria are given in the FAST Manual of Operations. The proposed study schedule is listed in the FAST Manual of Operations.

2.3 Specific Aims

2.3.1 Specific Aim 1

Primary Objective. The primary objective of the study is **to establish which immunosuppressive treatment, methotrexate or mycophenolate mofetil, results in a higher rate of corticosteroid-sparing treatment success, on an intent to treat basis.** Specifically, we will compare the fraction of subjects who achieve treatment success at six months (as defined in the Manual of Operations Section 2.5.1) between the two groups.

Primary Outcome. The primary outcome for Specific Aim 1 will be the difference in the proportion of patients assigned to mycophenolate mofetil vs. methotrexate who achieve treatment success (as defined in the Manual of Operations Section 2.5.1).

Patients who experience success at 6 months with the drug to which they were originally randomized (in Specific Aim 1) will continue on the same drug for an additional 6 months. This will be called Phase I (6-12 months). Patients will then be seen every 3 months (and will be examined at 9 months and at 12 months), until success at 12 months or treatment failure at any time. Patients who fail treatment before 12 months with the initial drug will be removed from the study and treated according to best medical judgment.

Secondary Objectives

- To determine whether patients exhibit a difference in time to control of inflammation within the first six months.
- To determine whether patients exhibit a difference in time to corticosteroid sparing control of inflammation within the first six months.
- To evaluate the proportion of patients achieving treatment success at 5 months and sustaining, for at least 28 days, to 6 months.
- To evaluate a difference in control in inflammation in the posterior/pan anatomic locations only, assessed at by six months.
- To evaluate a difference in control in inflammation in the interior and anterior/intermediate anatomic locations only, assessed at by six months.
- To determine whether there is a change in best spectacle-corrected visual acuity at six months.
- To determine whether patients treated exhibit a difference in health related quality of life at six months.
- To determine if there are differences in discontinuing treatment due to each of the following reasons: safety, intolerability or lack of efficacy at six months.
- To determine whether patients exhibit a difference in the proportion of patients having macular edema at 6 months
- To determine whether patients exhibit a difference in macular thickness at 6 months
- To determine whether patients exhibit a change in vitreous haze, assessed clinically by the NEI and Davis scales at 6 months
- To determine whether patients exhibit a change in vitreous haze, assessed by the photographic grading of haze by the NEI and Davis scales at 6 months.
- To determine the proportion of patients discontinuing due to intolerability at six months.
- To determine the rate of adverse events experienced at six months.
- To determine the proportion of patients discontinuing due to serious adverse events at six months.
- To determine whether patients exhibit a change in quality of life at six months.
- Tabulate the occurrence of dose reduction used in immunosuppressive treatment (see Manual of Operations Section 3.1 for dose reduction guidelines).
- To determine efficacy of treatment in Vogt-Koyanagi-Harada (VKH) patients at six months.

All the above analyses will be examined at the end of Phase I (6-12 months), in addition to the following:

- To determine whether patients exhibit a difference in the probability of controlling inflammation with complete discontinuation of steroids at twelve months in Phase I.

2.3.2 Specific Aim 2

Primary Objective. The primary objective of this aim is to evaluate the clinical efficacy of switching agents as rescue therapy after initial treatment failure.

Patients who experience treatment failure (as defined in the Manual of Operations) with the drug to which they were originally randomized in Specific Aim 1 will discontinue the current treatment and be administered rescue therapy with the second drug (in a masked fashion). This will be called Phase II. Treatment failure with the first drug is defined as the inability to continue taking the drug to which the patient has been randomized, either due to intolerability, safety concerns, or lack of efficacy.

Upon declaration of treatment failure, the patient will be automatically screened for Aim 2. If eligibility criteria are met, the second treatment will be administered and data will be collected for the Phase II baseline visit. Patients will then be seen every 4 weeks until 6 months or until treatment failure with the second drug. Treatment failure and success will be defined as in Aim 1. Patients who fail treatment before 6 months with the second drug will be removed from the study and treated according to best medical judgement.

Primary Outcome. The primary outcome is the fraction achieving treatment success at 6 months after starting Phase II. Treatment success is defined as in Aim 1 and described in the Manual of Operations Section 2.5.1.

Secondary Objectives. All of the secondary 6 month objectives listed for Aim 1 will be examined for Aim 2.

2.4 Randomization

2.4.1 Stratification between sites

Patients will be recruited from four sites: Aravind-Madurai, Aravind-Coimbatore, Casey Eye Institute-OHSU, and Proctor Foundation-UCSF (see Manual of Operations Section 2.2 for details). Patients will be randomized to two treatments (arms): methotrexate (X) or mycophenolate mofetil (Y). The treatment protocols are specified in the FAST Uveitis Trial Manual of Operations.

Within each site, assignments will be conducted using a block randomization scheme with randomly varying block sizes.

2.4.2 Randomization list

Lists of sequential randomization assignments will be prepared for each site. The randomization lists consist of a unique identifier for each patient, together with the assignments to treatment arms. The assignment of patient ID numbers and randomization is thus performed on enrollment.

The randomization lists for Aravind sites will be prepared by the Proctor site (see Section 10.3) and sent to the Chief Microbiologist, Dr. Lalitha Prajna, to be used only in case emergency unmasking is needed for patient safety. Similarly, the randomization lists will be sent to Dr. Christine Flaxel, a retinal specialist at Casey who will serve as the local contact in case emergency unmasking is needed. Dr. Stephen McLeod, Chairman Department of Ophthalmology, University of California San Francisco will serve as emergency contact for Proctor Foundation in San Francisco.

They will also be sent to hospital/clinic staff who are responsible for telling the study coordinators the treatment assignment for each patient after the patient is enrolled and the study ID has been assigned. These individuals are Ms. Sally Tsang, Clinic Manager Proctor Foundation, University of California, San Francisco, Mrs. T.S Chandravati, Aravind Eye Hospital Madurai, India, Mrs. A. Maglin Brinda Mary, Aravind Eye Hospital Coimbatore, India, and Dr. Jennifer Petrolatti, Casey Eye Institute, Oregon. Dr. Natalie Nardone will hold the randomization lists for Casey Eye Institute and Aravind Eye Hospitals. At these sites she will verify patient treatment assignment as a quality assessment.

A backup copy of the full randomization list for all four sites will be maintained by Tom Lietman, MD, (hereinafter TL). This list will be maintained as a hard copy stored in a locked file cabinet at the UCSF site, and to be used only in case emergency unmasking is needed for patient safety.

Distribution of the randomization list to Aravind and Casey will be accomplished using the University of California, San Francisco's encrypted email provision. Email is encrypted using the Advanced Encryption Standard (NIST FIPS 197) whenever the first four characters of the subject line are `PHI:` The sender is notified when the recipient receives a secure email; the recipient receives a notification of a secure email and can view it using the UCSF Secure Messenger website. We have successfully used this method in previous clinical trials (Steroids for Corneal Ulcers Trial, Mycotic Ulcer Treatment Trials). The randomization lists will each contain more randomization assignments than needed. Successively recruited patients will receive sequential assignments from the list. The long list provides a measure of added safety in case one of these sites recruits far more patients than expected relative to the other site.

As discussed below, the randomization lists will be provided as Excel® worksheets. No technical knowledge will be required to use these lists.

2.4.3 Block randomization

We will utilize a permuted block randomization scheme with a randomly varying block size (within each study site) to protect the integrity of the assignment masking.[3] Any particular block size will be unknown to the study investigators. We will choose randomly varying block sizes, picking a block of size 4 with probability 2/3 and a block of size 6 with probability 1/3.

Individuals have a higher probability of being in a block of size 6 because the blocks are larger. Many other choices would serve equally well. Given the block size, a random permutation of assignment orders will be generated.

2.4.4 Unique patient identifiers

Unique patient identifiers will be generated as follows. The first character will be a number: "1" for Proctor Foundation, "2" for Madurai, India, "3" for Coimbatore, India, and "4" for Casey Eye Institute. The next character is a checksum character, which will be a single letter. The last three characters will be sequential digits beginning at 001. An example identifier is 4J101; all identifiers have exactly five characters, and no other Aravind/Proctor study uses this format.

2.4.5 Random number generation

The choice of a random number seed determines the specific sequence of random numbers that will be produced by the random number generator. Once the seed is determined, the randomization assignments for all sites are determined. Details are given in the Appendix.

2.4.6 Provision of randomization list

Everyone to whom the randomization list should be provided (for each of the four sites) will receive it in the following format: a Microsoft Excel® spreadsheet containing the following columns: (1) the unique study identifier assigned to the patient (see Section 2.4.4), (2) an empty field into which the date of randomization may be entered (relevant only for the hospital/clinic staff holding the randomization lists), (3) the study drug assignment, written out in full as `Mycophenolate` or `Methotrexate`. Every other line will be tinted pale blue in the spreadsheet, to minimize errors in reading across. As discussed in Section 2.4.2, these lists will be treated confidentially.

2.4.7 Quality assurance

Three quality assurance steps for the randomization list preparation are conducted. First, the software will have been tested during previous studies (MUTT). Second, the software that generates the assignments verifies approximate balance of subjects in each group before writing the Microsoft Excel® files. Each file will contain the study site as the first line. Finally, the output files will be visually inspected. The software and procedures have already been developed and successfully used in previous studies.

2.4.8 Summary of disposition of randomization list

The following individuals will receive a copy of the randomization list:

Emergency Contact Personnel

- Dr. Lalitha Prajna *, Chief Microbiologist, Aravind Eye Hospital
- Dr. Christine Flaxel*, Retina specialist, Casey Eye Institute, Oregon
- Dr. Stephen McLeod*, Chairman Department of Ophthalmology, University of

California San Francisco

*Emergency contact persons who will consult the list only in case of an emergency in which unmasking is necessary for patient safety and Dr. Thomas Lietman on the DCC cannot be reached.

Data Coordinating Center (DCC) Personnel

- Dr. Thomas Lietman, Professor of Ophthalmology and Epidemiology, University of California San Francisco, Proctor Foundation
- Dr. Travis Porco, Principal statistician, Proctor Foundation
- Dr. Wayne Enanoria, Research epidemologist, Statistical programmer/analyst, Proctor Foundation

Clinic/Hospital Staff

- Ms. Sally Tsang, Clinic Manager Proctor Foundation, University of California, San Francisco
- Dr. Natalie Nardone, Study Coordinator, University of California, San Francisco
- Mrs. A. Maglin Brinda Mary, Aravind Eye Hospital Coimbatore, India
- Mrs. T.S Chandravati, Aravind Eye Hospital Madurai, India
- Dr. Jennifer Petrolatti, Casey Eye Institute, Oregon

2.5 Masking

The clinical examiners, refractionists, OCT technicians, fundus photographers and fundus graders will be masked to the treatment assignment. Note that only the individuals listed in Section 2.4.8 will have copies of the randomization list. Full details of procedures to maintain masking as well as for potential unmasking in the event it becomes necessary for safety reasons are provided in the Manual of Operations. Principal Investigator N. Acharya is masked.

3 Statistical Considerations

3.1 Baseline characteristics

At baseline, each eye (1) may be fully able to be assessed, (2) it may be possible assess part of the eye, but not be possible to assess the entire eye, or (3) it may not be possible to assess any of the eye. For each eye for which some assessment is possible, either (1) the eye showns no signs of uveitis, or the eye may show some signs of uveitis, but fail to meet the severity criteria (1+ anterior chamber cells, vitreous haze or no active retinal/choroidal lesions, as defined in the Manual of Operations), or (2) the eye meets the severity criteria as defined in the manual of operations. Some patients are monocular at baseline, one eye being either absent, or exhibiting such disease as to preclude the possibility of ever assessing the eye (i.e. phthisis).

For this trial, we summarize the above possibilities as follows. Each eye (OD or OS) may be

classified into one of the following types at baseline:

A. Eye fully assessible, does not meet the severity criteria as defined in the MOP
B. Eye partially assessible, does not meet the severity uveitis criteria in the assessible region
C. Eye fully assessible, meets severity criteria
D. Eye partially assessible, meets severity criteria in assessible region
E. Eye absent or too diseased to ever assess

Patients, not eyes, are the unit of assignment and of randomization. Thus, there are twenty-five possible types of patients. A patient is required to have at least one eye which meets severity criteria for uveitis, and which can be completely assessed. Eligibility is summarized in the following table; cells indicate the possibility of enrollment for a patient whose right eye classification corresponds to the row and whose left eye classification corresponds to the column (A-E being defined in the previous paragraph).

	OS: A	OS: B	OS: C	OS: D	OS: E
OD: A	Not eligible	Not eligible	**Eligible**	Not eligible	Not eligible
OD: B	Not eligible	Not eligible	**Eligible**	Not eligible	Not eligible
OD: C	**Eligible**	**Eligible**	**Eligible**	**Eligible**	**Eligible**
OD: D	Not eligible	Not eligible	**Eligible**	Not eligible	Not eligible
OD: E	Not eligible	Not eligible	**Eligible**	Not eligible	Not eligible

Assessment and follow-up depends on the status of the eye. Eyes classified as type E above are recorded as such at baseline, and never provide eye outcome related data. Because (a) inability to assess parts of the eye could be related to the progression of disease, but (b) inability to assess in the absence of signs of disease cannot be considered evidence of treatment failure, we use the following table to summarize how success at six months will be scored. In this table, the row corresponds to the status of an eye at baseline, and the column to the status of the eye considering the primary outcome of success at six months.

	Month 6: A	Month 6: B	Month 6: C	Month 6: D	Month 6: E
Baseline: A	Success	See below**	Fail	Fail	Fail*
Baseline: B	Success	See below**	Fail	Fail	Fail*
Baseline: C	Success	See below**	Fail	Fail	Fail*
Baseline: D	Success	See below**	Fail	Fail	Fail*
Baseline: E	NA	NA	NA	NA	NA

Specifically, note that an eye which is fully assessible at six months and which does not meet the specific criteria for failure of control is always considered a success. Eyes which are fully or partially assessible and which meet any of the criteria for failure are always considered to have failed. However, eyes which are only partially assessible but which meet no criteria in the assessible region may be scored successes or failures depending on their baseline status (see next paragraphs below). Eyes which were present at baseline but which are missing at the end of the study are listed as Fail* in the table; we propose to consider such eyes to have failed unless a specific reason demonstrates that the loss of the eye was completely unrelated to the presence or progression of disease.

The primary analysis is at the patient level. Both eyes must meet the success criteria for the patient to be considered a success.

****Incompletely assessible eyes.** Uveitis assessment for this purpose is based on (i) assessment of anterior cells, (ii) vitreous haze, and (iii) retinal or choroidal lesions. In the pilot study, (iv) assessment of vitreous cells was also used. We will have longitudinal measurements of inflammation according to the following schedule: anterior chamber cells, vitreous haze and active retinal/choroial lesions will be measured at Baseline, Week 2, Month 1 and every subsequent 4 weeks until the 6 month assessment (Phase I or Phase II) or 9 and 12 month assessment (Phase I 6-12 months).

Each of these (including the binary assessment of the presence of retinal or choroidal lesions) may be considered an ordinal variable, with relevant threshold values for each (used in determining eligibility for enrollment, or success in therapy).

Scoring of incompletely assessible eyes is governed by the following guiding principles:

1. In some patients, the front of the eye may be assessible, but the back of the eye cannot be examined and assessed clinically (even though the patient can still see out of the eye).
2. Worsening of uveitis may render it harder to assess the back of the eye, so that information cannot be considered missing at random in general.
3. Many uveitis patients have at least one eye which cannot be fully assessed, because of the progression of the disease itself. Excluding such patients or eyes completely is undesirable.
4. Treatment of uveitis will not reverse the damage which makes it difficult to assess all parts of the eye.
5. Worsening of cataracts may also cause an eye to become incompletely assessible, so that a change in assessibility status does not always indicate a worsening of uveitis or a failure of uveitis treatment.

We chose the following simple, but conservative, approach to scoring such eyes. For an incompletely assessible field (anterior cells, vitreous haze, or presence of retinal or choroidal lesions) at any time, the worst value seen until that time will be assigned for the unavailable measurement. Thus, a decreasing ability to assess regions of the eye—in the absence of evidence of inflammation or uveitis criteria—does not imply failure of therapy. Decreasing ability to assess eyes which had signs of uveitis will imply failure of therapy. It is understood that this procedure will misclassify some events such as: (i) an eye which had vitreous haze or a retinal or choroidal lesion at baseline, which resolved over the course of the six months, and for which a progressing cataract rendered the posterior of the eye impossible to assess, will be scored as a failure, or (ii) an eye for which the posterior region had no inflammation at baseline, which then became impossible to assess, and then which develops posterior inflammation which cannot be seen, will be scored as a success. We believe such misclassifications will be infrequent.

Selected secondary outcomes, including vision, macular edema, time to control of inflammation, will be analyzed at the eye level. All eyes that meet inclusion criteria of inflammation at baseline will be included in this analysis. Linear or generalized linear mixed modeling will be conducted (see below for details).

The following is a brief summary of general guiding principles.

- For the primary outcome, if any portion of the eye cannot be assessed at baseline, and it *still cannot be assessed at Visit 6 or Visit 12*, if all other markers of success are met, this portion of the eye would be considered to have had successful therapy.
- For the primary outcome, if any portion of the cannot be assessed *by Visit 6 or 12* and this same portion of the eye was *completely assessable* at baseline, if all other markers of success are met, then the last worst observation for this eye would be carried forward and used at the assessment of this eye portion.
- For the primary outcome, if an eye *becomes missing by Visit 6 or 12,* and it is related to uveitis (regardless of its disease status at baseline) if all other markers of success are met, this patient should be considered a failure.

3.1.1 Demographics and Patient History

All demographic and history variables (in particular, age, gender, occupation, and ethnicity/national origin) determined at enrollment will be summarized by counts and percentages tabulated by treatment assignment.

3.1.2 Prior and concurrent medication

We will present the oral and topical corticosteroid doses at presentation (specifically, the current daily dose at baseline) and other medications by randomization arm and study site.

3.1.3 Baseline comorbidities and history

Clinical variables at baseline (in particular, anatomical site) will be presented by gender, age, and study site. We will also tabulate the presence of associated systemic disease at baseline. Anatomical site will be classified at the patient level as site of most serious involvement. For example, if a patient has anterior inflammation in the right eye and panuveitis in the left, they would be classified as a panuveitis patient.

3.1.4 Compliance

Compliance is assessed through patient self-report and regular pill counts by study coordinators at each visit when patients bring in their medications.

3.2 Analysis

3.2.1 Summary of Principal Outcome Variables and Regression Variables

Variables

- Primary outcome: Patient treatment success by six months (see MOP, Section 2.6)
- Patient treatment success at twelve months (Phase I)
- Successful control of inflammation in both eyes by twelve months, with complete discontinuation of corticosteroids

- Best spectacle-corrected visual acuity, at baseline and at the time of failure or six months (two observations per patient)
- Time to corticosteroid sparing control of inflammation (6 months and 12 months)
- Change in health related quality of life subscores (PCS and MCS) from SF-36 and Vision Related Quality of Life from NEI-VFQ-25 and IND-VFQ at six months and twelve months
- Reason for discontinuation of therapy (if applicable) at six months and twelve months
- Macular thickness at baseline, and at six months and twelve months
- Presence of macular edema at six months and twelve months
- Vitreous haze assessed clinically by the NEI and Davis scales at baseline, six months, and twelve months
- Vitreous haze as assessed by the photographic grading of haze by the NEI and Davis scales at baseline six months and twelve months
- The proportion of patients discontinuing due to serious adverse events at six months and twelve months
- Tabulate the occurrence of dose redution used in immunosuppressive treatment.
- Treatment efficacy of VKH patients at six months and twelve months

Note that the presence of cataracts renders assessment of vitreous haze more difficult. Vitreous haze measurements in the presence of certain cataracts will be considered less reliable, and this will be considered in statistical modeling. Analyses will be repeated for differing assumptions about this bias. A maximum likelihood latent variable model will be considered, in which a true underlying vitreous haze level predicts an observed value. The observation model will include a higher probability of yielding a large observed value in the presence of a cataract.

Major independent variable of interest

- Treatment assignment (methotrexate or mycophenolate mofetil)

Additional regression variables used in selected analyses

- Anatomic location (coded dichotomously as either intermediate (code 0) or as being either posterior uveitis or panuveitis (code 1))
- Country
- Study site
- Gender
- Age
- Baseline quality of life (health and vision related)
- Baseline best spectacle-corrected visual acuity, vitreous haze, macular thickness

Inclusion of Data

- Data will be included for all outcomes within the window peroid of -2 weeks to +4 weeks around the 6 Month Visit date for Phase I (0-6 months).
- Data will be inlcuded for all outcomes within the window peroid of -2 weks to +4

weeks around the 12 Month Visit date for Phase I (6-12 months).

- Data will be included for all outcomes within the window period of -2 weeks to +4 weeks around the 6 Month Visit for Phase II (0-6 months).

3.2.2 Specific Aim 1

Primary Analysis.

The primary analysis will be a logistic regression model, predicting treatment success at 6 months based on treatment arm. Study country and site within country will be included as fixed effects; we wish to aggregate sites within countries, and countries within treatments provided we find no evidence of heterogeneity of sites or countries.

Specifically, the pre-specified primary analysis will be performed as follows. We denote the assignment group of patient i, i=1,...,N (where N is the number of subjects) by X_i^1, which equals 0 when the patient is in the methotrexate group and 1 when the patient is in the mycophenolate mofetil group. The outcome variable is Y_i, which is 1 if treatment success of patient i is achieved by six months, and 0 otherwise. The variable is missing if the patient is lost to follow-up or drops out of the study for reasons other than discontinuation due to intolerance or adverse events; if the patient discontinues the medication due to intolerance or adverse events such as abnormal laboratory findings, the value is 0.

The primary analysis is a logistic regression with the following statistical predictors: treatment arm and study site (within country). For this equation, we denote the study site (geographic site) for patient i using the following entirely standard encoding. The variables Z_{i1} indicates whether the patient was from study site 1 (i.e., is 0 if the patient was not from study site 1 and equals 1 if the patient is from study site 1). Similarly, Z_{i2}, Z_{i3}, and Z_{i4} indicate whether or not the patient is from study sites 2, 3, or 4 respectively; for each patient, one and only one of these variables equals 1, all others equaling 0. The regression coefficients for the study sites are denoted bji; the equation below gives site 1 as the baseline (this choice does not affect the conclusion, and we may choose to report the results in a different way). In particular, we propose to fit the following model:

$$\text{logit}(Y_i) = \beta_0 + \beta_1 X_i^1 + b_{2i} Z_{2i} + b_{3i} Z_{3i} + b_{4i} Z_{4i} \qquad (1)$$

(written for the case of two US and two Indian sites). The null hypothesis is $\beta_1 = 0$, which will be tested using a likelihood ratio test with one degree of freedom. Other exploratory models are discussed below. We will also fit the following models: (a) a model including drug, site, and drug × site interaction, (b) a model including only drug and site, (c) a model including only drug and country, and (d) a model with drug, country, drug-country interaction, and site within country. Provided there is no evidence of treatment × site interaction or treatment × country interaction, we will report pooled treatment effects and confidence intervals. In the event evidence suggests a difference between treatment sites, we will report treatment effects by site, and repeat the analysis excluding particular sites. Similarly, evidence of a treatment × country

interaction (as determined by the Wald T-test for the statistical significance of the test $b_{12} = 0$ in Equation (1)) will lead us to report treatment effects and confidence intervals by country.

Simulations suggest that use of a model containing interaction terms between site or country and treatment for the primary analysis is undesirable. Such a procedure results in modest loss of power unless the treatment effect is of opposite sign in different sites or countries.

The hypothesis test is to be two-sided with alpha of 0.05. We propose to compute the P-value by permutation testing, based on the block randomization scheme.

Prespecified Subgroup Analysis.

The prespecified subgroup analysis will test the hypothesis that there is a treatment effect separately in each anatomic group, using a logistic regression model.

We denote the two anatomic groups by X_i^3, which equals 0 when the patient is in the intermediate group and 1 when the patient is in the posterior/pan group. In the intermediate subgroup, we plan to determine whether there is evidence of a treatment effect (regardless of the effect in the posterior/pan group). Specifically, we will conduct this analysis in two ways: anatomical site at enrollment (split into three categories: anterior, anterior/intermediate or intermediate only, and posterior/panuveitis), and anatomical location by history (split into two categories: anterior/intermediate or intermediate only, and posterior/panuveitis). Anatomical location by history is considered the prespecified analysis; anatomical location at enrollment will supplement this finding.

We propose to proceed as follows. We propose to begin with Equation (1), adding terms $\beta_3 X_i^3 + \beta_{13} X_i^1 X_i^3$ for anatomic location and for treatment-location interaction. We wish to test the hypothesis that $\beta_{13} = 0$, i.e. that there is a difference in treatment efficacy between the anatomic locations, controlling for country. Alternative models will be fit in which the country, site, and treatment x country terms are omitted.

We will also report relative risks in each substratum, using relative risk regression.

Additional analyses will add gender and age to the predictors. The entire analysis will be repeated for each gender, and separately for each country (US and India) and anatomic location.

Planned Secondary Analyses.

Each of the following secondary analyses is designed to test the hypothesis that treatment assignment affects a given outcome, after controlling for selected covariates. All analyses will be repeated without controlling for covariates (i.e., using treatment assignment as the only predictor). In all cases, appropriate regression diagnostics and/or goodness of fit tests will be performed (further details are given below). In addition, we will compute jackknife influence statistics in each analysis, to determine whether or not any single observation (eyes or patients, as appropriate) have an undue effect on the final conclusion. All models with site effects will be repeated omitting this effect, and again repeated including a treatment-site effect, and with country and/or treatment by country interactions (i.e., pooling within countries when appropriate). When reporting findings, care will be taken to distinguish the single prespecified

test from supplemental tests (whether prespecified or unprespecified); exploratory analyses will always be labeled as such. All alpha levels are to be two-sided.

1. *Twelve-month endpoint for successes*. We propose to compare the proportion who maintain successful control for twelve months (i.e. the outcome is the proportion who have achieved control in all study eyes both at the six month visit and at the 12 month visit) between the two study arms. Per protocol, patients with successful control of their inflammation at 6 months remain on the same treatment until 12 months. We will use the same statistical model (and Wald procedure) as for the primary analysis. We test the hypothesis that the coefficients for treatment assignment and treatment assignment/anatomic location interaction both equal zero.

2. *Time to corticosteroid sparing control of inflammation*. We propose to use a Cox proportional hazard model with the outcome being the time to (1) first steroid-sparing control, and separately (2) first control of inflammation, with treatment assignment (and interaction) as the predictors. Time to first steroid-sparing control is the principal prespecified analysis here; alternative approaches will be conducted for additional insight and as sensitivity analyses. We will supplement this analysis with a parametric survival analysis using the Weibull distribution and also with a gamma distribution (note that individuals may drop out at any time, not just at the monthly visits), and with a method treating the time to success as interval censored. The outcome for this analysis is a single number for each patient (not for each eye). The primary statistical result will be the Wald test for the treatment assignment coefficient. We will repeat the analysis using study site as a fixed effect in this model (and as a sensitivity analysis, will explore random-effects survival analytic methods which are becoming available, see Pankratz et al.).[6] In supplementary analyses we will include age and anatomic location as additional covariates.

3. *Country and site within country.* We denote the country by X_i^2, which will be 0 for US locations and 1 for Indian locations; X_{i1}^4 is 1 only for patients in the second Indian site and 0 otherwise, while X_{i2}^4 is 1 only for patients in the second US site and 0 otherwise. As mentioned under the Primary analysis, we propose to fit models with country only, drug by country interaction, and with a drug by country interaction, including site within country as well. Analysis will be conducted within each site, then pooling the sites within country together.
Further details regarding pooling across centers are provided above under the main prespecified analysis.

4. *Best spectacle-corrected visual acuity (BSCVA).* The primary outcome variable for this secondary outcome will be the change in best spectacle-corrected visual acuity from baseline to final (as defined in the FAST MOP, i.e. for those who successfully control inflammation as defined in the MOP, or at the time of failure for those who fail; MOP, Section 2.6). Visual acuity change scores are available for both eyes for each patient.

The primary analysis will use a linear mixed-effects regression, where the outcome variable is the change in BSCVA in each eye, using treatment assignment as a statistical predictor (regressor, independent variable); a random effect will be used at the individual level, because of the possibility that changes in the two eyes from a given patient are correlated. In a supplementary analysis, we will include as predictors (independent variables) anatomic location of uveitis, interaction between anatomic location and treatment assignment, and the study site, together with a random effects for patient. We will fit these models using maximum likelihood

(R procedure `lmer`) and use likelihood-ratio tests to test the hypothesis that treatment assignment affects BSCVA change. Only eyes that are eligible and meet inflammation criteria at baseline will be included in this analysis. If at a given visit, vision cannot be assessed, we will carry the last observation forward. Additional sensitivity analyses for missing data will be used (including mixed effects models controlling for time, including all data from an individual).

Also, if there is no eye at Month 6 to assess, the patient will be given a logMAR value of 2.0.

Because of the possibility that the outcome variable (BSCVA change score) will exhibit non-normality, we will repeat the analyses using transformations of the outcome data (including power and log transformations, or more general monotone transformations).

Additional analyses will be performed using age, gender, ethnicity, and the steroid dose at each month as predictors.

An additional supplemental analysis will be conducted using final BSCVA (instead of the change score) as the outcome, and including baseline BSCVA in each eye as a predictor, using methods otherwise identical to those above.

5. *Quality of life.* We will also use a linear mixed model to assess health-related quality of life, measured by the SF-36 questionnaire (PCS and MCS scores) and vision related quality of life NEI-VFQ-25 and IND-VFQ at 6 months or at the time of failure, as described in the Manual of Operations. Predictors will be baseline quality of life, age, gender, ethnicity, study site (as a random effect), and treatment assignment, and we will test the hypothesis that the regression coefficient corresponding to treatment assignment equals zero using the Wald t-test. Similar assessments will be performed for vision-related quality of life questionnaires.

6. *Reason for discontinuation.* Individuals who discontinue study medication may do so due to inability to tolerate side effects, due to lack of efficacy, or for safety reasons. The outcome variable is whether the person discontinued due to intolerance, discontinued due to lack of efficacy, discontinued due to safety, or did not discontinue the medication. Because study site may be an important factor, we will use polytomous regression to model the discontinuation result as a function of treatment assignment (using a fixed effect for study site).[7] If evidence is found that treatment assignment influences discontinuation result, further analyses may be conducted to determine whether or not treatment assignment is associated with discontinuation due to intolerance, lack of efficacy, or to safety, or some combination of these. We propose to classify all individuals in a two by four table according to treatment assignment and discontinuation (not discontinued, discontinued due to intolerance, discontinued due to lack of efficacy, discontinued due to safety) and conduct the Fisher's exact test (in its $r \times c$ form). The use of an overall test prior to further analysis is designed to protect the overall error rate.

7. *Successful control of inflammation with complete discontinuation of steroids (Phase I 6-12 months).* Some individuals may be able to taper completely off of steroids while maintaining control of inflammation. The outcome variable is the fraction of individuals achieving such control in both eyes (out of the number of individuals starting therapy). We propose to compare this fraction between the two treatment groups using logistic regression. The statistical analysis will otherwise be identical to the primary analysis.

8. *Macular edema.* We wish to compare the fraction of patients with macular edema at 6 months, between the two treatment arms. This will be conducted using the Fisher exact test, with

a two-sided test at alpha of 0.05. Supplementary analyses will be based on logistic regression using the presence of macular edema as a binary outcome variable, with regressors ("independent variables") of treatment arm and anatomic location. Further analyses (including other baseline covariates or other subsets) will be labeled as exploratory.

9. *Change in Macular thickness.* We propose to test the hypothesis that macular thickness is different in the two treatment arms, at 6 months. We propose to model the macular thickness at 6 months using two regressors: treatment arm and baseline thickness. We will test the hypothesis that treatment arm is associated with final macular thickness, using the T-test of the regression coefficient for treatment arm in the model including baseline thickness as a second covariate (two sided using α=0.05). We will examine residuals for normality and homoskedasticity, and prepare residual vs fitted value plots. Standard transformations will be used in case of evidence that the assumptions have been violated.

We will also look at change in macular thickness in only patients who had macular edema at Baseline.

10. *Bayesian analysis.* Prior to data collection, we will elicit a Bayesian prior for the effect size (difference between the two treatment arms) from a group of uveitis experts, using methods our group has previously applied to the Steroids for Corneal Ulcers Trial. The likelihood function corresponding to Equation (1) will be used to yield a posterior distribution for the effect size. Quantiles of this distribution will be reported, together with sensitivity analyses (with respect to model choice, influential observations, and prior distribution).

11. *Alternative definitions for success*. Other definitions will be examined: (i) changing the algorithm for assigning values for unobservable uveitis examination fields (anterior cells, vitreous haze, retinal/choroidal lesions) so that any worsening of ability to assess the eye for any reason is scored a failure, or (ii) use of vitreous cells in the definition of uveitis.

12. *Change in vitreous haze* will be assessed using clustered polytomous logistic regression, using baseline vitreous haze as a covariate and follow-up time. Vitreous haze is an ordinal outcome variable. A random effect is needed because the two eyes of a given patient cannot be treated as statistically independent. Both the NEI and Davis scales will be analyzed, for both direct observations and photographic grading. Treatment assignment will be a covariate. Alternative methods will be examined, including a simple McNemar test in which we dichotomize vitreous haze assessments at baseline and at the final observation.

13. *Rate of adverse events* and the proportion of patients discontinuing due to adverse events will be tabulated by treatment assignment, age, and gender; confidence intervals will be reported.

14. *Treatment efficacy in VKH patients* will be assessed as a planned subgroup analysis. Note that anatomic location is also a planned subgroup analysis, as well as study site and study country (aggregating all sites within each country).

15. *Dose reduction* will be compared by arm using logistic regression based on treatment, and other covariates as needed.

16. Additional exploratory modeling will be conducted using clustered multinomial logistic regression using all time points and all observations of anterior chamber cells, vitreous haze, and retinal/choroidal lesions.

3.2.3 Specific Aim 2

Primary Analysis.

The primary analysis will compare the proportion of successes between (a) patients treated with mycophenolate mofetil following failure on methotrexate and (b) patients treated with methotrexate following failure on mycophenolate mofetil.

Specifically, we will conduct a logistic regression in which success or failure will be the outcome, and the predictors (regressors, independent variables) will be treatment group and reason for failure of the first drug (lack of efficacy vs any other reason). Supplementary analyses will include anatomic location (intermediate vs posterior/pan) and country. We will test the hypothesis that the coefficient for treatment group equals zero (i.e., that mycophenolate mofetil rescue after methotrexate failure has the same result as methotrexate rescue after mycophenolate mofetil failure). All alpha levels will be two-sided.

It is important to emphasize that estimation of the success rate of the second drug following the failure of the first is a central goal of the trial, arguably as or more important than the hypothesis test itself. The success rates and confidence intervals will be presented regardless of the results of the hypothesis test.

Secondary Analyses.

The following secondary analyses are planned.

We will also present the estimated success proportion in both treatment groups, together with the 95% confidence intervals. The two groups are the individuals who were undergoing methotrexate rescue therapy after mycophenolate mofetil, and those who were undergoing mycophenolate mofetil rescue therapy after methotrexate. Logistic regression will also be used to adjust for study site.

The second prespecified analysis will compare the rate of success between rescue patients and first-line patients, using logistic regression; we will test the hypothesis that the coefficient for rescue/initial equals zero. A supplemental variation of this analysis will include an additional predictor for whether the patient was on rescue therapy due to lack of efficacy, lack of safety or intolerance, or anatomic location. Two separate analyses are planned, each with an alpha of 0.05.

Exploratory and descriptive analyses of covariates such as reason for failure of the initial regime, age, disease (e.g., VKH), and affected region of the eye, will be presented.

3.3 Transformations and model adequacy

3.3.1 Primary Analysis

Sensitivity analyses based on modeling the individuals lost to follow-up will be conducted, however; we will determine how much of a treatment effect there would have had to have been in the patients lost to follow-up, for the results of the main hypothesis test to change.

3.3.2 Unspecified secondary analyses

Unprespecified analyses may be conducted following the primary analysis and will always be reported as such. Analyses will always be repeated including age and gender, in particular.

3.3.3 Model validation and sensitivity

In all cases, standard statistical procedures will always be followed to ensure that no evidence indicates a violation of the assumptions underlying the statistical models used. Specifically, we note the following: secondary analyses based on the use of age as a continuous predictor in logistic regression models with treatment success as an outcome will be assessed using the Hosmer-Lemeshow goodness-of-fit test. Linear models will always be assessed using residual plots (residuals vs. predicted values, and QQ plots), together with tests for normality (Anderson-Darling and Shapiro-Wilk procedures). For mixed models, we will examine marginal residuals, conditional residuals, and EBLUPs.[8] When modeling binary outcomes (using clustered logistic regressions), we will repeat analyses using a probit link as a check on robustness; we will also examine the Pearson goodness of fit statistic.[9] Jackknife influence estimates will be used in all analyses; single observations that could change the conclusions will always be reported. Analyses in which time to response is used as the outcome variable (in which Cox regression is conducted) will be supplemented with the Gill-Schumacher procedure for assessing the adequacy of the proportional hazards assumption for Cox regression.[10] Analyses in which our primary interest in in final outcomes will still be repeated using all available data (at all time points).

Failure of the modeling assumptions (such as normality) will result in conducting additional analyses. First, for continuous outcome variables, we will undertake normalizing or variance-stabilizing transformations of the outcome variable (such as power transformations). Second, robust procedures will be used to estimate the standard errors whenever possible. Third, the use of bootstrap procedures, when applicable, will be considered in estimation of standard errors.[11]

3.4 Sample Size Evaluation

3.4.1 Primary Calculation

The sample size for the trial will be 216 subjects, which we anticipate will provide approximately 80% power to detect a difference of 20% in the proportion of patients achieving control of inflammation at six months between the methotrexate and mycophenolate mofetil groups.

This sample size was determined based on the primary objective (superiority comparison of mycophenolate mofetil to methotrexate) and primary endpoint (treatment success). We assumed an effect size of 20%, as this was deemed to be clinically meaningful, and well within the distribution of the investigators' prior beliefs from published retrospective studies.

An approximate sample size is provided by the formula

$$2N=\frac{4(Z_\alpha+Z_\beta)^2\bar{p}(1-\bar{p})}{(p_c-p_i)^2} \quad (5)$$

(see Friedman et al. 2010), where α is the significance level (0.05, two sided), β is one minus the power (the desired power is 80%), p_c is in this case the probability of success in the methotrexate group (we estimate this at 0.4), p_i is the probability of success in the mycophenolate mofetil group (we estimate this at 0.6), and $\bar{p}$ is $\frac{1}{2}(p_i+p_c)$. We assume 10% will be lost to follow-up in the first six months; details are given in the full proposal. This yields approximately 108 patients in each of the two groups, for a total of $2\times108=216$ subjects.

A power table is provided below as a sensitivity analysis (to show how the detectable effect size changes with varying success rates.

	80% Power		90% Power	
Success rate with Drug A	**Detectable effect size**	**Success rate with Drug B**	**Detectable effect size**	**Success rate with Drug B**
20%	18%	38%	21%	41%
30%	20%	50%	23%	53%
40%	**20%**	**60%**	**23%**	**63%**
50%	20%	70%	22%	72%
60%	19%	79%	21%	81%

Simulation confirms that this method yields adequate sample sizes for the logistic regression (results not shown).

Note that for the final analysis, the critical value will be adjusted slightly because of the interim analysis.

Sample size readjustment

Simulation suggests that a baseline covariate which is associated with the outcome variable could modestly reduce the sample size needed for 80% power (simulation results are available upon request). Sample size readjustment based on baseline predictors will be considered, subject to approval by the DSMB. The guiding principle is (CHMP, Reflection Paper on methodological issues in confirmatory clinical trials planned with an adaptive design, 2007):

Analysis methods that control the type I error must be pre-specified. Whenever possible, methods for blinded sample size reassessment that properly control the type I error should be used, especially if the sole aim of the interim analysis is the re-calculation of sample size.

3.4.2 Power for Subgroup Analyses and Other Analyses

Subgroups in Specific Aim 1.

The prespecified subgroup analysis for specific aim 1 is to examine the difference between the methotrexate and mycophenolate mofetil groups within each anatomic location. Using Equation (5), we anticipate having in excess of 80% power to detect a difference of 25% in success rates.

The power for selected secondary outcomes is provided here.

Secondary Outcomes in Specific Aim 1.

1. Twelve-month endpoint for success. We assume an additional loss of 5% between 6 and 12 months (that is, in addition to the 10% already lost to follow-up in the first six months). **We expect approximately 78% power to detect a 20% difference in success rates at the 12-month endpoint.**

2. Time to corticosteroid sparing control of inflammation. For sample size planning, we use the approximate formula given in Friedman et al (2010) for the number in each group:

$$N = \frac{\left(Z_\alpha + Z_\beta\right)^2\left(\phi(\lambda_C)+\phi(\lambda_I)\right)}{\left(\lambda_I - \lambda_C\right)^2}$$

where λ_C is the hazard in the methotrexate group, λ_I is the hazard in the mycophenolate mofetil group, and $\phi(u) = \frac{u^2}{1-e^{-uT}}$

where T is the censoring time (6 months). Previous studies suggest a median success time of approximately 3.5 months for mycophenolate mofetil.[12] **Assuming a 10% loss to follow-up, 108 subjects in each group provides 80% power to detect a difference of 2.47 months in the expected difference.** (Note: λ_C is log(2)/3.5 mo. for this calculation.) We assume an alpha of 0.05 (two sided).

3. Change in BSCVA. For sample size planning, we assume a T-test comparing change scores between the two drugs, assuming a standard deviation of the change in visual acuity of 6.5 letters.[13, 14] The sample size of 108 will provide approximately 80% power to detect 2.63 letters of difference in the change score. In other words, **we expect to have 80% power to detect whether mycophenolate mofetil yields 2.63 letters more of improvement than methotrexate, and we will have greater power to detect greater differences.** The power formula is provided in Chow et al and is computationally implemented in R in the function `power.t.test` (which we used).[15]

4. Quality of life. For a power calculation, we consider the SF-36 questionnaire, which has two scales, the MCS and the PCS. The raw score standard deviation will be assumed to be 8.4 points;

we assume a correlation between baseline and six months of 0.6.[16, 17] Assuming that the baseline score will "explain" roughly 36% of the variance allows us to assume a corrected raw score standard deviation of 6.72 in a simplified calculation in which we treat the analysis as a T-test. The same power calculation formula used in (3) above reveals that **our sample size provides approximately 80% power to detect a raw score difference of roughly 2.57 between the two treatment groups.** This difference is roughly comparable to the small difference in scores found between intermediate uveitis patients and the general population[16], a difference we believe to be more than sufficient to detect clinically significant results. Note that the population mean of this score is standardized to 50 on 0 to 100 scale. Similar analyses will be conducted for the vision related quality of life (i.e. NEI-VFQ-25 and the IND-VFQ).

5. Rate of discontinuation. Based on retrospective studies, we expect approximately 13% to discontinue methotrexate due to tolerability and 5% to discontinue due to safety (laboratory abnormalities or other serious adverse events). We expect approximately 4% to discontinue mycophenolate mofetil due to tolerability and 5% to discontinue due to safety.[12, 18-20] For the purpose of the power calculation, we assume 10% loss to follow-up and consider only the comparison of discontinuation due to tolerability. **We use the power formula given in Freedman et al (p. 104) to calculate a power of 61% for this comparison.[3]**

6. Macular edema. Previous studies suggest approximately 38% of individuals with uveitis will manifest macular edema.[21] **We have approximately 80% power to detect a difference of a factor of two in the final proportion of macular edema (19% vs 38%).**

7. *Macular thickness.* **A sample size of 108 (before loss to follow-up) provides approximately 80% power to detect a 65 micron difference between the two treatment groups, assuming a standard deviation of 160 microns in the final macular thickness.**[22] This analysis is quite conservative, since a difference of 100 microns between these two groups is consistent with previous studies. Moreover, adjustment for variance explained by the baseline thickness (i.e. the use of a smaller effective standard deviation) would yield a still higher effective power.[22, 23]

Specific Aim 2.

In the primary comparison of Specific Aim 2, we will estimate the effectiveness of rescue therapy, controlling for treatment group and reason for failure.

The primary analysis is (a) to estimate the probability of success on mycophenolate mofetil following failure of methotrexate, with 95% confidence intervals, and (b) to estimate the probability of success on methotrexate following failure of mycophenolate mofetil, with 95% confidence intervals.

These results will also be reported by reason for failure of the first drug, by categories of (i) failure because of inability to tolerate the first drug, (ii) failure of the first drug to achieve control (efficacy), or (iii) failure due to safety.

One analysis of interest is to compare the success rates in these two groups, and we include the sample size considerations for this analysis below. For two drugs (mycophenolate mofetil and methotrexate), we conduct the sample size planning as follows (denoting two drugs simply as A and B). For treatment group j=0,1 (0 coding drug B rescue in patients failing drug A therapy,

1 coding drug A rescue in patients failing drug B therapy), we expect $n_j = N_0 r_1 (1 - s_j) r_2$ subjects to be available for Specific Aim 2 (where N_0 is the number of subjects randomized to each treatment, r_1 is the retention fraction in Specific Aim 1 (not lost to follow-up in Specific Aim 1), s_j is the expected success fraction for patients for initial treatment j, and r_2 is the retention fraction in Specific Aim 2.

Thus, the number of available patients for Specific Aim 2 are highly dependent on the results from Specific Aim 1. Scientifically, the result of rescue therapy is important regardless of the result in Aim 1. More power will be available for the primary comparison in Specific Aim 2 if treatment in Specific Aim 1 yielded relatively high and similar failure rates for both drugs. However, even if success rates are very different in Aim 1, the descriptive analyses will still provide important information to guide decision-making on second-line treatment.

Here, N_0=108, and r_1 is assumed to be 0.9 (10% loss to follow-up). For planning Specific Aim 2, we assume a success rate of 60% for patients treated with drug A Specific Aim 1, and a success rate of 40% for those treated with drug B. This is a conservative estimate of the difference expected based on retrospective studies[12, 18-20] and consistent with the pilot study. Finally, we are assuming an additional 5% loss to follow-up during Specific Aim 2 (in addition to the 10% already lost), so that r_2=0.95. The results are summarized in the following table, where the number enrolling does not include loss to follow-up, and the "expected complete" column has taken loss to follow-up into account (n_{jk}).

We anticipate the following:

Initial/Second Treatment	Expected Enrollment SA/2	Expected to Complete SA/2
B/A	58.3	55.4
A/B	38.9	36.9

Thus, we expect a total of n_1=58 patients (rounding down) to have failed one first-line therapy to be enrolled in rescue therapy. Similarly, we expect n_0=38 patients to be enrolled in the other rescue regimen.

Previous observational studies suggest a 42% success rate of mycophenolate mofetil in methotrexate-failing patients,.[24] A simple power analysis for comparing these proportions may be found from the formula (see Chao et al, p. 87):[15]

$$1-\beta = \Phi\left(\frac{|p_1 - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n_0} + \frac{p_1(1-p_1)}{n_1}}} - z_{\alpha/2} \right)$$

where p_0 is the probability of success with methotrexate rescue following mycophenolate mofetil failure, p_1 the probability of success with mycophenolate mofetil rescue following methotrexate failure, and Φ is the cumulative distribution function of the standard normal distribution. These assumptions yield a power of 0.87 if the rate of success with methotrexate rescue is 0.15. We

have approximately 80% power to detect a difference of 17% if the probability of success with mycophenolate mofetil is 0.42.

A power table for sensitivity analysis is provided. We chose selected scenarios of potential interest to show the wide range of scenarios for which we have sufficient power. The main scenario is the first row of the table; in other rows, we varied the number of patients or the success fractions for the first drug used. In particular, the results are not sensitive to the efficacy difference found in Specific Aim 1.

Power Table for Specific Aim 2

Drug A, then Drug B (number)	Drug B, then Drug A (number)	Success probability of Drug B in patients failing Drug A	Success probability of Drug A in patients failing Drug B	Approximate Power
58	38	0.42	0.15	87%
58	38	0.42	0.17	80%
58	38	0.15	0.42	83%
58	76	0.42	0.15	94%
116	38	0.42	0.15	96%
40	40	0.42	0.15	80%
58	38	0.40	0.15	80%

To summarize, the anticipated number of patients from Specific Aim 1 (58 enrolled in in Drug A, and 38 in Drug B) should provide approximately 80% power to detect a difference of 25% between the two groups, assuming a success probability of 42% and a two-tailed alpha of 0.05.

Secondary Outcomes in Specific Aim 2.

1. Confidence intervals for the probability of success will be reported for each rescue group and anatomic location (i.e. Patients receiving methotrexate or mycophenolate mofetil as first treatment versus receiving it as their second, rescue treatment). Note that in the event that there are insufficient numbers of patients available in one arm of Specific Aim 2 (for instance, far fewer patients available for methotrexate rescue than we anticipate), confidence intervals for estimating the proportion of success can still be computed for the anatomic locations in the other arm.
2. We propose, for each rescue group, to conduct logistic regression using success as an outcome, and reason for failure of the first drug as a categorical covariate (safety,

efficacy, tolerability). An overall likelihood ratio test for each will be conducted, with an alpha of 0.05/2=0.025.

3. An additional comparison will be undertaken between first-line and rescue patients with both methotrexate and mycophenolate mofetil.
4. Additionally, the same secondary outcomes assessed in Aim 1 will be analyzed using similar methods.

3.5 Missing data and loss to follow-up

Values of the primary study endpoint (treatment success at six months) cannot be analyzed when the individual is lost to follow-up. We distinguish information which is missing because of possible progression of the underlying condition we wish to treat from information which is lost for some other reason. Earlier, we discussed methods for handling missing values for specific uveitis fields in individuals. The discussion in this section applies only to loss to follow-up or to dropping out of the study. As emphasized in Carpenter & Kenward (2007), "there can be no universal analysis when data are missing". Our purpose is to vary the assumptions as well as the methods, to establish that the estimates of the treatment effect are robust as such assumptions are varied.

Our priority is the preservation of the intent to treat principle. We propose to report the results from all of the following methods:

1. The use of regression-based multiple imputation, based on all observed data for the patient.
2. Use of longitudinal generalized linear mixed effects regression, with visit as a covariate, and including a random effect for each person and for each eye within each person, using all the available measurements on each individual
3. Sensitivity analysis in which missing final outcome values are assigned success or failure, and the analysis conducted conditional on these assignments.
4. Analysis of complete cases only (individuals for which the six month follow-up is available)

However, we are proposing that **method 4 (complete case analysis) be considered the primary outcome**, based on recommendation by the DSMB. All other analyses are to be considered supplementary.

Multiple imputation will be conducted as follows. The following information will be used as regression covariates: (i) age, (ii) gender, (iii) inflammation assessments at all prior time points (anterior cells, vitreous haze, and retinal/choroidal lesions), (iv) steroid dose, (v) anatomic location (by patient, classified as anterior/intermediate or posterior/panuveitis), (vi) anatomic location by history, (vii) maximum steroid dose within the 90 days prior to enrollment, (vii) steroid dose at enrollment prior to randomization or study-related intervention, (ix) country, and (x) site within country. Any additional covariates must be prespecified. A regression model for the missing outcome information will be derived; specifically, a cross-validated procedure to yield the best prediction based on complete subjects will be derived, and ten multiple imputations

will be derived from it. The formula in Little and Rubin[25] will be used to derive the overall test statistic. All replications will be recorded and reported.

An alternative method (which we propose to used for sensitivity analysis) is hot deck multiple imputation (with ten replications).[25] Note that treatment assignment would never be missing. For definiteness, we choose the recursive random partitioning hot deck method used in the R package `rrp` with the default settings (command `rrp.impute`).

The possibility of data-driven modeling may render multiple imputation of an outcome variable undesirable to many reviewers as a primary outcome. An alternative method is to model the treatment success of person i at visit j, Y_{ij}, using generalized linear mixed models, with covariates being site, country, treatment assignment, country-assignment interaction, visit (1:6), and visit-drug interaction (method 2 above). Note that additional statistical modeling will be reported, in which we (a) omit visit-drug interaction, and/or country assignment interaction, (b) add visit-country interaction, or (c) add age or gender as covariates.

We believe carrying forward last observations to be particularly unhelpful in this study, because all patients are on a prescribed steroid taper. We also believe that differential loss to follow-up of well performing patients on one drug or the other could falsely make the poorer drug appear to give more favorable results, so that the complete case analysis must be interpreted with caution.

3.6 Pooling across sites

Approximately three-fourths of patients are expected to come from the Aravind sites, which are in the same hospital network in the geographic region serving the same patient population. UCSF/Proctor and Oregon/Casey serve slightly different populations, although we expect fewer cases overall in the U.S. sites.

3.7 Multiple comparisons

An alpha of 0.05 will be used for the primary analysis of Specific Aims 1 and the primary analysis of Specific Aim 2. The prespecified subgroup analyses of Specific Aim 1 will be conducted at an alpha level of 0.05 (as stated above) as well. However, the use of an overall test prior to subgroup analysis protects the overall type I error rate for the primary outcome, a procedure we apply within the analysis of each secondary outcome as well.

3.8 Interim Monitoring

The study will be monitored by a Data Safety Monitoring Committee (DSMC) appointed by the National Eye Institute. There will be one in-person meeting a year and additional phone calls as deemed necessary. The DSMC will be unmasked and received reports with information by treatment arm from the principal statistician.

3.9 Accrual Rate

Based on enrollment rates in previous trials and preliminary data (see proposal for details) we anticipate enrolling 7-8 subjects per month at all sites, for a total enrollment period of 2.5 years. If we conservatively assume we may only accrue 25% fewer subjects per month, then completion of enrollment would occur 3 years and 3 months after the start of the trial.

We will establish monthly recruitment goals for each of the 4 sites, taking into careful consideration local holidays which may cause recruitment rates to drop at certain times of the year. Careful monitoring of the recruitment process will enable us to determine whether one of our sites may be falling behind in recruitment, precursory to further investigation and intervention. Standard graphs of realized cumulative recruitment together with cumulative recruitment goals for (a) the study as a whole, and (b) for each of the four sites will be prepared, and provided to the Data and Safety Monitoring Committee at each meeting (or more frequently, if requested).

3.10 Interim Analysis

We propose to conduct two interim analyses, at approximately one-third and at approximately two-thirds of the way through the study. The exact fractions will be determined by availability of data and timing of DSMB meetings. We plan to examine the primary outcome variable using the same statistical model we plan for the final analysis. A flexible alpha spending function is specified in Section 6.

3.10.1 Stopping rules

Stopping rules for benefit, harm, and futility are discussed in Section 6.2. These rules or guidelines would be determined at the first meeting of the DSMC (see Section 6.2).

3.10.2 Execution of interim analysis

The principal statistician (TP) will conduct the interim analysis in an unmasked manner, subject to independent statistical review by the DSMC. Quality assurance will be conducted by database manager WE.

3.11 Final Analyses

The Primary Aim 1 analysis (and secondary objectives), identified in this Statistical Analysis Plan will be performed when all patients complete their 6 month assessment and the window period is completed. All other analyses will be completed after the 12 month visit for Phase I or 6 month visit for Phase II and window periods are complete.

3.12 Software

The standard software program R version 2.12 or higher (`http://www.r-project.org`) for the MacIntosh OS X will be used for all descriptive and inferential analyses.

4 Analysis Populations

4.1 Summary

The following analysis populations are planned for this study:

- The **screening population**, which is to include all patients who are screened for participation in the trial.
- The **safety population**, which is to include all patients who receive any amount of planned study medication (mycophenolate mofetil or methotrexate).
- The **intent-to-treat efficacy population**, which is to include all patients who are randomized. This is the primary population for the efficacy analyses.
- The **per-protocol efficacy population**, which is to include all patients in the intent-to-treat efficacy population, excluding patients with any of the following: (a) major protocol deviations, or (b) noncompliance with study medications (less than 50% of the study drug received by self report or pill counts at study visits).

4.2 Major protocol deviations

The incidence of deviations from the inclusion and exclusion criteria will be summarized using counts and percentages, and the treatment groups compared for the overall frequency of deviations using a $2\times N$ Fisher's exact test. Similar deviations will be grouped into general categories of deviations for a more condensed summary. A listing of deviations by participant will also be produced. Any major deviations from the protocol will be listed and/or summarized, including, but not limited to, participants who:

- never received study drug
- were subsequently found to be ineligible for the study
- never returned for a follow-up visit
- have follow-up visits outside the prescribed visit window

The number and percentage of randomized participants actually receiving study medication, or permanently discontinuing study drug (subdivided by reason) will be summarized. A summary of study participants randomized by site will also be provided. Treatment groups will be compared for the proportion and reason for study drug discontinuation using the chi-square test. A summary of participant status at the end of the study period will also be generated with categories including lost to follow-up.

5 Data Collection and Quality Assurance

5.1 Quality assurance and security

Data collection forms, training, security, and quality assurance are discussed in the Manual of Operations for the FAST Treatment Trial.

5.2 Analysis sets

Data sets for analysis will be produced at the Proctor central site by database manager WE. Each will be a Microsoft Excel® worksheet containing a single header line whose variable names match the Access database. Each analysis set will be in the form of a rectangular table in which each column corresponds to a single variable and each row to an observation. All missing values will be coded explicitly using the string `NA` (as used in the R software). Codes for categorical variables (such as 1 for male, and 2 for female) will be avoided in favor of self-documenting character strings (such as Male, Female) whenever possible. Automated checks will be made to ensure consistency and that each variable in the analysis set has in-range values (protecting against negative ages, spelling errors in categorical factors, and similar errors).

A detailed codebook will be prepared, containing for each variable, (a) the form from which the variable derived, (b) the text of the question, (c) all possible values for the variable, and (d) summary statistics for the variable. Note that all codes and character strings that represent categorical factors will be clearly defined in the codebook. Units for each continuous variable (e.g. central subfield thickness, logMAR) will be unambiguously indicated for each variable. Each release of the analysis set will be accompanied by the corresponding version of the codebook. Version numbering with dates will be strictly observed. Standard report-generation software included with the R statistical and data analysis package will be used to ensure consistency of the codebook and analysis set at all times.

5.3 Data monitoring reports

Data monitoring reports will be prepared based on analysis data sets. These will be prepared using report-generation software. Monitoring reports will include (a) recruitment reports for each site, (b) compliance reports, (c) retention reports, and (d) data quality reports. These will be reviewed at the central site on a monthly basis, and communicated to the study sites on a monthly basis.

6 Human Subjects

6.1 Summary of final dispositions

All subjects who provide informed consent will be accounted for in this study. The frequency of subjects in each population will be presented. We will also present the frequency of subjects in each subgroup, the frequency of withdrawal and loss to follow-up, and any major protocol violations.

6.2 Data and Safety Monitoring Committee

6.2.1 Scope

A Data and Safety Monitoring Committee (DSMC) will be empaneled by the NEI. We propose that this committee consist of 5-7 individuals, and should include (a) uveitis specialists, (b) an

independent biostatistician, (c) a bioethicist, and (d) a member to protect the interest of the Indian population. The committee will meet in person at least once per year. *Ad hoc* meetings as needed may also be convened. All study protocols will be subject to review and approval by Institutional Review Boards at UCSF, Oregon and Aravind, and by the DSMC.

The Data and Safety Monitoring Committee will meet to review the interim efficacy data when primary outcome data are available on approximately one third of the study subjects—approximately 6 months after the 72nd subject has been enrolled in the trial (as discussed above in Section 3.10), and when data are available on approximately 2/3 of subjects. The DSMC will make one of the following recommendations:

- Continue the trial without modifications
- Continue the trial with study modifications
- Terminate enrollment or treatment in the trial because of safety concerns
- Terminate enrollment or treatment in the trial because of efficacy

6.2.2 Meetings

All in-person and teleconference meetings of the DSMC and study personnel will consist of (a) "open" sessions, which may be attended as needed by masked study personnel, and (b) "closed" sessions, which may only be attended by unmasked study personnel (TP, WE, TL), and (c) "closed" sessions attended only by the DSMC personnel. Care will be taken so that *no* treatment assignments, data which would allow treatment assignments to be determined, or outcome data based on treatment assignments will be revealed during the open sessions.

The DSMC will be unmasked. Closed reports will tabulate baseline covariates, adverse events, and outcomes by treatment assignment and study site. Written closed reports will always use the labels `Treatment A` and `Treatment B` for increased information security. However, the DSMC will know which drug corresponds to which label.

Interim reports for the DSMC will be prepared by the central Proctor site (TP). These reports will include (a) recruitment overall, and by study site, (b) compliance, and (c) retention. The reports will also list study outcomes, and all adverse outcomes, including deaths. The DSMC will determine the database closure dates for each report in advance; archival copies of the (a) main REDCap database, and (b) study analysis file as they exist at the time of each report will be maintained.

All reports will be sent using secure email to the members of the DSMC two weeks prior to each meeting (or more, if desired by the DSMC).

Each printed (hard copy) interim report will be labeled clearly as confidential, bound so that the contents are not visible from the outside, and labeled with the name of each person authorized to receive it. Reports will be kept in possession of WE and TP and only delivered in person or by encrypted email; reports not delivered due to absences are to be shredded. In addition, redacted versions of the interim reports will be prepared which contain no masked study information, and which are suitable for restricted distribution to other personnel on an as-needed basis. All hard copies will be destroyed at the end of each meeting, except for a copy to be kept in a locked file cabinet accessible only to TP and WE.

6.2.3 Decisions

The DSMC will make decisions with the benefit of prespecified decision guidelines. These guidelines will be agreed upon at the initial meeting, and are expected to include (a) safety, (b) efficacy, (c) clinical importance, (d) effect of baseline covariates, or (e) validity.

Benefits. Unmasked interim analyses (See Section 3.10) will be conducted to determine whether or not sufficient evidence has accumulated to justify stopping the trial because one treatment is clearly superior (and therefore should be extended to all future cases). The guidelines for efficacy will use group sequential boundaries for judging the statistical significance of the primary outcome measure. The Lan and DeMets flexible alpha spending approach will be used.

Early discontinuation in this trial has the following disadvantages. First, early discontinuation will make it more difficult to assess homogeneity of study sites. In this trial, where the majority of planned enrollment is not from the US, discontinuation at time t=1/3 for instance would occur when only 15 American patients had been enrolled (under our enrollment projections), and at t=2/3, only 30 American patients. Reflection on these small numbers of American patients may limit the adoption of the results of the trial. Second, early discontinuation reduces the power to assess the secondary aims of Specific Aim 1, and for Specific Aim 2. For these reasons, we propose to use conservative stopping rules.

We propose to use a Hwang-Shih-deCani alpha spending function of the form

$$a^*(t) = \frac{a\left(1 - e^{-gt}\right)}{1 - e^{-g}}$$

with γ chosen to be equal to -5.623626 exactly. This value was chosen to make the alpha at t=1/3 approximately equal to 0.001. The resulting alpha at t=2/3 is approximately 0.0075. The R package `gsDesign` (v. 2.7-04 or higher) will be used for selected analyses.

The proposed plan is to have two interim looks, at approximately t=1/3 and t=2/3 (one third and two-thirds through the study), with the specific fractions to depend on the total available data at face to face DSMB meetings.

The use of a flexible alpha-spending function protects the 0.05 alpha level of the overall trial while allowing for additional interim analyses for efficacy (if needed), without specifying the number and timing of the analyses at the start of the study. We note that the alpha spending function, including the value of γ, cannot be changed once the trial has begun.

Harm. Stopping for harm will be done at the judgment of the DSMC. Several endpoints will be examined, including serious adverse events such as significant and sustained laboratory abnormalities as described in the protocol, or mortality. While the analysis would consider maldistribution of predictive factors such as age, it is recognized that ethical considerations require careful considerations of statistical tests as well as qualitative judgments in the light of experience. Any additional analyses required by the DSMC will be conducted by TP and WE, as needed.

Note that serious adverse events (SAE) are reported directly to the medical monitor (TM) within 24 hours of the time the study site learns of them, and the medical monitor will subsequently pass this information on to the DSMC Chair. The medical monitor will receive

notification of the event, the timing of the event, a medical narrative from the study site, the site location, and the patient identification number. The statistician will report the study treatment assignment to the DSMC Chair if deemed necessary by the DSMC. If use of either drug use clearly results in an unacceptable increase in the risk of treatment failures, then the study will be stopped. It is difficult to fully prescribe boundaries for monitoring safety because there need not be strong evidence to discontinue the study if it appears that the treatment is harmful.

Futility. Early discontinuation due to the unlikeliness of significant findings conditional on interim results would prevent the analysis of Specific Aim 2 and of the secondary aims of Specific Aim 1. No stopping rules based on futility or conditional power calculations are included in the trial plan.

7 Safety and tolerability

The analysis of safety in this study will include summaries of the following:

- Exposure
- Adverse events
    - Adverse events and serious adverse events (including deaths)
    - Adverse events leading to withdrawal
    - Any deaths

7.1 Exposure

Individuals are assumed to have exposure to the drug corresponding to the arm to which they were randomized.

7.2 Adverse Events

7.2.1 Individual events

Adverse event reporting procedures are described fully in the MOP. Non-serious adverse events (not requiring narrative form) are described in the MOP (Section 6.1). Serious non-ocular or ocular adverse events (which must be reported within 24 hours and which require a narrative form) are described in the MOP (Section 6.2). Adverse events will be reported in all presentations and publications according to Consort guidelines.

The proportion of subjects with safety-related events will be compared using logistic regression, using treatment assignment and age as predictors, and including enrollment site as a random effect. Descriptive tables of the number and frequency of adverse events will be broken down by treatment arm, age, gender, and known comorbidities. We will report total adverse events and serious adverse events, cross-tabulated by whether the adverse events were anticipated or unanticipated and by whether or not the adverse event led to discontinuation of medication.

In addition, we will compare the rate of each of the adverse events during the follow-up period using Poisson regression, which can take into account multiple instances of adverse events within

a single subject. Age will be included as a predictor as well as treatment group, and enrollment site will be included as a random effect.

The additional statistical analysis of adverse events we describe here is undertaken strictly to provide additional insight which may be useful to the DSMC and investigators. Interpretation of such findings must reflect the fact that unanticipated adverse events may occur and that we may have insufficient power to make inferences between the arms when considering rare events. Note that adverse events contribute to the outcome of the trial and specific analyses have been defined earlier.

7.2.2 Pooled adverse events

Adverse events will be analyzed according to four main categories:

- Proportion of subjects with *any ocular adverse event*
- Proportion of subjects with *any serious ocular adverse event*
- Proportion of subjects with *any systemic adverse event*
- Proportion of subjects with *any systemic serious adverse event*

The proportion of subjects with these events will be compared between the arms using Fisher's Exact Test. Poisson or negative binomial regression will be applied to compare the rates of overall adverse events, including recurrent events.

8 Reporting conventions

- All tables and data listings will be presented in landscape orientation, unless presented as part of the text of the final report.
- Figures will be presented in landscape orientation, unless the information is substantially easier to interpret in portrait orientation.
- Direct annotation of figures will be preferred to legends. All figures with more than one variable or item will contain either direct annotation or legends. All annotation will be unambiguously identifiable as such.
- Color will be used in figures only when needed to enhance clarity of communication. All color schemes will be evaluated for visual clarity for individuals with diminished color vision. All color encodings will be identified. Redundant encodings (such as the use of different plot symbols or line dash patterns) will be used in addition to color, so that all figures are interpretable after monochrome reproduction at 100 dots per inch. All dash patterns and line widths will be adequate to be distinguishable after monochrome reproduction at 100 dots per inch. Any distinction between plot symbols (circles, filled circles, diamonds, etc.) will remain clear after monochrome reproduction at 100 dots per inch.
- Fixed width sans serif fonts will be used for all labeling (Helvetica, Arial, or Futura).
- Boldface and italics will not be used unless substantial value is added.

- Decorative fonts and enhancements, including borders and shading, will not be used. Decorative presentation methods, such as ribbon graphs, will never be used.
- All information given in figures will also be presented in summary tables (perhaps only included in an Appendix or in supplementary materials).
- Only standard characters will be used in tables and data listings.
- All titles will be centered. The first title line will be the number of the table, figure, or listing. The second and possibly third lines will be the description of the table, figure, or data listing. The ICH numbering convention will be used for all.
- All footnotes will be left justified and at the page bottom. Footnotes will be used sparingly. Reference footnotes will be complete enough to locate any reference based on the information provided (Author, Journal, Pages, Date, or PubMed accession number).
- Missing values for numeric or character variables will be unambiguously identified as such using the special string `NA` (not available) in all settings; `NA` is the standard missing value code for our software. Each figure or table caption in which `NA` is used will indicate the meaning of `NA` in that figure or table. The abbreviation `NA` will never be used for any other purpose.
- All date values will presented in the form DDmmmYYYY format (e.g. 01jan2008), using four digit years. June will be encoded as `jne` (otherwise `jan` and `jun` would differ by only a single character), and July as `jly` (so that the lowercase letter `l`, easily confused with the digit `1`, will not be adjacent to any numerals).
- All tables, figures, and data listings will have the name of the program and a date/time stamp on the bottom of the output.

9 Abbreviations and acronyms

AES Advanced Encryption Standard
CAS Chemical Abstracts Service
DSMC Data and Safety Monitoring Committee
FAST First-line Antimetabolites as Steroid-sparing Treatment
FIPS Federal Information Processing Standard
ICH International Conference on Harmonization
logMAR log of minimum angle of resolution
MOP Manual of Operations and Procedures
MUTT Mycotic Ulcer Treatment Trial
NIST National Institute of Standards and Technology
SAP Statistical Analysis Plan
SCUT Steroids for Corneal Ulcers Trial
TM T. Margolis
TL T. Lietman
TP T. Porco
UCSF University of California, San Francisco
WE Wayne Enanoria

10 Appendix

All computations will be performed using the standard software package R (`http://www.r-project.org`). Statistician TP has twenty years of experience using R or very similar statistical computing environments (S, S-Plus).

Specification of the random number seed and pseudorandom number algorithm determines the entire randomization assignment (as is the case with any pseudorandom number generation method). Accordingly, the random number seed will be kept confidential, and the seed will be chosen carefully. In particular, easy-to-remember numbers or otherwise meaningful numbers (such as telephone numbers, birthdays, and so forth) are to be scrupulously avoided. The chosen seed will be used to generate the final randomization lists.

A printed copy of the randomization lists for all sites, the computer code used to generate them, and the random number seed will be maintained in a locked vault off site. The random number seed chosen will consist of at least eight digits, and a standard linear feedback shift-register algorithm will be used for pseudorandom number generation.[29]

11 Bibliography

1. Ethical Guidelines for Statistical Practice. Alexandria, Virginia: American Statistical Association, 1999.
2. Food and Drug Administration USA. International Conference on Harmonization, Guidance on Statistical Principles for Clinical Trials. U.S. Federal Register 1998;63:49583-49598.
3. Friedman LM, Furberg C, DeMets DL. Fundamentals of clinical trials, 4th ed. New York: Springer, 2010:xviii, 361 p.
4. ISBN User's Manual. Berlin: Staatsbibliothek zu Berlin — Preußischer Kulturbesitz, 2001.
5. Bolker BM, Brooks ME, Clark CJ, et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution 2009;24:127-135.
6. Pankratz VS, de Andrade M, Therneau TM. Random-effects Cox proportional hazards model: general variance components methods for time-to-event data. Genetic Epidemiology 2005;28:97-109.
7. Skrondal A, Rabe-Hesketh S. Multilevel logistic regression for polytomous data and rankings. Psychometrika 2003;68:267-287.
8. Nobre J, Singer J. Residual analysis for linear mixed models. Biometrical Journal 2007;49:863-875.
9. Tuerlinckx F, Rijmen F, Verbeke G, de Boeck P. Statistical inference in generalized linear mixed models: a review. British Journal of Mathematical and Statistical Psychology 2006;59:225-255.
10. Gill R, Schumacher M. A simple test of the proportional hazards assumption. Biometrika 1987;74:289-300.
11. Ocana J, El Halimi R, de Villa M, Sanchez J. Bootstrapping repeated measures data in a nonlinear mixed-models context. Universitat de Barcelona: Institut de Mathematica Preprint 367, 2005.
12. Galor A, Jabs DA, Leder HA, et al. Comparison of antimetabolite drugs as corticosteroid-sparing therapy for noninfectious ocular inflammation. Ophthalmology 2008;115:1826-32.
13. Baltatzis S, Tufail F, Yu EN, Vredeveld CM, Foster CS. Mycophenolate mofetil as an immunomodulatory agent in the treatment of chronic ocular inflammatory disorders. Ophthalmology 2003;110:1061-5.
14. Samson CM, Waheed N, Baltatzis S, Foster CS. Methotrexate therapy for chronic noninfectious uveitis: analysis of a case series of 160 patients. Ophthalmology 2001;108:1134-9.
15. Chow S-C, Shao J, Wang H. Sample size calculations in clinical research. New York: Marcel Dekker, 2003:ix, 358 p.
16. Murphy CC, Hughes EH, Frost NA, Dick AD. Quality of life and visual function in patients with intermediate uveitis. Br J Ophthalmol 2005;89:1161-5.
17. Brazier JE, Harper R, Jones NM, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. Bmj 1992;305:160-4.
18. Daniel E, Thorne JE, Newcomb CW, et al. Mycophenolate Mofetil for Ocular

Inflammation. Am J Ophthalmol 2009.
19. Gangaputra S, Newcomb CW, Liesegang TL, et al. Methotrexate for Ocular Inflammatory Diseases. Ophthalmology 2009.
20. Thorne JE, Jabs DA, Qazi FA, Nguyen QD, Kempen JH, Dunn JP. Mycophenolate mofetil therapy for inflammatory eye disease. Ophthalmology 2005;112:1472-7.
21. Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Sugar EA. The multicenter uveitis steroid treatment trial: rationale, design, and baseline characteristics. Am J Ophthalmol;149:550-561 e10.
22. Roesel M, Heimes B, Heinz C, Henschel A, Spital G, Heiligenhaus A. Comparison of retinal thickness and fundus-related microperimetry with visual acuity in uveitic macular oedema. Acta Ophthalmol 2009.
23. Grover S, Murthy RK, Brar VS, Chalam KV. Normative data for macular thickness by high-definition spectral-domain optical coherence tomography (spectralis). Am J Ophthalmol 2009;148:266-71.
24. Sobrin L, Christen W, Foster CS. Mycophenolate mofetil after methotrexate failure or intolerance in the treatment of scleritis and uveitis. Ophthalmology 2008;115:1416-21, 1421 e1.
25. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York: John Wiley, 2002.
26. Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. Statistics in Medicine 2001;20:3931-3946.
27. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. Stat Med 1994;13:1341-52; discussion 1353-6.
28. Lan KK, Wittes J. The B-value: a tool for monitoring data. Biometrics 1988;44:579-85.
29. Matsumoto M, Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation 1998;8:3-30.

FAST Uveitis Trial

Statistical Analysis Plan

Confidential

Version 1.5, March 2017

Prepared by:
FI Proctor Foundation
Data Coordinating Center

TABLE OF CONTENTS

1 Introduction

This document (Statistical Analysis Plan, SAP) describes the planned analysis and reporting for **the FAST (First-line Antimetabolites as Steroid-Sparing Treatment) Uveitis Trial**, University of California, San Francisco. It includes specifications for the statistical analyses and tables to be prepared for the final Clinical Study Report.

The proposed FAST Uveitis Trial is a block randomized, observer-masked, comparative effectiveness, Phase III clinical trial to compare the efficacy of mycophenolate mofetil (CAS 128794-94-5) to methotrexate (CAS 59-05-2) for the treatment of non-infectious uveitis requiring steroid-sparing therapy.

The content of this Statistical Analysis Plan meets the requirements stated by the US Food and Drug Administration and conforms to the American Statistical Association's Ethical Guidelines.[1, 2]

The following documents were reviewed in preparation of this Statistical Analysis Plan:

- FAST Uveitis Trial Manual of Operations
- ICH Guidance on Statistical Principles for Clinical Trials[2]
- Statistical Analysis Plan (prepared by C. McCulloch), Steroids for Corneal Ulcers Trial (T. Lietman and N. Acharya, principal investigators)
- Statistical Analysis Plan (prepared by T. Porco), Mycotic Ulcer Treatment Trial (T. Lietman and N. Acharya, principal investigators)

The planned analyses described in this Statistical Analysis Plan will be included in future manuscripts. Note, however, that exploratory analyses not necessarily identified in this Statistical Analysis Plan may be performed to support the analysis. All post-hoc or unplanned analyses which have not been delineated in this Statistical Analysis Plan will be clearly documented as such in the final Clinical Study Report, manuscripts, or any other document or submission.

The final SAP is subject to the approval of an appointed Data and Safety Monitoring Committee.

The following individuals contributed to this document: N. Acharya, T. Lietman, N. Nardone, and T. Porco.

2 Investigational Plan

2.1 Study Design

The proposed study is an international, multicenter, block randomized, observer-masked, comparative effectiveness clinical trial to determine which treatment, methotrexate or mycophenolate mofetil, is more effective as first-line corticosteroid-sparing treatment for patients with non-infectious intermediate, posterior and panuveitis requiring corticosteroid-sparing therapy.

Full details which specify the definition of treatment success are given in the FAST Manual of Operations.

2.2 Study Population

Eligible volunteers diagnosed with non-infectious uveitis who have given informed consent will be enrolled in this trial. Specific eligibility and exclusion criteria are given in the FAST Manual of Operations. The proposed study schedule is listed in the FAST Manual of Operations.

2.3 Specific Aims

2.3.1 Specific Aim 1

Primary Objective. The primary objective of the study is **to establish which immunosuppressive treatment, methotrexate or mycophenolate mofetil, results in a higher rate of corticosteroid-sparing treatment success, on an intent to treat basis.** Specifically, we will compare the fraction of subjects who achieve treatment success at six months (as defined in the Manual of Operations Section 2.5.1) between the two groups.

Primary Outcome. The primary outcome for Specific Aim 1 will be the difference in the proportion of patients assigned to mycophenolate mofetil vs. methotrexate who achieve treatment success (as defined in the Manual of Operations Section 2.5.1).

Patients who experience success at 6 months with the drug to which they were originally randomized (in Specific Aim 1) will continue on the same drug for an additional 6 months. This will be called Phase I (6-12 months). Patients will then be seen every 3 months (and will be examined at 9 months and at 12 months), until success at 12 months or treatment failure at any time. Patients who fail treatment before 12 months with the initial drug will be removed from the study and treated according to best medical judgment.

Secondary Objectives

- To determine whether patients exhibit a difference in time to control of inflammation within the first six months.
- To determine whether patients exhibit a difference in time to corticosteroid sparing control of inflammation within the first six months.
- To evaluate the proportion of patients achieving treatment success at 5 months and sustaining, for at least 28 days, to 6 months.
- To evaluate a difference in control of inflammation in the posterior/pan anatomic locations only, assessed at by six months.
- To evaluate a difference in control of inflammation in the interior and anterior/intermediate anatomic locations only, assessed at by six months.
- To determine whether there is a change in best spectacle-corrected visual acuity at six months.
- To determine whether patients treated exhibit a difference in health related quality of life at six months.
- To determine if there are differences in discontinuing treatment due to each of the following reasons: safety, intolerability or lack of efficacy at six months.
- To determine whether patients exhibit a difference in the proportion of patients having macular edema at 6 months
- To determine whether patients exhibit a difference in macular thickness at 6 months
- To determine whether patients exhibit a change in vitreous haze, assessed clinically by the NEI and Davis scales at 6 months
- To determine whether patients exhibit a change in vitreous haze, assessed by the photographic grading of haze by the NEI and Davis scales at 6 months.
- To determine the proportion of patients discontinuing due to intolerability at six months.
- To determine the rate of adverse events experienced at six months.
- To determine the proportion of patients discontinuing due to serious adverse events at six months.
- To determine whether patients exhibit a change in quality of life at six months.
- Tabulate the occurrence of dose reduction used in immunosuppressive treatment (see Manual of Operations Section 3.1 for dose reduction guidelines).
- To determine efficacy of treatment in Vogt-Koyanagi-Harada (VKH) patients at six months.
- To determine the proportion of patients beginning with at least 2+ inflammation in anterior chamber cells and experience at least a 2-step reduction (i.e. decreasing from 2+ to 0.5+; 3+ to 1+; 4+ to 2+).
- To determine the proportion of patients beginning with at least 2+ inflammation in vitreous haze and experience at least a 2-step reduction (i.e. decreasing from 2+ to 0.5+; 3+ to 1+; 4+ to 2+0.
- To evaluate a difference in treatment success controlling for vasculitis at baseline, assessed at six months.

- To explore the use of a dynamic process model (such as a Hidden Markov model) to assess differences in control of inflammation.

- VKH) patients at six months.

  - Proportion of patients who started with at least 1+ inflammation levels in anterior chamber cells who achieve a decrease to 0 level of inflammation in anterior chamber cells.
  - Proportion of patients who started with at least 1+ inflammation levels in vitreous haze who achieve a decrease to 0 level of inflammation in vitreous haze.

All the above analyses will be examined at the end of Phase I (6-12 months), in addition to the following:

- To determine whether patients exhibit a difference in the probability of controlling inflammation with complete discontinuation of steroids at twelve months in Phase I.

2.3.2 Specific Aim 2

Primary Objective. The primary objective of this aim is to evaluate the clinical efficacy of switching agents as rescue therapy after initial treatment failure.

Patients who experience treatment failure (as defined in the Manual of Operations) with the drug to which they were originally randomized in Specific Aim 1 will discontinue the current treatment and be administered rescue therapy with the second drug (in a masked fashion). This will be called Phase II. Treatment failure with the first drug is defined as the inability to continue taking the drug to which the patient has been randomized, either due to intolerability, safety concerns, or lack of efficacy.

Upon declaration of treatment failure, the patient will be automatically screened for Aim 2. If eligibility criteria are met, the second treatment will be administered and data will be collected for the Phase II baseline visit. Patients will then be seen every 4 weeks until 6 months or until treatment failure with the second drug. Treatment failure and success will be defined as in Aim 1. Patients who fail treatment before 6 months with the second drug will be removed from the study and treated according to best medical judgement.

Primary Outcome. The primary outcome is the fraction achieving treatment success at 6 months after starting Phase II. Treatment success is defined as in Aim 1 and described in the Manual of Operations Section 2.5.1.

Secondary Objectives. All of the secondary 6 month objectives listed for Aim 1 will be examined for Aim 2.

2.4 Randomization

2.4.1 Stratification between sites

Patients will be recruited from nine sites: Aravind Madurai-AEHM, Aravind Coimbatore-AEHC, Aravind Pondicherry-AEHP, Casey Eye Institute-OHSU, King Khaled Eye Specialist Hospital-KKESH, Centre for Eye Research Australia-CERA, Asociación Para Evitar La Ceguera en México-APEC, Northwestern University-NWU, and Proctor Foundation-UCSF (see Manual of Operations Section 2.2 for details). Patients will be randomized to two treatments (arms): methotrexate (X) or mycophenolate mofetil (Y). The treatment protocols are specified in the FAST Uveitis Trial Manual of Operations.

Within each site, assignments will be conducted using a block randomization scheme with randomly varying block sizes.

2.4.2 Randomization list

Lists of sequential randomization assignments will be prepared for each site. The randomization lists consist of a unique identifier for each patient, together with the assignments to treatment arms. The assignment of patient ID numbers and randomization is thus performed on enrollment.

The randomization lists for sites will be prepared by the Proctor site (see Section 10.3) and sent to the Emergency contact at each site to be used only in case emergency unmasking is needed for patient safety.

They will also be sent to hospital/clinic staff who are responsible for telling the study coordinators the treatment assignment for each patient after the patient is enrolled and the study ID has been assigned. At these sites she will verify patient treatment assignment as a quality assessment.

A backup copy of the full randomization list for all four sites will be maintained by Tom Lietman, MD, (hereinafter TL). This list will be maintained as a hard copy stored in a locked file cabinet at the UCSF site, and to be used only in case emergency unmasking is needed for patient safety.

Distribution of the randomization list to Aravind and Casey will be accomplished using the University of California, San Francisco's encrypted email provision. Email is encrypted using the Advanced Encryption Standard (NIST FIPS 197) whenever the first four characters of the subject line are `PHI:` The sender is notified when the recipient receives a secure email; the recipient receives a notification of a secure email and can view it using the UCSF Secure Messenger website. We have successfully used this method in previous clinical trials (Steroids for Corneal Ulcers Trial, Mycotic Ulcer Treatment Trials). The randomization lists will each contain more randomization assignments than needed. Successively recruited patients will receive sequential assignments from the list. The long list provides a measure of added safety in case one of these sites recruits far more patients than expected relative to the other site.

As discussed below, the randomization lists will be provided as Excel® worksheets. No technical knowledge will be required to use these lists.

Update: starting 18 January 2017, all patients will be randomized using the REDCap database. Coordinators can access only the randomization lists for their site. When patients are enrolled, the ID is logged in REDCap and the system provides the medication assignment. All emergency contacts have access to the REDCap database for their site in case of an emergency when unmasking is needed for patient safety. The REDCap database contains assignments for all previously enrolled patients. All previous randomization lists were deemed void and were destroyed. Each site submitted a certification of destruction form.

2.4.3 Block randomization

We will utilize a permuted block randomization scheme with a randomly varying block size (within each study site) to protect the integrity of the assignment masking.[3] Any particular block size will be unknown to the study investigators. We will choose randomly varying block sizes, picking a block of size 4 with probability 2/3 and a block of size 6 with probability 1/3. Individuals have a higher probability of being in a block of size 6 because the blocks are larger. Many other choices would serve equally well. Given the block size, a random permutation of assignment orders will be generated.

2.4.4 Unique patient identifiers

Unique patient identifiers will be generated as follows. The first character will be a number: "1" for CERA, "2" for UCSF, "3" for NWU, "4" for OHSU, "5" for APEC, "6" for KKESH, "7" for AEHM, "8" for AEHC, and "9" for AEHP. The next character is a checksum character, which will be a single letter. The last three characters will be sequential digits beginning at 001. An example identifier is 4J101; all identifiers have exactly five characters, and no other Aravind/Proctor study uses this format.

2.4.5 Random number generation

The choice of a random number seed determines the specific sequence of random numbers that will be produced by the random number generator. Once the seed is determined, the randomization assignments for all sites are determined. Details are given in the Appendix.

2.4.6 Provision of randomization list

Everyone to whom the randomization list should be provided (for each of the four sites) will receive it in the following format: a Microsoft Excel® spreadsheet containing the following columns: (1) the unique study identifier assigned to the patient (see Section 2.4.4), (2) an empty field into which the date of randomization may be entered (relevant only for the hospital/clinic staff holding the randomization lists), (3) the study drug assignment, written out in full as `Mycophenolate` or `Methotrexate`. As discussed in Section 2.4.2, these lists will be treated confidentially.

Update: As of 18 January 2017, all randomization lists are electronic on REDCap. Once a patient is enrolled, the study coordinator logs the patient in the database and the system reveals the drug assignment, written out in full.

2.4.7 Quality assurance

Three quality assurance steps for the randomization list preparation are conducted. First, the software will have been tested during previous studies (MUTT). Second, the software that generates the assignments verifies approximate balance of subjects in each group before writing the Microsoft Excel® files. Each file will contain the study site as the first line. Finally, the output files will be visually inspected. The software and procedures have already been developed and successfully used in previous studies.

2.4.8 Summary of disposition of randomization list

The following individuals will receive a copy of the randomization list:

Emergency Contact Personnel

OHSU	Maggie Ryan/ Christina Flaxel
CERA	Dr. Sukhpal Sandhu
UCSF	Dr. Stephen McLeod
AEHP	Dr. R. Venkatesh
AEHC	Dr. R. Revathi
AEHM	Dr. Lalitha Prajna
KKESH	Dr. Marco Mura

*Emergency contact persons who will consult the list only in case of an emergency in which unmasking is necessary for patient safety and Dr. Thomas Lietman on the DCC cannot be reached.

Data Coordinating Center (DCC) Personnel

- Dr. Thomas Lietman, Professor of Ophthalmology and Epidemiology, University of California San Francisco, Proctor Foundation
- Dr. Travis Porco, Principal statistician, Proctor Foundation
- Erica Browne, Data manager, Statistical programmer/analyst, Proctor Foundation

Clinic/Hospital Staff

- Ms. Sally Tsang, Clinic Manager Proctor Foundation, University of California, San Francisco
- Ms. Maya Rao*, Study Coordinator, University of California, San Francisco
- Ms. Erica Browne, Data Manager, University of California, San Francisco
- Mr. Sivaram, Study Coordinator, Aravind Eye Hospital Madurai, India
- Ms. V. Gracy Evangelin, Study Coordinator, Aravind Eye Hospital Madurai, India
- Ms. R. Srija, Study Coordinator, Aravind Eye Hospital Coimbatore, India
- Ms. R. Thilagavathi, Study Coordinator, Aravind Eye Hospital Pondicherry, India
- Ms. Tracy Giles, Study Coordinator, Casey Eye Institute, Oregon

- Ms. Yoko Burgoa, Study Coordinator, APEC, Mexico
- Ms. Hilda Hernández, Study Coordinator, APEC, Mexico
- Ms. Julie Morrison, Study Coordinator, CERA, Melbourne
- Ms. Tanya Pejnovic, Study Coordinator, CERA, Melbourne
- Ms. Julie A. Johnson, Study Coordinator, Northwestern University, Chicago
- Ms. Sara Al Nuwaysir, Study Coordinator, KKESH, Riyadh
- Mr. Abdulrahman Al Hommadi, Study Coordinator, KKESH, Riyadh

  *Ms. Rao will act as study coordinator to Proctor Foundation patients, as well as Coordinating Center Manager overseeing all other sites. She will have access to the randomization lists for other sites, in order to check patient treatment assignment as a quality assessment and manage distribution of medications to all sites.

2.5 Masking

The clinical examiners, refractionists, OCT technicians, fundus photographers and fundus graders will be masked to the treatment assignment. Note that only the individuals listed in Section 2.4.8 will have copies of the randomization list. Full details of procedures to maintain masking as well as for potential unmasking in the event it becomes necessary for safety reasons are provided in the Manual of Operations. Principal Investigator N. Acharya is masked.

3 Statistical Considerations

3.1 Baseline characteristics

At baseline, each eye (1) may be fully able to be assessed, (2) it may be possible assess part of the eye, but not be possible to assess the entire eye, or (3) it may not be possible to assess any of the eye. For each eye for which some assessment is possible, either (1) the eye showns no signs of uveitis, or the eye may show some signs of uveitis, but fail to meet the severity criteria (1+ anterior chamber cells, vitreous haze or no active retinal/choroidal lesions, as defined in the Manual of Operations), or (2) the eye meets the severity criteria as defined in the manual of operations. Some patients are monocular at baseline, one eye being either absent, or exhibiting such disease as to preclude the possibility of ever assessing the eye (i.e. phthisis).

For this trial, we summarize the above possibilities as follows. Each eye (OD or OS) may be classified into one of the following types at baseline:

A. Eye fully assessible, does not meet the severity criteria as defined in the MOP
B. Eye partially assessible, does not meet the severity uveitis criteria in the assessible region
C. Eye fully assessible, meets severity criteria
D. Eye partially assessible, meets severity criteria in assessible region
E. Eye absent or too diseased to ever assess

Patients, not eyes, are the unit of assignment and of randomization. Thus, there are twenty-five possible types of patients. A patient is required to have at least one eye which meets severity

criteria for uveitis, and which can be completely assessed. Eligibility is summarized in the following table; cells indicate the possibility of enrollment for a patient whose right eye classification corresponds to the row and whose left eye classification corresponds to the column (A-E being defined in the previous paragraph).

	OS: A	OS: B	OS: C	OS: D	OS: E
OD: A	Not eligible	Not eligible	**Eligible**	Not eligible	Not eligible
OD: B	Not eligible	Not eligible	**Eligible**	Not eligible	Not eligible
OD: C	**Eligible**	**Eligible**	**Eligible**	**Eligible**	**Eligible**
OD: D	Not eligible	Not eligible	**Eligible**	Not eligible	Not eligible
OD: E	Not eligible	Not eligible	**Eligible**	Not eligible	Not eligible

Assessment and follow-up depends on the status of the eye. Eyes classified as type E above are recorded as such at baseline, and never provide eye outcome related data. Because (a) inability to assess parts of the eye could be related to the progression of disease, but (b) inability to assess in the absence of signs of disease cannot be considered evidence of treatment failure, we use the following table to summarize how success at six months will be scored. In this table, the row corresponds to the status of an eye at baseline, and the column to the status of the eye considering the primary outcome of success at six months.

	Month 6: A	Month 6: B	Month 6: C	Month 6: D	Month 6: E
Baseline: A	Success	See below**	Fail	Fail	Fail*
Baseline: B	Success	See below**	Fail	Fail	Fail*
Baseline: C	Success	See below**	Fail	Fail	Fail*
Baseline: D	Success	See below**	Fail	Fail	Fail*
Baseline: E	NA	NA	NA	NA	NA

Specifically, note that an eye which is fully assessible at six months and which does not meet the specific criteria for failure of control is always considered a success. Eyes which are fully or partially assessible and which meet any of the criteria for failure are always considered to have failed. However, eyes which are only partially assessible but which meet no criteria in the assessible region may be scored successes or failures depending on their baseline status (see next paragraphs below). Eyes which were present at baseline but which are missing at the end of the study are listed as Fail* in the table; we propose to consider such eyes to have failed unless a specific reason demonstrates that the loss of the eye was completely unrelated to the presence or progression of disease.

The primary analysis is at the patient level. Both eyes must meet the success criteria for the patient to be considered a success.

****Incompletely assessible eyes.** Uveitis assessment for this purpose is based on (i) assessment of anterior cells, (ii) vitreous haze, and (iii) retinal or choroidal lesions. In the pilot study, (iv) assessment of vitreous cells was also used. We will have longitudinal measurements of inflammation according to the following schedule: anterior chamber cells, vitreous haze and active retinal/choroial lesions will be measured at Baseline, Week 2, Month 1 and every subsequent 4 weeks until the 6 month assessment (Phase I or Phase II) or 9 and 12 month

assessment (Phase I 6-12 months).

Each of these (including the binary assessment of the presence of retinal or choroidal lesions) may be considered an ordinal variable, with relevant threshold values for each (used in determining eligibility for enrollment, or success in therapy).

Scoring of incompletely assessible eyes is governed by the following guiding principles:

1. In some patients, the front of the eye may be assessible, but the back of the eye cannot be examined and assessed clinically (even though the patient can still see out of the eye).
2. Worsening of uveitis may render it harder to assess the back of the eye, so that information cannot be considered missing at random in general.
3. Many uveitis patients have at least one eye which cannot be fully assessed, because of the progression of the disease itself. Excluding such patients or eyes completely is undesirable.
4. Treatment of uveitis will not reverse the damage which makes it difficult to assess all parts of the eye.
5. Worsening of cataracts may also cause an eye to become incompletely assessible, so that a change in assessibility status does not always indicate a worsening of uveitis or a failure of uveitis treatment.

We chose the following simple, but conservative, approach to scoring such eyes. For an incompletely assessible field (anterior cells, vitreous haze, or presence of retinal or choroidal lesions) at any time, the worst value seen until that time will be assigned for the unavailable measurement. Thus, a decreasing ability to assess regions of the eye—in the absence of evidence of inflammation or uveitis criteria—does not imply failure of therapy. Decreasing ability to assess eyes which had signs of uveitis will imply failure of therapy. It is understood that this procedure will misclassify some events such as: (i) an eye which had vitreous haze or a retinal or choroidal lesion at baseline, which resolved over the course of the six months, and for which a progressing cataract rendered the posterior of the eye impossible to assess, will be scored as a failure, or (ii) an eye for which the posterior region had no inflammation at baseline, which then became impossible to assess, and then which develops posterior inflammation which cannot be seen, will be scored as a success. We believe such misclassifications will be infrequent.

Selected secondary outcomes, including vision, macular edema, time to control of inflammation, will be analyzed at the eye level. All eyes that meet inclusion criteria of inflammation at baseline will be included in this analysis. Linear or generalized linear mixed modeling will be conducted (see below for details).

The following is a brief summary of general guiding principles.

- For the primary outcome, if any portion of the eye cannot be assessed at baseline, and it *still cannot be assessed at Visit 6 or Visit 12*, if all other markers of success are met, this portion of the eye would be considered to have had successful therapy.
- For the primary outcome, if any portion of the cannot be assessed *by Visit 6 or 12* and this same portion of the eye was *completely assessable* at baseline, if all other markers of success are met, then the last worst observation for this eye would be carried forward and

used at the assessment of this eye portion.

- For the primary outcome, if an eye *becomes missing by Visit 6 or 12,* and it is related to uveitis (regardless of its disease status at baseline) if all other markers of success are met, this patient should be considered a failure.

3.1.1 Demographics and Patient History

All demographic and history variables (in particular, age, gender, occupation, and ethnicity/national origin) determined at enrollment will be summarized by counts and percentages tabulated by treatment assignment.

3.1.2 Prior and concurrent medication

We will present the oral and topical corticosteroid doses at presentation (specifically, the current daily dose at baseline) and other medications by randomization arm and study site.

3.1.3 Baseline comorbidities and history

Clinical variables at baseline (in particular, anatomical site and vasculitis) will be presented by gender, age, and study site. We will also tabulate the presence of associated systemic disease at baseline. Anatomical site will be classified at the patient level as site of most serious involvement. For example, if a patient has anterior inflammation in the right eye and panuveitis in the left, they would be classified as a panuveitis patient.

3.1.4 Compliance

Compliance is assessed through patient self-report and regular pill counts by study coordinators at each visit when patients bring in their medications.

3.2 Analysis

3.2.1 Summary of Principal Outcome Variables and Regression Variables

Variables

- Primary outcome: Patient treatment success by six months (see MOP, Section 2.6)
- Patient treatment success at twelve months (Phase I)
- Successful control of inflammation in both eyes by twelve months, with complete discontinuation of corticosteroids
- Best spectacle-corrected visual acuity, at baseline and at the time of failure or six months (two observations per patient)
- Time to corticosteroid sparing control of inflammation (6 months and 12 months)
- Change in health related quality of life subscores (PCS and MCS) from SF-36 and Vision Related Quality of Life from NEI-VFQ-25 and IND-VFQ at six months and twelve months
- Reason for discontinuation of therapy (if applicable) at six months and twelve months
- Macular thickness at baseline, and at six months and twelve months

- Presence of macular edema at six months and twelve months
- Vitreous haze assessed clinically by the NEI and Davis scales at baseline, six months, and twelve months
- Vitreous haze as assessed by the photographic grading of haze by the NEI and Davis scales at baseline six months and twelve months
- The proportion of patients discontinuing due to serious adverse events at six months and twelve months
- Tabulate the occurrence of dose redution used in immunosuppressive treatment.
- Treatment efficacy of VKH patients at six months and twelve months
- Treatment efficacy of patients with vasculitis at enrollment

Note that the presence of cataracts renders assessment of vitreous haze more difficult. Vitreous haze measurements in the presence of certain cataracts will be considered less reliable, and this will be considered in statistical modeling. Analyses will be repeated for differing assumptions about this bias. A maximum likelihood latent variable model will be considered, in which a true underlying vitreous haze level predicts an observed value. The observation model will include a higher probability of yielding a large observed value in the presence of a cataract.

Major independent variable of interest

- Treatment assignment (methotrexate or mycophenolate mofetil)

Additional regression variables used in selected analyses

- Anatomic location (coded dichotomously as either intermediate (code 0) or as being either posterior uveitis or panuveitis (code 1))
- Country
- Study site
- Gender
- Age
- Baseline quality of life (health and vision related)
- Baseline best spectacle-corrected visual acuity, vitreous haze, macular thickness
- Vasculitis

Inclusion of Data

- Data will be included for all outcomes within the window peroid of -2 weeks to +4 weeks around the 6 Month Visit date for Phase I (0-6 months).
- Data will be inlcuded for all outcomes within the window peroid of -2 weks to +4 weeks around the 12 Month Visit date for Phase I (6-12 months).
- Data will be included for all outcomes within the window period of -2 weeks to +4 weeks around the 6 Month Visit for Phase II (0-6 months).

3.2.2 Specific Aim 1

Primary Analysis.

The primary analysis will be a logistic regression model, predicting treatment success at 6 months based on treatment arm. We wish to aggregate sites within countries, and countries within treatments provided we find no evidence of heterogeneity of sites or countries.

Specifically, the pre-specified primary analysis will be performed as follows. We denote the assignment group of patient i, i=1,...,N (where N is the number of subjects) by X_i^1, which equals 0 when the patient is in the methotrexate group and 1 when the patient is in the mycophenolate mofetil group. The outcome variable is Y_i, which is 1 if treatment success of patient i is achieved by six months, and 0 otherwise. The variable is missing if the patient is lost to follow-up or drops out of the study for reasons other than discontinuation due to intolerance or adverse events; if the patient discontinues the medication due to intolerance or adverse events such as abnormal laboratory findings, the value is 0.

The primary analysis is a logistic regression with treatment arm as a predictor. For the primary analysis, we propose to use study site as a *random effect* (random intercept model). The null hypothesis is that the regression coefficient for treatment arm equals zero, which will be tested using a likelihood ratio test with one degree of freedom. We will also fit the following models: (a) a model including drug, site, and drug × site interaction, (b) a model including only drug and site, (c) a model including only drug and country, and (d) a model with drug, country, drug-country interaction, and site within country. Countries or sites with fewer than three observations will be pooled together. Provided there is no evidence of treatment × site interaction or treatment × country interaction, we will report pooled treatment effects and confidence intervals. In the event evidence suggests a difference between treatment sites, we will report treatment effects by site, and repeat the analysis excluding particular sites. Similarly, evidence of a treatment × country interaction will lead us to report treatment effects and confidence intervals by country.

Simulations suggest that use of a model containing interaction terms between site or country and treatment for the primary analysis is undesirable. Such a procedure results in modest loss of power unless the treatment effect is of opposite sign in different sites or countries.

The hypothesis test is to be two-sided with alpha of 0.05. We propose to compute the P-value by permutation testing, based on the block randomization scheme.

Prespecified Subgroup Analysis.

The prespecified subgroup analysis will test the hypothesis that there is a treatment effect separately in each anatomic group, using a logistic regression model.

We denote the two anatomic groups by X_i^3, which equals 0 when the patient is in the intermediate group and 1 when the patient is in the posterior/pan group. In the intermediate subgroup, we plan to determine whether there is evidence of a treatment effect (regardless of the effect in the posterior/pan group). Specifically, we will conduct this analysis in two ways: anatomical site at enrollment (split into three categories: anterior, anterior/intermediate or intermediate only, and posterior/panuveitis), and anatomical location by history (split into two

categories: anterior/intermediate or intermediate only, and posterior/panuveitis). Anatomical location by history is considered the prespecified analysis; anatomical location at enrollment will supplement this finding.

We propose to proceed as follows. We propose to begin with Equation (1), adding terms $\beta_3 X_i^3 + \beta_{13} X_i^1 X_i^3$ for anatomic location and for treatment-location interaction. We wish to test the hypothesis that $\beta_{13} = 0$, i.e. that there is a difference in treatment efficacy between the anatomic locations, controlling for country. Alternative models will be fit in which the country, site, and treatment x country terms are omitted.

We will also report relative risks in each substratum, using relative risk regression.

Additional analyses will add gender and age to the predictors. The entire analysis will be repeated for each gender, and separately for each country (US and India) and anatomic location.

Planned Secondary Analyses.

Each of the following secondary analyses is designed to test the hypothesis that treatment assignment affects a given outcome, after controlling for selected covariates. All analyses will be repeated without controlling for covariates (i.e., using treatment assignment as the only predictor). In all cases, appropriate regression diagnostics and/or goodness of fit tests will be performed (further details are given below). In addition, we will compute jackknife influence statistics in each analysis, to determine whether or not any single observation (eyes or patients, as appropriate) have an undue effect on the final conclusion. All models with site effects will be repeated omitting this effect, and again repeated including a treatment-site effect, and with country and/or treatment by country interactions (i.e., pooling within countries when appropriate). When reporting findings, care will be taken to distinguish the single prespecified test from supplemental tests (whether prespecified or unprespecified); exploratory analyses will always be labeled as such. All alpha levels are to be two-sided.

1. *Twelve-month endpoint for successes*. We propose to compare the proportion who maintain successful control for twelve months (i.e. the outcome is the proportion who have achieved control in all study eyes both at the six month visit and at the 12 month visit) between the two study arms. Per protocol, patients with successful control of their inflammation at 6 months remain on the same treatment until 12 months. We will use the same statistical model (and Wald procedure) as for the primary analysis. We test the hypothesis that the coefficients for treatment assignment and treatment assignment/anatomic location interaction both equal zero.

2. *Time to corticosteroid sparing control of inflammation*. We propose to use a Cox proportional hazard model with the outcome being the time to (1) first steroid-sparing control, and separately (2) first control of inflammation, with treatment assignment (and interaction) as the predictors. Time to first steroid-sparing control is the principal prespecified analysis here; alternative approaches will be conducted for additional insight and as sensitivity analyses. We will supplement this analysis with a parametric survival analysis using the Weibull distribution and also with a gamma distribution (note that individuals may drop out at any time, not just at the monthly visits), and with a method treating the time to success as interval censored. The outcome for this analysis is a single number for each patient (not for each eye). The primary statistical

result will be the Wald test for the treatment assignment coefficient. We will repeat the analysis using study site as a fixed effect in this model (and as a sensitivity analysis, will explore random-effects survival analytic methods which are becoming available, see Pankratz et al.).[6] In supplementary analyses we will include age and anatomic location as additional covariates.

3. *Country and site within country.* We denote the country by X_i^2, which will be 0 for US locations and 1 for Indian locations; X_{i1}^4 is 1 only for patients in the second Indian site and 0 otherwise, while X_{i2}^4 is 1 only for patients in the second US site and 0 otherwise. As mentioned under the Primary analysis, we propose to fit models with country only, drug by country interaction, and with a drug by country interaction, including site within country as well. Analysis will be conducted within each site, then pooling the sites within country together. Further details regarding pooling across centers are provided above under the main prespecified analysis.

4. *Best spectacle-corrected visual acuity (BSCVA).* The primary outcome variable for this secondary outcome will be the change in best spectacle-corrected visual acuity from baseline to final (as defined in the FAST MOP, i.e. for those who successfully control inflammation as defined in the MOP, or at the time of failure for those who fail; MOP, Section 2.6). Visual acuity change scores are available for both eyes for each patient.

The primary analysis will use a linear mixed-effects regression, where the outcome variable is the change in BSCVA in each eye, using treatment assignment as a statistical predictor (regressor, independent variable); a random effect will be used at the individual level, because of the possibility that changes in the two eyes from a given patient are correlated. In a supplementary analysis, we will include as predictors (independent variables) anatomic location of uveitis, interaction between anatomic location and treatment assignment, and the study site, together with a random effects for patient. We will fit these models using maximum likelihood (R procedure `lmer`) and use likelihood-ratio tests to test the hypothesis that treatment assignment affects BSCVA change. Only eyes that are eligible and meet inflammation criteria at baseline will be included in this analysis. If at a given visit, vision cannot be assessed, we will carry the last observation forward. Additional sensitivity analyses for missing data will be used (including mixed effects models controlling for time, including all data from an individual).

Also, if there is no eye at Month 6 to assess, the patient will be given a logMAR value of 2.0.

Because of the possibility that the outcome variable (BSCVA change score) will exhibit non-normality, we will repeat the analyses using transformations of the outcome data (including power and log transformations, or more general monotone transformations).

Additional analyses will be performed using age, gender, ethnicity, and the steroid dose at each month as predictors.

An additional supplemental analysis will be conducted using final BSCVA (instead of the change score) as the outcome, and including baseline BSCVA in each eye as a predictor, using methods otherwise identical to those above.

5. *Quality of life.* We will also use a linear mixed model to assess health-related quality of life, measured by the SF-36 questionnaire (PCS and MCS scores) and vision related quality of life NEI-VFQ-25 and IND-VFQ at 6 months or at the time of failure, as described in the Manual

of Operations. Predictors will be baseline quality of life, age, gender, ethnicity, study site (as a random effect), and treatment assignment, and we will test the hypothesis that the regression coefficient corresponding to treatment assignment equals zero using the Wald t-test. Similar assessments will be performed for vision-related quality of life questionnaires.

6. *Reason for discontinuation.* Individuals who discontinue study medication may do so due to inability to tolerate side effects, due to lack of efficacy, or for safety reasons. The outcome variable is whether the person discontinued due to intolerance, discontinued due to lack of efficacy, discontinued due to safety, or did not discontinue the medication. Because study site may be an important factor, we will use polytomous regression to model the discontinuation result as a function of treatment assignment (using a fixed effect for study site).[7] If evidence is found that treatment assignment influences discontinuation result, further analyses may be conducted to determine whether or not treatment assignment is associated with discontinuation due to intolerance, lack of efficacy, or to safety, or some combination of these. We propose to classify all individuals in a two by four table according to treatment assignment and discontinuation (not discontinued, discontinued due to intolerance, discontinued due to lack of efficacy, discontinued due to safety) and conduct the Fisher's exact test (in its $r \times c$ form). The use of an overall test prior to further analysis is designed to protect the overall error rate.

7. *Successful control of inflammation with complete discontinuation of steroids (Phase I 6-12 months).* Some individuals may be able to taper completely off of steroids while maintaining control of inflammation. The outcome variable is the fraction of individuals achieving such control in both eyes (out of the number of individuals starting therapy). We propose to compare this fraction between the two treatment groups using logistic regression. The statistical analysis will otherwise be identical to the primary analysis.

8. *Macular edema.* We wish to compare the fraction of patients with macular edema at 6 months, between the two treatment arms. This will be conducted using the Fisher exact test, with a two-sided test at alpha of 0.05. Supplementary analyses will be based on logistic regression using the presence of macular edema as a binary outcome variable, with regressors ("independent variables") of treatment arm and anatomic location. Further analyses (including other baseline covariates or other subsets) will be labeled as exploratory.

9. *Change in Macular thickness.* We propose to test the hypothesis that macular thickness is different in the two treatment arms, at 6 months. We propose to model the macular thickness at 6 months using two regressors: treatment arm and baseline thickness. We will test the hypothesis that treatment arm is associated with final macular thickness, using the T-test of the regression coefficient for treatment arm in the model including baseline thickness as a second covariate (two sided using α=0.05). We will examine residuals for normality and homoskedasticity, and prepare residual vs fitted value plots. Standard transformations will be used in case of evidence that the assumptions have been violated.

We will also look at change in macular thickness in only patients who had macular edema at Baseline.

10. *Bayesian analysis.* Prior to data collection, we will elicit a Bayesian prior for the effect size (difference between the two treatment arms) from a group of uveitis experts, using methods our group has previously applied to the Steroids for Corneal Ulcers Trial. The likelihood

function corresponding to Equation (1) will be used to yield a posterior distribution for the effect size. Quantiles of this distribution will be reported, together with sensitivity analyses (with respect to model choice, influential observations, and prior distribution).

11. *Alternative definitions for success*. Other definitions will be examined: (i) changing the algorithm for assigning values for unobservable uveitis examination fields (anterior cells, vitreous haze, retinal/choroidal lesions) so that any worsening of ability to assess the eye for any reason is scored a failure, or (ii) use of vitreous cells in the definition of uveitis.

12. *Change in vitreous haze* will be assessed using clustered polytomous logistic regression, using baseline vitreous haze as a covariate and follow-up time. Vitreous haze is an ordinal outcome variable. A random effect is needed because the two eyes of a given patient cannot be treated as statistically independent. Both the NEI and Davis scales will be analyzed, for both direct observations and photographic grading. Treatment assignment will be a covariate. Alternative methods will be examined, including a simple McNemar test in which we dichotomize vitreous haze assessments at baseline and at the final observation.

13. *Rate of adverse events* and the proportion of patients discontinuing due to adverse events will be tabulated by treatment assignment, age, and gender; confidence intervals will be reported.

14. *Treatment efficacy in VKH patients* will be assessed as a planned subgroup analysis. Note that anatomic location is also a planned subgroup analysis, as well as study site and study country (aggregating all sites within each country).

15. *Dose reduction* will be compared by arm using logistic regression based on treatment, and other covariates as needed.

16. If no difference is found for the primary outcome comparing treatment success between arms, we will assess whether methotrexate is non-inferior to mycophenolate mofetil, assuming a 10% non-inferiority margin. The non-inferiority margin of 10% is clinically meaningful and was based on investigator consensus. Methotrexate will be considered non-inferior to mycophenolate if the lower limit of the 95% CI for treatment success at 6 months is less than 10%. This analysis will be conducted because mycophenolate mofetil is much more expensive than methotrexate, so a determination that methotrexate is not inferior has clinical implications. We are interested in a one-sided comparison given the cost differential between methotrexate and mycophenolate mofetil.

17. Additional exploratory modeling will be conducted using clustered multinomial logistic regression using all time points and all observations of anterior chamber cells, vitreous haze, and retinal/choroidal lesions.

3.2.3 Specific Aim 2

Primary Analysis.

The primary analysis will compare the proportion of successes between (a) patients treated with mycophenolate mofetil following failure on methotrexate and (b) patients treated with methotrexate following failure on mycophenolate mofetil.

Specifically, we will conduct a logistic regression in which success or failure will be the outcome, and the predictors (regressors, independent variables) will be treatment group and reason for failure of the first drug (lack of efficacy vs any other reason). Supplementary analyses

will include anatomic location (intermediate vs posterior/pan) and country. We will test the hypothesis that the coefficient for treatment group equals zero (i.e., that mycophenolate mofetil rescue after methotrexate failure has the same result as methotrexate rescue after mycophenolate mofetil failure). All alpha levels will be two-sided.

It is important to emphasize that estimation of the success rate of the second drug following the failure of the first is a central goal of the trial, arguably as or more important than the hypothesis test itself. The success rates and confidence intervals will be presented regardless of the results of the hypothesis test.

Secondary Analyses.

The following secondary analyses are planned.

We will also present the estimated success proportion in both treatment groups, together with the 95% confidence intervals. The two groups are the individuals who were undergoing methotrexate rescue therapy after mycophenolate mofetil, and those who were undergoing mycophenolate mofetil rescue therapy after methotrexate. Logistic regression will also be used to adjust for study site.

The second prespecified analysis will compare the rate of success between rescue patients and first-line patients, using logistic regression; we will test the hypothesis that the coefficient for rescue/initial equals zero. A supplemental variation of this analysis will include an additional predictor for whether the patient was on rescue therapy due to lack of efficacy, lack of safety or intolerance, or anatomic location. Two separate analyses are planned, each with an alpha of 0.05.

Exploratory and descriptive analyses of covariates such as reason for failure of the initial regime, age, disease (e.g., VKH), and affected region of the eye, will be presented.

3.3 Transformations and model adequacy

3.3.1 Primary Analysis

Sensitivity analyses based on modeling the individuals lost to follow-up will be conducted, however; we will determine how much of a treatment effect there would have had to have been in the patients lost to follow-up, for the results of the main hypothesis test to change.

3.3.2 Unspecified secondary analyses

Unprespecified analyses may be conducted following the primary analysis and will always be reported as such. Analyses will always be repeated including age and gender, in particular.

3.3.3 Model validation and sensitivity

In all cases, standard statistical procedures will always be followed to ensure that no evidence indicates a violation of the assumptions underlying the statistical models used. Specifically, we note the following: secondary analyses based on the use of age as a continuous predictor in logistic regression models with treatment success as an outcome will be assessed using the Hosmer-Lemeshow goodness-of-fit test. Linear models will always be assessed using residual plots (residuals vs. predicted values, and QQ plots), together with tests for normality (Anderson-Darling and Shapiro-Wilk procedures). For mixed models, we will examine marginal residuals, conditional residuals, and EBLUPs.[8] When modeling binary outcomes (using clustered logistic regressions), we will repeat analyses using a probit link as a check on robustness; we will also examine the Pearson goodness of fit statistic.[9] Jackknife influence estimates will be used in all analyses; single observations that could change the conclusions will always be reported. Analyses in which time to response is used as the outcome variable (in which Cox regression is conducted) will be supplemented with the Gill-Schumacher procedure for assessing the adequacy of the proportional hazards assumption for Cox regression.[10] Analyses in which our primary interest in in final outcomes will still be repeated using all available data (at all time points).

Failure of the modeling assumptions (such as normality) will result in conducting additional analyses. First, for continuous outcome variables, we will undertake normalizing or variance-stabilizing transformations of the outcome variable (such as power transformations). Second, robust procedures will be used to estimate the standard errors whenever possible. Third, the use of bootstrap procedures, when applicable, will be considered in estimation of standard errors.[11]

3.4 Sample Size Evaluation

3.4.1 Primary Calculation

The sample size for the trial will be 216 subjects, which we anticipate will provide approximately 80% power to detect a difference of 20% in the proportion of patients achieving control of inflammation at six months between the methotrexate and mycophenolate mofetil groups.

This sample size was determined based on the primary objective (superiority comparison of mycophenolate mofetil to methotrexate) and primary endpoint (treatment success). We assumed an effect size of 20%, as this was deemed to be clinically meaningful, and well within the distribution of the investigators' prior beliefs from published retrospective studies.

An approximate sample size is provided by the formula

$$2N=\frac{4(Z_{\alpha}+Z_{\beta})^{2}\bar{p}(1-\bar{p})}{(p_c-p_i)^2} \tag{5}$$

(see Friedman et al. 2010), where α is the significance level (0.05, two sided), β is one minus the power (the desired power is 80%), p_c is in this case the probability of success in the

methotrexate group (we estimate this at 0.4), p_i is the probability of success in the mycophenolate mofetil group (we estimate this at 0.6), and $\bar{p}$ is $\frac{1}{2}(p_i+p_c)$. We assume 10% will be lost to follow-up in the first six months; details are given in the full proposal. This yields approximately 108 patients in each of the two groups, for a total of 2×108=216 subjects.

A power table is provided below as a sensitivity analysis (to show how the detectable effect size changes with varying success rates.

	80% Power		90% Power	
Success rate with Drug A	**Detectable effect size**	**Success rate with Drug B**	**Detectable effect size**	**Success rate with Drug B**
20%	18%	38%	21%	41%
30%	20%	50%	23%	53%
40%	**20%**	**60%**	**23%**	**63%**
50%	20%	70%	22%	72%
60%	19%	79%	21%	81%

Simulation confirms that this method yields adequate sample sizes for the logistic regression (results not shown).

Note that for the final analysis, the critical value will be adjusted slightly because of the interim analysis.

Sample size readjustment

Simulation suggests that a baseline covariate which is associated with the outcome variable could modestly reduce the sample size needed for 80% power (simulation results are available upon request). Sample size readjustment based on baseline predictors will be considered, subject to approval by the DSMB. The guiding principle is (CHMP, Reflection Paper on methodological issues in confirmatory clinical trials planned with an adaptive design, 2007): Analysis methods that control the type I error must be pre-specified. Whenever possible, methods for blinded sample size reassessment that properly control the type I error should be used, especially if the sole aim of the interim analysis is the re-calculation of sample size.

3.4.2 Power for Subgroup Analyses and Other Analyses

Subgroups in Specific Aim 1.

The prespecified subgroup analysis for specific aim 1 is to examine the difference between the methotrexate and mycophenolate mofetil groups within each anatomic location. Using Equation (5), we anticipate having in excess of 80% power to detect a difference of 25% in success rates.

The power for selected secondary outcomes is provided here.

Secondary Outcomes in Specific Aim 1.

1. Twelve-month endpoint for success. We assume an additional loss of 5% between 6 and 12 months (that is, in addition to the 10% already lost to follow-up in the first six months). **We expect approximately 78% power to detect a 20% difference in success rates at the 12-month endpoint.**

2. Time to corticosteroid sparing control of inflammation. For sample size planning, we use the approximate formula given in Friedman et al (2010) for the number in each group:

$$N = \frac{(Z_\alpha + Z_\beta)^2(\phi(\lambda_C) + \phi(\lambda_I))}{(\lambda_I - \lambda_C)^2}$$

where λ_C is the hazard in the methotrexate group, λ_I is the hazard in the mycophenolate mofetil group, and $\phi(u) = \frac{u^2}{1 - e^{-uT}}$

where T is the censoring time (6 months). Previous studies suggest a median success time of approximately 3.5 months for mycophenolate mofetil.[12] **Assuming a 10% loss to follow-up, 108 subjects in each group provides 80% power to detect a difference of 2.47 months in the expected difference.** (Note: λ_C is log(2)/3.5 mo. for this calculation.) We assume an alpha of 0.05 (two sided).

3. Change in BSCVA. For sample size planning, we assume a T-test comparing change scores between the two drugs, assuming a standard deviation of the change in visual acuity of 6.5 letters.[13, 14] The sample size of 108 will provide approximately 80% power to detect 2.63 letters of difference in the change score. In other words, **we expect to have 80% power to detect whether mycophenolate mofetil yields 2.63 letters more of improvement than methotrexate, and we will have greater power to detect greater differences.** The power formula is provided in Chow et al and is computationally implemented in R in the function `power.t.test` (which we used).[15]

4. Quality of life. For a power calculation, we consider the SF-36 questionnaire, which has two scales, the MCS and the PCS. The raw score standard deviation will be assumed to be 8.4 points; we assume a correlation between baseline and six months of 0.6.[16, 17] Assuming that the baseline score will "explain" roughly 36% of the variance allows us to assume a corrected raw score standard deviation of 6.72 in a simplified calculation in which we treat the analysis as a T-test. The same power calculation formula used in (3) above reveals that **our sample size provides approximately 80% power to detect a raw score difference of roughly 2.57 between the two treatment groups.** This difference is roughly comparable to the small difference in scores found between intermediate uveitis patients and the general population[16], a difference we believe to be more than sufficient to detect clinically significant results. Note that the population mean of this score is standardized to 50 on 0 to 100 scale. Similar analyses will be conducted for the vision related quality of life (i.e. NEI-VFQ-25 and the IND-VFQ).

5. Rate of discontinuation. Based on retrospective studies, we expect approximately 13% to discontinue methotrexate due to tolerability and 5% to discontinue due to safety (laboratory abnormalities or other serious adverse events). We expect approximately 4% to discontinue mycophenolate mofetil due to tolerability and 5% to discontinue due to safety.[12, 18-20] For the purpose of the power calculation, we assume 10% loss to follow-up and consider only the

comparison of discontinuation due to tolerability. **We use the power formula given in Freedman et al (p. 104) to calculate a power of 61% for this comparison.[3]**

6. Macular edema. Previous studies suggest approximately 38% of individuals with uveitis will manifest macular edema.[21] **We have approximately 80% power to detect a difference of a factor of two in the final proportion of macular edema (19% vs 38%).**

7. *Macular thickness.* **A sample size of 108 (before loss to follow-up) provides approximately 80% power to detect a 65 micron difference between the two treatment groups, assuming a standard deviation of 160 microns in the final macular thickness.**[22] This analysis is quite conservative, since a difference of 100 microns between these two groups is consistent with previous studies. Moreover, adjustment for variance explained by the baseline thickness (i.e. the use of a smaller effective standard deviation) would yield a still higher effective power.[22, 23]

Specific Aim 2.

In the primary comparison of Specific Aim 2, we will estimate the effectiveness of rescue therapy, controlling for treatment group and reason for failure.

The primary analysis is (a) to estimate the probability of success on mycophenolate mofetil following failure of methotrexate, with 95% confidence intervals, and (b) to estimate the probability of success on methotrexate following failure of mycophenolate mofetil, with 95% confidence intervals.

These results will also be reported by reason for failure of the first drug, by categories of (i) failure because of inability to tolerate the first drug, (ii) failure of the first drug to achieve control (efficacy), or (iii) failure due to safety.

One analysis of interest is to compare the success rates in these two groups, and we include the sample size considerations for this analysis below. For two drugs (mycophenolate mofetil and methotrexate), we conduct the sample size planning as follows (denoting two drugs simply as A and B). For treatment group j=0,1 (0 coding drug B rescue in patients failing drug A therapy, 1 coding drug A rescue in patients failing drug B therapy), we expect $n_j = N_0 r_1 \left(1 - s_j\right) r_2$ subjects to be available for Specific Aim 2 (where N_0 is the number of subjects randomized to each treatment, r_1 is the retention fraction in Specific Aim 1 (not lost to follow-up in Specific Aim 1), s_j is the expected success fraction for patients for initial treatment j, and r_2 is the retention fraction in Specific Aim 2.

Thus, the number of available patients for Specific Aim 2 are highly dependent on the results from Specific Aim 1. Scientifically, the result of rescue therapy is important regardless of the result in Aim 1. More power will be available for the primary comparison in Specific Aim 2 if treatment in Specific Aim 1 yielded relatively high and similar failure rates for both drugs. However, even if success rates are very different in Aim 1, the descriptive analyses will still provide important information to guide decision-making on second-line treatment.

Here, N_0=108, and r_1 is assumed to be 0.9 (10% loss to follow-up). For planning Specific Aim 2, we assume a success rate of 60% for patients treated with drug A Specific Aim 1, and a success rate of 40% for those treated with drug B. This is a conservative estimate of the difference expected based on retrospective studies[12, 18-20] and consistent with the pilot study. Finally, we are assuming an additional 5% loss to follow-up during Specific Aim 2 (in addition to

the 10% already lost), so that r_2=0.95. The results are summarized in the following table, where the number enrolling does not include loss to follow-up, and the "expected complete" column has taken loss to follow-up into account (n_{jk}).

We anticipate the following:

Initial/Second Treatment	Expected Enrollment SA/2	Expected to Complete SA/2
B/A	58.3	55.4
A/B	38.9	36.9

Thus, we expect a total of n_1=58 patients (rounding down) to have failed one first-line therapy to be enrolled in rescue therapy. Similarly, we expect n_0=38 patients to be enrolled in the other rescue regimen.

Previous observational studies suggest a 42% success rate of mycophenolate mofetil in methotrexate-failing patients,.[24] A simple power analysis for comparing these proportions may be found from the formula (see Chao et al, p. 87):[15]

$$1-\beta=\Phi\left(\frac{|p_1-p_0|}{\sqrt{\frac{p_0(1-p_0)}{n_0}+\frac{p_1(1-p_1)}{n_1}}}-z_{\alpha/2}\right)$$

where p_0 is the probability of success with methotrexate rescue following mycophenolate mofetil failure, p_1 the probability of success with mycophenolate mofetil rescue following methotrexate failure, and Φ is the cumulative distribution function of the standard normal distribution. These assumptions yield a power of 0.87 if the rate of success with methotrexate rescue is 0.15. We have approximately 80% power to detect a difference of 17% if the probability of success with mycophenolate mofetil is 0.42.

A power table for sensitivity analysis is provided. We chose selected scenarios of potential interest to show the wide range of scenarios for which we have sufficient power. The main scenario is the first row of the table; in other rows, we varied the number of patients or the success fractions for the first drug used. In particular, the results are not sensitive to the efficacy difference found in Specific Aim 1.

Power Table for Specific Aim 2

Drug A, then Drug B (number)	Drug B, then Drug A (number)	Success probability of Drug B in patients failing Drug A	Success probability of Drug A in patients failing Drug B	Approximate Power

58	38	0.42	0.15	87%
58	38	0.42	0.17	80%
58	38	0.15	0.42	83%
58	76	0.42	0.15	94%
116	38	0.42	0.15	96%
40	40	0.42	0.15	80%
58	38	0.40	0.15	80%

To summarize, the anticipated number of patients from Specific Aim 1 (58 enrolled in in Drug A, and 38 in Drug B) should provide approximately 80% power to detect a difference of 25% between the two groups, assuming a success probability of 42% and a two-tailed alpha of 0.05.

Secondary Outcomes in Specific Aim 2.

1. Confidence intervals for the probability of success will be reported for each rescue group and anatomic location (i.e. Patients receiving methotrexate or mycophenolate mofetil as first treatment versus receiving it as their second, rescue treatment). Note that in the event that there are insufficient numbers of patients available in one arm of Specific Aim 2 (for instance, far fewer patients available for methotrexate rescue than we anticipate), confidence intervals for estimating the proportion of success can still be computed for the anatomic locations in the other arm.
2. We propose, for each rescue group, to conduct logistic regression using success as an outcome, and reason for failure of the first drug as a categorical covariate (safety, efficacy, tolerability). An overall likelihood ratio test for each will be conducted, with an alpha of 0.05/2=0.025.
3. An additional comparison will be undertaken between first-line and rescue patients with both methotrexate and mycophenolate mofetil.
4. Additionally, the same secondary outcomes assessed in Aim 1 will be analyzed using similar methods.

3.5 Missing data and loss to follow-up

Values of the primary study endpoint (treatment success at six months) cannot be analyzed when the individual is lost to follow-up. We distinguish information which is missing because of possible progression of the underlying condition we wish to treat from information which is lost for some other reason. Earlier, we discussed methods for handling missing values for specific uveitis fields in individuals. The discussion in this section applies only to loss to follow-up or to dropping out of the study. As emphasized in Carpenter & Kenward (2007), "there can be no universal analysis when data are missing". Our purpose is to vary the assumptions as well as the methods, to establish that the estimates of the treatment effect are robust as such assumptions are varied.

Our priority is the preservation of the intent to treat principle. We propose to report the results from all of the following methods:

1. The use of regression-based multiple imputation, based on all observed data for the patient.
2. Use of longitudinal generalized linear mixed effects regression, with visit as a covariate, and including a random effect for each person and for each eye within each person, using all the available measurements on each individual
3. Sensitivity analysis in which missing final outcome values are assigned success or failure, and the analysis conducted conditional on these assignments.
4. Analysis of complete cases only (individuals for which the six month follow-up is available)

However, we are proposing that **method 4 (complete case analysis) be considered the primary outcome**, based on recommendation by the DSMB. All other analyses are to be considered supplementary.

Multiple imputation will be conducted as follows. The following information will be used as regression covariates: (i) age, (ii) gender, (iii) inflammation assessments at all prior time points (anterior cells, vitreous haze, and retinal/choroidal lesions), (iv) steroid dose, (v) anatomic location (by patient, classified as anterior/intermediate or posterior/panuveitis), (vi) anatomic location by history, (vii) maximum steroid dose within the 90 days prior to enrollment, (vii) steroid dose at enrollment prior to randomization or study-related intervention, (ix) country, and (x) site within country. Any additional covariates must be prespecified. A regression model for the missing outcome information will be derived; specifically, a cross-validated procedure to yield the best prediction based on complete subjects will be derived, and ten multiple imputations will be derived from it. The formula in Little and Rubin[25] will be used to derive the overall test statistic. All replications will be recorded and reported.

An alternative method (which we propose to used for sensitivity analysis) is hot deck multiple imputation (with ten replications).[25] Note that treatment assignment would never be missing. For definiteness, we choose the recursive random partitioning hot deck method used in the R package `rrp` with the default settings (command `rrp.impute`).

The possibility of data-driven modeling may render multiple imputation of an outcome variable undesirable to many reviewers as a primary outcome. An alternative method is to model the treatment success of person i at visit j, Y_{ij}, using generalized linear mixed models, with covariates being site, country, treatment assignment, country-assignment interaction, visit (1:6), and visit-drug interaction (method 2 above). Note that additional statistical modeling will be reported, in which we (a) omit visit-drug interaction, and/or country assignment interaction, (b) add visit-country interaction, or (c) add age or gender as covariates.

We believe carrying forward last observations to be particularly unhelpful in this study, because all patients are on a prescribed steroid taper. We also believe that differential loss to follow-up of well performing patients on one drug or the other could falsely make the poorer drug appear to give more favorable results, so that the complete case analysis must be interpreted with caution.

3.5.1 Injections

If a patient receives a corticosteroid injection 90 days after enrollment, it is not possible to truly assess the study drug's ability to manage inflammation at the Month 6 visit. Therefore, as a senstivity analysis, the primary outcome for these patients will be considered by the inflammation levels at the time of the injection. If the patient received the injection because of uncontrolled inflammation, the patient will be considered a treatment failure. If the patient met the definition of treatment success at the time of the injection, the patient will be considered a treatment success.

3.6 Pooling across sites

Approximately three-fourths of patients are expected to come from the Aravind sites, which are in the same hospital network in the geographic region serving the same patient population. UCSF/Proctor and Oregon/Casey serve slightly different populations, although we expect fewer cases overall in the U.S. sites.

3.7 Multiple comparisons

An alpha of 0.05 will be used for the primary analysis of Specific Aims 1 and the primary analysis of Specific Aim 2. The prespecified subgroup analyses of Specific Aim 1 will be conducted at an alpha level of 0.05 (as stated above) as well. However, the use of an overall test prior to subgroup analysis protects the overall type I error rate for the primary outcome, a procedure we apply within the analysis of each secondary outcome as well.

3.8 Interim Monitoring

The study will be monitored by a Data Safety Monitoring Committee (DSMC) appointed by the National Eye Institute. There will be one in-person meeting a year and additional phone calls as deemed necessary. The DSMC will be unmasked and received reports with information by treatment arm from the principal statistician.

3.9 Accrual Rate

Based on enrollment rates in previous trials and preliminary data (see proposal for details) we anticipate enrolling 7-8 subjects per month at all sites, for a total enrollment period of 2.5 years. If we conservatively assume we may only accrue 25% fewer subjects per month, then completion of enrollment would occur 3 years and 3 months after the start of the trial.

We will establish monthly recruitment goals for each of the sites, taking into careful consideration local holidays which may cause recruitment rates to drop at certain times of the year. Careful monitoring of the recruitment process will enable us to determine whether one of our sites may be falling behind in recruitment, precursory to further investigation and intervention. Standard graphs of realized cumulative recruitment together with cumulative recruitment goals for (a) the study as a whole, and (b) for each of the four sites will be prepared, and provided to the Data and Safety Monitoring Committee at each meeting (or more frequently, if requested).

3.10 Interim Analysis

We propose to conduct two interim analyses, at approximately one-third and at approximately two-thirds of the way through the study. The exact fractions will be determined by availability of data and timing of DSMB meetings. We plan to examine the primary outcome variable using the same statistical model we plan for the final analysis. A flexible alpha spending function is specified in Section 6.

3.10.1 Stopping rules

Stopping rules for benefit, harm, and futility are discussed in Section 6.2. These rules or guidelines would be determined at the first meeting of the DSMC (see Section 6.2).

3.10.2 Execution of interim analysis

The principal statistician (TP) will conduct the interim analysis in an unmasked manner, subject to independent statistical review by the DSMC. Quality assurance will be conducted by database manager WE.

3.11 Final Analyses

The Primary Aim 1 analysis (and secondary objectives), identified in this Statistical Analysis Plan will be performed when all patients complete their 6 month assessment and the window period is completed. All other analyses will be completed after the 12 month visit for Phase I or 6 month visit for Phase II and window periods are complete.

3.12 Software

The standard software program R version 2.12 or higher (`http://www.r-project.org`) for the MacIntosh OS X will be used for all descriptive and inferential analyses.

4 Analysis Populations

4.1 Summary

The following analysis populations are planned for this study:

- The **screening population**, which is to include all patients who are screened for participation in the trial.
- The **safety population**, which is to include all patients who receive any amount of planned study medication (mycophenolate mofetil or methotrexate).
- The **intent-to-treat efficacy population**, which is to include all patients who are randomized. This is the primary population for the efficacy analyses.
- The **per-protocol efficacy population**, which is to include all patients in the intent-to-treat efficacy population, excluding patients with any of the following: (a) major protocol deviations, or (b) noncompliance with study medications (less than 50% of the study drug received by self report or pill counts at study visits).

4.2 Major protocol deviations

The incidence of deviations from the inclusion and exclusion criteria will be summarized using counts and percentages, and the treatment groups compared for the overall frequency of deviations using a $2\times N$ Fisher's exact test. Similar deviations will be grouped into general categories of deviations for a more condensed summary. A listing of deviations by participant will also be produced. Any major deviations from the protocol will be listed and/or summarized, including, but not limited to, participants who:

- never received study drug
- were subsequently found to be ineligible for the study
- never returned for a follow-up visit
- have follow-up visits outside the prescribed visit window
- received a corticosteroid injection >90 days after enrollment for macular edema or at any time for inflammation

The number and percentage of randomized participants actually receiving study medication, permanently discontinuing study drug (subdivided by reason), and receiving injections >90 days past enrollment will be summarized. A summary of study participants randomized by site will also be provided. Treatment groups will be compared for the proportion and reason for study drug discontinuation using the chi-square test. A summary of participant status at the end of the study period will also be generated with categories including lost to follow-up.

5 Data Collection and Quality Assurance

5.1 Quality assurance and security

Data collection forms, training, security, and quality assurance are discussed in the Manual of Operations for the FAST Treatment Trial.

5.2 Analysis sets

Data sets for analysis will be produced at the Proctor central site by database manager WE. Each will be a Microsoft Excel® worksheet containing a single header line whose variable names match the Access database. Each analysis set will be in the form of a rectangular table in which each column corresponds to a single variable and each row to an observation. All missing values will be coded explicitly using the string `NA` (as used in the R software). Codes for categorical variables (such as 1 for male, and 2 for female) will be avoided in favor of self-documenting character strings (such as Male, Female) whenever possible. Automated checks will be made to ensure consistency and that each variable in the analysis set has in-range values (protecting against negative ages, spelling errors in categorical factors, and similar errors).

A detailed codebook will be prepared, containing for each variable, (a) the form from which the variable derived, (b) the text of the question, (c) all possible values for the variable, and (d) summary statistics for the variable. Note that all codes and character strings that represent categorical factors will be clearly defined in the codebook. Units for each continuous variable

(e.g. central subfield thickness, logMAR) will be unambiguously indicated for each variable. Each release of the analysis set will be accompanied by the corresponding version of the codebook. Version numbering with dates will be strictly observed. Standard report-generation software included with the R statistical and data analysis package will be used to ensure consistency of the codebook and analysis set at all times.

5.3 Data monitoring reports

Data monitoring reports will be prepared based on analysis data sets. These will be prepared using report-generation software. Monitoring reports will include (a) recruitment reports for each site, (b) compliance reports, (c) retention reports, and (d) data quality reports. These will be reviewed at the central site on a monthly basis, and communicated to the study sites on a monthly basis.

6 Human Subjects

6.1 Summary of final dispositions

All subjects who provide informed consent will be accounted for in this study. The frequency of subjects in each population will be presented. We will also present the frequency of subjects in each subgroup, the frequency of withdrawal and loss to follow-up, and any major protocol violations.

6.2 Data and Safety Monitoring Committee

6.2.1 Scope

A Data and Safety Monitoring Committee (DSMC) will be empaneled by the NEI. We propose that this committee consist of 5-7 individuals, and should include (a) uveitis specialists, (b) an independent biostatistician, (c) a bioethicist, and (d) a member to protect the interest of the Indian population. The committee will meet in person at least once per year. *Ad hoc* meetings as needed may also be convened. All study protocols will be subject to review and approval by Institutional Review Boards at UCSF, Oregon and Aravind, and by the DSMC.

The Data and Safety Monitoring Committee will meet to review the interim efficacy data when primary outcome data are available on approximately one third of the study subjects—approximately 6 months after the 72nd subject has been enrolled in the trial (as discussed above in Section 3.10), and when data are available on approximately 2/3 of subjects. The DSMC will make one of the following recommendations:

- Continue the trial without modifications
- Continue the trial with study modifications
- Terminate enrollment or treatment in the trial because of safety concerns
- Terminate enrollment or treatment in the trial because of efficacy

6.2.2 Meetings

All in-person and teleconference meetings of the DSMC and study personnel will consist of (a) "open" sessions, which may be attended as needed by masked study personnel, and (b) "closed" sessions, which may only be attended by unmasked study personnel (TP, WE, TL), and (c) "closed" sessions attended only by the DSMC personnel. Care will be taken so that *no* treatment assignments, data which would allow treatment assignments to be determined, or outcome data based on treatment assignments will be revealed during the open sessions.

The DSMC will be unmasked. Closed reports will tabulate baseline covariates, adverse events, and outcomes by treatment assignment and study site. Written closed reports will always use the labels `Treatment A` and `Treatment B` for increased information security. However, the DSMC will know which drug corresponds to which label.

Interim reports for the DSMC will be prepared by the central Proctor site (TP). These reports will include (a) recruitment overall, and by study site, (b) compliance, and (c) retention. The reports will also list study outcomes, and all adverse outcomes, including deaths. The DSMC will determine the database closure dates for each report in advance; archival copies of the (a) main REDCap database, and (b) study analysis file as they exist at the time of each report will be maintained.

All reports will be sent using secure email to the members of the DSMC two weeks prior to each meeting (or more, if desired by the DSMC).

Each printed (hard copy) interim report will be labeled clearly as confidential, bound so that the contents are not visible from the outside, and labeled with the name of each person authorized to receive it. Reports will be kept in possession of WE and TP and only delivered in person or by encrypted email; reports not delivered due to absences are to be shredded. In addition, redacted versions of the interim reports will be prepared which contain no masked study information, and which are suitable for restricted distribution to other personnel on an as-needed basis. All hard copies will be destroyed at the end of each meeting, except for a copy to be kept in a locked file cabinet accessible only to TP and WE.

6.2.3 Decisions

The DSMC will make decisions with the benefit of prespecified decision guidelines. These guidelines will be agreed upon at the initial meeting, and are expected to include (a) safety, (b) efficacy, (c) clinical importance, (d) effect of baseline covariates, or (e) validity.

Benefits. Unmasked interim analyses (See Section 3.10) will be conducted to determine whether or not sufficient evidence has accumulated to justify stopping the trial because one treatment is clearly superior (and therefore should be extended to all future cases). The guidelines for efficacy will use group sequential boundaries for judging the statistical significance of the primary outcome measure. The Lan and DeMets flexible alpha spending approach will be used.

Early discontinuation in this trial has the following disadvantages. First, early discontinuation will make it more difficult to assess homogeneity of study sites. In this trial, where the majority of planned enrollment is not from the US, discontinuation at time t=1/3 for instance would occur when only 15 American patients had been enrolled (under our enrollment projections), and at t=2/3, only 30 American patients. Reflection on these small numbers of American patients may

limit the adoption of the results of the trial. Second, early discontinuation reduces the power to assess the secondary aims of Specific Aim 1, and for Specific Aim 2. For these reasons, we propose to use conservative stopping rules.

We propose to use a Hwang-Shih-deCani alpha spending function of the form

$$a^*(t) = \frac{a\left(1 - e^{-g}\right)}{1 - e^{-g}}$$

with γ chosen to be equal to -5.623626 exactly. This value was chosen to make the alpha at t=1/3 approximately equal to 0.001. The resulting alpha at t=2/3 is approximately 0.0075. The R package `gsDesign` (v. 2.7-04 or higher) will be used for selected analyses.

The proposed plan is to have two interim looks, at approximately t=1/3 and t=2/3 (one third and two-thirds through the study), with the specific fractions to depend on the total available data at face to face DSMB meetings.

The use of a flexible alpha-spending function protects the 0.05 alpha level of the overall trial while allowing for additional interim analyses for efficacy (if needed), without specifying the number and timing of the analyses at the start of the study. We note that the alpha spending function, including the value of γ, cannot be changed once the trial has begun.

Harm. Stopping for harm will be done at the judgment of the DSMC. Several endpoints will be examined, including serious adverse events such as significant and sustained laboratory abnormalities as described in the protocol, or mortality. While the analysis would consider maldistribution of predictive factors such as age, it is recognized that ethical considerations require careful considerations of statistical tests as well as qualitative judgments in the light of experience. Any additional analyses required by the DSMC will be conducted by TP and WE, as needed.

Note that serious adverse events (SAE) are reported directly to the medical monitor (TM) within 24 hours of the time the study site learns of them, and the medical monitor will subsequently pass this information on to the DSMC Chair. The medical monitor will receive notification of the event, the timing of the event, a medical narrative from the study site, the site location, and the patient identification number. The statistician will report the study treatment assignment to the DSMC Chair if deemed necessary by the DSMC. If use of either drug use clearly results in an unacceptable increase in the risk of treatment failures, then the study will be stopped. It is difficult to fully prescribe boundaries for monitoring safety because there need not be strong evidence to discontinue the study if it appears that the treatment is harmful.

Futility. Early discontinuation due to the unlikeliness of significant findings conditional on interim results would prevent the analysis of Specific Aim 2 and of the secondary aims of Specific Aim 1. No stopping rules based on futility or conditional power calculations are included in the trial plan.

7 Safety and tolerability

The analysis of safety in this study will include summaries of the following:

- Exposure
- Adverse events
    - Adverse events and serious adverse events (including deaths)
    - Adverse events leading to withdrawal
    - Any deaths

7.1 Exposure

Individuals are assumed to have exposure to the drug corresponding to the arm to which they were randomized.

7.2 Adverse Events

7.2.1 Individual events

Adverse event reporting procedures are described fully in the MOP. Non-serious adverse events (not requiring narrative form) are described in the MOP (Section 6.1). Serious non-ocular or ocular adverse events (which must be reported within 24 hours and which require a narrative form) are described in the MOP (Section 6.2). Adverse events will be reported in all presentations and publications according to Consort guidelines.

The proportion of subjects with safety-related events will be compared using logistic regression, using treatment assignment and age as predictors, and including enrollment site as a random effect. Descriptive tables of the number and frequency of adverse events will be broken down by treatment arm, age, gender, and known comorbidities. We will report total adverse events and serious adverse events, cross-tabulated by whether the adverse events were anticipated or unanticipated and by whether or not the adverse event led to discontinuation of medication.

In addition, we will compare the rate of each of the adverse events during the follow-up period using Poisson regression, which can take into account multiple instances of adverse events within a single subject. Age will be included as a predictor as well as treatment group, and enrollment site will be included as a random effect.

The additional statistical analysis of adverse events we describe here is undertaken strictly to provide additional insight which may be useful to the DSMC and investigators. Interpretation of such findings must reflect the fact that unanticipated adverse events may occur and that we may have insufficient power to make inferences between the arms when considering rare events. Note that adverse events contribute to the outcome of the trial and specific analyses have been defined earlier.

7.2.2 Pooled adverse events

Adverse events will be analyzed according to four main categories:

- Proportion of subjects with *any ocular adverse event*
- Proportion of subjects with *any serious ocular adverse event*
- Proportion of subjects with *any systemic adverse event*

- Proportion of subjects with *any systemic serious adverse event*

The proportion of subjects with these events will be compared between the arms using Fisher's Exact Test. Poisson or negative binomial regression will be applied to compare the rates of overall adverse events, including recurrent events.

8 Reporting conventions

- All tables and data listings will be presented in landscape orientation, unless presented as part of the text of the final report.
- Figures will be presented in landscape orientation, unless the information is substantially easier to interpret in portrait orientation.
- Direct annotation of figures will be preferred to legends. All figures with more than one variable or item will contain either direct annotation or legends. All annotation will be unambiguously identifiable as such.
- Color will be used in figures only when needed to enhance clarity of communication. All color schemes will be evaluated for visual clarity for individuals with diminished color vision. All color encodings will be identified. Redundant encodings (such as the use of different plot symbols or line dash patterns) will be used in addition to color, so that all figures are interpretable after monochrome reproduction at 100 dots per inch. All dash patterns and line widths will be adequate to be distinguishable after monochrome reproduction at 100 dots per inch. Any distinction between plot symbols (circles, filled circles, diamonds, etc.) will remain clear after monochrome reproduction at 100 dots per inch.
- Fixed width sans serif fonts will be used for all labeling (Helvetica, Arial, or Futura).
- Boldface and italics will not be used unless substantial value is added.
- Decorative fonts and enhancements, including borders and shading, will not be used. Decorative presentation methods, such as ribbon graphs, will never be used.
- All information given in figures will also be presented in summary tables (perhaps only included in an Appendix or in supplementary materials).
- Only standard characters will be used in tables and data listings.
- All titles will be centered. The first title line will be the number of the table, figure, or listing. The second and possibly third lines will be the description of the table, figure, or data listing. The ICH numbering convention will be used for all.
- All footnotes will be left justified and at the page bottom. Footnotes will be used sparingly. Reference footnotes will be complete enough to locate any reference based on the information provided (Author, Journal, Pages, Date, or PubMed accession number).
- Missing values for numeric or character variables will be unambiguously identified as such using the special string `NA` (not available) in all settings; `NA` is the standard missing value code for our software. Each figure or table caption in which `NA` is used will indicate the

meaning of `NA` in that figure or table. The abbreviation `NA` will never be used for any other purpose.

- All date values will presented in the form DDmmmYYYY format (e.g. 01jan2008), using four digit years. June will be encoded as `jne` (otherwise `jan` and `jun` would differ by only a single character), and July as `jly` (so that the lowercase letter `l`, easily confused with the digit `1`, will not be adjacent to any numerals).
- All tables, figures, and data listings will have the name of the program and a date/time stamp on the bottom of the output.

9 Abbreviations and acronyms

AES Advanced Encryption Standard
CAS Chemical Abstracts Service
DSMC Data and Safety Monitoring Committee
FAST First-line Antimetabolites as Steroid-sparing Treatment
FIPS Federal Information Processing Standard
ICH International Conference on Harmonization
logMAR log of minimum angle of resolution
MOP Manual of Operations and Procedures
MUTT Mycotic Ulcer Treatment Trial
NIST National Institute of Standards and Technology
SAP Statistical Analysis Plan
SCUT Steroids for Corneal Ulcers Trial
TM T. Margolis
TL T. Lietman
TP T. Porco
UCSF University of California, San Francisco
WE Wayne Enanoria

10 Appendix

All computations will be performed using the standard software package R (`http://www.r-project.org`). Statistician TP has twenty years of experience using R or very similar statistical computing environments (S, S-Plus).

Specification of the random number seed and pseudorandom number algorithm determines the entire randomization assignment (as is the case with any pseudorandom number generation method). Accordingly, the random number seed will be kept confidential, and the seed will be chosen carefully. In particular, easy-to-remember numbers or otherwise meaningful numbers (such as telephone numbers, birthdays, and so forth) are to be scrupulously avoided. The chosen seed will be used to generate the final randomization lists.

A printed copy of the randomization lists for all sites, the computer code used to generate them, and the random number seed will be maintained in a locked vault off site. The random number seed chosen will consist of at least eight digits, and a standard linear feedback shift-register algorithm will be used for pseudorandom number generation.[29]

11 Document Revision History

13 January 2015

2.3.1 Specific Aim 1

- Added secondary analysis: To evaluate a difference in treatment success controlling for vasculitis at baseline, assessed at six months.
- Added secondary analysis: To explore the use of a dynamic process model (such as a Hidden Markov model) to assess differences in control of inflammation.
- Added secondary analysis: To determine the proportion of patients beginning with at least 2+ inflammation in anterior chamber cells and experience at least a 2-step reduction
- Added secondary analysis: To determine the proportion of patients beginning with at least 2+ inflammation in vitreous haze and experience at least a 2-step reduction
- Added secondary analysis: Proportion of patients who started with at least 1+ inflammation levels in anterior chamber cells who achieve a decrease to 0 level of inflammation in anterior chamber cells
- Added secondary analysis: Proportion of patients who started with at least 1+ inflammation levels in vitreous haze who achieve a decrease to 0 level of inflammation in vitreous haze

2.4.1 Stratification between sites

- Updated: Patients will be recruited from nine sites: Aravind Madurai-AEHM, Aravind Coimbatore-AEHC, Aravind Pondicherry-AEHP, Casey Eye Institute-OHSU, King Khaled Eye Specialist Hospital-KKESH, Centre for Eye Research Australia-CERA, Asociación Para Evitar La Ceguera en México-APEC, Northwestern University-NWU, and Proctor Foundation-UCSF (see Manual of Operations Section 2.2 for details).

2.4.4 Unique patient identifiers

- Updated for correct number of sites: Unique patient identifiers will be generated as follows. The first character will be a number: "1" for CERA, "2" for UCSF, "3" for NWU, and "4" for Casey Eye Institute OHSU, "5" for APEC, "6" for KKESH, "7" for Aravind Madurai AEHM, "8" for Aravind Coimbatore AEHC, and "9" for Aravind Pondicherry AEHP.

3.2.2 Planned Analysis of Primary Outcome

- Added that study site will be treated as a random effect in the primary analysis
- Added non-inferiority analysis with a prespecified non-inferiority limit if the primary outcome is no statistically significant.

3.5.1 Injections (New Section)

- Added section: If a patient receives a corticosteroid injection 90 days after enrollment, it is not possible to truly assess the study drug's ability to manage inflammation at the Month 6 visit. Therefore, as a senstivity analysis, the primary outcome for these patients will be considered by the inflammation levels at the time of the injection. If the patient received the injection because of uncontrolled inflammation, the patient will be considered a treatment failure. If the patient met the definition of treatment success at the time of the injection, the patient will be considered a treatment success.

4.2 Major protocol deviations

- Added: received a corticosteroid injection >90 days after enrollment for macular edema or at any time for inflammation

9 March 2017

2.4.2 Randomization List

- Updated: Starting 18 January 2017, all patients will be randomized using the REDCap database. Coordinators can access only the randomization lists for their site. When patients are enrolled, the ID is logged in REDCap and the system provides the medication assignment. All emergency contacts have access to the REDCap database for their site in case of an emergency when unmasking is needed for patient safety. The REDCap database contains assignments for all previously enrolled patients. All previous randomization lists were deemed void and were destroyed. Each site submitted a certification of destruction form.

2.4.6 Provision of randomization list

- Updated: As of 17 January 2017, all randomization lists are electronic on REDCap. Once a patient is enrolled, the study coordinator logs the patient in the database and the system reveals the drug assignment, written out in full.

12 Bibliography

1. Ethical Guidelines for Statistical Practice. Alexandria, Virginia: American Statistical Association, 1999.
2. Food and Drug Administration USA. International Conference on Harmonization, Guidance on Statistical Principles for Clinical Trials. U.S. Federal Register 1998;63:49583-49598.
3. Friedman LM, Furberg C, DeMets DL. Fundamentals of clinical trials, 4th ed. New York: Springer, 2010:xviii, 361 p.
4. ISBN User's Manual. Berlin: Staatsbibliothek zu Berlin — Preußischer Kulturbesitz, 2001.
5. Bolker BM, Brooks ME, Clark CJ, et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution 2009;24:127-135.
6. Pankratz VS, de Andrade M, Therneau TM. Random-effects Cox proportional hazards model: general variance components methods for time-to-event data. Genetic Epidemiology 2005;28:97-109.
7. Skrondal A, Rabe-Hesketh S. Multilevel logistic regression for polytomous data and rankings. Psychometrika 2003;68:267-287.
8. Nobre J, Singer J. Residual analysis for linear mixed models. Biometrical Journal 2007;49:863-875.
9. Tuerlinckx F, Rijmen F, Verbeke G, de Boeck P. Statistical inference in generalized linear mixed models: a review. British Journal of Mathematical and Statistical Psychology 2006;59:225-255.
10. Gill R, Schumacher M. A simple test of the proportional hazards assumption. Biometrika 1987;74:289-300.
11. Ocana J, El Halimi R, de Villa M, Sanchez J. Bootstrapping repeated measures data in a nonlinear mixed-models context. Universitat de Barcelona: Institut de Mathematica Preprint 367, 2005.
12. Galor A, Jabs DA, Leder HA, et al. Comparison of antimetabolite drugs as corticosteroid-sparing therapy for noninfectious ocular inflammation. Ophthalmology 2008;115:1826-32.
13. Baltatzis S, Tufail F, Yu EN, Vredeveld CM, Foster CS. Mycophenolate mofetil as an immunomodulatory agent in the treatment of chronic ocular inflammatory disorders. Ophthalmology 2003;110:1061-5.
14. Samson CM, Waheed N, Baltatzis S, Foster CS. Methotrexate therapy for chronic noninfectious uveitis: analysis of a case series of 160 patients. Ophthalmology 2001;108:1134-9.
15. Chow S-C, Shao J, Wang H. Sample size calculations in clinical research. New York: Marcel Dekker, 2003:ix, 358 p.
16. Murphy CC, Hughes EH, Frost NA, Dick AD. Quality of life and visual function in patients with intermediate uveitis. Br J Ophthalmol 2005;89:1161-5.
17. Brazier JE, Harper R, Jones NM, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. Bmj 1992;305:160-4.
18. Daniel E, Thorne JE, Newcomb CW, et al. Mycophenolate Mofetil for Ocular Inflammation. Am J Ophthalmol 2009.

19. Gangaputra S, Newcomb CW, Liesegang TL, et al. Methotrexate for Ocular Inflammatory Diseases. Ophthalmology 2009.
20. Thorne JE, Jabs DA, Qazi FA, Nguyen QD, Kempen JH, Dunn JP. Mycophenolate mofetil therapy for inflammatory eye disease. Ophthalmology 2005;112:1472-7.
21. Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Sugar EA. The multicenter uveitis steroid treatment trial: rationale, design, and baseline characteristics. Am J Ophthalmol;149:550-561 e10.
22. Roesel M, Heimes B, Heinz C, Henschel A, Spital G, Heiligenhaus A. Comparison of retinal thickness and fundus-related microperimetry with visual acuity in uveitic macular oedema. Acta Ophthalmol 2009.
23. Grover S, Murthy RK, Brar VS, Chalam KV. Normative data for macular thickness by high-definition spectral-domain optical coherence tomography (spectralis). Am J Ophthalmol 2009;148:266-71.
24. Sobrin L, Christen W, Foster CS. Mycophenolate mofetil after methotrexate failure or intolerance in the treatment of scleritis and uveitis. Ophthalmology 2008;115:1416-21, 1421 e1.
25. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York: John Wiley, 2002.
26. Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. Statistics in Medicine 2001;20:3931-3946.
27. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. Stat Med 1994;13:1341-52; discussion 1353-6.
28. Lan KK, Wittes J. The B-value: a tool for monitoring data. Biometrics 1988;44:579-85.
29. Matsumoto M, Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation 1998;8:3-30.