STATISTICAL ANALYSIS PLAN (SAP)

FOR

FLuorometholone as Adjunctive MEdical Therapy for Trachomatous Trichiasis Surgery (FLAME) Trial

Funded by: National Eye Institute/National Institutes of Health with grants
UG1EY030420 (Dr. Kempen) and UG1EY030419 (Dr. Ying).

Trial Registration Number: NCT04149210

Prepared by: Gui-shuang Ying, Ph.D.
Data Coordinating Center of the FLAME
Center for Preventive Ophthalmology and Biostatistics
Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania

Version: 4.1
Last Updated: August 10, 2024

Version Number	Author	Date	Revision Description
1.0	Gui-shuang Ying	11/08/2019	Initial version
2.0	Gui-shuang Ying	12/02/2022	Modified sample size calculation due to the lower than anticipated rate of bilateral surgery. The modified sample size was approved by the DSMC
3.0	Gui-shuang Ying	10/13/2023	Modified the primary analysis of the primary outcome. The primary outcome was changed to the cumulative incidence of postoperative TT by one year. The analysis was changed to the logistic regression model for the cumulative incidence of TT by one year. These changes were approved by DSMC
4.0	Gui-shuang Ying	07/18/2024	Based on the discussion with Drs. Maguire, Kempen, Burton, and Frick, treatment adherence (>75% vs. ≤75%) was modified by using a combination of data from bottle weight change, medication diary, and patient self-reported compliance. Modified the analysis plan for the treatment effect modifiers and predictors, and provided more details on the cost-effectiveness analysis
4.1	Gui-shuang Ying	08/10/2024	Clarified timepoint for secondary outcomes and safety outcomes, and added causal inference analysis for the treatment adherence.

1. OVERVIEW OF THE STUDY DESIGN

The **FL**uorometholone as **A**djuntive **ME**dical Therapy for Trachomatous Trichiasis (TT) Surgery (**FLAME**) Trial Trial is a prospective 1:1 randomized, parallel design, double-masked, placebo-controlled clinical trial of fluorometholone 0.1% eye drops vs. placebo in eyes with trachomatous trichiasis (TT) undergoing lid rotation surgery. Key aspects of the design and rationale that have a major bearing on the approach to data analysis, statistical issues, and data monitoring are noted below:

- There are 2 treatment groups.
  - Active treatment group: Fluorometholone 0.1% one drop two times daily for four weeks
  - The placebo control group: Placebo (artificial tears) one drop two times daily for four weeks
- The unit of randomization is person, and one or both eyes will be included in the study if eligible.
- The ratio of the number of patients assigned to the active treatment group to the placebo group is 1:1.
- Stratification: by surgeon.
- The duration of the treatment period is four weeks, and the total length of follow-up is one year after randomization.
- The primary outcome is cumulative incidence of postoperative TT by one year in study eyes as determined by the trained Field Team members at four weeks, six months, and one year. Postoperative TT is defined as the presence of one or more of the following: (1) one or more lashes in the upper eyelid touching the globe; (2) clinical evidence of epilation in the upper eyelid; (3) history of repeat TT surgery in the upper eyelid.
- Secondary efficacy outcome measures in study eyes are:
  - Cumulative incidence of reoperation for postoperative TT by one year
  - Number and location of trichiatic lashes from the upper eyelid touching the globe at one year
  - Entropion (presence and extent) of the upper eyelid at one year
  - Cost-effectiveness
- Safety outcome measures are:
  - Corneal opacity in study eyes at one year
  - Overcorrection in study eyes at one year
  - Eyelid notching/eyelid contour abnormalities in study eyes at one year
  - Lid closure defect in study eyes at one year
  - Granuloma in study eyes at one year
  - Pain score in study eyes at one year
  - IOP elevation in study eyes at week 4, and cumulative incidence of IOP elevation by one year. IOP elevation at a visit will be defined in two ways: (1) IOP increase of 10 mmHg or more from baseline; (2) IOP 30 mmHg or above.
  - Cumulative incidence of cataract surgery in study eyes by one year
  - Cumulative incidence of TT in the fellow eye by one year
  - Cumulative incidence of adverse events attributed to study treatment by one year

- Patient-reported outcomes:
  - Patient satisfaction at one year
  - Cosmetic outcome at one year
  - Health utility assessed by EQ-5D at one year
- Additional outcome
  - Visual acuity with presenting correction in study eyes at one year

2. STATISTICAL HYPOTHESES

For the primary outcome measure, the null hypothesis to be tested is:

$$H0: \pi_A - \pi_C = 0$$

where π is the cumulative proportion of eyes with TT recurrence by one year, **A** refers to the active treatment group using Fluorometholone 0.1% eye drop and **C** refers to the placebo control group.

3. SAMPLE SIZE DETERMINATION

Prior to the conduct of the trial, we determined the sample size using the assumptions based on previous studies.[1-5] During the conduct of the trial, we found the percent of patients with bilateral TT surgery was lower than assumed. Following the recommendation from the Data Safety Monitoring Committee (DSMC), the sample size was re-calculated based on the cumulated masked data from the trial. Here, we present the initial sample size calculated prior to the conduct of the trial, and the re-calculated sample size during the conduct of the trial.

3.1 INITIAL SAMPLE SIZE DETERMINED BEFORE THE CONDUCT OF THE TRIAL

Sample size calculation prior to the conduct of the trial is based on the following assumptions:

- Two-sided type 1 error rate of 5%.
- Statistical power is 90%. This trial is to be considered a definitive clinical trial for assessing the efficacy of Fluorometholone 0.1% eye drop for reducing the recurrence of TT; therefore, power is set higher than the traditional 80% level because missing a true treatment effect would be a serious error.
- Cumulative incidence of TT recurrence by one year in the placebo group is 20%. Our preliminary trial observed a recurrence rate of 25% (10/39),[5] while STAR trial reported a recurrence rate of 10%.[4]
- 25% lower risk of postoperative recurrence of TT in the fluorometholone 0.1% group. In our preliminary trial, we observed 29% lower risk of recurrence of TT (21/115 in treated group, and 10/39 in control group). We anticipate that an absolute difference of less than 5% would be insufficient to motivate uptake of the intervention at programmatic levels.
- Bilaterality of TT will be present in 75% based on our preliminary study.[5]
- Correlation for recurrence between two eyes from the same patient is 0.48 (Kappa). The inter-eye correlation is 0.15 in our preliminary trial,[5] 0.48 in BLTR-PLTR trial [3], 0.40 in the Doxy trial [2], and 0.48 used by Emily Gower's trial [1].
- The comparison of cumulative recurrence rate between active treatment group vs. placebo control group will be based on per eye analysis using generalized estimating equation (GEE) to adjust for the inter-eye correlation.

- Rate of loss to follow-up is conservatively estimated at 10%. In our preliminary phase 2 trial, 148 (95%) of 155 enrolled patients completed one-year follow up [5]. We assume 10% in this trial to be conservative as conducting this large multi-center clinical trial will be more challenging than our Phase 2 trial.

Postoperative TT incidence rate by one year		Risk ratio	80% power		85% power		90% power	
Placebo group	Treated group		Patients	Effective eyes*	Patients	Effective eyes*	Patients	Effective eyes*
20%	15%	0.75	1686	1812	2021	2172	**2254**	2424
15%	10%	0.67	1304	1402	1492	1604	1745	1876

*The number of independent eyes that would provide equivalent information (after adjustment for their inter-eye correlation).

Based on a comparison of proportions of TT recurrence in an eye by one year and these assumptions, 2424 effective eyes are needed to detect the 25% reduction (15% vs. 20% recurrence rate) in TT recurrence with fluorometholone treatment.

Based on the assumptions about inter-eye correlation, bilaterality of TT, and applying a 10% inflation of the sample size for loss to follow-up, the total number of participants = [(Eyes*1.48)/1.75]*1.1 = 2254 are needed (i.e., 1127 participants in each of the two groups). From these 2254 patients, total of 3944 eyes (i.e., 2254 * 1.75=3944 assuming 75% of bilateral patients with both eyes eligible for the study) will be enrolled into the trial.

3.2 RE-CALCULATED SAMPLE SIZE DURING THE CONDUCT OF THE TRIAL

The sample size was re-calculated using the following updated assumptions based on the interim analysis results from cumulated masked data by October 26, 2022 when 932 participants were enrolled and 508 of them completed the study:

- The rate of bilateral surgery is 35%, as observed in the 932 enrolled participants.
- The cumulative postoperative TT rate in study eyes by one year in the placebo group is 18-20%. Among the 684 study eyes from 508 participants who completed the one-year follow-up, the cumulative incidence rate of postoperative TT (in placebo group and fluorometholone 0.1% group combined) is 19.2% by one year.
- The inter-eye correlation in postoperative TT is 0.46. Among 176 bilateral participants who had both eyes enrolled into the study and completed the one-year follow-up, the inter-eye correlation in postoperative TT by one year is 0.46.
- The rate of loss to follow-up is approximately 2% as observed in the study.

Cumulative Postoperative TT incidence rate by one year				Number of Participants Enrolled	Number of Study Eyes
Placebo group	FML group	Rate Difference	% Reduction		
20.0%	15.0%	5.0%	25%	2190	2956
19.0%	14.3%	4.7%	25%	2383	3217
18.5%	13.9%	4.6%	25%	2434	3286
18.0%	13.5%	4.5%	25%	2484	3354

The sample size for the postoperative TT rate ranging from 18% to 20% in placebo group are shown in the above table. We targeted for enrolling approximately 2383 participants (3217 study eyes) to provide 90% power for detecting 25% reduction (from 19.0% to 14.3%) in cumulative postoperative TT rate with fluorometholone treatment.

No provision is made for "alpha-spending" in the sample size calculation based on our plan that no formal "looks" for efficacy will be conducted. This design derives largely from the study schedule, in that by the time one-year outcome data are available on 50% of the population, enrollment will be largely complete, and treatment lasts only four weeks after the last enrollment. Given that large programs would need to weigh expense against improved outcomes, lack of early stopping for efficacy will be beneficial for study impact by providing an optimally precise estimate of the treatment effect.

4. STATISTICAL ANALYSES

4.1 General Approaches to Statistical Analysis

Continuous measures will be summarized using mean, standard deviation (SD), and quantiles. Categorical measures will be summarized by proportions.

Analysis of variance will be used for comparing means and X^2 tests or Fisher Exact tests for comparing proportions.

The general guidelines for the comparison of primary and secondary outcomes between two treatment groups include:

- Analyses comparing treatment groups will follow the intention-to-treat (ITT) principle; that is, all patients will be analyzed in the group to which they are assigned regardless of the compliance of using the fluorometholone eye drops. However, exploratory sensitivity analyses will be performed using other approaches including per protocol analysis and imputation of missing data for those without postoperative TT outcome. In the rare situation that a participant was randomized but TT surgery was not performed, thus no study medication or placebo was provided, the participant will not be included in any statistical analysis of the study.
- The analysis of the primary outcome will be performed in both full analysis set (FAS) and per protocol set (PPS) as defined below. The analysis in FAS will be the primary analysis, and analysis in PPS will be the sensitivity analysis.
  - Full analysis set (FAS): The FAS will include all the participants who received TT surgery. The primary analysis will be based on the observed data. The sensitivity analysis will be based on the imputation of missing data in postoperative TT.
  - Per protocol set (PPS): The PPS will include participants who are in the FAS and have postoperative TT outcome data (i.e., completed the 12-month visit or have developed postoperative TT before one year) and without major protocol deviations that may influence primary outcomes. The participants who had poor treatment adherence (e.g., <75% of anticipated study medication or placebo) will be excluded from the per-protocol analysis. In this PPS cohort, the baseline characteristics will be compared between treatment groups and the substantially imbalanced baseline variables (e.g., absolute standardized mean difference between two treatment groups >0.25 [12]) will be included in the repeated measures logistic regression models when evaluating the treatment effect on postoperative TT. In addition, among all trial participants, the analysis for the causal inference of treatment effect using the approach of inverse probability weighting (IPW)[6, 7] will be performed as described in section 4.3.2.5.

- For the eye-specific primary or secondary outcomes, both eyes of bilateral patients (i.e., both eyes eligible and enrolled for the trial) and one eye of unilateral patients (i.e., only one eye eligible for the trial) will be included in the analyses. In addition, the outcome measures evaluated at multiple visits will all be included in the analyses as appropriate. The statistical methods that accommodate the correlation in outcomes between eyes and the correlation from repeated measures over time by using the generalized estimating equation (GEE) approach will be used [8-10]. In GEE, the time will be modelled as categorical (as we do not expect the relationship of outcome measures with time is linear), and the robust sandwich estimate of variance using an independent correlation structure will be used for calculating the odds ratio (95% confidence interval) and the p-value.
- Because the randomization was stratified by the surgeon, the surgeon will be modeled as a covariate in the GEE model. [11]

4.2 Baseline Descriptive Analysis

Tables will be generated and inspected to compare, by treatment group, the distribution of key baseline variables having descriptive and prognostic importance. These variables will include, but not be limited to, patient age, gender, severity of trachomatous trichiasis (TT), medication use, presence of epilation, presence of entropion, number of eyelashes touching the globe, number of eyelashes touching the cornea, location of trichiatic eyelashes, cornea opacity, visual acuity, ocular surface characteristics (presence of discharge, presence of conjunctivalization, presence of upper eyelid papillary hypertrophy, extent of follicles, extent of eyelid palpebral conjunctival scarring), etc.

Patient-level comparison of baseline characteristics between two treatment groups will use standard statistical techniques for comparing two independent groups: X^2 tests or Fisher's exact test (well count in a cell is less than 5) for equality of proportions, independent t-test for equality of means, Wilcoxon rank sum tests for skewed data. Eye-level comparison of baseline ocular characteristics will use the generalized estimating equations to account for the inter-eye correlation.[8] The distribution of continuous variables will be assessed by measures of normality and graphical displays so that non-parametric methods or data transformations can be applied when appropriate. In addition, the absolute standardized mean difference will be calculated to evaluate the balance of baseline characteristics between two treatment groups. [12]

4.3. Data Analyses of the Primary Outcome Variable

4.3.1. Primary analysis of primary outcome

The primary outcome measure is the cumulative incidence of post-operative trachomatous trichiasis (TT) in study eyes, defined as one or more eyelashes touching the globe in the upper eyelid, or clinical evidence of epilation in the upper eyelid, or a history of repeat trichiasis surgery in the upper eyelid of study eyes by one year after the baseline TT surgery.

In this trial, some patients will undergo concurrent surgery in both eyes with TT, and their treatment outcome (i.e., recurrence of TT) is likely to be correlated.[2] The primary assessment of efficacy will be the comparison of cumulative proportion of incident post-operative TT by one year post-surgery between the two treatment groups using a repeated measures logistic regression model, where the inter-eye correlation will be accounted for through generalized estimating equations (GEE) [3-5]. The repeated measures logistic regression model will be executed by using PROC GENMOD in SAS with the stratification factor Surgeon as a fixed-effect covariate, and the robust sandwich variance estimate will be calculated by using the option of TYPE=IND in the REPEATED Statement to account for the inter-eye correlation. The difference for cumulative incidence rate of post-operative TT by one year, the odds ratio and its 95% confidence intervals for the comparison between two treatment groups will be calculated from the repeated measures logistic regression model.

It is expected that the important baseline characteristics (those known to affect the risk of TT recurrence) will be balanced between the two treatment groups by stratified randomization (e.g. stratified by the

operating surgeon). If this is the case, no baseline covariates will be included in the logistic model. If two treatment groups are found to be substantially imbalanced with respect to baseline covariates (e.g., absolute standardized mean difference between two treatment groups >0.25 [12]), the imbalanced baseline variables will be included in the repeated measures logistic regression model.

4.3.2. Secondary analysis of primary outcome:

To fully evaluate the treatment effect on the primary outcome, we will perform secondary analyses of the primary outcome as described below.

4.3.2.1 Effect modification for primary outcome

To check the consistency of results over subgroups, we will assess the effect modification of the treatment on the primary outcome (cumulative TT recurrence over one year) with the following factors by including the treatment group indicator, subgroup indicator, and their interaction term in the repeated measures logistic regression model as described above. This analysis will be performed for each of the following potential effect-modifying variables separately. If we find any important interactions, stratum-specific recurrence rate by treatment group, their odds ratios (ORs), and 95% confidence intervals for treatment effect will be reported for each level of the effect-modifying variable.

a. Baseline upper eyelid trichiasis severity with two severity categories (severe: 6 or more in total number of upper eyelid lashes or epilation >⅓; not severe: 5 or less in the total number of upper eyelid lashes and epilation <1/3).

b. Baseline conjunctival (papillary) inflammation (presence or absence)

We will perform the statistical test for the effect modification from upper eyelid trichiasis severity and conjunction inflammation at a two-sided type I error rate of 0.025 (to correct for 2 tests of interaction). We do not expect the treatment effect modification from biological variables (age and gender), thus we will not perform formal statistical test for their effect modification. However, following NIH guidelines, we will present the primary outcome results for each age groups (<50 years, ≥50 years) and for gender (male, female).

4.3.2.2. Predictors for recurrence:

Predictors of TT recurrence to be evaluated will include:

a. Type of TT surgery (BLTR, PLTR)

b. Baseline upper eyelid trichiasis severity with 2 levels (severe: 6 or more in the total number of upper eyelid lashes or epilation >⅓; not severe: 5 or less in the total number of upper eyelid lashes or epilation <1/3)

c. Baseline severity of entropion in the upper eyelid with 3 levels (none/mild, moderate, severe/total)

d. Baseline corneal opacity with 3 levels (C0, C1, C2 or worse)

e. Baseline trachomatous trichiasis at lower eyelid (Yes, No)

f. Baseline trichiasis lashes touching the cornea (Yes, No)

g. Baseline conjunctival (papillary) inflammation (presence or absence)

h. Sex (male, female)

i. Age (as continuous)

j. Literacy with 2 levels (able to read, unable to read)

Univariable and multivariable logistic regression models will be used to evaluate the above predictors for recurrent trichiasis by one year. The predictors with $p<0.10$ from univariate analysis will be included in the initial multivariable model, which will go through backward variable selection by only keeping the statistically significant predictors with $p<0.05$ in the final multivariable model. In all these analyses, the inter-eye correlation will be accounted for using GEE.

4.3.2.3. Recurrence over time at four weeks, six months, and one year

Intention-to-treat analysis will be performed at each follow-up time point separately (four weeks, six months and one year) to assess the consistency of treatment effect over time on TT recurrence by using repeated measures logistic regression models. The stratification variable (surgeon) along with imbalanced baseline prognostic factors will be included as covariates in the repeated measures logistic regression model to estimate the adjusted odds ratio for comparing the TT recurrence between fluorometholone 0.1% and placebo. In addition, the data from all follow-up time points (four weeks, six months and one year) will be combined for longitudinal data analysis with and without adjustment of imbalanced baseline prognostic factors.

4.3.2.4. Analysis for early and late post-operative TT

If the primary analysis finds fluorometholone 0.1% eyedrops is effective in reducing the risk of post-operative TT, we will investigate whether the effect of the fluorometholone 0.1% changes over time by performing the above comparisons between two treatment groups for the incidence of early post-operative TT (defined as TT occurred within 6 months post-surgery) and later post-operative TT (defined as postoperative TT occurred after 6 months post-surgery) separately.

4.3.2.5. Analysis by treatment adherence

A analysis will be performed based on the level of adherence to randomized treatment with fluorometholone 0.1% or placebo. Adherence will be categorized into two levels (>75% adherence, ≤75% adherence) using the bottle weight change, medication diary, and self-reported adherence by applying the following hierarchical algorithm:
(1). If the bottle weight change is known and ≤75% of expected doses, then the patient has <75% adherence.
(2). If the bottle weight change >75% of the expected change then >75% adherence.
(3). If the medication diary indicates >75% and the bottle weight change is unknown, then >75% adherence.
(4). If the bottle weight is unknown and the medication diary is ≤75%, then <75% adherence.
(5). If the medication diary is unknown and the bottle weight is unknown, use the self-report adherence and only classify patients with self-reported compliance of “Very Good” as >75% adherence.

Treatment effect will be evaluated by comparing cumulative postoperative TT between active treatment group vs. placebo group among participants with treatment adherence level >75%, and also among participants with a treatment adherence level ≤75% as defined above. In these analyses, the baseline characteristics will be compared between treatment groups and the substantially imbalanced baseline variables (e.g., absolute standardized mean difference between two treatment groups >0.25 [10]) will be included in the repeated measures logistic regression models when evaluating the treatment effect on postoperative TT.

In addition, among all the participants (e.g., those with >75% adherence and those with ≤75% adherence combined), we will perform the analysis of causal inference for treatment effect using the approach of inverse probability weighting (IPW).[6, 7] For this analysis, we will first calculate the propensity score (PS) of ≤75% adherence for each study eye using the logistic regression model that include baseline demographics and clinical characteristics of TT as predictors. We will then calculate IPW use IPW=1/PS for study eyes with ≤75% adherence, and use IPW=1/(1-PS) for study eyes with >75% adherence. We will then use the IPW to estimate the average treatment effect among all study eyes.

4.3.2.6. Sensitivity analyses

Although every effort will be made to encourage patients to complete all the follow-up visits, we expect a small percent (<10%) of patients may be lost to follow-up or will not comply with the trial protocol during the study. Analyses will be performed to assess the robustness of the results with respect to dropouts and non-compliance with the eligibility criteria and the treatment protocol.

Analysis of primary outcome data from all patients who complete the 1-year follow-up (completed cases) will be performed with their treatment group assignment classified as assigned at randomization ("intent-to-treat"). Also, a "per protocol" analysis, including only those patients who met all eligibility criteria at baseline and completed the assigned treatment as specified in the protocol, will be performed.

Sensitivity analyses will be performed by using multiple imputation methods [13, 14] for those who dropped out of the trial. The propensity score method will be used to evaluate the impact of missing data. Further sensitivity analyses will be conducted using pattern mixture models for missing data if there are indications that data are not missing at random.[15]

4.4 Data Analyses of Secondary Efficacy Outcomes

Specific secondary efficacy outcome measures for the trial are: (1) cumulative incidence of reoperation for post-operative TT by one year; (2) the number and location of trichiatic lashes from the upper eyelid touching the globe at one year; and (3) entropion (presence and extent) of the upper eyelid at one year.

Descriptive statistics will be used for summarizing these secondary outcomes by treatment groups. For the statistical comparison of these eye-specific secondary outcomes between treatment groups, we will use the generalized linear model through generalized estimating equations [16, 17] to account for the inter-eye correlation. The stratification factor (operating surgeon) will be included as a covariate in the model [11]. These comparisons of secondary outcomes will be based on a Binomial model for the binary outcome (e.g., the incidence of post-operative TT), a Poisson model for the count outcome (number of lashes touching the globe), and multinomial models for the ordinal outcome (e.g., extent of epilation). The 95% confidence intervals for differences in these secondary outcomes between two treatment groups will be calculated from these model-based analyses.

4.5 Data Analyses of Safety Outcomes

The safety outcomes to be compared between treatment groups include corneal opacity in study eyes at one year, overcorrection in study eyes at one year, eyelid notching/eyelid contour abnormalities in study eyes at one year, lid closure defect in study eyes at one year, granuloma in study eyes at one year, pain score in study eyes at one year, the incidence of IOP elevation in study eyes at week 4, and cumulative incidence of IOP elevation by one year, the cumulative incidence of cataract surgery in study eyes by one year, the cumulative incidence of TT in the fellow eye by one year, and cumulative incidence of adverse events attributed to study treatment by one year. These safety outcomes include different types of data (categorical, ordinal, continuous). Descriptive statistics will be used for summarizing these safety outcomes by treatment groups. For safety outcomes evaluated at eye level over time, the statistical comparison between treatment groups will use the generalized linear model through generalized estimating equations [16, 17] to account for the inter-eye correlation. These comparisons will be based on a Binomial model for the binary safety outcome (e.g. presence of over-correction), multinomial models for the ordinal outcome (e.g., the severity level of cornea opacity, eyelid notching severity), and a Gaussian model for the normally-distributed continuous outcome (e.g., pain score).

For systemic adverse events evaluated at the patient level, the logistic regression will be performed. For ocular adverse events evaluated at eye level, the inter-eye correlation will be accounted for using generalized estimating equations. For the comparison of patient-specific systemic adverse events that

are uncommon (for uncommon event <5), we will use Fisher's exact tests for comparing between treatment groups for the incidence of adverse events, serious adverse events, and adverse events attributed to treatment.

The 95% confidence intervals for differences in safety outcomes between two treatment groups will be calculated from these model-based analyses when appropriate.

4.6 Data Analyses of Patient-Reported Outcomes

Patient-reported outcomes include: (1) patient overall satisfaction at 1 year, (2) cosmetic outcome at one year, and (3) Health utility assessed by EQ-5D at 1 year. Per-eye level reported outcomes (e.g. cosmetic outcome etc.) will be compared following the same approaches for eye-specific outcomes in sections 4.4 and 4.5. For patient-level reported outcomes (e.g., patient overall satisfaction score, health utility), the comparison between treatment groups will use the generalized linear models that include the surgeon as covariate, and 95% confidence intervals for differences in these patient-reported outcomes will be calculated from these regression models.

4.7 Data Analyses of Exploratory Outcome

The exploratory outcome include visual acuity with presenting correction at one year. We will check the distribution of visual acuity using histograms. The visual acuity in logMAR will be summarized using mean (SD) if normally distributed, and using median (inter-quartiles) if not normally distributed. The visual acuity also will be categorized into several clinically relevant levels (e.g., normal vision, vision impairment, blindness). The comparison of visual acuity (in LogMAR) between treatment groups will be performed using generalized linear models that account for the inter-eye correlation through generalized estimating equation.

4.8 Data Analyses for Cost-Effectiveness

We will perform an analysis comparing costs and effectiveness regardless of whether fluorometholone 0.1% is found to reduce the postoperative TT at a statistically significant level. The cost of the intervention itself will be calculated as the cost of the drops of fluorometholone 0.1% that are dispensed to patients. The price of the generic drug will be based on what its importers charge for fluorometholone 0.1% in Ethiopia. The cost of post-index surgery medications related to the eye will be derived from the surgeon-prescribed medications for each patient. The prices used also will be the costs to the program in Ethiopia. The cost of any additional medical care utilization (not including study visits) will include utilization based on self-report (including control of IOP elevation, hospitalization or care for other SAEs) and a price based on standard use of a clinic in Ethiopia. Costs of reoperations for TT specifically will be considered. All costs will be converted to US dollars based on the international exchange rate at the time the data were collected and averaging over the duration of the data collection; we will perform a sensitivity analysis using GDP-Purchasing Power Parity in the most recent year to convert to dollars. The outcomes will be cases of post-operative TT averted. If the proposed intervention is more effective and more expensive, the incremental cost-effectiveness ratio will be calculated as the difference in mean costs between two treatment groups. divided by the difference in TT cases between two treatment groups. If one intervention is more effective and less expensive then it will be described as dominating and is the obvious economic choice. We also will calculate quality adjusted life years (QALYs) based on the change of EQ-5D from baseline for the intervention and control groups and take the difference between the two treatment groups' means with the same specification for analyses depending on whether one alternative is more effective and more expensive or not. The incremental cost-effectiveness ratio will be expressed as dollars spent per QALY gained. Finally, we will conduct sensitivity analyses varying one variable at a time to determine whether any reasonable variation in costs or any observed variable being at the high or low end of a confidence interval rather than at the mean would change the interpretation of the economic value of the intervention (i.e., it would go from being cost-effective to not or vice versa.) For variables like prices, there will not be confidence intervals, but we can increase and decrease the price by 50% of the base case amount. If a change in one variable at a time or a change when making all variables either the most or least favorable to fluorometholone would change the qualitative interpretation of the results, we will

bootstrap the results to characterize the level of certainty about the conclusion using a cost-effectiveness acceptability curve. Dr. Kevin Frick, a well-known health economics and a professor of Carey Business School at Johns Hopkins University will oversee the DC for performing these cost-effectiveness analyses.

4.9 Handling Missing Data

Major efforts will be made by the entire study group to avoid loss to follow-up and subsequent missing data. However, despite these efforts some data for the primary and secondary outcome measures may be missing. The percentage of data missing for major analyses will be tabulated. The characteristics at baseline, and during follow-up, of patients who ultimately are unavailable for follow-up will be assessed by comparing distributions between those under follow-up to those who are lost to follow-up. When available, the reasons for loss to follow-up will be reviewed. If missing data account for more than a small percentage of expected data (>5%), key analyses will be performed not only with the actual observed data on patients under follow-up, but also using multiple imputation methods.[13, 18] The propensity score method will be used to evaluate the impact of missing data on the key analyses of the study. In the multiple imputation using the propensity score method, the conditional probability of missing outcome (i.e., the propensity score) will be first calculated for each subject from a logistic regression model with outcome data missing Yes/no as the dependent variable, and the baseline demographic and clinical variables as predictors. The missing outcome data will be then imputed based on the observed outcome data from subjects with similar propensity scores as those with missing outcome data.[18]

4.10 Identification of outliers, incorrectly collected data, and possibly fraudulent data

With each freeze of the database, a set of statistical and data analytic algorithms will be applied to detect data warranting further investigation and/or action. The identified outliers or fraud data will be investigated with study coordinators for data recording error, data entry error etc. As part of the preparation for any of the data analyses, continuous variables, including dates, are subjected to the techniques of exploratory data analysis in order to fully understand the distribution of the variable. If the outlier values are valid, statistical methods that minimize the impact of outliers will be used.

4.11 Software for Statistical Analysis

SAS/STAT software (SAS Institute, Inc., 100 SAS Campus Dr., Cary, NC, 27513-2414) will be used for performing most statistical analyses. When the application can be accommodated more easily by other software packages, Stata[19], R[20], or Mplus[21] will be used.

5. DATA MONITORING

An independent Data and Safety Monitoring Committee (DSMC) will monitor the trial following “NIH Policy For Data And Safety Monitoring” - release date: June 10, 1998) and the “National Eye Institute Guidelines for Data and Safety Monitoring of Clinical Trials” NOTICE: EY-01-002, release date March 2001. The first DSMC meeting(s) will be held prior to the start the trial to review and approve the trial protocol (with any needed revisions implemented prior to approval). DSMC meetings will be held semi-annually throughout 5-year trial to review trial protocol and the accumulated safety and efficacy data.

Because data on the primary outcome will not be available for 50% of the subjects until about the time that the last-enrolled subjects complete study treatment, the study protocol does not include “looks” for stopping for efficacy nor does the sample size calculation include “α-spending” concepts; nevertheless, the DSMC may take the actions they deem proper in carrying out their role. However, we will use the predictive power approach to consider stopping the trial early due to futility.[22]

There will be no formal statistical guidelines for stopping the trial early because of safety considerations. The magnitude of the difference in safety outcomes, as well as their severity will be considered in deciding whether the trial should be stopped.

REFERENCES

1. Gower, E.W., et al., *Trachomatous trichiasis clamp vs standard bilamellar tarsal rotation instrumentation for trichiasis surgery: results of a randomized clinical trial.* JAMA Ophthalmol, 2013. **131**(3): p. 294-301.
2. Habtamu, E., et al., *Oral doxycycline for the prevention of postoperative trachomatous trichiasis in Ethiopia: a randomised, double-blind, placebo-controlled trial.* Lancet Glob Health, 2018. **6**(5): p. e579-e592.
3. Habtamu, E., et al., *Posterior lamellar versus bilamellar tarsal rotation surgery for trachomatous trichiasis in Ethiopia: a randomised controlled trial.* Lancet Glob Health, 2016. **4**(3): p. e175-84.
4. Woreta, F., et al., *Three-year outcomes of the surgery for trichiasis, antibiotics to prevent recurrence trial.* Arch Ophthalmol, 2012. **130**(4): p. 427-31.
5. Kempen JH, T.-H.R., Hunduma L, Shifferaw MA, Pistilli M, Abashawl A, Lawrence SD, Alemayehu W, *Fluorometholone 0.1% as Ancillary Therapy for Trachomatous Trichiasis Surgery: Randomized Clinical Trial.* Am J Ophthalmol, 2018.
6. Robins, J., *A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period - Application to Control of the Healthy Worker Survivor Effect.* Mathematical Modelling, 1986. **7**(9-12): p. 1393-1512.
7. Chesnaye, N.C., et al., *An introduction to inverse probability of treatment weighting in observational research.* Clinical kidney journal, 2022. **15**(1): p. 14-20.
8. Liang, K.-Y. and S.L. Zeger, *Longitudinal Data Analysis Using Generalized Linear Models.* Biometrika, 1986. **73**(1): p. 13-22.
9. Ying, G.S., et al., *Tutorial on Biostatistics: Linear Regression Analysis of Continuous Correlated Eye Data.* Ophthalmic Epidemiol, 2017. **24**(2): p. 130-140.
10. Ying, G.S., et al., *Tutorial on Biostatistics: Statistical Analysis for Correlated Binary Eye Data.* Ophthalmic Epidemiol, 2018. **25**(1): p. 1-12.
11. Kernan, W.N., et al., *Stratified randomization for clinical trials.* J Clin Epidemiol, 1999. **52**(1): p. 19-26.
12. Stuart, E.A., B.K. Lee, and F.P. Leacy, *Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research.* J Clin Epidemiol, 2013. **66**(8 Suppl): p. S84-S90 e1.
13. Rubin, D.B., *Multiple Imputation for Nonresponse in Surveys*. 1987, New York: John Wiley & Sons.
14. DeSouza, C.M., A.T. Legedza, and A.J. Sankoh, *An overview of practical approaches for handling missing data in clinical trials.* J Biopharm Stat, 2009. **19**(6): p. 1055-73.
15. Carpenter, J.R., J.H. Roger, and M.G. Kenward, *Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation.* J Biopharm Stat, 2013. **23**(6): p. 1352-71.
16. Breslow, N.E. and D.G. Clayton, *Approximate Inference in Generalized Linear Mixed Models.* Journal of the American Statistical Association, 1993. **88**(421): p. 9-25.
17. Ten Have, T.R., A.R. Kunselman, and L. Tran, *A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering.* Stat Med, 1999. **18**(8): p. 947-60.
18. Lavori, P.W., R. Dawson, and D. Shera, *A multiple imputation strategy for clinical trials with truncation of patient data.* Statistics in Medicine, 1995. **14**(17): p. 1913-1925.
19. StataCorp., *Stata Statistical Software: Release 15.* 2017: College Station, TX.
20. R Development Core Team *R: A Language and Environment for Statistical Computing.* 2011.
21. Muthén, L.K., Muthén, B. O., *Mplus User's Guide. Sixth Edition*. 1998-2011, Los Angeles, CA.: Muthén & Muthén.
22. Snapinn, S., et al., *Assessment of futility in clinical trials.* Pharm Stat, 2006. **5**(4): p. 273-81.