MYOPIA TREATMENT STUDY
MTS1

Low-Dose Atropine for Treatment of Myopia

STATISTICAL ANALYSIS PLAN

Version 2.1
November 28, 2022
Based on Protocol Version v5.1 (April 06, 2020)

Revision History

VERSION NUMBER		AUTHOR	APPROVER	EFFECTIVE DATE	REVISION DESCRIPTION (INCLUDING SECTIONS REVISED)
SAP	Protocol				
1.0	V4.0 (Oct 4, 2019)	Rui Wu	Michele Melia	April 27, 2018	Initial version
2.0	V5.1 (April 6, 2020)	Rui Wu	Michele Melia	June 15, 2020	The following revisions from SAP Version 1.0 have been highlighted in this updated version. <u>Section 1</u> 1.It has been specified that if the fully-adjusted model failed to converge, or displayed evidence of estimate instability due to partial aliasing, the race/ethnicity and iris color variables would be combined as follows: East Asian (regardless of eye color), non-East Asian with brown eyes, non-East Asian with non-brown eyes. The number of East Asians with non-brown eye color would be tabulated and reported. 2.The maximum likelihood method is specified for the mixed model. 3.It is specified that the tipping point analysis will be conducted only if more than 10% of primary outcome data are missing. Details of implementing the tipping point analysis have been added.

VERSION NUMBER		AUTHOR	APPROVER	EFFECTIVE DATE	REVISION DESCRIPTION (INCLUDING SECTIONS REVISED)
SAP	Protocol				
					Section 3 The overall false discovery rate (FDR) in the secondary analyses will be controlled at the 5% level using the two-stage step-up FDR procedure. Section 4 1.It is specified that the subgroup analysis based on baseline age and baseline SER will be in 4 categories based on median split. 2.It is specified that an estimate of the treatment effect within each subgroup and a 95% confidence interval will be obtained for each of the times, 24 and 30 months, by adding a treatment by time by subgroup interaction into the primary analysis model and generating the appropriate contrasts. Each subgroup effect will be estimated separately, one subgroup per model. 3.The overall FDR in the subgroup analyses will be controlled at the 5% level using the two-stage step-up FDR procedure. Section 5 The overall FDR in the additional analyses will be controlled at the 5% level using the two-stage step-up FDR procedure. Section 6 1.In the previous version, it was specified that the average of the item responses at the 6-month visit would be calculated and compared using a Wilcoxon Rank Sum test for difference between the treatment groups. In the current version, it has been updated to comparing the average of the item responses at 24-month visit using a t-test to be consistent with the protocol. 2.The adjustment of FDR has been removed given that there is only one average score for the questionnaire. Section 7 1.Given that by the time sufficient data allowing for an analysis of efficacy or futility for the primary outcome would be possible, all or most subjects are expected to be within 1 year of completing the primary outcome, no formal statistical interim monitoring is proposed. 2.It has been added that the tabulated safety and efficacy data will be reviewed by the PEDIG Data and Safety Monitoring Committee at its biannual meetings,

VERSION NUMBER		AUTHOR	APPROVER	EFFECTIVE DATE	REVISION DESCRIPTION (INCLUDING SECTIONS REVISED)
SAP	Protocol				
					and if any unexpected safety or other issues arise, they can recommend to stop the study at any time.
2.1	V5.1 (April 6, 2020)	Rui Wu Rui Wu Digitally signed by Rui Wu DN: cn=Rui Wu, ou=North Wing Reason: I am the author of this document Location: Date: 2022-12-20 12:31+05:00	Michele Melia Michele Melia I am digitally signing this document. 2022-12-20 17:57-05:00	12-19-2022	The following revisions from SAP Version 2.0 have been highlighted in this updated version. <u>Section 1</u> Per discussion with M. Melia, the previous sensitivity analysis using multiple imputation has been replaced with an analysis of covariance (ANCOVA) model that compares change in SER between treatment groups at 24 months, while only adjusting for baseline factors as specified in the primary analysis. The rationale of this change was that the percentage of missing outcomes at the primary outcome visit (24 months) was low while, due to the virtual visits implemented for the intermediate visits (6, 12, and 18 months), the number of missing outcomes at those visits was higher. The imputation method previously specified would then primarily impute the outcomes at intermediate visits and would not be an effective check of the effect of missing data assumptions on the primary 24 months analysis. The new sensitivity analysis proposed will be used to assess if including the intermediate visits, which have the highest proportions of missing data, in the primary analysis longitudinal model affects the treatment group comparison of the primary outcome at 24 months. <u>Section 2</u> Based on the same rationale for the change in Section 1 above, the previous sensitivity analysis using multiple imputation has been replaced with an ANCOVA model that compares change in SER between treatment groups at 30 months, while only adjusting for baseline factors as specified in the primary analysis. <u>Section 3</u> 1.Per discussion with the study leads, the proportion of participants with >=0.5D, 1D, and 2D progression in myopia at 12, 24, and 30 months will now be tabulated. 2.Per discussion with M. Melia, the previous analysis calculating relative risk of >=1D (or 2D) progression in myopia has been replaced with treatment group comparison of proportions at 12, 24, and 30 months using Barnard's test. The originally proposed method of analysis, an analysis of time to progression using the proportional hazards model, was not feasible due to a relatively large proportion of missing data at interim visits due to COVID. While missing data is still a

VERSION NUMBER		AUTHOR	APPROVER	EFFECTIVE DATE	REVISION DESCRIPTION (INCLUDING SECTIONS REVISED)
SAP	Protocol				
					problem in the analyses of results of interim visits when using Barnard's test, the 24-month primary time point is largely unaffected, as the amount of missing data is minimal. 3.Per discussion with the study leads, to evaluate biologic activity of the drops, and the potential effect of dilute atropine on accommodation, as a post hoc analysis, mean binocular near point of accommodation at 6 months was compared between treatment groups using an ANCOVA model, adjusting for baseline binocular near point of accommodation, age, iris color (brown vs. not brown), and race (East Asian vs. non-East Asian).

20

The primary objective of MTS1 is to determine the efficacy of daily low-dose atropine (0.01%) for slowing myopia progression over a two-year treatment period in children aged 5 to less than 13 years with myopia -1.00D to -6.00D at the time of enrollment.

Participants are randomly assigned 2:1 to the following two treatment groups:

- Atropine Group: 0.01% atropine eyedrops administered 1 drop to each eye daily for 24 months, followed by 6 months off atropine eyedrops
- Placebo Group: Placebo eyedrops administered 1 drop to each eye daily for 24 months, followed by 6 months off placebo eyedrops

1. Primary Analysis: Refractive Error at 24 Months (On-Treatment)

The primary analysis will be a treatment group comparison of mean change from baseline to 24 months in spherical equivalent refractive error (SER), as measured by a masked examiner using cycloplegic autorefraction, using a longitudinal discrete time mixed model. The population-averaged method will be used to model the repeated measures on SER at 6, 12, 18, and 24-month visit. The time variable will be categorical to not impose assumptions regarding trend of SER over time. The correlations between SER measurements within person across visits will be estimated and a correlation structure will be selected accordingly. (Given that the follow-up visits are approximately equally spaced, an autoregressive covariance structure is considered likely. Other correlation structures will also be fitted, and the information criteria will be used to select the most appropriate covariance structure.) The model will include the interaction between time and treatment group and adjust for baseline SER, age, iris color (brown vs. non-brown), and race (East Asian vs. non-East Asian), to account for potential residual confounding, and improve power for the treatment comparison. The web data entry system mandated that baseline SER, age, and iris color could not be missing at enrollment. The value 'unknown' will be used in the mixed model for missing race/ethnicity value.

Given that at most a small number of participants of East Asian race are expected to have non-brown iris, there is a possibility of partial aliasing when including both race and iris color in the mixed model. If the fully-adjusted model fails to converge, or displays evidence of estimate instability, such as very large standard error associated with a partially-aliased covariate, the race and iris color variables will be combined as follows: East Asian (regardless of eye color), non-East Asian with brown eyes, non-East Asian with non-brown eyes. The number of East Asians with non-brown eye color will be tabulated and reported.

At baseline and all follow-up visits, including the 24-month visit, the mean of the three readings from autorefraction in each eye will be calculated and the mean of both eyes for each participant will be used for the analysis. If fewer than 3 readings are available in each eye, the mean of available readings will be used for each eye to obtain the mean of both eyes for each participant. If data from only one eye is available, the mean of readings on that eye will be used for analysis.

The mean change from baseline to 24 months in SER in each treatment group and the treatment group difference (atropine – placebo), together with their corresponding 95% confidence intervals, will be estimated using the mixed model with maximum likelihood estimation. Maximum likelihood estimation gives unbiased estimates of treatment effect in the presence of missing outcome data, as long as the missing data is missing at random (MAR) conditional on the variables included in the analysis model. The 2-sided null hypothesis of mean treatment difference equals zero (superiority hypothesis) will be tested at an alpha level of 0.05.

1.1.Principles to be Followed in Primary Analysis

Model assumptions for the longitudinal discrete time mixed model will be assessed, including linearity of the adjustment covariates (baseline SER and baseline age), and normality and homoscedasticity of the outcome distribution across the treatment groups. The linearity assumption of the baseline covariates of SER and age will be evaluated using descriptive scatterplots and by categorizing each of the baseline factors in the model to check for approximate linearity of the coefficients across ordered categories. A baseline covariate will be included as a continuous variable in the model if the assumptions for linearity are met for that covariate; otherwise it will be categorized. The median split will be used for categorization.

The primary analysis will follow the intent-to-treat principle; all randomized participants will be included in the analysis and analyzed according to their randomized treatment group regardless of whether the assigned treatment was actually received. However, only data from exams completed within a visit analysis window (±3 months from the expected visit date) will be included in the analysis. Given that a discrete time model will be used and the time points will be grouped into 6-month intervals (i.e. 6, 12, 18, or 24 months from randomization), if two consecutive visits are within 90 days from each other, only one of the two visits will be included in the analysis. The following principles will be used in choosing which visit to be included:

a. If a 24-month primary outcome visit is available, it will be included in the analysis.
b. Any other visit that is within 90 days of the 24-month visit will be excluded from the analysis.
c. If any two visits (other than 24-month visit) are within 90 days of each other, the visit closer to the expected visit date will be included in the analysis, and the other visit will be excluded.

There will be no explicit imputation of outcome data for exams not completed or completed outside the analysis window, as the mixed model will produce an unbiased estimate of treatment effect as long as missing outcome data are missing at random (MAR), and it is expected that including the baseline covariates and outcome data from interim follow-up exams in the analysis model is likely to meet MAR requirements, although this will not be verifiable. Hence, the sensitivity of results to the MAR assumption will be explored in sensitivity analyses (Section 1.2).

1.2.Sensitivity Analysis

A sensitivity analysis will be conducted to compare the mean change in SER from baseline to 24 months between the treatment groups using an analysis of covariance (ANCOVA) model, adjusting for baseline SER, age, iris color (brown vs. non-brown), and race (East Asian vs. non-East Asian). Possible partial aliasing of iris color and race will be handled as specified for the primary analysis model.

2. Secondary Objective: Efficacy off Atropine Treatment (30 Months)

The secondary objective is to determine the efficacy of atropine treatment for slowing progression of myopia after a period of 6 months off treatment. The same approach defined in Section 1 (including the sensitivity analysis) will be used to obtain a treatment group comparison of change from baseline to 30-months in SER, as measured by a masked examiner using cycloplegic autorefraction. The mean treatment group difference and the corresponding 95% confidence interval will be estimated at 30 months. However, the statistical testing of significance will be performed only if a statistically significant effect for treatment was found in the primary analysis at 24 months.

30-month visits will be included in the analysis as long as they are no earlier than 3 months and no later than 6 months from the expected visit date and same follow-up SER measurements selected for the primary analysis (at 6, 12, 18, or 24 months from randomization) using the principles specified in Section 1.1 will be included in the analysis.

A sensitivity analysis will be conducted to compare the mean change in SER from baseline to 30 months between the treatment groups using an ANCOVA model, adjusting for baseline SER, age, iris color (brown vs. non-brown), and race (East Asian vs. non-East Asian).

3. Secondary Outcomes

The overall false discovery rate (FDR)in the secondary analyses specified in Section 3 below will be controlled at the 5% level using the two-stage step-up false discovery rate procedure of Benjamini, Krieger, and Yekutieli.[1,2] This involves first applying the false discovery rate procedure of Benjamini and Hochberg[3] at alpha level = (overall α)/(overall α + 1) to estimate the number of true null hypotheses, and then applying the adaptive FDR adjustment[4] conditional on the number of true null hypotheses. This method generally has better power than the usual Benjamini-Hochberg FDR method.

3.1.Proportion of Participants with Progression >=0.5D, 1D, and 2D

The proportion of participants with progression >=0.5D, >=1D, and >=2D from baseline to 12, 24, and 30 months in each treatment group will be tabulated. The proportions will be compared between treatment groups using Barnard's test.

3.2.Change in Axial Length at 12 and 24 Months (On Treatment)

Axial length will be reported as the distributions of baseline length, 12-month length, 24-month length, and change in axial length from baseline to 12 and 24 months. A treatment group comparison of the change in axial length from baseline to 12 months and 24 months will be performed using a longitudinal discrete time mixed model with maximum likelihood estimation, which allows for interaction between time and treatment group, and adjusts for the same baseline covariates as the primary analysis. The same strategies specified in the primary analysis (Section 1, excluding sensitivity analysis) will be used to choose the appropriate covariance structure for the model.

At baseline and all follow-up visits, including the 12 and 24-month visits, the mean of the axial length readings in both eyes for each participant will be used for the analysis. If data from only one eye is available, the reading on that eye will be used for analysis. The treatment group differences (atropine – placebo) and a 95% confidence interval will be estimated using the mixed model.

3.3.Change in Axial Length at 30 Months (Off Treatment)

The same approach defined in Section 3.4 will be used to conduct a treatment group comparison of the change in axial length from baseline to 30 months.

3.4.Refractive Error at 12 Months (On Treatment)

The primary analysis specified in Section 1 (excluding the sensitivity analysis) will be used to obtain a treatment group comparison of change from baseline to 12-months in SER, as measured by a masked examiner using cycloplegic autorefraction.

3.5.Near Accommodation at 6 Months

As a post hoc analysis, mean binocular near point of accommodation at 6 months will be compared between treatment groups using an ANCOVA model, adjusting for baseline binocular near point of accommodation, age, iris color (brown vs. not brown), and race (East Asian vs. non-East Asian).

4. Subgroup Analysis

The treatment group difference for change in SER from baseline to 24 and 30 months within the following subgroups will be explored:

- East Asian vs. non-East Asian race
- Brown iris versus non-brown iris
- Baseline younger versus older age (based on median split)
- Baseline lower versus higher myopia level by SER (based on median split)
- Baseline age and baseline SER (4 categories based on median splits)

An estimate of the treatment effect within each subgroup and a 95% confidence interval will be obtained for each of the times, 24 and 30 months, by adding a treatment by time

by subgroup interaction into the primary analysis model and generating the appropriate contrasts. Each subgroup effect will be estimated separately, one subgroup per model.

Previous studies have suggested that race and/or eye color may interact with the treatment effect of atropine. Several studies have consistently found atropine to be effective in East Asian populations and a meta-analysis has suggested that atropine might be more effective in Asian populations than in white children, although this conclusion is limited by the lack of studies in non-Asian populations. Other research has shown that atropine is rapidly taken up by melanocytes and released over time, leading to a longer time to achieve mydriasis and a prolonged mydriatic effect in heavily-pigmented eyes as atropine is released over time, potentially leading to an increased treatment effect. Conversely, if the mechanism by which atropine slows myopic progression is through local retinal effects, one might speculate that higher melanocyte density might prevent the atropine from reaching the retina, which might result in brown eyes having less treatment effect than non-brown eyes.

Atropine might be expected to be more effective in children with lower amounts of myopia given more potential for suppression of myopia progression. Likewise, atropine might be hypothesized to have a greater treatment effect in younger than in older children given they are earlier in the course of myopic progression and have more room for potential suppression.

For each time point, the planned subgroup analyses also will be conducted using a continuous time longitudinal model, if linearity assumptions with time are met, to obtain p-values for the subgroup effect. Even if change in SER is not precisely linear with time, if it is monotonically decreasing, the continuous model is expected to have higher power than the discrete time primary analysis model, and will be favored over the discrete time model for obtaining p-values. The baseline factor and the baseline factor by treatment interaction will be included as terms in the model, and the 3-way subgroup, time, and treatment interaction will be used to determine whether there is a significant subgroup effect. The false discovery rate for the subgroup analyses will be controlled using the two-stage step up FDR procedure to control the overall FDR at 5%. Subgroup analyses will be interpreted with caution, particularly if the corresponding overall analysis does not demonstrate a significant treatment group difference.

5. Additional Analyses

The overall false discovery rate for the additional analyses specified in Section 5.1 and 5.2 below will be controlled at the 5% level using the two-stage step up false discovery rate (FDR) procedure.

5.1.Treatment Effect Over First Year of Treatment

The treatment effect on change in SER from baseline through the first year will be compared with the treatment effect on change in SER from end of first year through the second year, by constructing the appropriate contrasts in the primary analysis model.

5.2.Exploratory Analyses of Additional Ocular Biometric Parameters

As exploratory analyses at 24 and 30 months, change in flat corneal radius, anterior chamber depth, and lens thickness from baseline will each be compared between treatment groups using a longitudinal discrete time mixed model with maximum likelihood estimation, including the interaction between time and treatment group, and adjusting for the baseline covariates from the primary analysis model. The same strategies specified in the primary analysis (Section 1, excluding sensitivity analysis) will be used to choose the appropriate covariance structure for the model.

At baseline and all follow-up visits, including the 12 and 24-month visits, the mean of the readings in both eyes for each participant will be used for the analysis. If data from only one eye is available, the reading on that eye will be used for analysis. The treatment group differences (atropine – placebo) and a 95% confidence interval will be estimated using the mixed model.

5.3.Data Tabulations

The following tabulations will be performed both according to treatment group and among overall sample:

- Baseline demographics and clinical characteristics
- A flow chart accounting for all participants for all visits and phone calls
- Visit and phone contact completion rates for each follow-up visit
- Protocol deviations
- Proportion of participants needing bifocals by 24 months (i.e. during treatment)

5.4.Compliance

Compliance with study medication will be assessed at the 6-month, 12-month, 18-month, and 24-month outcome exams. For each of these exams, the proportion of calendar days that study medication was reported used and the proportion of unused study medication ampules will be tabulated in each of the two treatment groups.

Compliance with refractive correction will be assessed at the 6-month, 12-month, 18-month, and 24-month outcome exams. After discussion with the parent and child, study personnel will classify the proportion of time refractive error was worn will be described as excellent (76% to 100%), good (51% to 75%), fair (26% to 50%), or poor (≤ 25%). The distribution of refractive correction compliance will be tabulated in each of the two treatment groups.

6. Safety Analyses

6.1.Adverse Events of Eye Drops

An eye drops questionnaire will be administered at randomization and at each follow-up visit. The distribution of scores on each survey item will be summarized by treatment group at the time of randomization and at each follow-up exam up until and including the 24-month visit. The average of the item responses at the 24-month visit will be calculated and compared with a t-test for difference in distributions between treatment groups.

6.2.Visual Acuity

The proportion of participants with loss of best corrected distance vision >1 logMAR line at 30 months in either eye will be compared between treatment groups using Barnard's test. The proportion of participants with loss of best corrected near binocular vision >1 logMAR line at 6 months will be tabulated by treatment groups and compared using Barnard's test. The Bonferroni correction will be used to account for multiplicity and control the type I error rate.

7. Interim Monitoring

By the time sufficient data allowing for an analysis of efficacy or futility for the primary outcome would be possible, all or most subjects are expected to be within 1 year of completing the primary outcome; hence, no formal statistical interim monitoring is proposed.

Tabulated safety and efficacy data will be reviewed by the PEDIG Data and Safety Monitoring Committee (DSMC) at its biannual meetings, and if any unexpected safety or other issues arise, they can recommend stopping the study at any time.

During the DSMC meeting in October 2019, the DSMC approved the proposal of NOT performing interim monitoring analysis for futility based on the rationale that by the time 50% of the 24-month data are available (May 2021), the recruitment will have been finished and the remaining participants will have between <1 to 7 months of remaining time on treatment before all participants have 24-month data (December 2021). The DSMC also approved the proposal of NOT performing interim monitoring analysis for efficacy based on the rationale that even if efficacy was found in an interim analysis for the 24-month on-treatment primary outcome analysis, the 30-month off-treatment secondary analysis is needed to understand whether the benefit persists after treatment is discontinued.

References

1. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika.* 2006;93(3):491-507.
2. SAS Institute Inc. SAS/STAT® 14.3 User's Guide 2017. In: Cary, NC: SAS Institute Inc:6564-6567.
3. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological).* 1995;57(1):289-300.
4. Benjamini Y, Hochberg Y. On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. *Journal of Educational and Behavioral Statistics.* 2000;25(1):60-83.
5. Newman SC. Commonalities in the classical, collapsibility and counterfactual concepts of confounding. *Journal of clinical epidemiology.* 2004;57(4):325-329.