Statistical Analysis Plan

for

AZ202001

Azura Ophthalmics

A Multicenter, Vehicle-controlled, Randomized Study to Evaluate the Safety, Tolerability and Pharmacodynamics of AZR-MD-001 in Patients with Meibomian Gland Dysfunction (MGD)

AZ202001

Version 1.0, dated 25JAN2021

Author					
Name	Function	Signature	Date		
			dd	mmm	yy
Allie Ferreira	Lead Biostatistician	DocuSigned by: Allie Ferreira Signer Name: Allie Ferreira Signing Reason: I approve this document Signing Time: 26-Jan-2021 | 3:31:48 PM PST 7E9BCA56B1664BECBE5379C4E9A322D7	26	Jan	21

SIGNATURE PAGE

The undersigned hereby jointly declare that they have reviewed the Statistical Analysis Plan, and agree to its form and content.

Approval by					
Name	**Function**	**Signature**	**Date**		
			dd	mmm	yy
Charles Bosworth	Sponsor Representative	DocuSigned by: Charles Bosworth Signer Name: Charles Bosworth Signing Reason: I approve this document Signing Time: 26-Jan-2021 | 8:32:02 PM PST 256C76D0E93C4B929533E1A523A7E742	26	01	21
Audra Wilson	Project Manager	DocuSigned by: Audra Wilson Signer Name: Audra Wilson Signing Reason: I approve this document Signing Time: 26-Jan-2021 | 8:41:25 PM PST F115CB0AB167461CB8B5BFA9E44BBE3B	21	01	21
Elizabeth Kane	Director, Analytical Services	DocuSigned by: Elizabeth Kane Signer Name: Elizabeth Kane Signing Reason: I approve this document Signing Time: 26-Jan-2021 | 11:44:37 AM PST F410DEF5F6684C45B1DD943845226FF9	26	JAN	21
Venita DePuy	Independent Biostatistician	DocuSigned by: Venita DePuy Signer Name: Venita DePuy Signing Reason: I have reviewed this document Signing Time: 26-Jan-2021 | 12:00:35 PM PST 42F90E5CCD51467A85A120D428B369DB	26	JAN	21
Allie Ferreira	Lead Biostatistician	DocuSigned by: Allie Ferreira Signer Name: Allie Ferreira Signing Reason: I approve this document Signing Time: 26-Jan-2021 | 3:32:18 PM PST 7E9BCA56B1664BECBE5379C4E9A322D7	26	Jan	21

TABLE OF CONTENTS

DOCUMENT HISTORY

Revision Date	Author	Version	Reason for Change
25JAN2021	*Allie Ferreira*	1.0	NA – Initial Version

1 ABBREVIATIONS

Abbreviation	Definition
AE	Adverse Event
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
BCVA	Best-Corrected Visual Acuity
BMI	Body Mass Index
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CRF	Case Report Forms
CSR	Clinical Study Report
DBP	Diastolic Blood Pressure
FDA	Food and Drug Administration
IOP	Intraocular Pressure
ITT	Intent-To-Treat Population
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
MGD	Meibomian Gland Dysfunction
MGS	Meibum Gland Secretion Score
MGYLS	Meibomian Glands Yielding Liquid Secretion
MICD	Most Important Clinical Difference
MITT	Modified Intent-To-Treat Population
MITT2	Modified Intent-To-Treat 2 Population
OSDI	Ocular Surface Disease Index
PT	Preferred Term
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SBP	Systolic Blood Pressure
SD	Standard Deviation
SOC	System Organ Class
SOP	Standard Operating Procedure
SPEED	Standard Patient Evaluation of Eye
TBUT	Tear Break-Up Time
TEAE	Treatment Emergent Adverse Event
VAS	Visual Analogue Scale

2 SUMMARY

TITLE	A Multicenter, Vehicle-controlled, Randomized Study to Evaluate the Safety, Tolerability and Pharmacodynamics of AZR-MD-001 in Patients with Meibomian Gland Dysfunction (MGD)
PREFACE	This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for Azura protocol AZ202001 (Multicenter, Vehicle-controlled, Randomized Study to Evaluate the Safety, Tolerability and Pharmacodynamics of AZR-MD-001 (Selenium disulfide Agonist) in Patients with Meibomian Gland Dysfunction (MGD). This study is being completed to assess the safety, tolerability, and pharmacodynamics of AZR-MD-001 in patients with meibomian gland dysfunction (MGD). The following documents were reviewed in preparation of this SAP: • Clinical Research Protocol AZ202001 issued 12MAR2020 • Case report forms (CRFs) issued 25JUN2020 for AZ202001
PURPOSE	The purpose of this SAP is to outline the planned analyses in support of the Clinical Study Report (CSR) for protocol AZ202001. Exploratory analyses not necessarily identified in this SAP may be performed to support the clinical development program. Any post-hoc, or unplanned, analyses not identified in this SAP will be clearly identified in the respective CSR.
STUDY OBJECTIVES	To evaluate the safety, tolerability, and pharmacodynamics of AZR-MD-001 ointment/semi-solid drug applied to the lower lid twice weekly for up to 3 months compared to its vehicle in patients with meibomian gland dysfunction (MGD).
STUDY DESIGN	Multicenter, investigator-masked, vehicle-controlled, randomized, parallel group study comparing AZR-MD-001 (0.5%) ointment/semi-solid drug and AZR-MD-001 vehicle dosed twice-weekly in the evening.
ENDPOINTS	The study's **primary efficacy sign for MGD** is change from baseline to month 3 in meibum gland secretion score (MGS) (0 to 45 scale). The study's **primary efficacy symptom for MGD** is change from baseline to month 3 in total OSDI. **Other Efficacy Measures:** • Change from Baseline to day 14, month 1 and month 1.5 in MGS (0 to 45 scale) • MGS score (0 to 45 scale) at each visit • Proportion of patients with an MGS score > 12 at each visit • Change from Baseline to day 14, month 1, month 1.5 and month 3 in the number of Meibomian Glands Yielding Liquid Secretion (MGYLS)

	(0 to 15 scale) • MGYLS (0 to 15 scale) at each visit • Change from Baseline to day 14, month 1, month 1.5, and month 3 in TBUT • TBUT at each visit • Proportion of patients with a TBUT score > 5 at each visit • Change from Baseline to day 14, month 1, month 1.5 and month 3 in Standard Patient Evaluation of Eye Dryness (SPEED) • SPEED at each visit • Proportion of patients with a SPEED < 6 at each visit • Change from Baseline to day 14, month 1, month 1.5 and month 3 in average visual analogue scale (VAS) • Average VAS at each visit • Change from Baseline to day 14, month 1, month 1.5 and month 3 in worst VAS • Worst VAS at each visit • Change from Baseline to day 14, month 1 and month 1.5 in Total OSDI • Total OSDI, OSDI sub-scales, and individual items from the OSDI across visits • Proportion of patients with a Total OSDI change from baseline > 4.5, the known minimally important clinical difference (MICD) for early to moderate disease, at month 3 • Proportion of patients with a Total OSDI < 13 at each visit • Number of expressible glands yielding clear meibum at day 14, month 1, month 1.5, and month 3 • Eyelid margin erythema/telangiectasias at day 14, month 1, month 1.5, and month 3 • Corneal and conjunctival staining (0 to 5 scale) at each visit **Safety Endpoints:** • Adverse events • Vital signs • Study medication tolerability as measured by the Ocular Comfort Questionnaire • Urine pregnancy test • Best-corrected visual acuity (BCVA; Logarithmic visual acuity chart) • Biomicroscopy

	• Ophthalmoscopy • Intraocular pressure (IOP)
INTERIM ANALYSES	No interim analyses are planned for this study.
FINAL ANALYSES	All final planned analyses identified in this SAP will be completed after the last participant has completed 3-months follow up.

3 STUDY OBJECTIVES AND ENDPOINTS

3.1 STUDY OBJECTIVE

To evaluate the safety, tolerability, and pharmacodynamics of AZR-MD-001 ointment/semi-solid drug applied to the lower lid twice-weekly for up to 3 months compared to its vehicle in patients with meibomian gland dysfunction (MGD).

3.2 STUDY ENDPOINTS

3.2.1 PRIMARY EFFICACY ENDPOINTS

The study's **primary efficacy sign for MGD** is change from baseline to month 3 in meibum gland secretion score (MGS) (0 to 45 scale).

The study's **primary efficacy symptom for MGD** is change from baseline to month 3 in total OSDI.

3.2.2 OTHER EFFICACY ENDPOINTS

- Change from Baseline to day 14, month 1 and month 1.5 in MGS (0 to 45 scale)
- MGS score (0 to 45 scale) at each visit
- Proportion of patients with an MGS score > 12 at each visit
- Change from Baseline to day 14, month 1, month 1.5 and month 3 in the number of Meibomian Glands Yielding Liquid Secretion (MGYLS) (0 to 15 scale)
- MGYLS (0 to 15 scale) at each visit
- Change from Baseline to day 14, month 1, month 1.5, and month 3 in TBUT
- TBUT at each visit
- Proportion of patients with a TBUT score > 5 at each visit
- Change from Baseline to day 14, month 1, month 1.5 and month 3 in Standard Patient Evaluation of Eye Dryness (SPEED)
- SPEED at each visit
- Proportion of patients with a SPEED < 6 at each visit
- Change from Baseline to day 14, month 1, month 1.5 and month 3 in average visual analogue scale (VAS)

- Average VAS at each visit
- Change from Baseline to day 14, month 1, month 1.5 and month 3 in worst VAS
- Worst VAS at each visit
- Change from Baseline to day 14, month 1 and month 1.5 in Total OSDI
- Total OSDI, OSDI sub-scales, and individual items from the OSDI across visits
- Proportion of patients with a Total OSDI change from baseline > 4.5, the known minimally important clinical difference (MICD) for early to moderate disease, at month 3
- Proportion of patients with a Total OSDI < 13 at each visit
- Number of expressible glands yielding clear meibum at day 14, month 1, month 1.5, and month 3
- Eyelid margin erythema/telangiectasias at day 14, month 1, month 1.5, and month 3
- Corneal and conjunctival staining (0 to 5 scale) at each visit

3.2.3 Safety Endpoints

- Adverse events
- Vital signs
- Study medication tolerability as measured by the Ocular Comfort Questionnaire
- Urine pregnancy test
- Best-corrected visual acuity (BCVA; Logarithmic visual acuity chart)
- Biomicroscopy
- Ophthalmoscopy
- Intraocular pressure (IOP)

4 Sample Size

Estimates for sample size calculations are from the first interim analysis from Azura Clinical Protocol AZ201801. The standard deviation for MGS was 3.35 units for an ineffective dose and 9.48 units for the high dose. A sample size of 12 subjects per group will have 90% power to detect a difference of 10.4 units between the active treatment group and the vehicle group using a two-sample t-test at a significance level of 0.05. The standard deviation for Total OSDI was 13.01 units for an ineffective dose and 6.79 units for the high dose. A sample size of 11 subjects per group will have 90% power to detect a difference of 15.9 units between the active treatment group and the vehicle group using a two-sample t-test at a significance level of 0.05.

The total number of randomized patients for the study will be up to approximately 30. Approximately 15 patients should have a baseline MGS score of < 6 and approximately 15 patients should have a baseline

MGS score ≥6 and ≤ 12. Based upon data from ongoing study AZ201801 and the simplified inclusion/exclusion criteria for this study a screen failure rate of ~ 40% is expected. Thus, ~42 patients will need to be screened to achieve ~30 patients randomized to treatment.

5 Sequence of Planned Analyses

5.1 Interim Analyses

There are no planned Interim Analyses for this study.

5.2 Final Analyses and Reporting

All final planned analyses identified in this SAP will be completed after the last participant has completed month 3 follow up. Any post-hoc, exploratory analyses completed to support planned study analyses, which were not identified in this SAP, will be documented and reported as necessary. Any results from these unplanned analyses will also be clearly identified in the final study report.

6 Analysis Populations

6.1 Intent to Treat Population (ITT)

The intent-to-treat (ITT) population for this study includes all enrolled and randomized patients.

6.2 Modified Intent to Treat Population (mITT)

The modified intent-to-treat (ITT) population for this study includes all patients randomized and, who have values at randomization, and at least 1 post-randomization value for MGS at a regularly scheduled visit (i.e., Day 14 or Month 1). All patients in the mITT population will be analyzed by the treatment received. This population will be used for the primary and secondary efficacy analyses.

6.3 Modified Intent to Treat 2 Population (mITT2)

The modified intent-to-treat 2 (mITT2) population will be comprised of patients who are included in the mITT and have the randomization MGS score in the study eye ≥6 and ≤ 12. The mITT2 population will be analyzed by the treatment received. The primary efficacy endpoints related to MGD will be repeated using this mITT2 population as sensitivity analyses.

6.4 Safety Population

The safety population includes all treated patients. This is the primary analysis population for safety variables; subjects are analyzed under the treatment received.

7 General Issues for Statistical Analysis

Descriptive statistics for continuous variables will include the sample size, mean, standard deviation (SD), median, minimum, and maximum. Percent and frequencies will be provided for categorical variables. Results will be generated by treatment group.

7.1 Analysis Software

Analysis data sets, statistical analyses and associated output generated by Avania will be generated using SAS® Software version 9.4 or later, or R version 3.3.2 or later.

7.2 Disposition of Subjects and Withdrawals

The number and percent of subjects in each analysis population will be presented, with percentages based on the ITT Population All subjects who provide written informed consent will be accounted for. A summary of the reasons for screen failure among subjects who signed an Informed Consent will be provided.

The frequency and percent of enrolled subjects who completed each scheduled assessment will be presented for the ITT, mITT and mITT2 populations. The number and percentage of patients exposed (i.e. initiated procedure), prematurely terminated (overall and by reason of premature termination) and completed patients will be summarized.

7.3 Methods for Withdrawals and Missing Data

The method of Last Observation Carried Forward (LOCF) will be used for efficacy on the mITT population. In these analyses, non-missing values recorded at visit 3 (i.e., Day 14) or later will be used to replace missing data at visits where data are not recorded.

If more than 5% of patients are missing data for MGS score or OSDI at 3 months prior to LOCF, then multiple imputation will be used for a sensitivity analysis for these endpoints. Missing MGS score at 3 months (prior to LOCF) will be imputed using a monotone linear regression multiple imputation approach for continuous outcome data. However, assuming the missing data pattern will not be completely monotone at first, then a Markov Chain Monte Carlo (MCMC) imputation will first be carried out using MGS scores at all time points where it is collected (prior to 3 months) to make a monotone missing data pattern. There will be 50 datasets generated in this manner to create 50 datasets with a

monotone missing data pattern. For each of these 50 data sets, missing MGS score at 3 months will then be imputed once from a monotone multiple imputation linear regression model with independent variables: age, gender, duration of disease, and TBUT at month 3. Also included will be baseline and post-baseline non-missing MGS scores. The ANCOVA model will then be carried out on each of the resulting 50 complete datasets, with the ANCOVA results being combined across the 50 complete datasets using standard multiple imputation theory to obtain one overall p-value assessing the treatment on reduction in MGS score after accounting for missing data.

The same analysis will be conducted for missing OSDI data at 3 months if more than 5% of subjects are missing this assessment, using baseline and post-baseline non-missing OSDI in the imputation model.

7.4 Protocol Deviations

Protocol deviations will be summarized in the CSR for the ITT population. This summary will include the number and percent of participants (overall and by site) with each deviation type. A listing of protocol deviations will be presented and will include deviation number, date deviation occurred, visit, type of deviation, description of deviation, and action taken.

7.5 Multiple Comparisons and Multiplicity

To address the multiple primary endpoints defined in this study, the primary endpoints have been prioritized into a hierarchical structure. In order to test the primary symptom endpoint of Total OSDI, the primary sign endpoint of MGS must be statistically significantly higher in the 0.5% AZR-MD-001 treatment group compared to the vehicle treatment group using a two-sided significance level of 0.05. Using this strategy, the family-wise Type I error rate will be maintained at the 0.05 significance level for the two primary endpoints.

There will be no statistical adjustment of the significance level for multiple testing for the other efficacy endpoints as these comparisons will be descriptive in nature.

7.6 Assessment of Homogeneity

Tests of homogeneity across sites will be done to determine if the sites have reasonably homogenous responses. An analysis of covariance (ANCOVA) model will be performed for both the primary efficacy sign for MGD and primary efficacy symptom for MGD. For the primary efficacy sign for MGD, change in MGS score will be the dependent variable, baseline MGS score and duration of disease will be covariates, and the independent variables will consist of treatment group, site and a site*treatment group interaction term. For the primary efficacy symptom for MGD, change in OSDI will be the dependent variable, and baseline OSDI and duration of disease will be the covariates. A site*treatment

group interaction that is significant at a 0.15 level will require further investigation of results by site to assess if sites are still poolable despite being significant at the liberal 0.15 level of significance.

In study sites with small numbers of subjects, it will be difficult to evaluate a site by treatment interaction. For this reason, study sites with fewer than 5 subjects may be combined into pseudosites. Pseudosites will be used for all multivariate analyses including the analysis to determine a site by treatment interaction for pooling. All pseudosite categorizations will be determined before the database is locked.

8 Demographics and Other Baseline Characteristics

In general, continuous data will be summarized with descriptive statistics (number of patients, mean, standard deviation, median, minimum, and maximum) and will be analyzed using analysis of variance (ANOVA) techniques or 2-sample t-tests for between-group comparisons, and paired t-tests for within group comparisons. Categorical variables will be summarized by sample size (N), frequency count, and percent, and they will be analyzed using Pearson's chi-square test or Fisher's exact test (if the expected cell count is less than 5 in 25% or more of the cells). Ordinal variables will be analyzed using the Cochran-Mantel-Haenszel (CMH) or the Wilcoxon rank-sum test for between-treatment comparisons and the sign-rank test for within-treatment comparisons.

8.1 Demographics and Vital Signs

Participant demographics for the ITT, mITT, mITT2, and safety population will be summarized in a table. Sex, race, duration of MGD, and ethnicity will be summarized with frequency and percent. Age, height, weight, BMI, SBP, DBP and heart rate will be summarized with number of observations, mean, median, minimum, maximum and standard deviation.

8.2 Baseline Non-Ophthalmic Medical History

The frequency and percent of patients with relevant medical and surgical history for each body system (e.g. General, Respiratory, Cardiovascular) will be presented for the mITT population. A listing of medical history information will be presented with information regarding body system, relevant medical condition/surgery, and start and end date. The listing will also include information regarding if a urine pregnancy test was performed and the result.

8.3 Baseline Ophthalmic History

The frequency and percent of patients with relevant ophthalmic and surgical history by ocular disease / condition / procedure by will be presented for the mITT population. A listing of ophthalmic history will be presented with information regarding the eye/eyelid, the ocular disease / condition / procedure, any related current medication for the condition, and start and end date.

8.4 Prior and Concurrent Medications

All relevant prior and concomitant medications will be presented in a listing for the mITT population. Patients must have discontinued and be willing to remain off all other ophthalmic preparations including artificial tears during the study. In the event that rescue medication is required for clinically relevant worsening of keratitis, patients will be provided a concomitant regimen the investigator/treating clinician.

9 Efficacy Analyses

9.1 Primary Efficacy Sign for MGD

The primary sign for MGD is based on change from baseline in meibum gland secretion (MGS) score (0 to 45 scale) at the 3-month follow-up between AZR-MD-001 group versus vehicle group. The primary efficacy sign will be evaluated in the mITT and mITT2 populations and will be analyzed at month 3 using an analysis of covariance (ANCOVA) model with baseline MGS score and duration of disease stratum (< 5 years or ≥ 5 years) as covariates and treatment (AZR-MD-001 ointment/semi-solid drug or vehicle) as factors in the model. The difference between treatment groups will be evaluated using t-tests of the least square means from this model. The two-sided confidence interval (95%) will be provided for the difference between treatments.

The following pseudocode will be used:

```
PROC GENMOD data=xxx;
BY visit;
CLASS treatment (ref = 'AZR-MD-001 VEHICLE') duration;
MODEL chg = treatment base duration;
LSMEANS treatment;
run;
```

9.2 Primary Efficacy Symptom for MGD

The primary symptom for MGD is based on change from baseline in total OSDI at the 3-month follow-up between AZR-MD-001 group versus vehicle group. The primary efficacy symptom will be evaluated in

the mITT and mITT2 populations and will be analyzed at month 3 using an analysis of covariance (ANCOVA) model with baseline OSDI, baseline MGS score stratum (<6 or ≥6) and duration of disease stratum (< 5 years or ≥ 5 years) as covariates and treatment group as factors in the model. The difference between treatment groups will be evaluated using t-tests of the least square means from this model. The two-sided confidence interval (95%) will be provided for the difference between treatments.

The following pseudocode will be used:

```
PROC GENMOD data=xxx;
BY visit;
CLASS treatment (ref = 'AZR-MD-001 VEHICLE') duration baseMGS;
MODEL chg = treatment base duration baseMGS;
LSMEANS treatment;
run;
```

9.3 Other Efficacy Analyses

All other efficacy endpoints will be summarized and analyzed in the mITT population unless otherwise stated.

9.3.1 MGS Score

The change from baseline to day 14, month 1, and month 1.5 in MGS score (0 to 45 scale) will be summarized using N, mean, standard deviation, median, minimum, and maximum.

The change from baseline will be analyzed at each visit in the same manner as described in Section 9.1.

9.3.2 MGS Score at Each Visit

MGS score (0 to 45 scale) at each visit will be summarized in the mITT population using N, mean, standard deviation, median, minimum, and maximum.

The MGS score will be analyzed at each visit using an ANCOVA model with baseline MGS score and duration of disease stratum (< 5 years or ≥ 5 years) as covariates and treatment (AZR-MD-001 ointment/semi-solid drug or vehicle) as factors in the model. The difference between treatment groups will be evaluated using t-tests of the least square means from this model. The two-sided confidence interval (95%) will be provided for the difference between treatments.

The following pseudocode will be used:

```
PROC GENMOD data=xxx;
BY visit;
CLASS treatment (ref = 'AZR-MD-001 VEHICLE') duration;
MODEL score =treatment base duration;
LSMEANS treatment;
run;
```

9.3.3 Proportion of Patients with MGS Score > 12 at Each Visit

The proportion of patients with an MGS score > 12 at each visit will be summarized using shift tables displaying the frequency and percent of patients in each category (yes/no) by treatment group.

Logistic regression will be performed to compare treatments with respect to the proportion of patients achieving MGS score >12 at each visit, adjusting for baseline MGS score stratum and duration of disease stratum. The odds ratio, along with the 95% CI, for the treatment effect will be presented. Pearson's chi-square test, or Fisher's exact test (if the expected cell count is less than 5 in 25% or more of the cells), will also be performed.

The following pseudocode will be used:

```
PROC LOGISTIC data=xxx;
BY visit;
CLASS treatment (ref = 'AZR-MD-001 VEHICLE') duration baseMGS;
MODEL response =treatment base duration baseMGS;
LSMEANS treatment;
run;

PROC FREQ data=xxx;
BY visit;
TABLE treatment * response/EXACT;
ODS OUTPUT CHISQ = outc (where=(STATISTIC='Chi-Square'))
FISHERSEXACT=outf (where=(NAME1='XP2_FISH'));
Run;
```

9.3.4 Change from Baseline in Number of MGYLS

The change from baseline to day 14, month 1, month 1.5, and month 3 in the number of Meibomian Glands Yielding Liquid Secretion (MGYLS) (0 to 15 scale) will be summarized using N, mean, standard deviation, median, minimum, and maximum.

An ANCOVA test will be performed to assess the difference of change from baseline in number of MGYLS between treatment groups with baseline MGYLS, baseline MGS score stratum and duration of disease stratum as covariates. The difference between treatment groups will be evaluated using t-tests

of the least square means from this model. The two-sided confidence interval (95%) will be provided for the difference between treatments.
The following pseudocode will be used:

```
PROC GENMOD data=xxx;
BY visit;
CLASS treatment (ref = 'AZR-MD-001 VEHICLE') duration baseMGS;
MODEL chg =treatment base duration baseMGS;
LSMEANS treatment;
run;
```

9.3.5 MGYLS AT EACH VISIT

MGYLS (0 to 15 scale) at each visit will be summarized using N, mean, standard deviation, median, minimum, and maximum.

An ANCOVA test will be performed to assess the difference of MGYLS between treatment groups with baseline MGYLS, baseline MGS score stratum and duration of disease stratum as covariates. The difference between treatment groups will be evaluated using t-tests of the least square means from this model. The two-sided confidence interval (95%) will be provided for the difference between treatments.

The following pseudocode will be used:

```
PROC GENMOD data=xxx;
BY visit;
CLASS treatment (ref = 'AZR-MD-001 VEHICLE') duration baseMGS;
MODEL score =treatment base duration baseMGS;
LSMEANS treatment;
run;
```

9.3.6 CHANGE FROM BASELINE IN TBUT

The change from baseline to day 14, month 1, month 1.5, and month 3 in TBUT will be summarized and analyzed using the same methods as described in Section 9.3.4.

9.3.7 TBUT AT EACH VISIT

TBUT at each visit will be summarized and analyzed using the same methods as described in Section 9.3.5.

9.3.8 PROPORTION OF PATIENTS WITH TBUT > 5

The proportion of patients with a TBUT score > 5 at each visit will be summarized and analyzed using the same methods as described in Section 9.3.3.

9.3.9 Change from Baseline in SPEED

The change from baseline to day 14, month 1, month 1.5, and month 3 in SPEED will be summarized and analyzed using the same methods as described in Section 9.3.4.

9.3.10 SPEED at Each Visit

SPEED at each visit will be summarized and analyzed using the same methods as described in Section 9.3.5.

9.3.11 Proportion of Patients with SPEED < 6

The proportion of patients with a SPEED < 6 at each visit will be summarized and analyzed using the same methods as described in Section 9.3.3.

9.3.12 Change from Baseline in Average VAS

The change from baseline to day 14, month 1, month 1.5, and month 3 in average VAS will be summarized and analyzed using the same methods as described in Section 9.3.4.

9.3.13 Average VAS at Each Visit

Average VAS at each visit will be summarized and analyzed using the same methods as described in section 9.3.5.

9.3.14 Change from Baseline in Worst VAS

The change from baseline to day 14, month 1, month 1.5, and month 3 in worst VAS will be summarized and analyzed using the same methods as described in Section 9.3.4.

9.3.15 Worst VAS at Each Visit

Worst VAS at each visit will be summarized will be summarized and analyzed using the same methods as described in section 9.3.5.

9.3.16 Change from Baseline in Total OSDI

The change from baseline to day 14, month 1 and month 1.5, in total OSDI will be summarized and analyzed using the same methods as described in Section 9.3.4.

9.3.17 OSDI Across Visits

Total OSDI, OSDI sub-scales, and individual items from the OSDI across visits will be summarized and analyzed using the same methods as described in section 9.3.5.

9.3.18 Proportion of Patients with Total OSDI Change from Baseline > 4.5 at Month 3

The proportion of patients with a total OSDI change from baseline > 4.5, the known minimally important clinical difference (MICD) for early to moderate disease, at month 3 will be summarized and analyzed using the same methods as described in Section 9.3.3.

9.3.19 Proportion of Patients with Total OSDI < 13

The proportion of patients with a total OSDI < 13 at each visit will be summarized and analyzed using the same methods as described in Section 9.3.3.

9.3.20 Number of Expressible Glands Yielding Clear Meibum

The number of expressible glands yielding clear meibum at day 14, month 1, month 1.5, and month 3 will be summarized and analyzed using the same methods as described in section 9.3.5.

9.3.21 Eyelid Margin Erythema and Telangiectasias

Eyelid margin erythema and telangiectasias scores at day 14, month 1, month 1.5, and month 3 will be summarized using N, mean, standard deviation, median, minimum, and maximum. Actual values and shifts from baseline will be summarized by parameter and time point. Changes from baseline will be analyzed at each time point using Cochran-Mantel-Haenszel (CMH) statistics to test for a treatment effect.

The following pseudocode will be used:

```
PROC FREQ data=xxx;
BY visit;
TABLE treatment * base * result / CMH;
ODS output CMH=outcmh (where=(ALTHYPOTHESIS='General Association'));
run;
```

9.3.22 Corneal and Conjunctival Staining at Each Visit

Corneal and conjunctival staining scores (0 to 5 scale) at each visit will be summarized by frequency of each score and shift from baseline. Changes from baseline will be analyzed at each time point using Cochran-Mantel-Haenszel (CMH) statistics as described in section 9.3.21.

10 Safety Analyses

Safety measures will be analyzed using the safety population and shift tables will be presented for change from baseline variables.

10.1 Ocular Comfort Questionnaire / Study Medication Tolerability

Ocular comfort questionnaire measures (drug felt comfortable, drug felt soothing, drug felt moistening/lubricating, drug enhanced clear vision, drug stickiness, drug blur, drug burning/stinging, and drug discomfort), will be summarized by treatment group for screening, baseline (Day 0), Day 14, Month 1, Month 1.5, Month 3. Summary will include N, mean, standard deviation, median, minimum, and maximum response (measured in %) at each visit, as well as change from baseline.

10.2 Best-Corrected Visual Acuity (BCVA)

A summary of raw visual acuity (number of letters read) by eye (study eye and non-study eye), treatment group, and visit will be provided in tables for screening, Baseline (Day 0), Day 14, Month 1, Month 1.5, and Month 3. Summary will include N, mean, standard deviation, median, minimum, and maximum response (measured in %) at each visit, as well as change from baseline.

The proportion of patients with a change from baseline BCVA ≤5 and >5 and a change from baseline ≤10 and >10 at each visit will be summarized by study eye and non-study eye using shift tables displaying the frequency and percent of patients in each category (≤5 and >5 / ≤10 and >10) by treatment group and eye.

10.3 Biomicroscopy

Biomicroscopy will be summarized by treatment group, eye (study eye and non-study eye), and visit for screening, Baseline (Day 0), Day 14, Month 1, Month 1.5, and Month 3 by treatment group. The summary will include frequency and percent of each response. A shift table (shift form baseline) will be provided by treatment.

10.4 Ophthalmoscopy

Ophthalmoscopy results (lens status, cataract, vitreous and fundus) will be summarized by treatment group, eye (study eye and non-study eye), and visit for screening (dilated) and month 3 (not dilated) using a shift table (shift form baseline).

10.5 Intraocular Pressure (IOP)

A summary of raw IOP will be provided by treatment group and eye (study eye and non-study eye) at baseline and Month 3. Summary will include N, mean, standard deviation, median, minimum, and maximum response (measured in %) at each visit, as well as change from baseline.

11 Adverse Events

All adverse events (AEs) will be coded using the standardized Medical Dictionary for Regulatory Activities (MedDRA) central coding dictionary, version 23.1 or greater.

11.1 All Adverse Events

Summaries of incidence rates of individual TEAEs by System Organ Class (SOC) and Preferred Term (PT) will be prepared. Because a subject may experience more than one AE, summaries will provide both the number of subjects experiencing at least one event and the number of events within a reporting period. Percentages provided will be the percent of subjects experiencing one or more adverse events. In addition, incidence of TEAEs will be presented by severity (mild, moderate, severe) and by relationship (unrelated, unlikely, possible, probable, very likely/certain) to investigational treatment. Subjects experiencing an event within a given PT and SOC more than once will be counted under the maximum severity/relationship experienced.

A TEAE is any AE that is new in onset or was aggravated in severity or frequency following the application of first study drug, up to 30 days of receiving the last study drug. If the start date of any AE is incomplete or missing, the event will be assumed to be a TEAE, unless the incomplete start date (month and/or year) or the stop date (complete or incomplete) clearly indicates that the event started prior to the study drug.

A listing of all adverse events will include the subject number, AE number, the AE SOC and PT, the severity of AE, whether or not the AE is classified as serious (SAE), the relationship of the AE to the study intervention, the action taken, the outcome, and the adjudication status.

11.2 Adverse Events Leading to Withdrawal

A summary of incidence rates (frequencies and percentages) of AEs leading to study withdrawal, by SOC will be prepared for the safety population. A data listing of AEs leading to withdrawal will also be provided, displaying details of the event(s) captured on the CRF.

11.3 Serious Adverse Events

Summaries of incidence rates and relationship to the investigational drug/procedure of individual SAEs by SOC and PT will be prepared. Summaries will provide both the number of subjects and the number of events within a reporting period. Percentages provided will be the percent of subjects experiencing one or more SAEs. A data listing of SAEs will also be provided, displaying details of the event(s) captured on the CRF.

11.4 Drug or Procedure Related Adverse Events

Summaries of incidence rates of drug and procedure related AEs by SOC and PT will be prepared. Related events include those classified as possible, probable, very likely/certain. Events classified as unrelated or unlikely will be considered not related. Summaries will provide both the number of subjects and the number of events within a reporting period. Percentages provided will be the percent of subjects experiencing one or more study intervention related adverse events. Data listings of drug and procedure related AEs will also be provided, displaying details of the event(s) captured on the CRF.

11.5 Deaths

Should any subjects die during the course of the trial, relevant information will be supplied in a data listing.

12 Other Planned Analyses

12.1 Planned Subgroup Analyses

12.1.1 Sex

Analyses of the primary efficacy sign and symptom will be repeated by gender.

12.1.2 MGD diagnosis and MGS score

Patients will be stratified by duration of MGD diagnosis (i.e., < 5 years or ≥ 5 years) and baseline MGS score for the qualified eye (i.e., the eye meeting the inclusion/exclusion criteria). If the eyes have the same MGS score the right eye will be selected as the study eye.

Thus, subgroup analyses of the primary efficacy sign and symptom are planned for the 4 groups defined by the 2 stratification factors:

1. MGD diagnosis < 5 years and MGS score for the qualified eye < 6
2. MGD diagnosis < 5 years and MGS score for the qualified eye ≥ 6 and ≤ 12

3. MGD diagnosis ≥ 5 years and MGS score for the qualified eye < 6
4. MGD diagnosis ≥ 5 years and MGS score for the qualified eye ≥ 6 and ≤ 12

13 Summary of Changes from the Protocol

The following table provides a list of changes from the protocol to the SAP, and the justification for each change.

Section	Description	Justification
9 Efficacy Analyses	Removed reference to pairwise comparisons between treatment groups.	Throughout the protocol, there is reference to performing pairwise comparisons between treatment groups which is thought to be leftover from the AZ201801 study where there were more than 2 treatment groups. For this study, we will use LSMEANS to compare the two treatment groups, as applicable.

14 Reporting Conventions

All reporting will meet the standards of applicable SOPs.