TITLE PAGE

Protocol Title:	A Randomized, Double-Masked, Vehicle-Controlled, Dose Ranging Study to Assess the Efficacy and Safety of Voclosporin Ophthalmic Solution (VOS) in Subjects with Dry Eye Syndrome
Protocol Number:	AUR-VOS-2019-01
Compound Number:	Voclosporin
Short Title:	AUDREY (AUrinia DRy EYe)
Sponsor Name:	Aurinia Pharmaceuticals Inc.
Legal Registered Address:	#1203 - 4464 Markham Street Victoria, BC V8Z 7X8 Canada

Regulatory Agency Identifier Number(s)

Registry	ID
IND	102,279
NCT Number:	NCT04147650

APPROVAL SIGNATURES

Prepared By:

[redacted]

Signature | Date (dd-mmm-yyyy)

Approved By:

[redacted]

Signature | Date (dd-mmm-yyyy)

[redacted]

Signature | Date (dd-mmm-yyyy)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

VERSION HISTORY

This Statistical Analysis Plan (SAP) for study AUR-VOS-2019-01 is based on protocol Version 3.0 dated 26th June 2020.

Table 1 SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1		Not Applicable	Original version

1. INTRODUCTION

This document details the planned statistical analysis of safety and efficacy data from the Aurinia Dry Eye (AUDREY) protocol (AUR-VCS-2019-01) titled "A Randomized, Double-Masked, Vehicle-Controlled, Dose Ranging Study to Assess the Efficacy and Safety of Voclosporin Ophthalmic Solution (VOS) in Subjects with Dry Eye Syndrome".

The proposed analyses are based on the contents of the final version of the protocol (Version 3.0) dated 26th June 2020.

A complete list of proposed outputs is available in a separate document (AUR-VCS-2019-01-TFL).

1.1. Objectives and Endpoints

To assess the efficacy and safety of three different concentrations of VOS (0.05%, 0.10%, and 0.20%) when administered twice a day (BID) in both eyes (OU) over 12 weeks compared to vehicle in subjects with Dry Eye Syndrome (DES).

The clinical hypothesis for this study is that at least one of three concentrations of VOS (0.05%, 0.10%, and 0.20%) when administered OU BID over 4 weeks is significantly different compared to vehicle alone for the following DES sign endpoint:

- Proportion of subjects with a ≥10 mm increase from baseline in Schirmer Tear Test (STT) at Week 4 in the study eye (Section 5.1.2)

Table 2 Estimand Details

Objective Clinical Category	Statistical Category	Estimand/Variable
Primary Objective To assess the efficacy of three different concentrations of VOS (0.05%, 0.10%, and 0.20%) when administered BID OU over 12 weeks compared to vehicle in subjects with DES. Due to the Coronavirus (COVID-19) pandemic a number of subjects failed to attend Visit 4 where the primary efficacy endpoint was assessed. As detailed in Section 6.2, the primary assessment of efficacy will be made using the modified intent-to-treat (mITT) population with the intent-to-treat (ITT) being considered a supplementary population.		

Objective Clinical Category	Statistical Category	Estimand/Variable
Efficacy Category 1	Primary	Variable: Proportion of subjects with ≥10 mm increase from baseline in STT at Week 4 in the study eye Multiple Comparisons Procedure (MCP): Overall alpha: 5%, 3 treatment comparisons vs vehicle. Statistical significance claimed where p<0.0166 Population: Modified Intent-to-Treat (mITT) Intercurrent event(s) strategy (IES): Subjects withdrawing from study before Week 4 and/or failing to provide a Week 4 STT assessment (for reasons other than Coronavirus) will be analyzed as non-responders Population Level Summary (PLS): Odds ratios (OR) for each active treatment group compared to vehicle (Logistic regression) Analysis: Logistic regression model including covariates: Investigator site, Baseline Eye Dryness level (<60mm or ≥60 mm), and Baseline STT
	Sensitivity	Analysis: Logistic regression model without covariates
	Supplementary	Population: Per Protocol (PP) Analysis: Logistic regression model including covariates: Investigator site, Baseline Eye Dryness level, and Baseline STT

Objective Clinical Category	Statistical Category	Estimand/Variable	
	Supplementary	Population:	PP
		Analysis:	Logistic regression model without covariates
	Supplementary	Population:	Intent-to-Treat (ITT)
		Analysis:	Logistic regression model including covariates: Investigator site, Baseline Eye Dryness level, and Baseline STT
	Supplementary	Population:	ITT
		Analysis:	Logistic regression model without covariates
	Supplementary	Population:	mITT
		IES:	Analysis of observed STT response only.
		Analysis:	Logistic regression model including covariates: Investigator site, Baseline Eye Dryness level, and Baseline STT
	Supplementary	Population:	mITT
		IES:	Analysis of observed STT response only.
		Analysis:	Logistic regression model without covariates

Objective Clinical Category	Statistical Category	Estimand/Variable	
	Supplementary	Population:	ITT
		IES:	Analysis of observed STT response only.
		Analysis:	Logistic regression model including covariates: Investigator site, Baseline Eye Dryness level, and Baseline STT
	Supplementary	Population:	ITT
		IES:	Analysis of observed STT response only.
		Analysis:	Logistic regression model without covariates
	Supplementary	Population:	PP
		IES:	Analysis of observed STT response only.
		Analysis:	Logistic regression model including covariates: Investigator site, Baseline Eye Dryness level, and Baseline STT
	Supplementary	Population:	PP
		IES:	Analysis of observed STT response only.
		Analysis:	Logistic regression model without covariates

Objective Clinical Category	Statistical Category	Estimand/Variable	
Efficacy Category 2	Key Secondary	Variable:	Mean change from baseline to Week 4 in Eye Dryness Visual Analogue Scale (VAS)
		MCP:	Overall alpha: 5%, 3 treatment comparisons vs vehicle. Statistical significance claimed where $p<0.0166$
		Population:	mITT subjects with an Eye Dryness VAS at baseline ≥60 mm
		IES:	Subjects withdrawing from study before Week 4 and/or failing to provide a week 4 Eye Dryness VAS will not contribute to this analysis
		PLS:	Least Squares mean (LSM) differences between active treatment groups and vehicle
		Analysis:	Analysis of Covariance (ANCOVA) model including covariates for Investigator site and baseline Eye Dryness score.
	Supplementary	Population:	PP subjects with an Eye Dryness VAS at baseline ≥60 mm
		Analysis:	ANCOVA model including covariates for Investigator site and baseline Eye Dryness score.
	Supplementary	Population:	ITT subjects with an Eye Dryness VAS at baseline ≥60 mm
		Analysis:	ANCOVA model including covariates for Investigator site and baseline Eye Dryness score.

Objective Clinical Category	Statistical Category	Estimand/Variable	
Efficacy Category 3	Secondary	Variable:	Proportion of subjects with ≥10 mm increase from baseline in STT in the study eye analyzed at Weeks 2, 8 and 12
		MCP:	No adjustment
		Population:	ITT
		IES:	All observed data contribute to the analysis.
		PLS:	OR for each active treatment group compared to vehicle (Logistic regression)
		Analysis:	Logistic regression model including covariates: Investigator site, Baseline Eye Dryness level and Baseline STT
	Secondary	Variable:	Mean change from baseline in Fluorescein Corneal Staining (FCS) score for the 5 individual regions of the cornea to Weeks 2, 4, 8 and 12 in the study eye
		MCP:	No adjustment
		Population:	ITT
		IES:	All observed data contribute to the analysis. Mixed Model Repeated Measures (MMRM) used for imputation of missing data.
		PLS:	LSM differences between treatment groups at each time point
		Analysis:	MMRM model for each of the 5 regions including covariates: Investigator site, Baseline Eye Dryness level, Baseline FCS region score, Visit, Treatment group and Visit*Treatment interaction

Objective Clinical Category	Statistical Category	Estimand/Variable	
	Secondary	Variable:	Mean change from baseline in Total FCS score to Weeks 2, 4, 8 and 12 in the study eye
		MCP:	No adjustment
		Population:	ITT
		IES:	All observed data contribute to the analysis. MMRM used for imputation of missing data.
		PLS:	LSM differences between treatment groups at each time point
		Analysis:	MMRM model including covariates: Investigator site, Baseline Eye Dryness level, Baseline total FCS, Visit, Treatment group and Visit*Treatment interaction
	Secondary	Variable:	Mean change from baseline in STT score to Weeks 2, 4, 8 and 12 in the study eye
		MCP:	No adjustment
		Population:	ITT
		IES:	All observed data contribute to the analysis. MMRM used for imputation of missing data.
		PLS:	LSM differences between treatment groups at each time point
		Analysis:	MMRM model including covariates: Investigator site, Baseline Eye Dryness level, Baseline STT, Visit, Treatment group and Visit*Treatment interaction

Objective Clinical Category	Statistical Category	Estimand/Variable	
	Secondary	Variable:	Mean change from baseline in Eye Dryness VAS score to Weeks 2, 4, 8 and 12
		MCP:	No adjustment
		Population:	ITT
		IES:	All observed data contribute to the analysis. MMRM used for imputation of missing data.
		PLS:	LSM differences between treatment groups at each time point
		Analysis:	MMRM model including covariates: Investigator site, Baseline Eye Dryness, Visit, Treatment group and Visit*Treatment interaction
	Secondary	Variable:	Mean change from baseline in Ocular Discomfort VAS score to Weeks 2, 4, 8 and 12
		MCP:	No adjustment
		Population:	ITT
		IES:	All observed data contribute to the analysis. MMRM used for imputation of missing data.
		PLS:	LSM differences between treatment groups at each time point
		Analysis:	MMRM model including covariates: Investigator site, Baseline Eye Dryness level, Baseline Ocular Discomfort, Visit, Treatment group and Visit*Treatment interaction

Objective Clinical Category	Statistical Category	Estimand/Variable	
	Secondary	Variable:	Mean change from baseline in Burning/Stinging, Itching, Photophobia, Eye Pain, Foreign Body Sensation and Blurred Vision VAS scores to Weeks 2, 4, 8 and 12
		MCP:	No adjustment
		Population:	ITT
		IES:	All observed data contribute to the analysis. MMRM used for imputation of missing data.
		PLS:	LSM differences between treatment groups at each time point
		Analysis:	MMRM model for each scale including covariates: Investigator site, Baseline Eye Dryness level, Individual Baseline Symptom Score, Visit, Treatment group and Visit*Treatment interaction
	Secondary	Variable:	Mean change from baseline in the sum of Individual Symptom VAS score to Weeks 2, 4, 8 and 12
		MCP:	No adjustment
		Population:	ITT
		IES:	All observed data contribute to the analysis. MMRM used for imputation of missing data.
		PLS:	LSM differences between treatment groups at each time point
		Analysis:	MMRM model including covariates: Investigator site, Baseline Eye Dryness level, Baseline Total Symptom Score, Visit, Treatment group and Visit*Treatment interaction

Objective Clinical Category	Statistical Category	Estimand/Variable	
	Secondary	Variable:	Mean change from baseline in Symptom Assessment in Dry Eye (SANDE) frequency and severity scores to Weeks 2, 4, 8 and 12
		MCP:	No adjustment
		Population:	ITT
		IES:	All observed data contribute to the analysis. MMRM used for imputation of missing data.
		PLS:	LSM differences between treatment groups at each time point
		Analysis:	MMRM model including covariates: Investigator site, Baseline Eye Dryness level, Baseline SANDE frequency or severity Score, Visit, Treatment group and Visit*Treatment interaction
Safety Objective To assess the safety of three different concentrations of VOS (0.05%, 0.10%, and 0.20%) when administered twice a day (BID) in both eyes (OU) over 12 weeks compared to vehicle in subjects with DES.			
Safety Category 1	AEs (Adverse Events) and SAEs (Serious AEs)	Incidence rates to be calculated for ocular and non-ocular AEs, SAEs and various subgroups.	

Objective Clinical Category	Statistical Category	Estimand/Variable	
	Best Corrected Visual Acuity (BCVA)	Variable:	Mean Change from Baseline in BCVA over time
		Population:	Safety
		Analysis:	Descriptive
	Slip-Lamp Biomicroscopy	Variable:	Changes from Baseline in Slit-Lamp Biomicroscopy over time
		Population:	Safety
		Analysis:	Descriptive
	Dilated Ophthalmoscopy	Variable:	Changes from Baseline in Dilated Ophthalmoscopy over time
		Population:	Safety
		Analysis:	Descriptive

1.2. Study Design

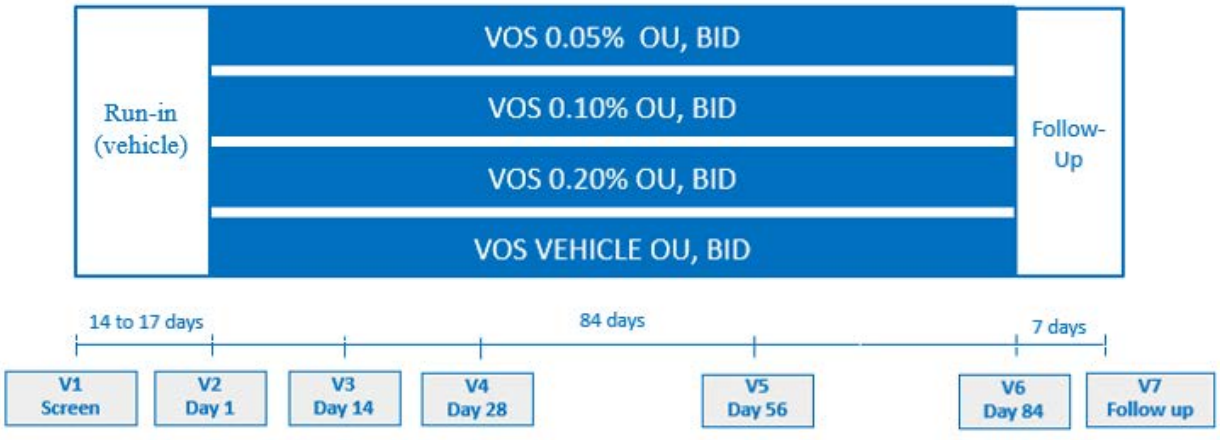
This is a Phase 2/3, multi-center, randomized, double-masked, vehicle-controlled study to assess the efficacy and safety of three different concentrations of VOS when administered OU, BID over 12 weeks in subjects with mild to moderate DES. Subjects will undergo a 2-week run-in period in which VOS vehicle will be self-administered OU, BID. Subjects will be re-assessed to confirm that all participants meet all of the inclusion criteria and none of the exclusion criteria. It was estimated the study will enroll 480 subjects (updated to 505 subjects due to COVID-19) across 9 study centers. Eligible subjects will be randomized in a 1:1:1:1 ratio to one of the following study treatment arms after the 14- to 17-day run-in period:

- 0.05% VOS BID
- 0.10% VOS BID
- 0.20% VOS BID
- VOS Vehicle BID

Subjects will be stratified by eye dryness VAS score ≥60 mm and <60 mm.

See Figure 1 for the AUDREY (AUR-VOS-2019-01) study schematic.

Figure 1 AUDREY (AUR-VOS-2019-01) Study Schematic



Notes: V = Visit; VOS = Voclosporin ophthalmic solution

2. STATISTICAL HYPOTHESES

The primary endpoint will be tested under the following hypotheses with statistical significance declared for any of the three comparisons versus vehicle where a p-value <0.0166 is observed:

H_{01}: There is no difference between the proportion of subjects with a ≥10 mm change from baseline in STT to Week 4 in the study eye for those receiving VOS (0.05%, 0.10%, or 0.20%) versus vehicle.

H_{11}: There is a difference between the proportion of subjects with a ≥10 mm change from baseline in STT to Week 4 in the study eye in for those receiving VOS (0.05%, 0.1%, or 0.2%) versus vehicle.

Following a significant result in the primary analysis the three pairwise comparisons for the key secondary endpoint will be tested using alpha=0.0166. The hypotheses to be tested are:

H_{02}: In subjects with baseline eye dryness VAS ≥60 mm, there is no difference between the mean change from baseline to Week 4 in the Eye Dryness VAS for those receiving VOS (0.05%, 0.10%, or 0.20%) versus vehicle.

H_{12}: In subjects with baseline eye dryness VAS ≥60 mm, there is a difference between the mean change from baseline to Week 4 in the eye dryness VAS for those receiving VOS (0.05%, 0.10%, or 0.20%) versus vehicle.

For significance to be declared in the key secondary endpoint analysis, the same treatment group should show improvement over vehicle in both the primary analysis and the key secondary analysis.

In order to maintain the familywise error rate at 5%, the Bonferroni correction will be used with statistical significance of the 3 voclosporin vs vehicle comparisons of the primary and key secondary endpoints being declared when p<0.0166

No adjustment is made for multiple secondary or safety endpoints; statistical significance will be declared when p<0.05.

3. SAMPLE SIZE DETERMINATION

This study was expected to enroll 480 subjects (amended to 505 subjects due to COVID-19) into 4 treatment arms. A two-group continuity corrected Chi square test with a 0.0166 two-sided significance level (adjusted for three treatment comparisons versus placebo) will have at least 80% power to detect the difference between a vehicle response rate of 20% and a VOS response rate of 40% when the sample size in each group is 120 (total N=480). Response is defined as an increase of ≥10 mm in STT from baseline to Week 4.

While the effect of withdrawals will be investigated, subjects withdrawing prior to Week 4 STT assessment for any reason will be counted as non-responders in the primary analysis and therefore no adjustment of sample size for withdrawals is necessary.

This sample size provides greater than 90% power to detect a significant change from baseline within any one of the three active treatment groups in the Eye Dryness VAS assuming the standard deviation of changes is 40 mm and the mean change from baseline is 20 mm (two-sided alpha=0.0166).

The key secondary endpoint is analyzed within the population of subjects with an Eye Dryness VAS ≥60 mm at baseline. It is expected that this will reduce the sample size for this analysis by 25% (from 120 per group to 90 per group). Assuming a standard deviation of 30 mm for change in eye dryness score and an improvement of any active arm compared to placebo (alpha=0.0166) of 15 mm, a sample size of 90 subjects per group provides at least 80% power to detect a significant difference.

The study was ongoing at the time of the Coronavirus (COVID-19) pandemic. The impact of the pandemic was continually being assessed and it became apparent that a number of subjects were unable to attend the primary efficacy visit at Week 4 due to site closure. To account for the higher incidence of missing data and ensure data from at least 480 subjects were available for analysis of the primary endpoint (STT response at Week 4), the target sample size was increased to approximately 505 subjects.

4. POPULATIONS FOR ANALYSIS

The following analysis populations will be considered:

- Screened Population: all subjects who sign informed consent
- Run-In Population: all subjects who received at least one dose of run-in medication.
- Intent-to-Treat Population: the ITT population includes all randomized subjects. Subjects in the ITT population will be analyzed as randomized.
- Modified Intent-to-Treat Population: the mITT population is based on the ITT population and excludes all subjects who were unable to attend their primary endpoint assessment at the Week 4 visit due to Coronavirus-related site closure. Due to block enrolment, this applies to all subjects at Site 36. Thus, these subjects will be excluded from the mITT population. Subjects in this population will be analyzed as randomized. As detailed in Section 6.2, the mITT population will be used for the primary assessment of efficacy.
- Per Protocol Population: The PP population includes subjects in the mITT population who do not have significant protocol deviations prior to their primary endpoint assessment at Week 4 and who complete the Week 4 STT. Protocol deviations will be assessed prior to database lock and unmasking. The PP population will be analyzed as randomized.
- Safety Population: The safety population includes all randomized subjects who have received at least one dose of randomized study treatment. The safety population will be analyzed for all safety assessments. Subjects in the safety population will be analyzed as treated.

The run-in population will be used to describe all subjects who entered the run-in period.

Baseline data will be summarized using the safety, ITT, PP and mITT populations. Safety data will be summarized using the safety population, and efficacy analysis will use the ITT, PP and mITT populations.

Assignment of treatment group for the safety population will be made using an individual subject's highest dose received.

5. STATISTICAL ANALYSES

5.1. General Considerations

All statistical analyses and reporting will be performed using the SAS® System Version 9.4 or later.

Unless otherwise specified, continuous variables will be summarized with descriptive statistics (n, mean, median, standard deviation, standard error, minimum, and maximum), and categorical variables will be summarized with counts and percentages.

All statistical models, unless otherwise specified, will include terms for the randomization strata (Eye Dryness), investigator site, appropriate baseline measure and treatment group. Where baseline score and randomization strata provide the same data (e.g., in the analysis of Eye Dryness), only baseline score will be used in the model. The eye dryness stratification level used in the modeling process will be a subject's actual baseline eye dryness level (<60mm or ≥60mm) even if this is different to the level recorded at randomization.

5.1.1. Decision Criteria

All tests will be 2-sided with statistical significance being declared at the nominal 5% level. For the 5% error rate to be maintained in the primary and subsequently in the key secondary analyses, each voclosporin vs vehicle comparison will be tested at the 0.0166 alpha level. Confidence intervals (CI) will be labeled as nominal 95% CI and will be constructed as 98.33% CI.

Intervals for secondary and safety analyses will be labeled and constructed as 95% CIs.

5.1.2. Unit of Analysis

Safety endpoints will be analyzed for each eye. For efficacy symptom-related endpoints, the unit of analysis is both eyes. For efficacy sign-related endpoints, the unit of analysis will be the study eye as defined by the following:

Study Eye: The study eye will be defined as the qualifying eye that achieves the lowest STT score at baseline (Visit 2). Should both eyes be qualifying eyes with identical STT scores at baseline, the eye with the worst FCS score will be used. Should these scores also be equal, the right eye will be used.

5.1.3. Calculation of Study Day

Study day will be calculated as the number of days from first dose of study drug (Day 1):

- Date of event – date of first dose of study drug + 1, for events on or after first dose
- Date of event – date of first dose of study drug, for events before first dose

To this end, Day 0 remains undefined.

5.1.4. Baseline

For on-treatment comparisons, baseline is defined as the last non-missing value (either scheduled or unscheduled) following run-in and before the subject receives the first dose of randomized study drug on Day 1. As subjects progress from the end of run-in directly to the start of randomized study treatment, baseline data will come from Visit 2 (Day 1).

Visual acuity assessments are occasionally repeated as unscheduled visits a few minutes after the visit assessment (for example, due to dirty glasses). Such unscheduled assessments on the same day as the baseline visit will be available to be selected as the baseline value.

For summaries of changes over the run-in period, baseline will be defined as the screening visit with changes *from* the screening visit (Visit 1) *to* the baseline visit (Visit 2) being described.

5.1.5. Visit Windowing for Analysis

All data collected will be summarized and analyzed as belonging to the closest protocol-scheduled visit.

Subjects with multiple sets of data within a single visit window will contribute their data collected closest to the scheduled visit day. Should two sets of data be equidistant, the later shall be used.

All data will be listed.

5.1.6. Analysis Timelines

All statistical analysis will be undertaken following database lock.

5.1.7. Use of Study Site in Analysis Models

By default, study site will be included in all analysis models to adjust for any differences that may exist. Should the addition of study site to the models cause issue (e.g., convergence of a logistic model), sites will be combined in the following iterative manner prior to analysis:

- The site causing issue will be combined with the site with the lowest number of subjects and the model re-run.
- If issues persist, the combining of sites will continue in a similar fashion.

5.1.8. Handling of Missing Data

Response endpoints will be analyzed at specific time points. Subjects with missing responses will be included as non-responders in the primary analysis.

Endpoints measured on the continuous scale (including VASs) and collected / analyzed at multiple visits will be analyzed using MMRM in order to account for incomplete data.

Missing and partial start and stop dates will be imputed for analysis purposes as follows (medical history partial dates will not be imputed):

- Should the data available be sufficient to deduce that the event or medication started prior to first dose of randomized study medication, the date will be imputed as being as late as possible (as close to randomized study medication start as possible). For example, if only the year 2017 is known and randomized study medication started in 2020, 31st Dec 2017 will be imputed.

- Should the data available be sufficient to deduce that the event or medication started subsequent to the first dose of randomized study medication, the date will be imputed as being as early as possible (as close to randomized study medication start as possible). For example, if only the year 2020 is known and randomized study medication started in 2019, 1st Jan 2020 will be imputed.

- Should the data available be ambiguous as to whether the event or medication started during randomized dosing then the date of the first dose of randomized study medication will be imputed. For example, if only December 2019 is known and randomized study medication started on 15th December 2019, 15th Dec 2019 will be imputed.

- Stop dates will never be imputed to be earlier than their corresponding start date.

5.2. Participant Dispositions

Subject disposition will be summarized as follows:

The number of subjects who failed screening and the reasons for failure will be tabulated for the Screened Population.

The number of subjects who entered the vehicle run-in period, completed run-in or the reason for failure will be tabulated for the run-in population.

Numbers of subjects randomized and who are in the safety population will be summarized by treatment group and overall for the safety population.

Numbers of subjects randomized and who are in each analysis population (ITT, mITT and PP) will be summarized by treatment group and overall. This summary will be repeated for individual levels of the stratification variable.

Reasons for withdrawal from the study (including number of subjects completing) will be summarized by treatment group and overall.

Reasons for exclusion from the PP analysis population will be summarized by treatment group and overall.

Visit participation will be summarized following the visit windowing algorithm.

5.3. Primary Endpoint Analysis

The primary endpoint is the proportion of subjects with ≥10 mm increase from baseline in STT at Week 4 in their study eye (see Table 2).

Due to COVID-19, the primary assessment of efficacy will be made using the mITT population.

5.3.1. Definition of endpoint

The primary endpoint is a response endpoint with subjects who have increased their STT score by ≥10 mm between baseline and week 4 (following visit windowing) being defined as responders. All subjects who have insufficient data to ascertain a response will be analysed as non-responders.

5.3.2. Main analytical approach

The response for each subject in the mITT population will be analyzed using a logistic regression model including covariates for investigator site, baseline eye dryness level and baseline STT score.

Response rates for each arm will be displayed along with Odds Ratios for each active voclosporin arm compared to the vehicle arm. Nominal 95% CIs and p-values for each comparison will be provided.

5.3.3. Sensitivity analysis

To assess the impact of covariates in the primary model, the primary analysis will be repeated using a logistic model without covariates.

5.3.4. Supplementary analyses

To assess the impact of subjects who violated key aspects of the protocol, the primary model will be run using the PP population (with and without covariates).

To assess the impact of omitting subjects who failed to provide a valid STT at Week 4 due to COVID-19 reasons, the primary model will be run using the ITT population (with and without covariates).

To assess the impact of assuming all subjects with missing data are non-responders, the primary model will be run using a dataset of observed responses only for the ITT, mITT and PP populations (with and without covariates).

5.4. Secondary Endpoints Analysis

5.4.1. Key/Confirmatory secondary endpoint

The key secondary endpoint is the mean change from baseline to Week 4 in the Eye Dryness VAS. The population will be all mITT subjects with a baseline Eye Dryness VAS ≥60 mm (see Table 2).

5.4.1.1. Definition of endpoint

This endpoint is defined as the difference between Week 4 Eye Dryness VAS and the baseline Eye Dryness VAS (after visit windowing). Subjects without a Week 4 Eye Dryness VAS score will not contribute to this analysis.

5.4.1.2. Main analytical approach

This data will be analyzed using an ANCOVA model including covariates for Investigator site, baseline Eye Dryness score and treatment group.

Mean changes for each arm will be displayed along with least squares mean changes and differences between least squares mean changes for each active voclosporin arm compared to the vehicle arm. Nominal 95% CIs and p-values for each comparison will be provided.

Statistical significance of a treatment group comparison will only be claimed following significance in the primary endpoint analysis for the same dose level.

5.4.1.3. Sensitivity analysis

No sensitivity analyses are planned for this endpoint.

5.4.1.4. Supplementary analyses

To assess the impact of subjects who violated key aspects of the protocol, the analysis will be run using the ITT and PP populations.

5.4.2. Supportive secondary endpoint(s)

Other supportive secondary endpoints are described in Table 2. Details of their analysis are below.

5.4.2.1. Schirmer Tear Test at other visits

Observed STT responses at Weeks 2, 8 and 12 will be analyzed in a similar fashion to the primary endpoint using the ITT population only.

5.4.2.2. Total FCS score

Mean change in Total FCS score at Weeks 2, 4, 8 and 12 will be analyzed using MMRM analysis. Under the missing at random assumption the MMRM model estimates the mean treatment effect assuming that, following withdrawal, subjects would have continued in a similar fashion to other subjects on the same arm who have similar covariates and similar data (up to the point of withdrawal).

The model will include terms for investigator site, baseline Eye Dryness level, treatment, visit, treatment by visit interaction and baseline Total FCS score.

Mean changes for each visit and each arm will be displayed along with least squares mean changes and differences between least squares mean changes for each visit (and overall) for each active voclosporin arm compared to the vehicle arm. Associated 95% CIs and p-values for each comparison will be provided.

Mean change over the run-in period will be summarized.

5.4.2.3. FCS score for the 5 individual regions

The FCS scores for each individual region will be analyzed in a similar fashion to the Total FCS score.

5.4.2.4. Schirmer Tear Test Score

The STT score will be analyzed in a similar fashion to the Total FCS score.

5.4.2.5. Eye Dryness Score

Eye dryness scores will be analyzed in a similar fashion to the Total FCS score.

Eye Dryness scores will additionally be assessed with respect to changes from baseline within treatment group. P-values and associated 95% CIs for changes from baseline for each treatment group will be provided from the MMRM model. Note that as baseline Eye Dryness will be a covariate in the model, Eye Dryness baseline level will not be included in the model.

5.4.2.6. Ocular Discomfort Score

Ocular discomfort score will be analyzed in a similar fashion to the Total FCS score.

5.4.2.7. Individual Symptom Severity Scores

Individual symptom severity scores (Burning/Stinging, Itching, Photophobia, Eye Pain, Foreign Body Sensation and Blurred Vision) will be analyzed in a similar fashion to the Total FCS score.

5.4.2.8. Total Symptom Severity Score

Total symptom severity score will be analyzed in a similar fashion to the Total FCS score.

5.4.2.9. SANDE Severity and Frequency Scores

SANDE severity and frequency scores will be analyzed in a similar fashion to the Total FCS score.

5.5. Tertiary/Exploratory Endpoints Analysis

There are no further efficacy endpoints defined for analysis.

5.6. Safety Analyses

Safety summaries and analyses will comprise study drug exposure, adverse events, Best Corrected Visual Acuity, Slit-Lamp Biomicroscopy and Dilated Ophthalmoscopy. Each is described in the following sections.

5.6.1. Extent of Compliance and Exposure

Compliance will be summarized separately for the run-in period and the randomized treatment period.

Compliance will be calculated for the run-in period and the treatment period as follows:

$$Compliance\ \% = 100 * \frac{Total\ Capsules\ Dispensed - Total\ Capsules\ Returned}{(Days\ in\ Period) * 2}$$

Note: Days in the run-in period will not include the day of randomization (Day 1)

Compliance will be summarized using the safety, ITT, mITT and PP populations using the following categories:

- <80%
- 80 to 120%
- >120%

Exposure to study drug will be summarized as a duration.

Duration will be calculated as: Last Dose Day – First Dose Day + 1. It will be summarized using the following categories:

- <1 week
- 1 to <4 weeks
- 4 to <8 weeks
- ≥8 weeks

5.6.2. Adverse Events

A treatment-emergent adverse event is defined as any AE with an onset date on or after the first dose of randomized study treatment (Day 1).

Any AE with an onset date between the date of first dose of run-in medication and the day prior to the first dose of randomized study treatment is defined as a run-in AE.

Adverse event summary tables will be produced separately for ocular events and non-ocular events.

Summary tables will include proportions of subjects experiencing events as well as counts of the event in question. For summaries by severity, where only the most severe events are counted, summaries will not include event counts.

Summaries of treatment-emergent adverse events (TEAE) by decreasing frequency of System Organ Class, decreasing frequency of Preferred Term and treatment group will include:

- All events
- All events by severity
- Treatment-related events

- Disease-related events
- Events Resulting in study drug interruption
- Events Resulting in study drug discontinuation
- Events Resulting in death
- Above summaries (other than by severity and Resulting in death) repeated for serious TEAEs

An overall summary table with the numbers of subjects and events in each of the above categories will be created.

A table of all TEAEs by decreasing frequency of Preferred Term will be produced.

A summary of all run-in events will be produced by decreasing frequency of System Organ Class and decreasing frequency of Preferred Term. Summaries of treatment-related, disease-related and serious events will also be produced for the run-in period.

Separate listings of all serious TEAEs, all TEAEs resulting in death and all serious non-TEAEs will be produced.

Adverse event coding will use the Medical Dictionary for Regulatory Activities (MedDRA) with the version used being specified on relevant outputs.

5.6.3. Best Corrected Visual Acuity

The BCVA logMAR (logarithm base 10 of the minimal angle of resolution) score will be summarized for each eye by visit as absolute values and change from baseline. 95% CIs for change from baseline within each treatment group will be provided. Treatment comparisons will be provided for descriptive purposes.

Mean change over the run-in period will be summarized.

5.6.4. Slit-Lamp Biomicroscopy

The proportions of subjects with normal or abnormal results in each category for each eye will be summarized by visit.

A shift table will also be produced showing the changes from baseline to on-treatment results for each category.

5.6.5. Dilated Ophthalmoscopy

The proportions of subjects with normal or abnormal results in each category for each eye will be summarized by each visit.

A shift table will also be produced showing the changes from baseline to on-treatment result for each category.

5.7. Other Analyses

Not Applicable.

5.8. Interim Analyses

No interim analysis is planned.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1 List of Abbreviations

AE	Adverse Event
ANCOVA	Analysis of Covariance
AUDREY	Aurinia Dry Eye
BID	Twice daily
CI	Confidence Interval
DES	Dry Eye Syndrome
FCS	Fluorescein Corneal Staining
IES	Intercurrent Events Strategy
ITT	Intent-to-Treat
LSM	Least Squares Mean
MCP	Multiple Comparisons Procedure
mITT	Modified Intent-to-Treat
MMRM	Mixed Model Repeated Measures
OR	Odds Ratio
OD	Right Eye
OS	Left Eye
OU	Both Eyes
PLS	Population Level Summary
STT	Schirmer Tear Test
TEAE	Treatment-Emergent Adverse Event
VAS	Visual Analogue Scale
VOS	Voclosporin Ophthalmic Solution

6.2. Appendix 2: Changes to Protocol-Planned Analyses

All analysis will be performed as specified in Protocol version 3.0 dated 26th June 2020.

6.3. Appendix 3: Other Data

6.3.1. Study performance

A table showing the earliest and latest date of informed consent, earliest and latest randomization and the latest last contact date will be produced.

6.3.2. Baseline Characteristics and Demography

Demographic variables (age, sex, race and ethnicity) will be summarized by treatment group and overall for the Safety, ITT, mITT and PP populations. Demography tables will be run for each study site individually.

6.3.3. Medical History

Medical history will be summarized for the Safety and the ITT populations. Ocular and non-ocular histories will be summarized separately by System Organ Class and Preferred Terms.

6.3.4. Prior and Concomitant Medications

Prior and concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) level 2 and Preferred Terms for the safety and ITT populations. Tables will be produced separately for ocular and non-ocular medications for:

- Prior Medications (defined as medications that start and stop prior to the start of the first dose of study medication).
- Prior and Concomitant Medications (defined as medications that are being taken at the time of the first dose of study medication).
- Concomitant Medications (defined as those medications that start following the first dose of study medication).

All medical procedures will be listed.

6.3.5. Subject Data Listings

All data collected on the database will be included in subject data listings.

7. REFERENCES

None.