STATISTICAL ANALYSIS PLAN

PROTOCOL: NGF0221

A 4-week, Phase III, multicenter, double-masked, vehicle-controlled clinical study to evaluate safety and efficacy of Cenegermin (Oxervate®) 20 mcg/mL ophthalmic solution versus vehicle, in patients with severe Sjogren's dry eye disease under treatment with Cyclosporine A (PROTEGO-2 study).

APPROVAL PAGE

Document Information	
Protocol Number:	NGF0221
Version:	Final 1.0
Document Date:	31 March 2023

The Statistical Analysis Plan has been completed and reviewed and the contents are approved for use for the analysis.

Author Statistician details:	
Name:	PPD
Job Role:	PPD
Company:	PPD
Signature:	PPD
Date of signature:	31-Mar-2023 | 12:03 CEST

Lead Statistician details:	
Name:	PPD
Job Role:	PPD
Company:	PPD
Signature:	PPD
Date of signature:	31-Mar-2023 | 13:36 CEST

Sponsor Approver details:	
Name:	PPD
Job Role:	PPD
Company:	PPD
Signature:	PPD
Date of signature:	03-Apr-2023 | 15:26 CEST

Contents

Index of Tables

Abbreviations

ADR	Adverse Drug Reaction
AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomical Therapeutic Chemical
BCDVA	Best Corrected Distance Visual Acuity
BDRM	Blind Data Review Meeting
CsA	Cyclosporine A
CSR	Clinical Study Report
eCRF	electronic Case Report Form
ENR	Enrolled set
EU	European Union
FAS	Full Analysis set
KFR	Key First efficacy Results
IDEEL	Impact of dry eye on everyday life
IMP	Investigational Medicinal Product
MAR	Missing at Random
MCMC	Markov chain Monte Carlo
MedDRA	Medical Dictionary for regulatory activities
MI	Multiple Imputation
MI-RD	Multiple Imputation using retrieve dropouts
MMRM	Mixed Model for Repeated Measurements
MNAR	Missing Not at Random
NEI	National Eye Institute
PP	Per protocol set
PT	Preferred Term
rhNGF	recombinant human Nerve Growth Factor
RND	Randomized set
SAE	Serious Adverse Event
SAF	Safety set
SANDE	Symptom Assessment in Dry Eye
SAP	Statistical Analysis Plan
SEM	Standard error of the mean
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TFBUT	Tear Film Break-Up Time
TID	Ter in die (three times a day)
US	United States

1. Revision History

Table 1: Revision history

Document Version	Changes Made	Document Date
Final 1.0	First release	31 March 2023

2. Introduction

This document outlines the statistical methods to be implemented in the analysis of the data of NGF0221 Clinical Trial. The purpose of this plan is to provide general guidelines from which the analysis will proceed, containing a more technical and detailed elaboration of the principal features of the analysis described in the protocol. Any changes to the protocol or Case Report Form (CRF) may necessitate updates to the Statistical Analysis Plan (SAP). In case of deviations from this updated SAP, explanations will be provided in the Clinical Study Report (CSR).

This SAP is based on study protocol Version No. 5.0 – 16 June 2022 [1], Patient diary Version No. 3.0 – 07 February 2022 [2] and Case Report Form Version No. MSC004 – 02 March 2023 [3].

3. Study Design

3.1 General design and plan

This is a 4-weeks phase III, multicenter, double-masked, vehicle-controlled clinical study to evaluate safety and efficacy of cenegermin ophthalmic solution at 20 mcg/mL solution versus vehicle, in patients with severe Sjogren's dry eye disease under treatment with Cyclosporine A (Cyclosporine A or other topical ophthalmic treatment of the same class).

3.2 Study Objectives and endpoint

The study objective is to assess the efficacy and safety of cenegermin ophthalmic solution at 20 mcg/mL concentration administered three times daily for 4 weeks in patients with severe Sjogren's dry eye disease under treatment with Cyclosporine A (CsA).

3.2.1 Primary

Objective(s)	Endpoint(s)
• To compare the efficacy of cenegermin vs. vehicle in Schirmer I test (without anesthesia) >10mm/5min at week 4 by testing the superiority.	• Schirmer I test (without anesthesia) >10mm/5min at week 4.
• To compare the efficacy of cenegermin vs. vehicle in Symptom Assessment in Dry Eye questionnaire (SANDE) global score at week 12 by testing the superiority.	• Change from baseline in SANDE global score at week 12

3.2.2 Secondary

Objective(s)	Endpoint(s)
• To compare the efficacy of cenegermin vs. vehicle in Schirmer I test at week 4, 8, 12 and 16 by testing the superiority.	• Change from baseline in Schirmer I test [Time frame: week 4, 8, 12 and 16].

Objective(s)	Endpoint(s)
• To compare the efficacy of cenegermin vs. vehicle in Cornea and conjunctiva vital staining with fluorescein (National Eye Institute [NEI] scales) at week 4, 8, 12 and 16 by testing the superiority.	• Change from baseline in Cornea and conjunctiva vital staining with fluorescein (NEI scale) [Time frame: week 4, 8, 12 and 16].
• To compare the efficacy of cenegermin vs. vehicle in Tear Film Break-Up Time (TFBUT) at week 4, 8, 12 and 16 by testing the superiority.	• Change from baseline in TFBUT [Time frame: week 4, 8, 12 and 16].
• To compare the efficacy of cenegermin vs. vehicle in SANDE scores at week 8, 12 and 16 by testing the superiority.	• Change from baseline in SANDE global scores, and for severity and frequency [Time frame: week 8, 12 and 16].
• To compare the efficacy of cenegermin vs. vehicle in worsening in symptom scores (SANDE) and/or NEI score at week 4 by testing the superiority.	• Number of patients experienced a worsening in symptom scores (SANDE global score) and/or NEI score ≥ 50% assessed at week 4.
• To compare the efficacy of cenegermin vs. vehicle in impact of dry eye on everyday life (IDEEL) questionnaire at week 4, 8, 12 and 16 by testing the superiority.	• IDEEL questionnaire [Time frame: week 4, 8, 12 and 16].

3.2.3 Exploratory

Objective(s)	Endpoint(s)
• To evaluate the use of preservative free artificial tears use during the treatment period.	• Proportion and frequency of preservative free artificial tears use (n° drops/day) during the treatment period.
• To evaluate the use of preservative free artificial tears use during the treatment period.	• Frequency of preservative free artificial tears use (n° drops/day) during the follow-up period.
• To compare the efficacy of cenegermin vs. vehicle in Schirmer I test at week 2 by testing the superiority.	• Change from baseline in Schirmer I test (without anesthesia) Vs week 2.
• To compare the efficacy of cenegermin vs. vehicle in SANDE scores at week 2 and 4 by testing the superiority.	• Change from baseline in SANDE global scores, and for severity and frequency [Time frame: week 2 and 4].
• To compare the efficacy of cenegermin vs. vehicle in Cornea and conjunctiva vital staining with fluorescein (NEI scale) at week 2 by testing the superiority.	• Change from baseline in Cornea and conjunctiva vital staining with fluorescein (NEI scale) Vs week 2.
• To compare the efficacy of cenegermin vs. vehicle in TFBUT at week 2 by testing the superiority.	• Change from baseline in TFBUT Vs week 2.
• To compare the efficacy of cenegermin vs. vehicle in worsening in symptom scores (SANDE) and/or NEI score at week 2 by testing the superiority.	• Number of patients experienced a worsening in symptom scores (SANDE global score) and/or NEI score ≥ 50% assessed at week 2.

Objective(s)	Endpoint(s)
• To compare the efficacy of cenegermin vs. vehicle in Schirmer II test at week 4 by testing the superiority.	• Change from baseline Schimer test II (with topical Anesthesia) vs Week 4.
• To evaluate the Best Corrected Distance Visual Acuity (BCDVA) throughout the study.	• Change from baseline in BCDVA.

3.2.4 Safety

Objective(s)	Endpoint(s)
• To evaluate the safety and tolerability of cenegermin vs. vehicle, as measured by Treatment-Emergent Adverse Events (TEAEs).	• Incidence and frequency of TEAEs, assessed throughout the study.

3.3 Schedule of evaluations

During the screening (day-8) all procedures for inclusion will be performed. From the day of screening the patients will stop any kind of further treatment, except CsA and commercially available preservative free artificial tears provided by Sponsor for a period of 8 days and 10 days as maximum: day -8 until the baseline visit (day 1+2). At the end of the wash out period (day -8 to baseline), patients meeting the entry criteria for this study will be randomized 1:1 and treated for 4 weeks with either cenegermin ophthalmic solution 20 mcg/mL TID or vehicle TID.

In addition to topical CsA eye drops (both groups will continue with topical CsA eye drops, or other topical ophthalmic treatment of the same class, as previously prescribed), during the 4 weeks of masked treatment only the administration of IMP is allowed. Nevertheless, if strictly needed, the patient can take preservative free artificial tears (provided by the Sponsor). The use (n° drops/day) of preservative free artificial tears will be clearly documented in a patient's diary and in the eCRF.

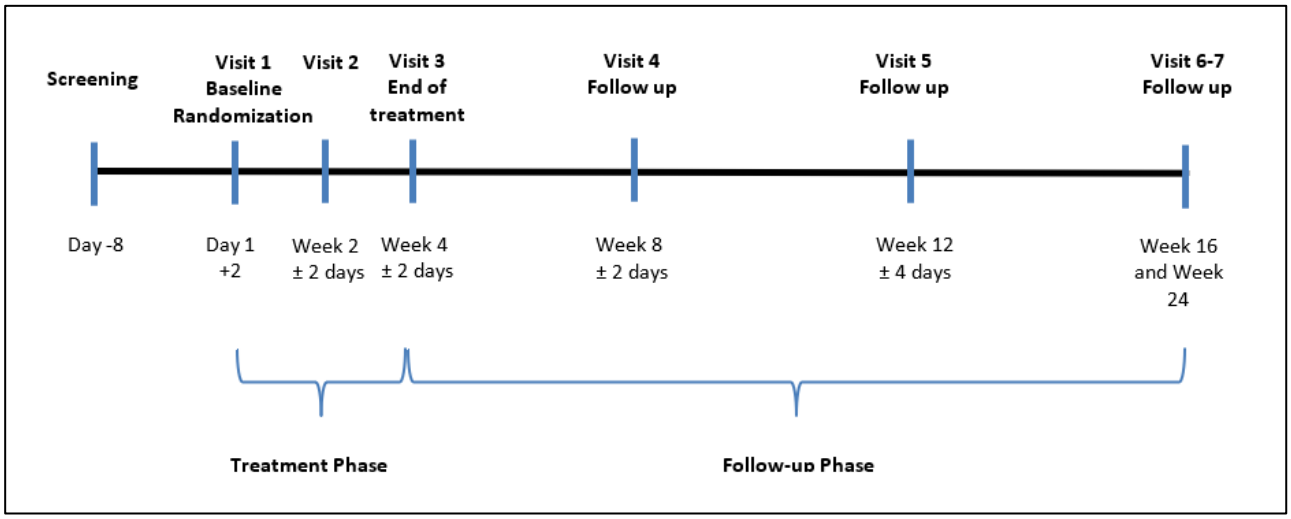
Eight weeks post treatment with no further topical ocular treatment except CsA and commercially available preservative-free artificial tears provided by the Sponsor TID will follow the completion of the double-blind treatment. During this follow up period, the patient can administer additional preservative-free artificial tear eye drops, provided by Sponsor, only if strictly needed, and must document in the patient's diary the number of additional drops administered for each eye.

Patients will then be followed up for efficacy and safety endpoints until week 16 and for safety endpoints until week 24. During the follow up period after week 12 until week 24 other topical treatments, in addition to CsA, will be allowed at discretion of the treating physician. Such treatments will be recorded by the investigators.

Throughout the study, the patient can continue any systemic treatment as required to control their autoimmune disease, and any change in systemic treatments must be recorded by the investigators.

The total duration of the study is 25 weeks including 1 week of screening.

Figure 1: Study duration



4. General definitions

4.1 Investigational drug and study treatment

Oxervate®, an ophthalmic solution containing cenegermin 20 mcg/mL, which is a recombinant human Nerve Growth Factor (rhNGF); reference product is vehicle. Test and reference will be instilled in both eyes for 28 consecutive days according to the following scheme:

- Group 1: one drop of cenegermin 20 mcg/mL will be instilled in both eyes TID (every 6 hours, e.g. 7:00 am, 01:00 pm; 07:00 pm).
- Group 2: one drop of vehicle will be instilled in both eyes TID (every 6 hours, e.g. 7:00 am, 01:00 pm; 07:00 pm).

Both groups will continue with topical CsA eye drops (or other topical ophthalmic treatment of the same class) as previously prescribed at least 15 minutes apart. The IMP will be provided in a monthly box containing 28 daily vials. Together with the IMP monthly box, patients will be provided with a sufficient number of pipettes and adaptors for the administration of the IMP.

4.2 Date of first and last administration of study drug

The date of first administration of study drug is defined as the baseline date.

The date/time of last administration of study drug is defined as the Date of last treatment administration as per End of treatment eCRF form. This value has to be consistent with IMP Administration information recorded on the corresponding eCRF forms. In case of different dates, the latest will be used as treatment end.

4.3 Study day

The study day describes the day of the event or assessment date, relative to the reference start date which is the date of Visit 1 - Baseline.

The study day is defined as:

- The date of the event (visit date, onset date of an event, assessment date etc.) – reference start date + 1 if event is on or after the reference start date.
- The date of the event (visit date, onset date of an event, assessment date etc.) – reference start date if event precedes the reference start date.

4.4 Visit Schedule and Visit Windows

Assessments and study visits will be performed as listed in Paragraph no. 2 ("Schedule of evaluations") and in Paragraphs no. 6 ("Study Procedure and Assessments") of the Study Protocol [1] containing additional details about scheduled visits and time windows.

Scheduled visits and time windows are:

- Screening (Day -8)
- Visit 1 - Baseline (Day 1 + 2 days)
- Visit 2 - Week 2 (Day 14 ± 2 days)
- Visit 3 - End of treatment Week 4 (Day 28 ± 2 days)
- Visit 4 - follow-up Week 8 (Day 56 ± 2 days)
- Visit 5 - follow-up Week 12 (Day 84 ± 4 days)
- Visit 6 - follow-up Week 16 (Day 112 ± 7 days)
- Visit 7 - follow-up Week 24 (Day 168 ± 7 days)

For efficacy and safety endpoints, data will be evaluated according to the Visit at which they have been collected.

4.5 Baseline

Baseline is defined as the last visit prior to and including date of the randomization visit (Visit 1). Unless otherwise specified, baseline values are defined as the measurements taken during this visit. In case of multiple measurements during baseline visit, the last assessment will be considered as baseline evaluation before start of treatment.

4.6 Post-baseline

For safety and efficacy evaluations all assessments after the first administration of study drug are considered as "post-baseline" assessment.

4.7 Eligible eye

Assuming that all the inclusion/exclusion criteria are met in both eyes, the worse eye (eligible eye) will be determined at the baseline visit using a stepladder approach, as follows:

1. Schirmer's Test: worse eye determined as eye with lower Schirmer I score.
2. NEI score: if Schirmer I score is the same in both eyes, worse eye will be determined by NEI score (cornea + conjunctival staining).

If all the above are identical, determination of the worse eye will be based on the investigator's judgement, otherwise the right eye will be used as the eligible eye.

5. Sample size justification

The sample size of the study is calculated based on results from previous studies.

Sample size was calculated according to the following assumptions:

- randomization ratio 1:1.
- a difference of 42% in Schirmer I rate of responders (value >10mm/5min at week 4.) in favor of rhNGF treatment at Week 4, with a responder rate of 33% for vehicle.
- 10 points in Global SANDE score improvement at Week 12, with a standard deviation of 12 points.

Based on these assumptions, a total sample size of 72 evaluable patients (80 pts considering 10% of patients not evaluable after enrollment) allows to achieve an overall power of 90% to show superiority of rhNGF eye drops solution vs vehicle in terms of primary endpoint, considering a one-sided alpha of 0.025.

Specifically:

- A power of 96.3% is achieved for Schirmer I primary endpoint.
- A power of 94.1% is achieved for SANDE primary endpoint.

6. Randomization and blinding

6.1.1 Randomization

Eligible patients will be randomized in a 1:1 ratio to either cenegermin ophthalmic solution 20 mcg/mL TID or vehicle ophthalmic solution TID.

Randomization will be stratified by site to ensure balanced assignment across treatment groups. The stratified permuted block randomization list will be generated with a computer procedure by an independent statistician not involved in the conduct of the study.

Randomization will be performed through IRS. Each Patient Kit number will be randomly associated with a treatment group. The randomization list will be provided to the facility responsible for IMP packaging/labelling for the purpose of IMP preparation. Each randomized patient will be allocated with randomization number according to the stratified randomization list. Dropouts after randomization will not be replaced.

The enrollment of patients will be scheduled in order to assure an inclusion of approximately 80 patients to ensure at least 72 patients assuming up to 10% of patients not evaluable after enrollment.

6.1.2 Unblinding

The identity of the treatments will remain unknown to the patient, investigator, site staff and Sponsor's clinical research personnel until the completion of the study (after full data base lock) except in case of specific events that will require unmasking of the patient.

The vials containing cenegermin (20 mcg/mL) or vehicle will be identical in appearance, and the contents of the vials will be indistinguishable. All staff directly involved in the analysis of study results will remain masked to treatment assignments while the study is in progress.

The investigator will be provided with a protected access to the randomization list so only in case of a medical emergency the investigator can open the treatment allocation for a specific patient.

Also Dompé's Pharmacovigilance contact person will be provided with a protected access to randomization list, so if required by the Pharmacovigilance activities the Pharmacovigilance contact person can open the treatment allocation for a specific Patient.

In the event of a medical emergency where the knowledge of patient treatment is required to provide the patient with appropriate care, investigators will have the possibility to unmask the treatment assignment for a specific patient. The investigators are encouraged to contact the Contract Research Organization staff before becoming unmasked if there is sufficient time.

If the investigator becomes unmasked for any reason, this information will be recorded on source data and in the eCRF of the study, specifying the date and the reason.

Unmasking events will be recorded and reported in the final study report.

6.2 Overview of planned statistical analyses

Analyses will be conducted when all enrolled subjects have completed the study and the study database has been locked and unblinded. Key first efficacy results (KFR) will be provided close to the database closure in order to identify efficacy findings in a short timeframe.

The list of tables, listings, and figures to be provided at each analysis is reported in appendix 15.

7. Statistical Analysis

7.1 General

All patient data collected during the study will be listed by patient and site.

Appropriate descriptive statistics will be produced by treatment arms according to the nature of the variable.

For continuous data, number of observations, mean, standard deviation, median, Q1-Q3, range (minimum and maximum) and 95% confidence intervals will be presented.

For categorical data, frequency distributions and percentages with 95% confidence intervals (Wilson method) per category will be presented.

For primary analysis, the threshold for claiming statistical significance will 0.025 one-sided. For other analysis, unless otherwise specified, the significance level for statistical testing will be 0.05 and two-sided tests.

The statistical analyses (both descriptive and inferential) of some endpoints will be done on the eligible eye when applicable. Details will be provided in the statistical analysis outputs as notes.

Additional post-hoc analysis may be produced to further allow comparison between treatment and control, according to the results obtained. Any deviations from the original statistical plan (including unplanned analyses) will be documented in the CSR.

7.2 Analysis sets

7.2.1 Screened set

The Screened set will consist of all patients with signed written informed consent.

7.2.2 Enrolled set (ENR)

The ENR set will consist of all patients with signed written informed consent and fulfillment of eligibility criteria (i.e., not reported as screening failure).

7.2.3 Randomized set (RND)

The RND set will consist of all patients in the ENR set who are randomized to the study, regardless of whether they receive the IMP or not.

7.2.4 Full Analysis set (FAS)

The FAS population will consist of all randomized patients who received at least one dose of the investigational product. FAS population will be analyzed according to intention-to-treat (ITT) principle, i.e. by treatment allocation regardless happening of intercurrent events (treatment policy strategy). The FAS population will be used for the primary analysis of the study and to present results on efficacy data.

7.2.5 Safety set (SAF)

The SAF set will consist of all randomized patients who received at least one dose of the investigational product. SAF set will be analyzed according to the actual treatment received. The SAF population will be used to present results on safety data.

7.2.6 Per Protocol set (PP)

The PP set will consist of all randomized patients who received at least one dose of the investigational product and do not have Major Protocol Deviations. The PP population will be used for sensitivity analyses.

7.3 Primary estimand

The primary endpoints will be analyzed using treatment policy estimands, defined by the following:

Primary endpoint:

- Population: Adult patients with severe Sjogren's dry eye disease under treatment with Cyclosporine A, as defined by the inclusion-exclusion criteria of the study (and matching the FAS definition);
- Variable: Schirmer I test at week 4;
- Intercurrent event: The occurrence of an intercurrent event is irrelevant. All observed values will be used regardless of occurrence of an intercurrent event. Retrieved drop-outs will be used for data imputation of missing data;
- Population-level summary: Difference in proportion of patients reaching a value of Schirmer I test (without anesthesia) >10mm/5min at week 4.

Coprimary endpoint:

- Population: Adult patients with severe Sjogren's dry eye disease under treatment with Cyclosporine A, as defined by the inclusion-exclusion criteria of the study (and matching the FAS definition);
- Variable: SANDE global score at week 12;

- Intercurrent event: The occurrence of an intercurrent event is irrelevant. All observed values will be used regardless of occurrence of an intercurrent event. Retrieved drop-outs will be used for data imputation of missing data;
- Population-level summary: Difference in means of change from baseline in the global SANDE score at week 12.

7.4 Usage of analysis sets

The usage of the analysis sets for the creation of tables and figures is illustrated in Table 2. Unless otherwise specified, all listings will be done for RND set. All listings will report:

- planned and actual treatment names included,
- the flag(s) of the analysis set(s) used to analyze the information of the listing (according to Table 2).

Table 2: Usage of analysis set

Analysis	ENR	RND	FAS	SAF	PP
Subject enrolment and disposition	X				
Protocol deviations		X			
Study discontinuations		X			
Demographics and baseline characteristics			X		
Ocular and systemic Medical History and Concomitant Diseases			X		
Prior and concomitant medications			X	X	
Dry Eye History			X		
Other baseline characteristics			X		
Evaluation of Treatment Compliance and Exposure			X	X	
Evaluation of Efficacy			X		
Analysis of primary efficacy endpoint			X		
Sensitivity analyses			X		
Supportive analyses			X		X
Analysis of secondary efficacy endpoints			X		
Evaluation of Safety				X	

7.5 Sub-group analyses

Subgroup analyses of primary and secondary endpoints will be performed on the following subgroups of baseline characteristics:

- Age class (≤Median, >Median).
- Race.
- Ethnicity.
- Gender.

Within each subgroup level, descriptive in nature analyses will be performed on:

- Schirmer I test.
- SANDE global, frequency and severity scores.
- NEI scale.

- TFBUT.
- IDEEL questionnaire.

at each available timepoints by means of descriptive statistics.

In case of continuous measures, analyses will be provided for baseline visit, each post-baseline visit, and their change from baseline. Comparisons between treatments will be performed using two-sample t-test (assuming unequal variances) or, if the required assumptions are not met, the two-sample Mann–Whitney U test will be used. Assumption of normal distribution of endpoints for continuous data will be assessed by a visual inspection of distribution.

In case of categorical variables, the Fisher's Exact test will be provided in addition to summary description.

Statistical tests for interaction (between subgroup variable and treatment arm) will be performed before investigating further subgroups: analyses of primary endpoints will be performed using regression models including treatment, subgroup variable and their interaction as covariate. If the interaction between subgroup variable and treatment arm is statistically significant at 15% nominal level in at least one regression model, then subgroup analysis will be performed.

Variables that may be evaluated after test for interaction are:

- Region (Europe, US),
- Sjögren's syndrome (primary, secondary), *where "primary" is defined as a confirmed diagnosis of Sjögren's syndrome while "secondary" as other autoimmune disease known to induce Sjögren's Dry Eye Disease*,
- Medical history,
- Concomitant medication.

Subgroup levels may be pooled to have an adequate number of subjects per strata to perform the analysis. Decision on pooling will be taken during the Blind Data Review Meeting (BDRM) before data base lock.

7.6 Interim analysis

No interim analysis is planned.

7.7 Data Monitoring Committee

No Data Monitoring Committee has been established.

7.8 Handling of missing and incomplete data

All reasonable efforts will be made to reduce the rate of missing data. Investigators will be trained about the importance of patient retention and full data capture. Also, any reasonable attempts should be made by the investigators to emphasize continued subject's participation for the full duration of the trial.

Patients who discontinue the treatment will not be withdrawn from the study by default but will be asked to complete safety and efficacy observations as per the protocol, unless otherwise they withdraw their consent. Patients who discontinue study treatment and decide to remain in the study

by following the schedule of assessments and continuing to adhere to protocol requirements are defined as "retrieved drop-out" patients.

For the primary and key secondary analyses, missing data will be handled by means of Multiple Imputation using retrieve dropouts (MI-RD), as detailed in sections 10.1.2 and 10.2.2.

In addition, to further minimize missing data, if a patient cannot refer to the site for a planned follow-up visit, the investigator will try to obtain any relevant information from the patients, including documents/laboratory results available from local medical care.

The number of subjects with missing data will be presented under the "Missing" category. Missing values will not be included in the denominator count when computing percentages. When continuous data will be summarized, only the non-missing values will be evaluated for computing summary statistics.

For the analyses of parameters for which an independent blinded reading center may be involved per protocol, the following rules will be used:

1. If assessment from blinded reading center is available, such value will be used for the analysis.
2. If assessment form blinded reading center is not available, the corresponding evaluation (if any) from the site investigator will be used.
3. If both assessments (blinded reading center and site investigator) are not available, a statistical imputation method will be applied as described in this SAP.

Rules for handling missing data of IDEEL questionnaire are reported in section 12.3, according to the dedicated manual.

7.9 Changes in the planned analysis

The following changes have been included, compared to the protocol [1]:

- In the secondary endpoint on SANDE, the global score has been included (section 3.2.2) in addition to severity and frequency scores.
- A clarification on the use of data from blinded reading center and site investigator has been added in section 7.8.
- The first key secondary endpoint in section 10.2.1 has been reworded in alignment with the definition of the Schirmer I test primary endpoint.
- A clarification on the use of a Mixed Model for Repeated Measurements (MMRM), extension of ANCOVA for repeated measurements, after imputation through MI-RD for IDEEL, NEI scale and TFBUT endpoints has been added in section 10.2.2, in alignment with treatment policy estimand used for primary endpoints.
- Considering the progress of data collection, the overview of planned statistical analyses (section 6.2) has been updated compared to protocol. A single database lock will be performed at the end of the study (instead of two locks), followed by the release of the KFR and by the final analyses. This decision represents a conservative choice in favor of the integrity of the study.

7.10 Blind Data Review Meeting (BDRM)

BDRMs will be held before DB locks. Any other details will be provided in the BDRM documentation (Pre-Analysis Review).

7.11 Software

All statistical analyses and data processing will be performed using Statistical Analysis Systems (SAS®) Software (release 9.4 or later).

8. Evaluation of Demographic and Baseline Characteristics

8.1 Subject enrolment and disposition

All presentations of subject disposition will be by treatment group, and overall.

For describing the subject disposition, the following populations will be summarized:

1. Subjects screened overall (N).
2. Subjects enrolled overall (N, 100%).
3. Subjects enrolled but not randomized and reasons for non-allocation overall (N, %).
4. Subjects randomized by treatment group, and overall (N, %).
5. Subjects randomized but not treated by treatment group, and overall (N, %).
6. Subjects in each analysis set (FAS, SAF, PP) and reasons for exclusion by treatment group, and overall (N, %).

For points 1 - 3, the percentage denominator will be the number of ENR subjects; for points 4 - 6, the percentage denominator will be the number of randomized subjects within each arm.

Listings will be provided based on ENR set.

8.2 Protocol deviations

All the protocol deviations will be discussed case by case before unblinding of the treatment code with the clinical team during the BDRM and described in the BDRM Report. Any deviation from these protocol procedures will be reported in the study-specific Protocol Deviation form.

Number of occurrences and of subjects with at least one major and minor protocol deviation will be summarized for each treatment and overall.

8.3 Study discontinuations

The following information will be summarized for the randomized patients by treatment and overall:

- Trial completers.
- Subjects who discontinued the trial prematurely (and reasons).
- Number of subjects who completed each planned visit.
- Subjects who discontinued the IMP (and reasons).
- Subjects who discontinued the IMP but completed the study.

- Subjects who discontinued the IMP and discontinued the trial prematurely.
- Broken randomization code (and reasons).

If more than 30% of randomized subjects discontinue the study prematurely, the distribution of the time from randomization to discontinuation will be summarized using the Kaplan-Meier method. The degree of uncertainty will be expressed with 95% confidence limits (calculated per the method proposed by Greenwood). Comparison of curves among arms will be performed with the log-rank test. Kaplan-Meier graphs will be presented along with the number of patient-at-risk at exact time points. Subjects who have not prematurely discontinued the trial will be censored at study termination.

8.4 Demographics and baseline characteristics

The baseline demographic characteristics will be summarized by treatment and overall, by means of descriptive statistics. No statistical testing will be carried out.

The following demographic and baseline characteristics will be reported:

- Geographic region of the site (Europe, US) ,
- Site,
- Age (years),
- Sex (Male, Female),
- Race,
- Ethnicity,
- Potential childbearing,
- Contraception method,
- Pregnancy test performed and result and reasons for no pregnancy test.

Potential childbearing, Contraception methods, and Pregnancy test results will be provided for female patients only.

8.5 Ocular and systemic Medical History and Concomitant Diseases

A disease is defined:

- “Previous disease” if it is not ongoing at screening visit (“ongoing” box is ticked as “No”).
- “Concomitant disease” if it is ongoing at screening visit (“ongoing” box is ticked as “Yes”).

Previous and/or concomitant diseases will be coded using Medical Dictionary for regulatory activities (MedDRA) dictionary and reported in separate tabulations. Frequency distributions and percentages will be summarized by treatment, type (Ophthalmic and Non-Ophthalmic), by System Organ Class (SOC) and Preferred Term (PT).

Counts will be given for both SOC and PT by subject. Subjects experiencing more than one previous/concomitant disease event will be counted only once within each SOC and PT.

Subjects experiencing more than one disease classified in the same category will be counted only once.

8.6 Prior and concomitant medications

A medication is defined:

- "Prior medication" if taken and stopped prior to first dose of study treatment.
- "Concomitant medication" if taken after the first dose of study treatment.

according to the available parts of the medication dates (see Table 7). In case of missing information not directly allowing allocation to either of the two above categories of medications, the medication will be considered as concomitant.

Prior and/or concomitant medications will be coded using World Health Organization Drug Dictionary and reported in separate tabulations. Frequency distributions and percentages will be summarized by treatment, type (Ophthalmic and Non-Ophthalmic), by Anatomical Main Group (1st level of the Anatomical Therapeutic Chemical (ATC) classification), Chemical Subgroup (4th level of the ATC classification) and Preferred Name.

Subjects taking more than one medication classified in the same category will be counted only once.

8.7 Dry Eye History

The dry eye history characteristics will be summarized by treatment and overall, by means of descriptive statistics. No statistical testing will be carried out. The following information will be reported:

- Eye affected by Dry Eye.
- Time since first diagnosis of Dry Eye.
- Has the patient received prior treatment for dry eye?
- Eye diagnosed with Severe Sjögren's Dry Eye.
- Time since first diagnosis of Severe Sjögren's Dry Eye.

8.8 Other baseline characteristics

No statistical testing will be carried out for comparing baseline characteristics.

8.8.1 Eligible eye

Eye which is defined as eligible eye will be reported.

8.8.2 SANDE

Baseline SANDE scores (frequency, severity, and overall) will be descriptively summarized by treatment and overall as a continuous variable. Information on the SANDE assessment (done/not done) will be reported, with reasons in case of not execution. Rules for derivation of SANDE scores are reported in section 12.2.

8.8.3 IDEEL

Baseline IDEEL scores will be descriptively summarized by treatment and overall as a continuous variable. Information on the IDEEL assessment (done/not done) will be reported, with reasons in case of not execution. Rules for derivation of IDEEL scores are reported in section 12.3.

8.8.4 BCDVA

Baseline SNELLEN equivalent scores will be summarized by treatment and overall as an ordinal variable for the right eye, the left eye and eligible eye. Information on the BCDVA assessment (done/not done) will be reported, with reasons in case of not execution.

8.8.5 External ocular examination

The results of the external ocular examination will be summarized by treatment and overall, by means of descriptive statistics. Information on the assessment (done/not done) will be reported, with reasons in case of not execution. The following information will be reported for the eligible eye:

- Any clinically significant abnormalities and types.
- Any evidence of punctual occlusion.
    - if yes, by what mechanism was this seen.
- Any clinically significant abnormalities in the function of the eyelid muscles.
    - if yes, specification.
- Any abnormalities in the conjunctiva and types.

The same information will be provided, in a separate tabulation, for the non-eligible eye.

8.8.6 Schirmer test I and Schirmer test II

Baseline Schirmer test I and Schirmer test II will be descriptively summarized by treatment and overall as continuous variables. Information on the assessments (done/not done) will be reported, with reasons in case of not execution.

8.8.7 Slit Lamp Ophthalmic Exam

The results of the Slit Lamp Ophthalmic Exam will be summarized by treatment and overall, by means of descriptive statistics. Information on the assessment (done/not done) will be reported, with reasons in case of not execution. The following information will be reported for the eligible eye:

- Evaluation.
    - if Abnormal, comments.
- Horizontal diameter of the cornea (mm).

8.8.8 Fluorescein Staining (NEI Scale)

The results of the Fluorescein Staining will be summarized by treatment and overall, by means of descriptive statistics. Information on the assessment (done/not done) will be reported, with reasons in case of not execution. The following information will be reported for the eligible eye:

- Has Ocular surface staining performed?
- Cornea – Result.
- Conjunctiva – Result.
- Total Cornea Score.
- Total Conjunctiva Score.
- NEI Score

Rules for derivation of NEI scores are reported in section 12.4.

8.8.9 Tear Film Break-up Time (TFBUT)

Baseline TFBUT results will be summarized by treatment and overall as a continuous variable for the right eye, the left eye and eligible eye. Information on the TFBUT assessment (done/not done) will be reported, with reasons in case of not execution.

9. Evaluation of Treatment Compliance and Exposure

9.1 Compliance to IMP

The assessment of patients' compliance to the IMP will be made by determining the number of drops administered to the eligible eye. On a per patient basis, the evaluation of the compliance will be done using the following formula:

$$Compliance\ (\%) = \frac{total\ number\ of\ drops\ in\ the\ eligible\ eye\ taken\ during\ the\ treatment\ period}{total\ number\ of\ drops\ in\ the\ eligible\ eye\ scheduled\ during\ the\ treatment\ period} \times 100$$

Where derivations rules for "*total number of drops in the eligible eye taken during the treatment period*" and for "*total number of drops in the eligible eye scheduled during the treatment period*" are provided in Table 3.

Compliance will be summarized by treatment and overall, by means of summary statistics. In addition, compliance to IMP will also be presented for the following categories: <80%, 80%-120%, >120%.

9.2 Exposure to IMP

The extent of exposure to IMP in days will be summarized with descriptive statistics by treatment group and overall. The extent of exposure (days) will be calculated using the formula reported in Table 3.

9.3 Use of preservative free artificial tears

The use (n° drops/day) of preservative free artificial tears will be summarized with descriptive statistics by treatment group and overall, by study period (treatment / follow-up) and overall according to the following formulas:

$$Use_{\ TREAT} = \frac{total\ number\ of\ drops\ of\ the\ preservative\ free\ artificial\ tears\ during\ the\ \text{treatment}\ period}{total\ number\ of\ days\ of\ the\ \text{treatment}\ period} \times 100$$

$$Use_{\ F-UP} = \frac{total\ number\ of\ drops\ of\ the\ preservative\ free\ artificial\ tears\ during\ the\ \text{follow}-\text{up}\ period}{total\ number\ of\ days\ of\ the\ \text{follow}-\text{up}\ period} \times 100$$

Details are provided in Table 3.

10. Evaluation of Efficacy

10.1 Analysis of primary efficacy endpoint

10.1.1 Testing strategy and multiplicity adjustment

The following null hypothesis is defined on the first primary endpoint: the proportion of patients reaching a value of Schirmer I test (without anesthesia) >10mm/5min at week 4 in cenegermin (rhNGF) is lower or equal than control:

H_{01}: $T_{rhNGF} \leq T_{CONTROL}$

H_{11}: $T_{rhNGF} > T_{CONTROL}$

where T_{rhNGF} and $T_{CONTROL}$ are the proportion of patients reaching a value of Schirmer I test (without anesthesia) >10mm/5min at week 4 for cenegermin and control groups, respectively. The null hypothesis H_{01} will be rejected if the associated primary analysis p-value will be lower than 0.025.

Similarly, the following null hypothesis is defined on the co-primary endpoint: the change from baseline (reduction) in the global SANDE score at week 12 in cenegermin is lower or equal than control:

H_{02}: $\mu_{rhNGF} \leq \mu_{CONTROL}$

H_{12}: $\mu_{rhNGF} > \mu_{CONTROL}$

where μ_{rhNGF} and $\mu_{CONTROL}$ are the change from baseline (reduction) in the global SANDE score at week 12 for cenegermin and control groups, respectively. The null hypothesis H_{02} will be rejected if the associated primary analysis p-value will be lower than 0.025.

Both null hypotheses H_{01} and H_{02} must be rejected to claim superiority of cenegermin over control. Consequently, no multiplicity correction of type I error will be applied on primary endpoints analysis.

10.1.2 Analysis details

Both co-primary endpoints will be analysed by means of MI-RD to handle missing data. Retrieved drop-out patients are defined as patients who discontinue study treatment and decide to remain in the study by following the schedule of assessments and continuing to adhere to protocol requirements. Consequently, for each primary endpoint, MI will be performed based on the subjects' allocated treatment arm and observed values (baseline and intermediate) as covariates in a regression model using data from Retrieved drop-out patients that have the primary endpoint assessment done.

10.1.2.1 Schirmer I test >10mm/5min at week 4

This primary endpoint will be analyzed by means of a logistic regression model adjusting by pre-defined baseline factors (gender, age class, baseline Schirmer I test value as fixed effects and site as random effect) and a one-sided test will be used to test for differences between treatment groups.

For the imputation of missing data of Schirmer I test, a regression model will be created by including the baseline Schirmer I test value, gender, age class, and Schirmer I test at week 2. Two hundred data sets will be generated, and the random seed number will be 180404. MI will be implemented in the following steps:

1. Partial imputation assuming Missing at Random (MAR) will be carried out to impute intermittent (non-monotone) missing data of Schirmer I test up to week 4 based on a

multivariate joint Gaussian imputation model using the Markov chain Monte Carlo (MCMC) method. A separate imputation model will be used for each treatment arm. In case of non-convergence or non-estimability issues, a single imputation model will be considered (with treatment arm added as explanatory variable to the model). A total of 200 datasets will be created. These datasets will be utilized in Step #2.
2. The remaining monotone missing data will be imputed using a MI regression including the same explanatory variables of the previous step. According to MI-RD approach, only non-missing values from retrieve dropouts will be used to inform the regression model.
3. Each of the 200 imputed datasets will be analyzed using observed and imputed data. Rubin's rule will be used for combining results to draw inference.

If there are not enough retrieved dropouts for convergence of MI-RD regression model, a reference-based Multiple Imputation (MI) approach will be adopted to consider a missing not at random (MNAR) mechanism for missing data: imputation of values in the rhNGF arm (above step #2) will be done using the non-missing values from the control group (this approach will be referred as "Copy Reference"). This approach does not assume benefits for rhNGF in case of discontinuation and limits a post-discontinuation clinical effect to that of vehicle. The final decision on the use of the MI-RD vs Copy-Reference for this endpoint will be done at the time of the analysis and reported in the CSR.

Whether for MI-RD or Copy-Reference approach, the adjusted estimated treatment differences in proportions between cenegermin and vehicle at week 4 will be displayed together with the corresponding two-side 95% confidence intervals and p-value.

10.1.2.2 Change from baseline in the global SANDE score at week 12

This co-primary endpoint will be analyzed by means of an Analysis of Covariance (ANCOVA) adjusting by pre-defined baseline factors (gender, age class, baseline global SANDE score as fixed effects and site as random effect) and a one-sided test will be used to test for differences between treatment groups. Rules for derivation of SANDE scores are reported in section 12.2.

For the imputation of missing data, a regression model will be created by including the baseline global SANDE score, gender, age class, and intermediate global SANDE scores up to week 12. Two hundred data sets will be generated, and the random seed number will be 252404. MI will be implemented in the following steps:

1. Partial imputation assuming MAR will be carried out to impute intermittent (non-monotone) missing data of global SANDE score based on a multivariate joint Gaussian imputation model using the MCMC method. A separate imputation model will be used for each treatment arm. In case of non-convergence or non-estimability issues, a single imputation model will be considered (with treatment arm added as explanatory variable to the model). A total of 200 datasets will be created. These datasets will be utilized in Step #2.
2. The remaining monotone missing data will be imputed using a MI regression including the same explanatory variables of the previous step. According to MI-RD approach, only non-missing values from retrieve dropouts will be used to inform the regression model.
3. Each of the 200 imputed datasets will be analyzed using observed and imputed data. Rubin's rule will be used for combining results to draw inference.

If there are not enough retrieved dropouts for convergence of MI-RD regression model, a Copy Reference approach will be adopted to consider a MNAR mechanism for missing data. The final decision on the use of the MI-RD vs Copy-Reference for this endpoint will be done at the time of the analysis and reported in the CSR.

Whether for MI-RD or Copy-Reference approach, the adjusted estimated treatment differences in means for global SANDE score between cenegermin and vehicle at week 12 will be displayed together with the corresponding two-side 95% confidence intervals and p-value.

10.1.3 Sensitivity analyses

The following sensitivity analyses are defined to assess the robustness of results on each primary endpoint versus assumptions used in the statistical model for the main estimators:

- The comparison between treatment and control will be performed by means of MI under MAR assumption instead of MNAR. MI will be implemented in two steps:
  - Intermittent missing data up to primary time point (Week 4 for Schirmer I test, and Week 12 for SANDE) will be imputed using MCMC methods assuming MAR. A separate imputation model will be used for each treatment arm. The imputation models will include the observed values (baseline and intermediate), gender, and age class as covariates. In case of non-convergence or non-estimability issues, a single model will be considered with treatment arm added as explanatory variable to the model.
  - The remaining missing values with a monotone missing data pattern will be imputed based on data of corresponding treatment group.
  - Each imputed dataset will be analyzed using observed and imputed data. Rubin's rule will be used for combining results to draw inference. The adjusted estimated treatment differences between cenegermin and vehicle (difference in proportions for Schirmer's test and differences in means for SANDE) will be displayed together with the corresponding two-side 95% confidence intervals and p-values.
- A tipping point strategy will be used as a sensitivity analysis for missing data for assessment of superiority (if shown) of rhNGF for each endpoint. Tipping point analysis will assess how departures from MI under MNAR assumptions must be to overturn conclusions from the primary superiority analysis. Tipping point will be based on iterative application of MI. In the first iteration, MI-RD method is assumed as described in section 10.1.2. In successive iterations, the imputed values for the rhNGF arm are shifted by a constant to represent a worse effect in each iteration. This can be achieved by using the N completed datasets obtained under MI-RD and shift the imputed values. The tipping point is the shift at which the p-value becomes non-significant. Outline of tipping point:
  - Use MI-RD to impute missing values.
  - Assuming that the tipping point is between -10 and 0, the following actions will be performed for $\Delta = 0, -0.1, -0.2, -0.3, \ldots, -10$:
    - Add Δ to the imputed values at the rhNGF arm.
    - Analyze the completed data sets using the same method outlined in section 10.1.2.
    - the plot of all p-values for Δ between -10 and 0 will be created.
  - based on the plot, the approximate value of the tipping point Tp_{approx} is identified considering a threshold for the p-values of 0.02500.
  - after identification of Tp_{approx}, finest research of the tipping point will be performed in the range $Tp_{approx} \pm 1$ by incrementing Δ of 0.01 and following the usual steps:
    - Add Δ to the imputed values at the rhNGF arm.
    - Analyze the completed data sets using the same method outlined in section 10.1.2.
  - The tipping point is the smallest Δ at which p-value ≥ 0.02500.

10.1.4 Supportive analyses

For each primary endpoint, the regression models detailed in section 10.1.2 will be performed for supportive purposes:

- by considering complete cases only (i.e. without considering patients with missing primary endpoint).
- on the PP set instead of FAS.

10.2 Analysis of secondary efficacy endpoints

10.2.1 Testing strategy and multiplicity adjustment

In case the analysis of the primary endpoints leads to rejection of null hypothesis, some secondary endpoints (defined as key secondary endpoints) will be tested in a conditional sequential manner to show superiority of cenegermin versus control (at alpha one-sided 0.025) according to the following ranking:

1. Proportion of patients reaching a value of Schirmer I test (without anesthesia) >10mm/5min at week 8.
2. Change from baseline in SANDE scores for severity at week 12.
3. Change from baseline in SANDE scores for frequency at week 12.
4. Change from baseline in IDEEL modules (Quality of Life [27 questions], Dry eye Treatment satisfaction & Bother [10 questions], and Dry eye Symptom bother [20 questions]) at week 12 and at week 4 (in this order).
5. Change from baseline in Cornea and conjunctiva vital staining with fluorescein NEI scale at week 4, week 8 and at week 12 (in this order).
6. Change from baseline in TFBUT at week 4, week 8 and at week 12 (in this order).

This hierarchical test strategy protects the family-wise false positive error rate at the overall one-sided 0.025 level.

10.2.2 Analysis details

Analysis of key secondary endpoints are detailed below in sections 10.2.2.1 - 10.2.2.6 and p-values to be evaluated for comparison between treatments in a sequential order are the following:

1. P-value of treatment variable from logistic regression model on proportion of patients reaching a value of Schirmer I test (without anesthesia) >10mm/5min at week 8 (section 10.2.2.1),
2. P-value of treatment variable from ANCOVA regression model on change from baseline in SANDE score for severity at week 12 (section 10.2.2.2),
3. P-value of treatment variable from ANCOVA regression model on change from baseline in SANDE score for frequency at week 12 (section 10.2.2.3),

4.1 P-value of treatment variable from MMRM regression model on IDEEL Quality of Life module at week 12 (section 10.2.2.4),

4.2 P-value of treatment variable from MMRM regression model on IDEEL Dry eye Treatment satisfaction & Bother module at week 12 (section 10.2.2.4),

4.3 P-value of treatment variable from MMRM regression model on IDEEL Dry eye Symptom bother module at week 12 (section 10.2.2.4),

4.4 P-value of treatment variable from MMRM regression model on IDEEL Quality of Life module at week 4 (section 10.2.2.4),

4.5 P-value of treatment variable from MMRM regression model on IDEEL Dry eye Treatment satisfaction & Bother module at week 4 (section 10.2.2.4),

4.6 P-value of treatment variable from MMRM regression model on IDEEL Dry eye Symptom bother module at week 4 (section 10.2.2.4),

5.1 P-value of treatment variable from MMRM regression model on Change from baseline in Cornea and conjunctiva vital staining with fluorescein NEI scale at week 4 (section 10.2.2.5),

5.2 P-value of treatment variable from MMRM regression model on Change from baseline in Cornea and conjunctiva vital staining with fluorescein NEI scale at week 8 (section 10.2.2.5),

5.3 P-value of treatment variable from MMRM regression model on Change from baseline in Cornea and conjunctiva vital staining with fluorescein NEI scale at week 12 (section 10.2.2.5),

6.1 P-value of treatment variable from MMRM regression model on Change from baseline in TFBUT values at week 4 (section 10.2.2.6),

6.2 P-value of treatment variable from MMRM regression model on Change from baseline in TFBUT values at week 8 (section 10.2.2.6),

6.3 P-value of treatment variable from MMRM regression model on Change from baseline in TFBUT values at week 12 (section 10.2.2.6).

10.2.2.1 Schirmer I test >10mm/5min at week 8

The proportion of patients reaching a value of Schirmer I test (without anesthesia) >10mm/5min at week 8 will be analyzed as described in section 10.1.2.1 for the primary efficacy endpoints.

10.2.2.2 Change from baseline in SANDE score for severity at week 12

The change from baseline in SANDE score for severity at week 12 will be analyzed as described in section 10.1.2.2 for the primary efficacy endpoints.

10.2.2.3 Change from baseline in SANDE scores for frequency at week 12

The change from baseline in SANDE score for frequency at week 12 will be analyzed as described in section 10.1.2.2 for the primary efficacy endpoints.

10.2.2.4 Change from baseline in IDEEL modules (Quality of Life, Dry eye Treatment satisfaction & Bother, and Dry eye Symptom bother) at week 12 and at week 4

Rules for derivation of IDEEL scores are reported in section 12.3. The change from baseline of each IDEEL module will be analyzed by means of an MMRM adjusting by pre-defined factors (gender, age class, IDEEL module baseline value, treatment, visit, and treatment by visit interaction; subject will be considered as a random effect). The covariance matrix used will be "unstructured". Comparisons versus vehicle TID will be provided using least square means at each visit and overall and a one-sided test will be used to test for differences between treatment groups.

For the imputation of missing data, a regression model will be created by including the baseline value, gender, age class, and intermediate scores up to week 12. Two hundred data sets will be generated, and the random seed number will be 225670. MI will be implemented in the following steps:

1. Partial imputation assuming MAR will be carried out to impute intermittent (non-monotone) missing data of IDEEL module score based on a multivariate joint Gaussian imputation model using the MCMC method. A separate imputation model will be used for each treatment arm. In case of non-convergence or non-estimability issues, a single imputation model will be considered (with treatment arm added as explanatory variable to the model). A total of 200 datasets will be created. These datasets will be utilized in Step #2.
2. The remaining monotone missing data will be imputed using a MI regression including the same explanatory variables of the previous step. According to MI-RD approach, only non-missing values from retrieve dropouts will be used to inform the regression model.
3. Each of the 200 imputed datasets will be analyzed using observed and imputed data. Rubin's rule will be used for combining results to draw inference.

If there are not enough retrieved dropouts for convergence of MI-RD regression model, a Copy Reference approach will be adopted to consider a MNAR mechanism for missing data. The final decision on the use of the MI-RD vs Copy-Reference for this endpoint will be done at the time of the analysis and reported in the CSR.

Whether for MI-RD or Copy-Reference approach, the adjusted estimated treatment differences in means for IDEEL module scores between cenegermin and vehicle at each available timepoint will be displayed together with the corresponding two-side 95% confidence intervals and p-value.

10.2.2.5 Change from baseline in Cornea and conjunctiva vital staining with fluorescein NEI scale at week 4, week 8 and at week 12

Rules for derivation of NEI scores are reported in section 12.4. The change from baseline in Cornea and conjunctiva vital staining with fluorescein NEI scale will be analyzed by means of an MMRM adjusting by pre-defined factors (gender, age class, NEI scale baseline value, treatment, visit, and treatment by visit interaction; subject will be considered as a random effect). The covariance matrix used will be "unstructured". Comparisons versus vehicle TID will be provided using least square means at each visit and overall and a one-sided test will be used to test for differences between treatment groups.

For the imputation of missing data, a regression model will be created by including the baseline value, gender, age class, and intermediate scores up to week 12. Two hundred data sets will be generated, and the random seed number will be 569888. MI will be implemented in the following steps:

1. Partial imputation assuming MAR will be carried out to impute intermittent (non-monotone) missing data of NEI scale score based on a multivariate joint Gaussian imputation model using the MCMC method. A separate imputation model will be used for each treatment arm. In case of non-convergence or non-estimability issues, a single imputation model will be considered (with treatment arm added as explanatory variable to the model). A total of 200 datasets will be created. These datasets will be utilized in Step #2.
2. The remaining monotone missing data will be imputed using a MI regression including the same explanatory variables of the previous step. According to MI-RD approach, only non-missing values from retrieve dropouts will be used to inform the regression model.
3. Each of the 200 imputed datasets will be analyzed using observed and imputed data. Rubin's rule will be used for combining results to draw inference.

If there are not enough retrieved dropouts for convergence of MI-RD regression model, a Copy Reference approach will be adopted to consider a MNAR mechanism for missing data. The final decision on the use of the MI-RD vs Copy-Reference for this endpoint will be done at the time of the analysis and reported in the CSR.

Whether for MI-RD or Copy-Reference approach, the adjusted estimated treatment differences in means for NEI scale scores between cenegermin and vehicle at each available timepoint will be displayed together with the corresponding two-side 95% confidence intervals and p-value.

10.2.2.6 Change from baseline in TFBUT at week 4, week 8 and at week 12

The change from baseline in TFBUT values will be analyzed by means of a MMRM adjusting by pre-defined factors (gender, age class, TFBUT baseline value, treatment, visit, and treatment by visit interaction; subject will be considered as a random effect). The covariance matrix used will be "unstructured". Comparisons versus vehicle TID will be provided using least square means at each visit and overall and a one-sided test will be used to test for differences between treatment groups.

For the imputation of missing data, a regression model will be created by including the baseline value, gender, age class, and intermediate scores up to week 12. Two hundred data sets will be generated, and the random seed number will be 101149. MI will be implemented in the following steps:

1. Partial imputation assuming MAR will be carried out to impute intermittent (non-monotone) missing data of TFBUT values based on a multivariate joint Gaussian imputation model using the MCMC method. A separate imputation model will be used for each treatment arm. In case of non-convergence or non-estimability issues, a single imputation model will be considered (with treatment arm added as explanatory variable to the model). A total of 200 datasets will be created. These datasets will be utilized in Step #2.
2. The remaining monotone missing data will be imputed using a MI regression including the same explanatory variables of the previous step. According to MI-RD approach, only non-missing values from retrieve dropouts will be used to inform the regression model.
3. Each of the 200 imputed datasets will be analyzed using observed and imputed data. Rubin's rule will be used for combining results to draw inference.

If there are not enough retrieved dropouts for convergence of MI-RD regression model, a Copy Reference approach will be adopted to consider a MNAR mechanism for missing data. The final decision on the use of the MI-RD vs Copy-Reference for this endpoint will be done at the time of the analysis and reported in the CSR.

Whether for MI-RD or Copy-Reference approach, the adjusted estimated treatment differences in means for NEI scale scores between cenegermin and vehicle at each available timepoint will be displayed together with the corresponding two-side 95% confidence intervals and p-value.

10.2.3 Descriptive Analysis

Independently of results on primary and key secondary endpoints, descriptive in nature analyses will be performed on all secondary endpoints (section 3.2.2) at each available timepoints by means of descriptive statistics.

In case of continuous measures, analyses will be provided for baseline visit, each post-baseline visit, and their change from baseline. Comparisons between treatments will be performed using two-sample t-test (assuming unequal variances) or, if the required assumptions are not met, the two-

sample Mann–Whitney U test will be used. Assumption of normal distribution of endpoints for continuous data will be assessed by a visual inspection of distribution.

In case of categorical variables, shift tables versus baseline will be summarized for all post-baseline visits, and the Fisher's Exact test will be performed.

The following graphical representations will be provided by treatment:

- Schirmer I test Observed Mean (±SEM) and Mean change from baseline (±SEM) at each timepoint.
- Observed percentage of patients with Schirmer I test >10mm/5min at each timepoint.
- SANDE global score Observed Mean (±SEM) and Mean change from baseline (±SEM) at each timepoint.
- Cornea and conjunctiva vital staining total score Observed Mean (±SEM) and Mean change from baseline (±SEM) at each timepoint.
- TFBUT Observed Mean (±SEM) and Mean change from baseline (±SEM) at each timepoint.

11. Evaluation of Safety

11.1 Adverse events

Adverse Events (AEs) started before administration of study treatment will be considered as pre-treatment; any AE started on or after the date of the first dose of study medication or started prior to first dose and worsened in severity after the first dose will be considered as Treatment-Emergent Adverse Event (TEAE).

In case of missing or incomplete dates not allowing a direct allocation to any of the two categories of AEs (pre-treatment/TEAE), an allocation will be done according to the available parts of the onset and the end dates (see Table 6). In case of TEAE, the event can be further classified as:

- On Treatment period, or
- On Follow-up period

according to the available parts of the onset and the end dates (see Table 6).

All AEs will be coded by SOC and PT according to MedDRA thesaurus. In addition, each AE will be graded to capture the relationship to IMP and severity (see protocol for details).

Pre-treatment AEs will be presented in the listings only.

TEAE summaries will be presented, displaying frequencies and percentages of patients reporting AEs within each SOC in decreasing order of total frequency. Along with AEs, number of events will be reported. On each of these summaries, patients will be counted only once per SOC and, withing each SOC, patients will be counted only once per PT.

“Possible”, “Probable” and “Highly Probable” or missing relationships will be considered as related to study drug (ADR – Adverse Drug Reaction) for the summary tables. “None” and “unlikely” relationships will be considered not related to study drug.

The following table will be presented by treatment group and overall:

- A summary overview (patient with at least one event, percentage, and number of events) of:
  - TEAEs,
  - Ophthalmic TEAEs,
  - Non-Ophthalmic TEAEs,
  - severe TEAEs,
  - severe Ophthalmic TEAEs,
  - severe Non-Ophthalmic TEAEs,
  - serious TEAEs,
  - serious Ophthalmic TEAEs,
  - serious Non-Ophthalmic TEAEs,
  - non-serious TEAEs,
  - ADRs,
  - Ophthalmic ADRs,
  - Non- Ophthalmic ADRs,
  - serious ADRs,
  - TEAEs leading to discontinuation of IMP,
  - TEAEs leading to discontinuation of study,
  - TEAEs leading to death.

The following tables will be presented by treatment group:

- Summary of Ophthalmic TEAEs by primary SOC and PT and by study period (on treatment/follow-up and overall).
- Summary of Non-Ophthalmic TEAEs by primary SOC and PT and by study period (on treatment/follow-up and overall).
- Summary of TEAEs by primary SOC, PT and Severity.
- Summary of Serious TEAEs by Primary SOC and PT and by study period (on treatment/follow-up and overall).
- Summary of ADRs by Primary SOC and PT and by study period (on treatment/follow-up and overall).
- Summary of ADRs by Primary SOC and PT and Severity.
- Summary of TEAEs leading to IMP Discontinuation by Primary SOC and PT and by study period (on treatment/follow-up and overall).
- Summary of TEAEs leading to study Discontinuation by Primary SOC and PT and by study period (on treatment/follow-up and overall).
- Summary of TEAEs leading to Death by Primary SOC and PT.
- Listing of Deaths.

12. Derivations and date conventions

12.1 Variable derivation

Table 3: Variable derivation rules

Parameter	Calculation
Screened	A patient is considered screened if (s)he has "Has written Informed Consent been obtained?" = Yes AND "Date of Informed Consent Signature at screening" filled
Screening failure	A patient is considered screening failure if (s)he has "Screen Failure date" filled
Enrolled	A patient is considered enrolled if (s)he has been screened and (s)he is not a screening failure.
Eligible eye	Assuming that all the inclusion/exclusion criteria are met: • If only one eye fulfills the criteria of Severe Sjögren's Dry Eye ("Which eye is diagnosed with Severe Sjögren's Dry Eye?" is answered with "Right Eye" or "Left Eye") than the selected eye will be the "eligible eye". • If both eyes fulfill the criteria of Severe Sjögren's Dry Eye ("Which eye is diagnosed with Severe Sjögren's Dry Eye?" is answered with "Both") than the eligible eye will be selected according to the following rank: i. the eye with lower Schirmer I score, otherwise ii. the eye with worse NEI score (cornea+conjunctival staining), otherwise iii. the eye with worse investigator's judgement ("Were there any clinically significant abnormalities noted?" Equal to yes for a specific eye), otherwise iv. the right eye.
Change from baseline	Each change from baseline will be defined as difference between the value at each post-baseline assessment and the baseline value.
Time since first diagnosis of Dry Eye	Randomization date - Date of diagnosis for eligible eye + 1
Time since first diagnosis of Severe Sjögren's Dry Eye	Randomization date - Date of diagnosis of Severe Sjögren's Dry Eye for eligible eye + 1
Total number of drops in the eligible eye taken during the treatment period	Sum of Morning/Afternoon/Evening Number of times the IMP has been administered for the eligible eye between Randomization visit date and End of Treatment date, considering 1 drop each time (morning, afternoon, evening). (from the eCRF form "IMP Administration" forms)

Total number of drops in the eligible eye scheduled during the treatment period	3*(Date of End of Treatment visit – Date of Randomization + 1) – (assessments not done on Randomization visit + assessments not done on End of Treatment visit)
The extent of exposure (days)	Date of last administration of IMP – Date of first administration of IMP + 1
Total number of drops of the preservative free artificial tears during the treatment period	Sum of Morning/Afternoon/Evening Number of drops taken for eligible eye between Randomization visit date and End of Treatment date. (from the eCRF form "Frequency of preservative free artificial tears use" forms)
Total number of days of the treatment period	End of Treatment date – Randomization date + 1
Total number of drops of the preservative free artificial tears during the follow-up period	Sum of Morning/Afternoon/Evening Number of drops taken after End of Treatment date. (from the eCRF form "Frequency of preservative free artificial tears use" forms)
Total number of days of the follow-up period	Last follow-up available visit* date – End of Treatment date + 1 (* with assessment of Frequency of preservative free artificial tears use)
Time from randomization to study discontinuation (days)	End of Study date – Randomization date + 1
Primary and Secondary Sjögren's	For the subgroup analyses, a patient will be classified as primary Sjögren's syndrome if the patient experienced any previous and/or concomitant ocular with diagnosis='Sjogren Syndrome'. Otherwise the patient will be classified as secondary Sjögren's syndrome
Conversion of Time Intervals	If a time interval was calculated in minutes, hours or days and needs to be converted into months or year the following conversion factors will be used: • 1 hour = 60 minutes • 1 day = 24 hours • 1 week = 7 days • 1 month = 30.4375 days • 1 year = 365.25 days
General rule	For calculation of a time interval, in case time is available, it is considered in the formula. Time intervals will be expressed in hours in if less than 24 hours and days if more than 24 hours.

12.2 SANDE derivation

The SANDE questionnaire is comprised of two questions:

- How often do your eyes feel dry and/or irritated?
- How severe you feel your symptoms of dryness and/or irritation are?

This questionnaire uses a 100 mm horizontal line for each question to assess ocular discomfort and/or dryness experienced by the patients. In the SANDE questionnaire, frequency of symptoms ranges from “rarely” to “all of the time” and the severity of symptoms ranges from “very mild” to “very severe”. Patients are asked to place a mark on the two given lines based on the extent of their symptoms.

The locations of the marks made by the patients on the 100 mm horizontal lines will be measured in mm from left to right and recorded as frequency and severity score, respectively.

The global SANDE score will be calculated by multiplying the frequency score by the severity score and obtaining the square root.

12.3 IDEEL derivation

IDEEL is a 57-item questionnaire developed by Abetz et al. [4], that that assess the impact of dry eye symptoms on everyday life. The possible range of each IDEEL score was 0 (complete disability) to 100 (no disability). This questionnaire consists of 3 modules:

- Dry eye Treatment satisfaction & Bother (10 questions) comprising two sections: “Treatment – In General” and “Treatment- Eye drops”.
- Dry eye Quality of Life (27 questions) comprising three sections: “Daily activities”, “Feelings” and “Work”.
- Dry eye Symptom Bother (20 questions).

Rules for the calculation of the IDEEL dimension scores are reported in Table 4 and Table 5.

Table 4: Calculation of IDEEL dimension scores

Parameter	Calculation
Item scaling	• Dry Eye Treatment Satisfaction module: o Items 1-5 are scored on a 5-point Likert-like scale from 0 "none of the time" to 4 "all of the time" o Items 6, 8-10 are scored on a 5-point Likert-like scale from 0 "all of the time" to 4 "none of the time" o Item 7 is scored on a dichotomous scale: 1 "Yes", 0 "No". This item is not included in the calculation of the scores. • Dry Eye Impact on Daily Life module: o Items 1-9 are scored on a 5-point Likert-like scale from 1 "all of the time" to 5 "none of the time". Patients can also answer "I can no longer do this activity due to my dry eyes" (scored 0) or "I did not do this activity for reasons other than my dry eyes / Not applicable" (coded 6 but scored 5 for calculation of the scores) o Items 10-21, 23-27 are scored on a 5-point Likert-like scale from 0 "all of the time" to 4 "none of the time" o Item 22 is scored on a dichotomous scale: 1 "Yes", 0 "No". This item is not included in the calculation of the scores • Dry Eye Symptom-Bother module: o Item 1 is scored on a 5-point Likert-like scale from 0 "none of the time" to 4 "all of the time" o Items 2-20 are scored on a 4-point Likert-like scale from 1 "not at all" to 4 "very much". Patients can also answer "I did not have this symptom / Not applicable" (scored 0).
Weighting of items	All items are given equal weight in the calculation of the scores.
Range of scores	0-100 scales.
Scoring procedure	• Dry Eye Treatment Satisfaction module: o Satisfaction with Treatment Effectiveness score: If patients never use treatment for their dry eyes (answer is 0 "none of the time" to item 1) then items 2-5 are missing, and the Satisfaction with Treatment Effectiveness score is missing. The Satisfaction with Treatment Effectiveness score is calculated if at least 50% (2 items) of the 4 items within the dimension are completed, non-missing; otherwise the score is set to missing. The Satisfaction with Treatment Effectiveness score is calculated as the mean of the non-missing item scores 2-5 multiplied by 25. o Treatment-Related Bother score: If patients never use treatment for their dry eyes (answer is 0 "none of the time" to item 1) then item 6 is missing. If patients never use eye drops to treat their dry eyes (answer is 0 "No" to item 7) then items 8-10 are missing, and the Treatment-Related Bother score is missing. The Treatment-Related Bother score is calculated if at least 50% (2 items) of the 4 items within the dimension are completed, non-missing; otherwise the score is set to missing. The Treatment-Related Bother score is calculated as the mean of the non-missing item scores 6, 8-10 multiplied by 25. • Dry Eye Impact on Daily Life module: o Impact on Daily Activities score: Response "I did not do this activity for reasons other than my dry eyes / Not applicable" for items 1-6 is re-coded 5 instead of 6. Other responses are scored with the original response code. The Impact on Daily Activities score is calculated if at least 50% (3 items) of the 6 items within the dimension are completed, non-missing; otherwise the score is set to missing. The Impact on Daily Activities score is calculated as the mean of the non-missing item scores 1-6 multiplied by 20. o Emotional Impact score: The Emotional Impact score is calculated if at least 50% (6 items) of the 11 items within the dimension are completed, non-missing; otherwise, the score is set to missing. The Emotional Impact score is calculated as the mean of the non-missing item scores 10-20 multiplied by 25.

Parameter	**Calculation**
	○ Impact on Work score: If patients are not currently working (answer is 0 "No" to item 22) then items 23-27 are missing, and the Impact on Work score is missing. The Impact on work score is calculated only when patients answer "Yes" to "Are you currently working?". The Impact on Work score is calculated if at least 50% (3 items) of the 5 items within the dimension are completed, non-missing; otherwise the score is set to missing. The Impact on Work score is calculated as the mean of the non-missing item scores 23-27 multiplied by 25. • Dry Eye Symptom-Bother module: ○ The Symptom Bother score is calculated if at least 50% (10 items) of the 20 items within the dimension are completed, non-missing; otherwise the score is set to missing. The Symptom Bother score is calculated as the mean of the non-missing item scores 1-20 multiplied by 25.
Interpretation and method of treating missing data	All the scores are calculated only if at least 50% of the items of the dimension are completed; otherwise, the scores are missing.

Table 5: Summary of dimensions and clusters of items of the IDEEL questionnaire

Module	Dimensions	Number of items	Cluster of items	Direction of dimension scores
Dry Eye Impact on Daily Life	Impact on Daily Activities	6	1-6	Higher score = less impact on daily activities
	Emotional Impact due to Dry Eye	11	10-20	Higher score = less emotional impact
	Impact on Work due to Dry Eye	5	23-27	Higher score = less work impact
Dry Eye Symptom-Bother	Dry Eye Symptom-Bother	20	1-20	Higher score = greater symptom bother
Dry Eye Treatment Satisfaction	Satisfaction with Treatment Effectiveness	4	2-5	Higher score = greater satisfaction with treatment effectiveness
	Treatment- Related Bother / Inconvenience	4	6, 8-10	Higher score = less treatment related bother

12.4 Corneal and Conjunctiva Vital Staining with Fluorescein (NEI scale)

Corneal staining total score is defined as the sum of scores from 5 corneal areas: Central, Superior, Temporal, Nasal, and Inferior. The score for each area ranges from 0 to 3. Thus, the corneal staining total score can range from 0 to 15.

The conjunctiva is divided into a superior paralimbal area, an inferior paralimbal area and a peripheral area with a grading scale of 0–3 and with a maximal score of 9 for the nasal and temporal conjunctiva. Thus, the conjunctival staining total score can range from 0 to 18.

Corneal and conjunctiva vital staining total score is the sum of corneal staining total score and conjunctiva total score.

Information from blinded reading center and site investigators will be used as described in section 7.8.

12.5 Partial date conventions

Table 6: Algorithm for Treatment Emergence of Adverse Events

AE START DATE	AE STOP DATE	RULE for TEAE definition	RULE for "Treatment"/"Follow-up" study period definition for TEAE summaries
Known	Known, Partial or Missing	If AE start date < IMP start date, then not TEAE If AE start date >= IMP start date, then TEAE	*In case of Treatment Discontinuation:* If start date > Date of last IMP intake, then Follow-up study period for TEAE occurrence. Otherwise, Treatment study period for TEAE occurrence. *In case of NO Treatment Discontinuation:* If start date ≥ Date of "Day 21" visit, then Follow-up study period for TEAE occurrence. Otherwise, Treatment study period for TEAE occurrence.
Partial, but known components show that it cannot be on or after IMP start date	Known, Partial or Missing	Not TEAE	Not applicable

AE START DATE	**AE STOP DATE**	**RULE for TEAE definition**	**RULE for "Treatment"/"Follow-up" study period definition for TEAE summaries**
Partial, could be on or after IMP start date	Known	If AE stop date < IMP start date, then not TEAE If AE stop date >= IMP start date, then TEAE	If TEAE, then Treatment study period for TEAE occurrence.
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If AE stop date < IMP start date, then not TEAE If AE stop date >= IMP start date, then TEAE	
	Missing	Assumed TEAE	
Missing	Known	If AE stop date < IMP start date, then not TEAE If AE stop date >= IMP start date, then TEAE	If TEAE, then Treatment study period for TEAE occurrence.
	Partial	Impute AE stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If AE stop date < IMP start date, then not TEAE If AE stop date >= IMP start date, then TEAE	
	Missing	Assumed TEAE	

NOTE: Assignment to "Treatment" or "Follow-up" study period is applicable only for TEAEs.

Table 7: Algorithm for Prior/Concomitant medications

MEDICATION START DATE	MEDICATION STOP DATE	RULE for prior or concomitant categorization
Known	Known	If medication stop date < date of first dose of IMP, assign as prior If medication stop date >= date of first dose of IMP, assign as concomitant
	Partial	Impute medication stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If medication stop date < date of first dose of IMP, assign as prior If medication stop date >= date of first dose of IMP, assign as concomitant
	Missing	Assign as concomitant
Partial or Missing	Known	If medication stop date < date of first dose of IMP, assign as prior If medication stop date >= date of first dose of IMP, assign as concomitant
	Partial	Impute medication stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If medication stop date < date of first dose of IMP, assign as prior If medication stop date >= date of first dose of IMP, assign as concomitant
	Missing	Assign as concomitant

13. Tables, Figures and Listings

13.1 Table Conventions

- If the categories of a parameter are ordered, then all categories between the maximum and minimum category are presented in the table even in case of frequency equal to 0.
- If the categories are not ordered (e.g., Medical History), then only those categories for which there is at least 1 subject represented in 1 or more groups are included.
- Unless otherwise specified, the estimated mean and median for a set of values are printed out to 1 more significant digit than the original values, and standard deviations.
- Missing descriptive statistics or p-values which cannot be estimated are reported as "-".

13.2 Listing Conventions

- Listings will be sorted for presentation in order of subject number, treatment groups (planned), and visit.
- All listings will contain the planned and actual treatment.
- Dates are printed in SAS DATE9.format ("ddMMMyyyy": 01JUL2000). Missing portions of dates are represented on subject listings as dashes (--JUL2000).

14. References

1. NGF0221 Clinical Study Protocol "A 4 week, Phase III, multicenter, double-masked, vehicle-controlled study to evaluate safety and efficacy of Oxervate® (cenegermin) 20 mcg/mL ophthalmic solution versus vehicle, in patients with severe Sjogren's dry eye disease" Version No. 5.0 – 16 June 2022.
2. Patient diary Version No. 3.0 – 07 February 2022.
3. Case Report Form Version No. MSC004 – 02 March 2023.
4. Abetz L, Rajagopalan K, Mertzanis P, Begley C, Barnes R, Chalmers R; Impact of Dry Eye on Everyday Life (IDEEL) Study Group. Development and validation of the impact of dry eye on everyday life (IDEEL) questionnaire, a patient-reported outcomes (PRO) measure for the assessment of the burden of dry eye on patients. Health Qual Life Outcomes. 2011 Dec 8;9:111. doi: 10.1186/1477-7525-9-111. PMID: 22152125; PMCID: PMC3269387.

15. Appendices

15.1 Tables

Number - Title - Population	KFR	Final analysis
Table 14.1.1 Subject enrolment and disposition. Screened set	x	x
Table 14.1.2.1 Major protocol deviations. Randomized set		x
Table 14.1.2.2 Minor protocol deviations. Randomized set		x
Table 14.1.3.1 Study discontinuation and primary reasons for withdrawal. Randomized set		x
Table 14.1.4.1 Summary of demographic characteristics. Full analysis set	x	x
Table 14.1.4.2 Summary of baseline characteristics. Full analysis set	x	x
Table 14.1.5.1 Medical History. Full analysis set		x
Table 14.1.5.2 Concomitant diseases. Full analysis set		x
Table 14.1.6.1 Prior medication. Full analysis set		x
Table 14.1.6.2 Concomitant medication. Full analysis set		x
Table 14.1.6.3 Prior medication. Safety analysis set		x
Table 14.1.6.4 Concomitant medication. Safety analysis set		x
Table 14.1.7 Dry Eye History. Full analysis set		x
Table 14.1.8.1 Other baseline characteristics: Eligible eye, SANDE scores, IDEEL scores, Schirmer test I and Schirmer test II. Full analysis set		x
Table 14.1.8.2 Other baseline characteristics by eye: BCDVA, External ocular examination, Slit Lamp Ophthalmic Exam, NEI scale and TFBUT. Full analysis set		x
Table 14.2.1.1 Schirmer I test (without anesthesia) >10mm/5min at week 4 - Logistic regression model - Multiple imputation. Full analysis set	x	x
Table 14.2.1.2 Change from baseline in the global SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Multiple imputation. Full analysis set	x	x
Table 14.2.2.1 Sensitivity analysis: Schirmer I test (without anesthesia) >10mm/5min at week 4 - Logistic regression model - Multiple imputation. Full analysis set		x
Table 14.2.2.2 Sensitivity analysis: Change from baseline in the global SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Multiple imputation. Full analysis set		x

Number - Title - Population	KFR	Final analysis
Table 14.2.2.3 Sensitivity analysis: Schirmer I test (without anesthesia) >10mm/5min at week 4 - Logistic regression model – Tipping point. Full analysis set		x
Table 14.2.2.4 Sensitivity analysis: Change from baseline in the global SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Tipping point. Full analysis set		x
Table 14.2.3.1 Supportive analysis: Schirmer I test (without anesthesia) >10mm/5min at week 4 - Logistic regression model – Complete cases only. Full analysis set		x
Table 14.2.3.2 Supportive analysis: Change from baseline in the global SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Complete cases only. Full analysis set		x
Table 14.2.3.3 Supportive analysis: Schirmer I test (without anesthesia) >10mm/5min at week 4 - Logistic regression model – Multiple imputation. Per Protocol analysis set		x
Table 14.2.3.4 Supportive analysis: Change from baseline in the global SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Multiple imputation. Per Protocol analysis set		x
Table 14.2.4.1 Schirmer I test (without anesthesia) >10mm/5min at week 8 - Logistic regression model - Multiple imputation. Full analysis set		x
Table 14.2.4.2 Change from baseline in the severity SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Multiple imputation. Full analysis set		x
Table 14.2.4.3 Change from baseline in the frequency SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Multiple imputation. Full analysis set		x
Table 14.2.4.4.1.1 Change from baseline in the IDEEL Quality of Life module (Daily activities) up to week 12 - MMRM - Multiple imputation. Full analysis set		x
Table 14.2.4.4.1.2 Change from baseline in the IDEEL Quality of Life module (Feelings) up to week 12 - MMRM - Multiple imputation. Full analysis set		x
Table 14.2.4.4.1.3 Change from baseline in the IDEEL Quality of Life module (Work) up to week 12 - MMRM - Multiple imputation. Full analysis set		x
Table 14.2.4.4.2.1 Change from baseline in the IDEEL Dry eye Treatment satisfaction & Bother module (Treatment – In General) up to week 12 - MMRM - Multiple imputation. Full analysis set		x
Table 14.2.4.4.2.2 Change from baseline in the IDEEL Dry eye Treatment satisfaction & Bother module (Treatment – Eye drops) up to week 12 - MMRM - Multiple imputation. Full analysis set		x
Table 14.2.4.4.3 Change from baseline in the IDEEL Dry eye Symptom bother module up to week 12 - MMRM - Multiple imputation. Full analysis set		x
Table 14.2.4.5 Change from baseline in the Cornea and conjunctiva vital staining with fluorescein NEI scale up to week 12 - MMRM - Multiple imputation. Full analysis set		x
Table 14.2.4.6 Change from baseline in the TFBUT up to week 12 - MMRM - Multiple imputation. Full analysis set		x

Number - Title - Population	KFR	Final analysis
Table 14.2.5.1 Absolute values and change from baseline in Schirmer I test at each timepoint. Full analysis set	x	x
Table 14.2.5.2.1 Absolute values and change from baseline in SANDE global scores at each timepoint. Full analysis set	x	x
Table 14.2.5.2.2 Absolute values and change from baseline in SANDE severity scores at each timepoint. Full analysis set	x	x
Table 14.2.5.2.3 Absolute values and change from baseline in SANDE frequency scores at each timepoint. Full analysis set	x	x
Table 14.2.5.3 Number of patients experienced a worsening in symptom scores (SANDE global score) and/or NEI score ≥ 50% assessed at week 4. Full analysis set		x
Table 14.2.5.4.1 Absolute values and change from baseline in IDEEL Quality of Life module at each timepoint. Full analysis set	x	x
Table 14.2.5.4.2 Absolute values and change from baseline in IDEEL Dry eye Treatment satisfaction & Bother module at each timepoint. Full analysis set	x	x
Table 14.2.5.4.3 Absolute values and change from baseline in IDEEL Dry eye Symptom bother module at each timepoint. Full analysis set	x	x
Table 14.2.5.5 Absolute values and change from baseline in Cornea and conjunctiva vital staining with fluorescein (NEI scale) at each timepoint. Full analysis set	x	x
Table 14.2.5.6 Absolute values and change from baseline in TFBUT at each timepoint. Full analysis set	x	x
Table 14.2.6.1.1 Subgroup evaluation: Schirmer I test (without anesthesia) >10mm/5min at week 4 - Treatment by region interaction- Logistic regression model - Multiple imputation. Full analysis set		x
Table 14.2.6.1.2 Subgroup evaluation: Schirmer I test (without anesthesia) >10mm/5min at week 4 - Treatment by Sjögren's syndrome interaction- Logistic regression model - Multiple imputation. Full analysis set		x
Table 14.2.6.1.3 Subgroup evaluation: Schirmer I test (without anesthesia) >10mm/5min at week 4 - Treatment by medical history interaction- Logistic regression model - Multiple imputation. Full analysis set		x
Table 14.2.6.1.4 Subgroup evaluation: Schirmer I test (without anesthesia) >10mm/5min at week 4 - Treatment by concomitant medication interaction- Logistic regression model - Multiple imputation. Full analysis set		x
Table 14.2.6.2.1 Subgroup evaluation: Change from baseline in the global SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Treatment by region interaction- Multiple imputation. Full analysis set		x
Table 14.2.6.2.2 Subgroup evaluation: Change from baseline in the global SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Treatment by Sjögren's syndrome interaction- Multiple imputation. Full analysis set		x
Table 14.2.6.2.3 Subgroup evaluation: Change from baseline in the global SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Treatment by medical history interaction- Multiple imputation. Full analysis set		x
Table 14.2.6.2.4 Subgroup evaluation: Change from baseline in the global SANDE score at week 12 - Analysis of Covariance (ANCOVA) - Treatment by concomitant medication interaction- Multiple imputation. Full analysis set		x
Table 14.2.6.3.1 Absolute values and change from baseline in Schirmer I test at each timepoint by age class. Full analysis set		x

Number - Title - Population	KFR	Final analysis
Table 14.2.6.4.1 Absolute values and change from baseline in Cornea and conjunctiva vital staining with fluorescein (NEI scale) at each timepoint by age class. Full analysis set		x
Table 14.2.6.5.1 Absolute values and change from baseline in TFBUT at each timepoint by age class. Full analysis set		x
Table 14.2.6.6.1.1 Absolute values and change from baseline in SANDE global scores at each timepoint by age class. Full analysis set		x
Table 14.2.6.6.2.1 Absolute values and change from baseline in SANDE severity scores at each timepoint by age class. Full analysis set		x
Table 14.2.6.6.3.1 Absolute values and change from baseline in SANDE frequency scores at each timepoint by age class. Full analysis set		x
Table 14.2.6.7.1.1 Absolute values and change from baseline in IDEEL Quality of Life module at each timepoint by age class. Full analysis set		x
Table 14.2.6.7.2.1 Absolute values and change from baseline in IDEEL Dry eye Treatment satisfaction & Bother module at each timepoint by age class. Full analysis set		x
Table 14.2.6.7.3.1 Absolute values and change from baseline in IDEEL Dry eye Symptom bother module at each timepoint by age class. Full analysis set		x
Table 14.2.6.y.x Repeated set of tables form 14.2.6.3.1 to 14.2.6.2.7.3.1 for all subgroups		x
Table 14.3.1.1 Summary of treatment emergent adverse events (TEAE). Safety set		x
Table 14.3.1.2 Summary of Ophthalmic TEAEs by primary SOC and PT by study periods. Safety set		x
Table 14.3.1.3 Summary of Non-Ophthalmic TEAEs by primary SOC and PT by study periods. Safety set		x
Table 14.3.1.4 Summary of TEAEs by primary SOC and PT by severity. Safety set		x
Table 14.3.1.5 Summary of Serious TEAEs by primary SOC and PT by study periods. Safety set		x
Table 14.3.1.6 Summary of ADRs by primary SOC and PT by study periods. Safety set		x
Table 14.3.1.7 Summary of ADRs by primary SOC and PT by severity. Safety set		x
Table 14.3.1.8 Summary of TEAEs leading to IMP discontinuation by primary SOC and PT by study periods. Safety set		x
Table 14.3.1.9 Summary of TEAEs leading to study discontinuation by primary SOC and PT by study periods. Safety set		x
Table 14.3.1.10 Summary of TEAEs leading to death by primary SOC and PT. Safety set		x
Table 14.3.2.1.1 Compliance to IMP. Full analysis set		x
Table 14.3.2.1.2 Compliance to IMP. Safety set		x
Table 14.3.2.2.1 Extent of exposure. Full analysis set		x
Table 14.3.2.2.2 Extent of exposure. Safety set		x
Table 14.3.2.3.1 Use of preservative free artificial tears. Full analysis set		x
Table 14.3.2.3.2 Use of preservative free artificial tears. Safety set		x

15.2 Listings

Number - Title - Population	KFR	Final analysis
Listing 16.1.7 Randomisation Scheme. Randomized set		x
Listing 16.2.1 Patient disposition in analysis data sets and primary reasons for withdrawal. Enrolled set		x
Listing 16.2.2 Protocol deviations. Randomized set		x
Listing 16.2.3.1 Study discontinuation and reasons for withdrawal. Randomized set		x
Listing 16.2.3.2 Study Visits. Randomized set		x
Listing 16.2.4.1 Demographic data. Safety set		x
Listing 16.2.4.2 Baseline Characteristics. Safety set		x
Listing 16.2.4.3 Medical history. Safety set		x
Listing 16.2.4.4 Concomitant Diseases. Safety set		x
Listing 16.2.4.5 Prior medication and therapy. Safety set		x
Listing 16.2.4.6 Concomitant medication and therapy. Safety set		x
Listing 16.2.4.7 Dry Eye History. Safety set		x
Listing 16.2.5.1 Compliance to IMP. Safety set		x
Listing 16.2.5.2 Exposure to IMP. Safety set		x
Listing 16.2.5.3 Use of preservative free artificial tears. Safety set		x
Listing 16.2.6.1 Schirmer test I by patient and visit. Safety set		x
Listing 16.2.6.2 Schirmer test II by patient and visit. Safety set		x
Listing 16.2.6.3 SANDE scores by patient and visit. Safety set		x
Listing 16.2.6.4 NEI scale by patient and visit. Safety set		x
Listing 16.2.6.5 TFBUT by patient and visit. Safety set		x
Listing 16.2.6.6 IDEEL scores by patient and visit. Safety set		x
Listing 16.2.7.1 Listing of all AEs by Patient. Safety set		x
Listing 16.2.7.2 Listing of all AEs leading to IMP discontinuation. Safety set		x
Listing 16.2.7.3 Listing of all AEs leading to study discontinuation. Safety set		x
Listing 16.2.7.4 Listing of SAEs by Patient. Safety set		x

Number - Title - Population	KFR	Final analysis
Listing 16.2.7.5 Listing of ADR by Patient. Safety set		x
Listing 16.2.7.6 Listing of Deaths. Safety set		x

15.3 Figures

Number - Title - Population	KFR	Final analysis
Figure 14.1.3.2 Time from randomization to discontinuation (Kaplan-Meier) . Randomized set		x
Figure 14.2.3.5 Tipping point Shirmer test. Full analysis set		x
Figure 14.2.3.6 Tipping point plot SANDE. Full analysis set		x
Figure 14.2.5.1.1 Schirmer I test Observed Mean (±SEM) at each timepoint. Full analysis set		x
Figure 14.2.5.1.2 Schirmer I test Mean change from baseline (±SEM) at each timepoint. Full analysis set		x
Figure 14.2.5.1.3 Observed percentage of patients with Schirmer I test >10mm/5min at each timepoint. Full analysis set		x
Figure 14.2.5.2.1.1 SANDE global score Observed Mean (±SEM) at each timepoint. Full analysis set		x
Figure 14.2.5.2.1.2 SANDE global score Mean change from baseline (±SEM) at each timepoint. Full analysis set		x
Figure 14.2.5.5.1 Cornea and conjunctiva vital staining total score Observed Mean (±SEM) at each timepoint. Full analysis set		x
Figure 14.2.5.5.2 Cornea and conjunctiva vital staining total score Mean change from baseline (±SEM) at each timepoint. Full analysis set		x
Figure 14.2.5.6.1 TFBUT Observed Mean (±SEM) at each timepoint. Full analysis set		x
Figure 14.2.5.6.2 TFBUT Mean change from baseline (±SEM) at each timepoint. Full analysis set		x

15.4 SAS Code for the primary analysis

15.4.1 Schirmer I test >10mm/5min at week 4

Standard	**Alternative code in case of no convergence**
PROC MI DATA= EFFICACY01 NIMPUTE=200 SEED=180404 OUT=EFFICACY_MONO1; VAR BASVAL SCHIRMER_W2 AVAL; MCMC CHAIN=SINGLE NBITER=200 NITER=100 IMPUTE=MONOTONE; WHERE TRT=1; RUN; PROC MI DATA=EFFICACY01 NIMPUTE=200 SEED=180404 OUT=EFFICACY_MONO2; VAR BASVAL SCHIRMER_W2 AVAL; MCMC CHAIN=SINGLE NBITER=200 NITER=100 IMPUTE=MONOTONE; WHERE TRT=2; RUN; DATA EFFICACY_MONO; SET EFFICACY_MONO1 EFFICACY_MONO2; RUN;	PROC MI DATA= EFFICACY01 NIMPUTE=200 SEED=180404 OUT=EFFICACY_MONO; VAR BASVAL TRT SCHIRMER_W2 AVAL; MCMC CHAIN=SINGLE NBITER=200 NITER=100 IMPUTE=MONOTONE; RUN;
PROC SORT; BY _IMPUTATION_ SUBJID; RUN;	
PROC MI DATA=EFFICACY_MONO SEED=200582 NIMPUTE=1 OUT=EFFICACY_IMP; BY _IMPUTATION_; CLASS TRT AGE SEX EOTSTT; MONOTONE REG(AVAL = TRT AGEc SEX BASVAL / DETAILS); MNAR MODEL(AVAL / MODELOBS= (EOTSTT='NOT COMPLETED')); VAR TRT AGE SEX BASVAL SCHIRMER_W2 AVAL; RUN;	PROC MI DATA=EFFICACY_MONO SEED=200582 NIMPUTE=1 OUT=EFFICACY_IMP; BY _IMPUTATION_; CLASS TRT AGE SEX EOTSTT; MONOTONE REG(AVAL = TRT AGE SEX BASVAL / DETAILS); MNAR MODEL(AVAL / MODELOBS= (TRT=2 /*Vehicle*/)); VAR TRT AGE SEX BASVAL SCHIRMER_W2 AVAL; RUN;

Standard	**Alternative code in case of no convergence**
DATA EFFICACY_IMP; SET EFFICACY_IMP; IF AVAL>10 THEN AVALC='Y'; ELSE AVALC='N'; RUN; ODS OUTPUT DIFFS=GLIMMIX_DIFFS; ODS OUTPUT PARAMETERESTIMATES=ESTIMATES; PROC GLIMMIX DATA=EFFICACY_IMP METHOD = QUAD; BY _IMPUTATION_; CLASS TRT(REF="2") AGE(REF="<=Median") SEX(REF="F") SITEID; MODEL AVALC(EVENT='Y')=TRT AGEc SEX BASVAL /S DIST=BINARY LINK=LOGIT OR; RANDOM INTERCEPT / SUBJECT=SITEID; LSMEANS TRT / ILINK ODDSRATIO CL DIFF=control("2"); RUN; ODS OUTPUT PARAMETERESTIMATES=POOLED_ESTIMATES; PROC MIANALYZE PARMS = ESTIMATES ; CLASS TRT AGE SEX; MODELEFFECTS INTERCEPT TRT AGE SEX BASVAL ; RUN; ODS OUTPUT PARAMETERESTIMATES=POOLED_DIFFS; PROC MIANALYZE DATA = GLIMMIX_DIFFS ; MODELEFFECTS ESTIMATE; STDERR STDERR; RUN; DATA POOLED_ODDS_RATIO; SET POOLED_DIFFS; ODDSRATIO=EXP(ESTIMATE); ODDS_95LL=EXP(LCLMEAN); ODDS_95UL=EXP(UCLMEAN); RUN;	

15.4.2 Change from baseline in the global SANDE score at week 12

Standard	**Alternative code in case of no convergence**
PROC MI DATA= EFFICACY01 NIMPUTE=200 SEED=252404 OUT=EFFICACY_MONO1; VAR BASVAL SANDE_W2 SANDE_W4 SANDE_W8 AVAL; MCMC CHAIN=SINGLE NBITER=200 NITER=100 IMPUTE=MONOTONE; WHERE TRT=1; RUN; PROC MI DATA=EFFICACY01 NIMPUTE=200 SEED=252404 OUT=EFFICACY_MONO2; VAR BASVAL SANDE_W2 SANDE_W4 SANDE_W8 AVAL; MCMC CHAIN=SINGLE NBITER=200 NITER=100 IMPUTE=MONOTONE; WHERE TRT=2; RUN; DATA EFFICACY_MONO; SET EFFICACY_MONO1 EFFICACY_MONO2; RUN;	PROC MI DATA= EFFICACY01 NIMPUTE=200 SEED=252404 OUT=EFFICACY_MONO; VAR BASVAL TRT SANDE_W2 SANDE_W4 SANDE_W8 AVAL; MCMC CHAIN=SINGLE NBITER=200 NITER=100 IMPUTE=MONOTONE; RUN;
PROC SORT; BY _IMPUTATION_ SUBJID; RUN;	
PROC MI DATA=EFFICACY_MONO SEED=200582 NIMPUTE=1 OUT=EFFICACY_IMP; BY _IMPUTATION_; CLASS TRT AGE SEX EOTSTT; MONOTONE REG(AVAL = TRT AGE SEX / DETAILS); MNAR MODEL(AVAL / MODELOBS= (EOTSTT='NOT COMPLETED')); VAR TRT AGE SEX BASVAL SANDE_W2 SANDE_W4 SANDE_W8 AVAL; RUN;	PROC MI DATA=EFFICACY_MONO SEED=200582 NIMPUTE=1 OUT=EFFICACY_IMP; BY _IMPUTATION_; CLASS TRT AGE SEX EOTSTT; MONOTONE REG(AVAL = TRT AGE SEX / DETAILS); MNAR MODEL(AVAL / MODELOBS= (TRT=2 /*Vehicle*/)); VAR TRT AGE SEX BASVAL SANDE_W2 SANDE_W4 SANDE_W8 AVAL; RUN;

Standard	**Alternative code in case of no convergence**
DATA EFFICACY_IMP; SET EFFICACY_IMP; CHG=AVAL-BASE; RUN; ODS OUTPUT DIFFS=DIFFS; ODS OUTPUT PARAMETERESTIMATES=ESTIMATES; PROC MIXED DATA=EFFICACY_IMP; BY _IMPUTATION_; CLASS TRT(REF="2") AGE(REF="<=Median") SEX(REF="F") SITEID; MODEL CHG=TRT AGE SEXC BASVAL /S; RANDOM INTERCEPT / SUBJECT=SITEID; LSMEANS TRT / CL DIFF =control("2"); RUN; ODS OUTPUT PARAMETERESTIMATES=POOLED_ESTIMATES; PROC MIANALYZE PARMS = ESTIMATES ; CLASS TRT AGE SEX; MODELEFFECTS INTERCEPT TRT AGE SEX BASVAL ; RUN; ODS OUTPUT PARAMETERESTIMATES=POOLED_DIFFS; PROC MIANALYZE DATA = DIFFS ; MODELEFFECTS ESTIMATE; STDERR STDERR; RUN;	