Protocol No: ALM-488-001

CP Project ID: ALU19001

Phase 1/2 Trial of ALM-488 in Subjects Undergoing Head & Neck Surgery

Statistical Analysis Plan

Draft Version: 2.0

Date: 25JAN2021

Table of Contents

1 Abbreviations

AE	Adverse Event
ATC	Anatomical Therapeutic Class
DLT	Dose Limiting Toxicity
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ENR	Enrolled analysis set
FAS	ALM-488 Activity Full Analysis Set
FL	Fluorescence
ICF	Informed consent form
IDMC	Independent Data Monitoring Committee
IRR	Infusion-related reaction
IV	Intravenous
MedDRA	Medical Dictionary for Regulatory Activities
NCI-CTCAE	National Cancer Institute – Common Terminology Criteria for Adverse events
PAS	PK Analysis Set
PK	Pharmacokinetics
PT	Preferred Term
RMANOVA	Repeated Measures Analysis of Variance
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SOC	System Organ Class
SRC	Safety Review Committee
TEAE	Treatment Emergent Adverse Event
TLF	Tables, Listings, and Figures
WHO	World Health Organization
WLR	White Light Reflectance

2 Introduction

2.1 Preface

The purpose of this document is to detail the statistical methodology to be used for the final statistical analysis of study ALM-488-001.

This Statistical Analysis Plan (SAP) covers the statistical summaries, analyses and methods of all the non-pharmacokinetic (PK) data. All details of the statistical summaries, analyses and methods for the PK data will be specified in a separate PK SAP. This SAP should be used in conjunction with the protocol. If there are any discrepancies between the protocol and SAP, this SAP will prevail. Any deviations from this SAP that are implemented in the final analysis will be documented with sound clinical and statistical rationale in the Clinical Study Report (CSR).

A separate document contains the table, listing and figure specifications and any example programming codes.

2.2 Timing of Statistical Analyses

- The following statistical analyses are planned for this study:
- Dose Escalation or De-Escalation
- Dose Timing Determination

3 Modification History

3.1 Changes to the Planned Analyses

The statistical analyses as specified in this SAP are mostly consistent with the statistical analyses as specified in the study protocol version Amendment 1, 24 August 2020.

We propose to include two additional specific analysis sets ENR, and DLT. They will allow to better describe the different steps of the study. They are not defined in the protocol. The additional analysis sets are defined below in section 5.

This SAP has been updated based on changes associated with protocol amendment 1.0.

4 Study Design

This is a Phase 1/2, open-label study in subjects undergoing head and neck surgery. The study will include up to 6 Dose Escalation/De-Escalation cohorts, followed by 2 pre-planned additional Dose Timing cohorts. Dose cohorts will start at 100 mg and follow the escalation/de-escalation described in the protocol.

Indication	Parotid Neoplasm, Thyroid Neoplasms, Head and Neck Neoplasms, Head and Neck Surgery
Design	Open-label study: Dose Escalation/Dose De-Escalation and Dose Timing cohorts

Phases	Phase ½
Primary Objective	To evaluate the safety of ALM-488 administered by intravenous (IV) infusion to subjects undergoing head and neck surgery.
Secondary Objectives	• To characterize the pharmacokinetics (PK) of ALM-488 in subjects undergoing head and neck surgery. • To determine the recommended dose of ALM-488 needed to generate an adequate fluorescence signal in nerves. • To evaluate the effect of timing of administration of ALM-488 relative to surgery (e.g., up to 5 hours before surgery) on adequacy of fluorescence nerve labeling.
Exploratory Efficacy Objectives	• To evaluate the ability of imaging techniques (i.e. camera systems) to distinguish nerve tissue from non-nerve tissue. • To evaluate the intensity of signal of fluorescence relative to nerve diameter. • To evaluate the intensity of signal of fluorescence when there is overlying tissue. • To assess the depth of fluorescence penetration when there is overlying non-nerve tissue. • To evaluate surgeon confidence and operating time with a surgeon rating scale.
Treatment	ALM-488 Sterile Solution is an intravenously administered, synthetic, peptide dye conjugate indicated for the real-time intraoperative fluorescence detection and localization of nerve tissue.
Number of subjects	9-36 subjects in the Dose Escalation/De-Escalation cohorts 12 subjects in the Dose Timing Cohorts Total 21-48 subjects
Interim analysis	No formal interim analysis is planned for this study
Planned enrolment	Three sites in the United States.

The aim is to determine the lowest possible dose that achieves adequate imaging performance (nerve identification) with an acceptable toxicity profile. The Safety Review Committee (SRC) will determine the Optimal Dose based on 1) the safety data and 2) adequacy of fluorescence observed in each of the dose groups. Following determination of the Optimal Dose, Dose Timing Cohorts will

commence at this Optimal Dose. The Dose Timing Cohorts will compare fluorescence characteristics at different ALM-488 administration times before surgery (3-5 hours, 1-3 hours).

Please refer to Section 5 of the protocol for full details on the study design.

4.1 Sample Size Estimation

The primary objective of this study is to evaluate the safety of ALM-488 administered by intravenous (IV) infusion to subjects undergoing head and neck surgery.

At each dose level, a cohort of 3 subjects will be enrolled for safety and tolerability evaluation. If there are no subjects with a DLT at a given dose level, new subjects will be enrolled at the next dose escalation level. If there are ≥1 subjects having any DLT, the enrollment to that specific dose and dose escalation will stop. New subjects will be enrolled at the next dose de-escalation level. There will be a minimum of 3 cohorts. Given that there will be 3-6 subjects for each dose level, 9-36 subjects will be needed for the dose escalation/de-escalation cohorts.

It is anticipated that the proposed dose range (100-1000mg) is expected to yield optimal imaging performance, with an initial starting dose of 100mg that has been selected based on supportive nonclinical testing. If the 100 mg dose is found to provide adequate fluorescence imaging of nerves, a dose de-escalation schedule will be followed to ensure that the lowest effective dose is determined. If imaging at starting first de-escalated dose (75 mg) becomes less adequate compared to 100mg as determined by the SRC, then imaging at dose escalation schedule will resume - beginning at 200mg. The Optimal Dose will be determined by the SRC as the lowest possible dose and toxicity profile that achieves adequate imaging performance of nerves using WLR combined with FL.

The Dose Timing Cohort will use the Optimal Dose as defined by the SRC following completion of the Dose Escalation/De-Escalation cohorts. Subjects in the Dose Timing phase will be administered ALM-488 at two planned operative times before surgery to assess the impact of timing of administration on fluorescence image quality. This phase will consist of 2 cohorts, 6 subjects/cohort: (1) dose administration 3-5 hours in advance of surgery, (2) dose administration 1-3 hours in advance of surgery. A total of 12 subjects will be needed for the Dose Timing phase.

In summary, 9-36 subjects may be enrolled into the dose escalation/dose de-escalation cohorts and 12 subjects may be enrolled into the dose timing cohort. A total of 21-48 subjects may be enrolled into this study.

4.2 Randomization, blinding and unblinding procedures

No randomization, blinding or unblinding procedures are planned for this study.

5 Analysis Sets and Cohorts

Analysis Sets are described as follows:

- Enrolled analysis set (ENR) defined as all subjects who provided informed consent.
- Safety Analysis Set (SAF), defined as all subjects who received at least one dose of ALM-488

- Dose-Limiting Toxicity (DLT) Evaluable Set, defined as subjects who receive ALM-488 and either 1) complete the safety follow-up through the DLT evaluation period (i.e., Day 15 ± 8) or 2) experiences a DLT at any time during the DLT evaluation period.
- Activity Analysis or Full Analysis Set (FAS), defined as all subjects in SAF who also received at least one evaluable image.
- Pharmacokinetics (PK) Analysis Set defined as all subjects who completed a PK blood sampling period without any major protocol violations which would render the data unreliable. The PK will be the primary analysis set for the pharmacokinetic analyses. PK of analyses will be fully described in the PK SAP.

SAF will be used for all safety analyses. FAS will be used for all efficacy analyses. The PK Analysis Set will be used for all PK analyses; ENR set will be the primary analysis set for disposition and listings. PP set will constitute the primary analysis set for early efficacy. DLT Evaluable Analysis Set will be used to assess the tolerability of ALM-488 in the dose escalation phase.

The Dose Escalation/De-Escalation cohort (Cohort A) consists of all dose escalation/de-escalation cohorts: Cohort A1, A2, A3, A4, A5, A6. In case of DLT rate = 0 or special FDA requests additional cohorts may be added to the tables.

The Dose Time Determination cohort (Cohort B) consists of: Cohort B1 - dose administration 3-5 hours in advance of surgery, and Cohort B2 - dose administration 1-3 hours in advance of surgery.

If not otherwise stated in the respective section, the statistical analyses will be performed for the following analysis sets:

Analyses	ENR	FAS	DLT	SAF	PK
Disposition	✓				
Demographics and baseline characteristics				✓	
Medical history				✓	
Exposure and compliance				✓	
Previous and concomitant therapies				✓	
Early Efficacy		✓			
Pharmacokinetic					✓
Safety				✓	
Dose-Limiting Toxicity			✓		
Listings		✓		✓	

6 Objectives

6.1 Primary Objective

To evaluate the safety of ALM-488 administered by intravenous (IV) infusion to subjects undergoing head and neck surgery.

6.2 Secondary Objectives

- To characterize the PK of ALM-488 in subjects undergoing head and neck surgery.
- To determine the dose of ALM-488 needed to generate an adequate fluorescence signal in nerves.
- To evaluate the effect of timing of administration of ALM-488 relative to surgery (e.g., up to 5 hours before surgery) on adequacy of fluorescence nerve labeling.

6.3 Exploratory Objectives

Exploratory objectives will evaluate early evidence of efficacy. These are:

- To evaluate the ability of imaging techniques (i.e. camera systems) to distinguishing nerve tissue from non-nerve tissue.
- To evaluate the intensity of signal of fluorescence relative to nerve diameter.
- To evaluate the intensity of signal of fluorescence when there is overlying tissue.
- To assess the depth of fluorescence penetration when there is overlying non nerve tissue.
- To evaluate surgeon confidence and operative time with a surgeon rating scale.

7 Endpoints

7.1 Safety Endpoints

The endpoints to assess the safety of ALM-488 include:

- Adverse Events (TEAEs)
- Vital signs
- Clinical laboratory evaluations
- ECGs
- Physical examinations

Summaries of AEs will be based on Treatment Emergent Adverse Events (TEAEs) and will be summarized separately for each device and each of the Cohorts A1-A6, B1 and B2, as well as overall for Cohort A and Cohort B.

Vital signs, labs, ECGs, and physical exams will be summarized separately for each of the Cohorts A1-A6, B1 and B2, as well as overall for Cohort A and Cohort B.

7.2 Secondary Endpoints

The endpoints to characterize the PK of ALM-488 will be fully described in the PK SAP.

The endpoints to determine the dose of ALM-488 needed to generate an adequate fluorescence signal in nerves include the following:

- Timing of administration of ALM-488 relative to surgery

- Dose
- Intensity (signal-to-background ratio- SBR)
- Surgeon rating scale

The endpoints to determine the dose of ALM-488 needed to generate an adequate fluorescence signal in nerves will be summarized for each of Cohorts A1-A6, as well as overall for Cohort A.

The endpoints to evaluate the effect of timing of administration of ALM-488 relative to surgery on adequacy of fluorescence nerve labeling include the following:

- Timing of administration of ALM-488 relative to surgery
- Time from injection to skin incision
- Intensity (SBR)
- Surgeon rating scale

The endpoints to evaluate the effect of timing of administration of ALM-488 relative to surgery on adequacy of fluorescence nerve labeling will be summarized for each of Cohorts B1 and B2, as well as overall for Cohort B.

7.3 Exploratory Endpoints

To evaluate the ability of imaging techniques (i.e. camera systems) to distinguishing nerve tissue from non-nerve tissue, the following endpoints will be evaluated:

- Ability to detect fluorescence will be studied by determing an optimal cut-off using Youden index and receiver operating characteristic ROC curve analysis.

This endpoint will be summarized separately for each device and each of the Cohorts A1-A6, B1 and B2, as well as overall for Cohort A and Cohort B.

To evaluate the intensity of signal of fluorescence relative to nerve diameter, the following endpoints will be evaluated:

- Intensity (signal-to-background ratio SBR)
- nerve diameter

These endpoints will be plotted against each other, separately for each device and each of the Cohorts A1-A6, B1 and B2, as well as overall for Cohort A and Cohort B.

To evaluate the intensity of signal of fluorescence when there is overlying tissue, the following endpoints will be evaluated:

- Signal-to-background ratio)

This endpoint will be summarized separately for images and each of the Cohorts A1-A6, B1 and B2 as well as overall for Cohort A and Cohort B.

To evaluate surgeon confidence and operative time with a surgeon rating scale, the following endpoints will be evaluated:

- Surgeon confidence will be captured via the surgeon rating scale (1=do not see nerve compared to WLR, 2=see nerve less well compared to WLR, 3=See nerve equally well

compared to WLR, 4=See nerve better compared to WLR, 5=See nerve much better compared to WLR) for each image.

- For each surgery, the following yes/no question will be asked: "Outside of a clinical trial setting with all the required extra documentation, would this technology have saved you operative nerve-dissecting time compared to WLR?"

Surgeon confidence will be summarized by image, while time saving (yes/no) will be summarized by patient.

Surgeon confidence will be summarized for each device and each of the Cohorts A1-A6, B1 and B2, as well as overall for Cohort A and Cohort B.

Time saving (yes/no) will be for each of the Cohorts A1-A6, B1 and B2, as well as overall for Cohort A and Cohort B.

8 General Statistical Methods and Definitions

8.1 General statistical methods

Summary tables will be structured with a column for each device and/or cohort.

The statistical analysis will be based on separate analysis populations, defined in section 4 Analysis Sets and Cohorts. Summary tables will usually be structured with a column for each cohort and the dose associated (ALM-488 100mg, ALM-488 XXXmg ,…, ALM-488 XXXmg for Dose Escalation/De-Escalation, and as well as for Dose Timing. If dose modification occurs or if fluorescence signal has not been achieved after 6 dose cohorts the modified or the additional dose escalation cohorts will be presented on the tables.

Additional cohorts beyond dose cohort 6 may be enrolled if an adequate fluorescence signal has not been achieved after 6 dose cohorts have been evaluated. Additional dose escalation cohorts may be added following discussion with FDA if further improvement of fluorescence is needed for nerve identification and if the DLT rate is 0

In general, continuous variables will be summarised using descriptive statistics, i.e. generally displaying number of subjects in the respective analysis population, number of subjects with data, number of subjects with missing values, mean, standard deviation, minimum, lower quartile, median, upper quartile, and maximum.

Categorical variables will be summarised by using frequency counts and percentages. For events that can occur more than once, the number of events may also be summarized. In addition, the number of subjects with missing values will be displayed.

Means and medians will be presented by 1 additional decimal place and standard deviation will be presented by 2 additional decimal places than the standard presentation level of the respective data. Minimum and maximum values will be presented using the same number of decimal places as the data. Percentages will be presented to 1 decimal place if not otherwise stated.

If the number of subjects in a category is 0, then percentage will not be displayed, and only a count of 0 will be shown.

P-values will be reported to 4 decimal places at least. Values less than 0.0001 will be displayed as <0.0001. Values above 0.9995 will be displayed as 1.0000.

In listings, data will be sorted by dose Cohort, site, and subject ID, and when appropriate by image, visit or other identifiers for sequence or type of observation.

Factor 365.2425 is used for duration conversions between days and years. Factor 30.436875 is used for duration conversions between days and months.

If not otherwise specified, all statistical tests will be two-sided, and confidence intervals will be two-sided and will be at the 95% confidence level.

8.2 Covariates and strata

No stratified analyses are planned for this study.

Study sites, countries, and regions

The three sites are in the Unites States.

- 101 - UCSD
- 102 Harvard/MEE
- 103 - Stanford

8.3 Subgroups

For the efficacy analysis, the subgroup of images with overlying tissue will be explored. In this subgroup, the depth will be measured via the amount of overlying tissue that there can be to still maintain the ratio of nerve to non-nerves. Other subgroup analyses will be presented by cohorts and by cameras (OnLume or Zeiss). Both cameras are able to capture WLR and FL lights.

Results for the different subgroups will be presented descriptively.

8.4 Missing data

Only observed data will be summarized for all Cohorts. Except for partial or missing dates, no imputation for missing data is planned for the summary or analyses of these study data.

The missing component(s) of incomplete dates (e.g. start and/or stop dates of AE, concomitant medication, medical history) will be assumed as the most conservative value possible. For example, if the start date has a missing day value, the first day of the month will be imputed for study day computations, etc. If day is missing for an end date, the last day of the month will be imputed. If the start date has a missing month value, the first month of the year will be imputed for study day computations, etc. If month is missing for an end date, the last month of the year will be imputed. For determination of treatment-emergent status, the start date will be imputed as the date of the first dose of study drug, unless there is clear evidence (through comparison of partial dates/times) to suggest otherwise.

Date imputation will only be used for computational purposes such as treatment-emergent status, etc. Actual data values, as they appear in the original eCRFs, will be presented in the subject data listings.

8.5 Rescreening

Note, for subjects who are rescreened, their Day 1 labs will be used as their Screening and Day 1 values.

8.6 Baseline Definition

For all variables, baseline will be considered to be the last value obtained prior to the date and time of incision. Change from baseline will be the difference between the value at a given time point and the baseline value. The baseline value corresponds to the screening visit in the study schedule.

8.7 Study Day, Visit Window, and COVID-19

For each patient, the day of surgery will be considered Study Day 1. Each assessment will be assigned a study day. The calculation for study day is dependent on whether the actual date of assessment is before or after the date of study Day 1 and is calculated as follows:

Before study Day 1:

study day = (date of assessment – Date of first dose of study medication Day 1).

On or after study Day 1:

study day = (date of assessment – Date of first dose of study medication Day 1) +1.

Days within the screening period will be numbered with negative numbers until day -1 defined as the day before AML-488 treatment (Day 1).
If the full date of the assessment is not known or not imputable based on the guidelines in Section 7.3, then no study day will be assigned. Study Day, and any corresponding durations will appear partial or missing in the listings.

For each patient, the end of the study is defined as the date of their last contact in this study.

Data will be analyzed based on the categorized visit captured in the database. No dates will be used to reassign visits or define visit windows. Although Day 15 or Early Termination is defined with +/-8 days, extensions to the visit window may be allowed due to impacts related COVID-19. Although Day 28/End of Study is defined with +/- 5 days, extensions to the visit window may be allowed due to impacts related COVID-19.

Any impacts due to COVID-19 will be noted in the study documentation.

The visits captured in the database include:

- Screening (Day -28 to Day -1)
- Baseline (Day 1) (pre- and post-operative)
- Day 15 or Early Termination

- Day 28/End of Study

9 Patient Accounting and Disposition

9.1 Patient accounting

The number and relative frequencies of subjects in each Cohort will be presented overall, by Analysis Sets, and individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, and B2), including individual reasons of exclusion from the respective analysis set.

A list of subjects' first ALM-488 infusion date time start, date time end, time to discontinuation and reasons for discontinuation will be presented.

9.2 Disposition and withdrawals

The number and relative frequency will be presented for

- Subjects completing the study
- Subjects who prematurely discontinued the study
- Reasons for study discontinuation (Adverse Event, death, lost to follow-up, physician decision, pregregnancy, progressive disease, site or study terminated by sponsor, technical problems, withdrawal by subject, adverse device effect or other)
- Subjects completing surgery
- Subjects completing treatment

Completing the study is defined as having completed the infusion, surgery, and end of study telephone call.

A list of subjects' trial end date, time to trial discontinuation and reason for trial discontinuation will be created.

9.3 Minor and Major Protocol deviations

The number and relative frequency of subjects with major protocol deviations will be presented by treatment cohort and study site for FAS.

Listings of all subjects with protocol deviations as specified above violating in- and exclusion criteria and allocation to trial populations will be created.

10 Demographics and Baseline Characteristics

Demographic and baseline characteristics as specified in detail below will be presented descriptively overall, by Analysis Sets, and by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

10.1 Demographics

The following demographic characteristics will be presented by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

- Sex (male/female)
- Age (years)
- Age group (≥18 and <65, ≥65 and <85, ≥85 years)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Unknown, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not Reported, Unknown)
- Childbearing potential or less than 2 years menopausal (yes/no);
- Acceptable contraception (yes/no)
- Surgery type (Thyroidectomy, Parotidectomy, Neck Dissection)

Date of informed consent, protocol version, informed consent form (ICF) type and ICF version number will be listed.

10.2 Surgery Type

The type of surgery (Thyroidectomy, Parotidectomy, Neck Dissection) will be summarized overall in the Demographics and Baseline Characteristics and further listed overall, and by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

10.3 Medical history

The diseases are coded according to the latest available Medical Dictionary for Regulatory Activities (MedDRA® version 23.1). The frequency of diseases recorded from medical history will be presented after classification into previous and concomitant conditions by system organ class (SOC) as well as the frequencies of preferred terms (PT) within each SOC. If subjects have more than one disease within an SOC or PT they will be counted only once for the respective SOC or PT.

A listing of medical history will be provided.

10.4 Exposure and Compliance

The number and percentage of subjects will be presented by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2) on the SAF analysis set.

The infusion duration will be calculated in minutes as:

When there are no interruptions in infusions:

Infusion duration (minutes) = (time of infusion end – time of infusion start) + 1

When infusion interruptions will be reported the time of the interruption need to be subtracted from time of infusion.

Descriptive statistics for the treatment duration will be presented.

The infusion completed per protocol, details if deviated as well as infusion reactions IR will be described.

Study ALM-488 dose and timing of administration and the summary of exposure will be listed by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2) for the SAF Analysis Set

10.5 Dosage and Timing of Administration

Descriptive statistics for the dose and timing will be presented overall and by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2) for the SAF Analysis Set.

11 Previous and Concomitant Therapies

Previous and concomitant medications are coded according to the World Health Organization drug dictionary (WHO-Drug version B3 March 2020) and stored with Anatomical Therapeutic Class (ATC) codes and generic names.

Therapies will be classified as previous if the stop date and/or time was before the date of first dose of study medication. All other medications are defined as concomitant. Missing or partly missing stop dates will be imputed using the rules defined in Section 7.3.

The number and frequency of previous and concomitant medications will be given per ATC level 2. If a patient has received more than 1 drug within an ATC class, he/she will be counted only once for this ATC class.

Previous and Concomitant Therapies will be listed overall, by Analysis Sets, and by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

12 Determination of Optimal Dose

The optimal dose will be the lowest possible dose and toxicity profile that achieves adequate imaging performance: (1) DLT rate is 0 subjects for a 3-patient cohort (2) adequate fluorescence signal of nerves compared to background non-nerve tissue. A dot plot will be presented to display Cohorts A1 through A6. Cohorts A1 through A6 will be displayed on the x-axis, with a cluster of 3 to 6 dots for each cohort. The dose will be displayed on the y-axis. Dots will be open circles for subjects who do not experience a DLT. Solid red dots will be used for subjects who experience a DLT. A listing of the DLTs will also be provided.

Image intensity ratios will also be plotted with time from injection to skin incision on the x-axis, image intensity ratio on the y-axis, and a separate color for each dose level.

13 Efficacy

The data on the signal-to-background ratio for FL and WLR will be provided by Stanford University for the different regions of interest ROIs. All efficacy analyses will be completed on the FAS Analysis Set and will be repeated in PP analysis set if the latest is different from FAS Analysis Set. The analyses using signal to background ratio SBR for cohort A will only use images produced by Zeiss camera wihle both Zeiss and OnLume devices will be used for timing cohort B.

13.1 Timing, Dose, Size and Intensity

The efficacy endpoints pertaining to timing of administration, dose, and intensity (signal-to-background ratio SBR), along with the surgeon rating scale will be assessed throughout enrollment into Cohorts A1-A6 to determine the final dose for use in Cohorts B1 and B2. These efficacy endpoints will also be used to evaluate the effect of timing of administration of ALM-488 by comparison of Cohort B1 and B2. The following analyses are planned for these efficacy endpoints:

- For each Cohort, SBR will be averaged to get a single mean ratio per patient; then a paired t-test or Wilcoxon signed rank test will be used, as appropriate, to compare SBR between FL and WLR.
- SBR will be also listed by cohort type, and surgery type and device for both FL with WLR.
- Image intensity ratios will also be analyzed with repeated measures analysis of variance (RMANOVA) with timing, image, method (FL /WLR), and device (Zeiss/OnLume) as the repeated factors in the model.
- SBR for FL and WLR measurements will be plotted. Values to the right of the line will indicate that there is improved visualization with FL images and values to the left of the line will indicate that there is improved visualization with WLR images(Equal performance of FL versus WLR will be indicated by dashed line (slope = 1)].
- To further assess timing, Cohorts B1 and B2 will be compared. Image ratios will be averaged to get a single mean ratio per patient; then an unpaired t-test or Wilcoxon rank sum test will be used, as appropriate, separately for FL and WLR.

Paired t-tests will be used unless markedly non-normal data is observed including bimodal or other severe assumption violations; minor skewness is robust to the paired t-test and RMANOVA assumptions. If extreme skewness with outliers, or bimodal distributions are observed, then a Wilcoxon signed rank test will be used instead of a paired t-test and RMANOVA will be performed on the ranked data.

Similarly, unpaired t-tests will be used unless markedly non-normal data is observed including bimodal or other severe assumption violations; minor skewness is robust to the unpaired t-test assumptions. If extreme skewness with outliers, or bimodal distributions are observed, then a Wilcoxon rank sum test will be used instead.

The optimal cut-off points of SBR and the ability to detect the nerve will be studied using ROC (Receiver Operating Characteristics) curves analyses through logistic regression. In the logistic regression the outcome variable will be the categorized surgeon rating scale 1 : I do not see the nerve versus 4 to 5 : the nerve is visible. The Youden J index given by J= Sensitivity - (1 - Specificity) will be used to select the optimal predicted probability cut-off. It is the maximum vertical distance between ROC curve and diagonal line. This will allow to maximize the difference between True Positive and False Positive. All possible SBR values and the corresponding sensitivity and specificity will be obtained using proc logistic in SAS [1]. The optimal cut-off will be determined overall and by device and individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

Image intensity ratios will be plotted against nerve diameter, with ratios on the y-axis and diameters on the x-axis. Plots will be displayed overall and by device and individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

Image intensity ratios will be summarized separately for images with and without overlying tissue, overall and by device and individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

In the subset of images with overlying non-nerve tissue, depth will be measured via the amount of overlying tissue that there can be to still maintain the ratio of nerve to non-nerves. Depth will be summarized overall and by device and individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

13.2 Surgical Imaging Worksheet/ Rating Scale

Surgeon confidence will be captured via the surgeon rating scale (1=do not see nerve , 2=see nerve less well compared to WLR, 3=see nerve equally well compared to WLR, , 4=see nerve better compared to WLR, 5=see nerve much better compared to WLR) for each image. Surgeon confidence will be summarized overall and by device, individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2) as well surgery type.

Image intensity ratios will also be correlated with the surgeon rating scale using Pearson correlation coefficients unless markedly non-normal data is observed including bimodal or other severe assumption violations; otherwise Spearman correlation coefficients will be used. Correlation coefficients will be presented overall and by device and by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

A listing will be provided.

13.3 Clinical Benefit ALM-488 Questionnaire

Clinical benefit for intraoperative nerve visualization questionnaire will be listed by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2) for each patients and type of surgery.

13.4 Nerve Dissecting Time

For each surgery, the following yes/no question will be asked: Outside of a clinical trial setting with all the required extra documentation, would this technology have saved you operative nerve-dissecting time compared to WLR alone (yes/no)?

Counts and percentages will be summarized overall and by device and individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

14 Safety

All safety analyses will be completed on the SAF Analysis Set.

14.1 Adverse Events

All information obtained on AEs will be displayed by treatment and patient. The number and percentage of subjects with AEs will be tabulated by System Organ Class (SOC) and Preferred Term (PT) with a breakdown by treatment. A patient with multiple AEs events within a SOC is only counted once towards the total of this SOC.

AEs will be coded using MedDRA version 23.1 or higher and presented by SOC and PT.

The analysis will focus on the treatment-emergent AEs (TEAE), i.e., AEs which started or worsened on or after the study ALM-488 infusion and before the end of the study.

Number and frequencies of subjects with TEAEs as well as the number of events will be given by SOC and by PT within each SOC overall and by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2) for the following:

- Any AE
- AE type
- At least one TEAE
- Serious TEAEs
- TEAE by CTCAE
- Infusion reactions IR overall and by CTCAE
- Related TEAEs
- Adverse Device Effects
- All TEAEs by maximum intensity

The counts and percentages of subjects with TEAEs as well as number of events will be presented by SOC and by PT and by decreasing frequency for:

- All TEAEs
- IR
- Serious TEAEs.
- Related TEAEs
- Serious Related.
- TEAEs leading to study discontinuation (if occurring in more than 5 subjects, otherwise a listing will be sufficient)
- Adverse Device Effects
- All TEAEs by intensity.
- All related TEAEs by intensity
- All TEAEs by outcome

Additionally, number and frequencies of subjects with TEAEs and number of events will be presented by PT and by decreasing frequency for

- All TEAEs
- Serious TEAEs
- Related TEAEs.

Finally, a summary of Serious TEAEs and deaths by individual Cohorts and for all patients

Listings will be created for

- All AEs
- Serious AEs (SAEs)
- AEs leading to discontinuation of study medication
- AEs leading to discontinuation from study
- AEs leading to death

and will also include subjects screened but not infused with ALM-488.

If severity is missing, the event will not be included in the frequency tables presenting events by intensity. If relationship to study drug is missing, the event will be assessed as unrelated if it started before start of study medication; in all other cases it will be assumed to be related.

14.2 Vital Signs

The actual values and the change from baseline in vital signs will be summarized over time with n, mean, standard deviation, and median, minimum, and maximum.

Vital signs i.e. height, weight, systolic blood pressure, diastolic blood pressure, heart rate, and respiratory rate will be summarized descriptively over time by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

Furthermore, the number and percentage of subjects based on reference ranges will be described according to the derived categories (i.e. Low, Normal, High) for vital signs.

A listing of vital signs will be presented for all measurements

14.3 Clinical Laboratory evaluations

Hematology, Clinical Chemistry, and Coagulation, Urinalysis

All laboratory data will be listed by individual cohort, patient and time point and if normal ranges are available abnormalities will be flagged. Summary statistics will be provided by study dose and time point.

The Hematology parameters include: red blood cells, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, white blood cells, neutrophils, eosinophils, basophils, lymphocytes, monocytes, %neutrophils, %eosinophils, %basophils, %lymphocytes, %monocytes.

The clinical chemistry parameters include: alkaline phosphatase, AST/SGOT, ALT/SGPT, total bilirubin, calcium, phosphorus, blood urea nitrogen, creatinine, GFR, total protein, albumin, glucose, potassium, sodium, chloride, magnesium, bicarbonate, uric acid, lactate dehydrogenase.

Coagulation parameters include: PT, aPTT, INR.

The urinalysis analysis will include protein, blood, leukocytes, glucose, ketones and pH.

Quantitative data will be examined for trends using descriptive statistics (number of patients with data, number of patients with missing values, mean, standard deviation, median, minimum, and maximum) of actual values and changes from baseline to each visit/time point. This will be presented by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

Quantitative data based on reference ranges will be described according to the NCI-CTCAE version 5.0 grade at each visit over time and by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

Qualitative data will be described according to the categories (i.e. Normal; Abnormal, Not clinically significant; Abnormal, Clinically significant; Missing) and by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

Shift tables of quantitative data showing changes with respect to the normal range and clinical significance between baseline and endpoint post-baseline visits as well as the worst value at any post-baseline visit.

Shift tables of qualitative data showing changes between baseline and post-baseline visits.

Data from unplanned determinations, i.e. usually determinations where the investigator felt follow-up (FU) was necessary, will be included in the number and frequency counts of clinically significant values. They will also be included in the data listings.

A listing of laboratory values will be presented for all measurements and for abnormal findings, i.e. measurements outside of reference ranges. Laboratory values that are outside of reference ranges will be flagged in the data listings, along with corresponding reference ranges.

Shift tables showing changes with respect to toxicity grade according to NCI-CTCAE version 5.0 will be listed using the worst grade.

Laboratory values will be listed by timepoint (screening, prior to ALM-488 infusion (baseline), after ALM-488 infusion, Day 15, Day 28), overall and by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2) and for all patients.

Data from unplanned visits will not be summarized but will be included in the data listings.

14.4 Echocardiograms (ECGs)

Values of heart rate, PR, QRS, QT and QTcF, the interpretation and the clinical significance of the ECG will be summarized by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2) using descriptive statistics for 12-Lead ECG/ Lead II Rhythm Strip. In addition, QTcF will be summarized using categories as defined in ICH E14 (i.e., QTcF (ms): >450-480; >480-500; >=500).

The Bazett's Correction (QTcB) and Fridericia's Correction (QTcF) are derived as follows:

Bazett's Correction (QTcB) $$QTc_b = \frac{QT}{\sqrt{RR}}$$

Fridericia's Correction (QTcF) $QTc_f = \frac{QT}{\sqrt[3]{RR}}$

where: RR = RR-interval measured in seconds.

ECG results will be listed including clinically significant data for the 12-lead ECG.

14.5 Physical Examinations

Results from physical examinations will be summarized by visit/timepoint (screening, prior to ALM-488 infusion (baseline), after ALM-488 infusion, Day 15, Day 28), change from baseline by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

PE will be provided in a data listing by subject.

14.6 Blood alergy

Results from blood allergy parameters will be summarized by visit/timepoint (pre-infusion, end of infusion, thirty minutes after end of infusion and 22 hours after the end of infusion), by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

Blood alergy will be provided in a data listing by subject.

14.7 Pregnancy Tests

All available data will be listed by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

14.8 Serum Samples

All available data will be listed by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

14.9 Chest X-Ray/CT

All available data will be listed by individual Cohorts (i.e., Cohorts: A, A1, A2, A3, A4, A5, A6, B, B1, B2).

15 Pharmacokinetic Analyses (PK)

The PK analyses will be described in a separate PK SAP and will use the PP and PK analysis set.

16 Interim Analyses

An interim analysis is not planned for this study.

17 Statistical Analyses for Safety Monitoring

Whenever a cohort is completed or when needed based on safety and toxicity information, selected outputs from the final analysis will be created for safety monitoring by the IDMC. The content of these safety analyses is specified in a separate IDMC Charter.

18 Statistical Analyses for Clinical Trial Registries

The following additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements:

Both Serious Adverse Events and Non-Serious Adverse Events will be summarized by MedDRA preferred term.

- An adverse event is considered 'Serious' whether or not it is a TEAE
- An adverse event is considered in the 'Non-Serious' category if it is both a TEAE and is not serious

For each Serious AE and Non-Serious AE, for each term and treatment group, the following are provided:

- the number of subjects at risk of an event;
- the number of subjects who experienced each event term;
- the number of events experienced.

19 Software

The data will be analysed using SAS Version 9.4 or higher.

20 Tables, Listings, and Figures

20.1 Tables

Below is Table of Contents for the Tables.

20.2 Listings

Below is Table of Contents for the Listings.

20.3 Figures

Below is Table of Contents for the Figures.

21 References

1. Manoj Pandey (2016) ROC Curve: Making way for correct diagnosis SP11 - PharmaSUG 2016 Ephicacy Lifescience Analytics Pvt. Ltd., Bangalore, India Abhinav Jain, Ephicacy Consulting Group Inc., Iselin, NJ https://www.pharmasug.org/proceedings/2016/SP/PharmaSUG-2016-SP11.pdf
2. Hassan T. et al. (2019): Vital signs in adults. National quality improvement project, national report 2018/19. The Royal College of Emergency Medicine.

22 Appendices

Appendix I: Normal ranges for vital signs, oxygen saturation and body temperature

Parameter	LLN	ULN
Blood pressure (diastolic)	50 mmHg	90 mmHg
Blood pressure (systolic)	90 mmHg	140 mmHg
Pulse rate	60 bpm	100 bpm
Respiratory rate	10 bpm	20 bpm
Oxygen saturation	92 %	100 %
Body temperature	35.0 °C	38.0 °C

LLN = lower limit of normal, ULN = upper limit of normal, bpm = beats per minute.

LLNs and ULNs for pulse rate, respiratory rate, oxygen saturation and body temperature are taken from [2].