Sponsor: CEL-SCI Corporation

Study title: A Phase III, Open-label, Randomised, Multi-center Study of the Effects of Leucocyte Interleukin, Injection [Multikine] Plus Standard of Care (Surgery + Radiotherapy or Surgery + Concurrent Chemoradiotherapy) in Subjects with Advanced Primary Squamous Cell Carcinoma of the Oral Cavity / Soft Palate Versus Standard of Care Only

Study number: CS001P3

ICON Study Number 3140/0001

SAP version: Final: 14Feb2014

Amendment 1: 27Nov2020

Yaping Cai

Yaping Cai
27 Nov 2020 14:31:032+0000

REASON: I approve this document

5847fe6f-1cee-4c47-ac55-feff592129f2

Signature Date

Blinded Biostatistician:

Yaping Cai
Principal Biostatistician
ICON Clinical Research

Sponsor's Representatives:

Eyal Talor

Eyal Talor
27 Nov 2020 15:09:022+0000

REASON: I approve this document

5a7c1b19-989b-4b38-8dd9-45225e5fa404

Signature Date

Eyal Talor, PhD
Chief Scientific Officer
CEL-SCI Corporation

John Cipriano

John Cipriano
27 Nov 2020 15:29:050+0000

REASON: I approve this document

1ab20719-df7a-48a9-9eff-f62f73358e01

Signature Date

John Cipriano

Senior Vice President - Regulatory Affairs

CEL-SCI Corporation

Confidentiality statement:

- The information provided in this document is strictly confidential.
- The Recipients of the SAP must not disclose the confidential document and any related information to other persons.
- In addition, they must keep this confidential document in a controlled environment which prevents unauthorized access to the document.
- This document and information herein is the confidential property of CEL-SCI Corporation, and may not be duplicated, reproduced, transmitted or disclosed to any third party except as authorized by CEL-SCI Corporation

DocUUID : 6ad6803a-25bd-4f41-a70e-51e4a1baec22

TABLE OF CONTENTS

Abbreviations

AE	adverse event
BDRM	Biostatistics Data Review Meeting
BSA	Body surface area
CIZ	Cyclophosphamide, Indomethacin, Zinc
CR	complete response
CRTx	Chemoradiotherapy
CTCAE	Common Toxicity Criteria for Adverse Events
DCFs	data clarification forms
DDT	disease-directed therapy
DM	data management
DMC	Data Monitoring Committee
DSUR	Data Safety Update Report
eCRF	electronic Case Report Form
EGFR	Epidermal growth factor receptor
eITT	Evaluable ITT
EMA	European Medicines Agency
EORTC 30	European Organization for Research and Treatment of Cancer QLQ (Questionnaire)
EORTC H&N 35	European Organization for Research and Treatment of Cancer Head and Neck QLQ (Questionnaire)
EOS	End Of Study
ERC	Evaluability Review Committee
FDA	Food and Drug Administration
FFLP-LRC	Freedom from Local Progression
GCP	Good Clinical Practice
GEE	generalized estimating equations
GLP	Good Laboratory Practice
GMP	Good Manufacturing Practice
H&N	Head and Neck
HIPAA	Health Insurance Portability and Accountability Act of 1996
HNC	head and neck cancer
IC	informed consent
ICH	International Conference on Harmonization
IEC	Independent Evaluation Committee
IND	Investigational New Drug Application
IRB	Institutional Review Board
ITT	Intent to Treat

IU	International unit
IV	Intravenous
IVRS	Interactive Voice Response System
KPS	Karnofsky performance status
LAFU	Last Alive Follow Up
LCLLC	Lavin Consulting LLC
LD	Longest Diameter
LRC	Loco-regional control
MCAR	missing completely at random
MK	Multikine
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
OS	overall survival
OTR	Overall tumor response
pCR	Pathology complete response
PD	progressive disease
PFS	progression-free survival
PHI	Protected Health Information
PR	partial response
QOL	Quality of life
QRL	Query Resolution Log
RANTES	An 8kD Protein belonging to the PF4 (Platelet Activating Factor 4) Super Family of chemo-attractants, attracting CD4+, CD45RO+ T-cells and Monocytes at the inflammatory site
RECIST	Response Evaluation Criteria in Solid Tumors
RTx	Radiotherapy
RTOG	Radiation Therapy Oncology Group
SAE	serious adverse event
SAP	Statistical Analysis Plan
SCC	squamous cell carcinoma
SCCHN	squamous cell carcinoma of the head and neck
SD	stable disease
SDV	source data verification
SEER	Surveillance, Epidemiology, and End Results
SOC	standard of care
SOPs	standard operating procedures
SSP	Study Specific Procedure
TEAE	Treatment emergent adverse event
TCI	time-critical interval

TLF	table, listing, and figure
TNM	Tumor, nodes, metastases
TTP	Time To Progression
US	United States

1.0 Version Control and GCP/ICH Governance

This Statistical Analysis Plan (SAP) is based on the protocol amendment 3, version 4.0, dated 27 Jun 2014.

The statistical analysis will be conducted under the following ICON SOPs:

- BST002-SOP Evaluability of Subjects for Analysis
- BST004-SOP Statistical Analysis Plan
- BST005-SOP Statistical Review of Statistical Analysis packages.

and in compliance with the EU GDPR Law for all study subjects in the EU.

2.0 Objectives

2.1 *Primary Efficacy Objective*

The primary objective is to compare overall survival in the Multikine + CIZ + SOC group to that in the SOC alone group for superiority of the former.

2.2 *Secondary Efficacy Objectives*

The following secondary comparisons are also planned:

(1) Overall survival (OS) in Multikine + SOC vs. SOC
(2) Loco-regional Control (LRC) in Multikine + CIZ + SOC vs. SOC
(3) Progression Free Survival (PFS) in Multikine + CIZ + SOC vs. SOC
(4) QOL in Multikine + CIZ + SOC vs. SOC
(5) Examination of the histopathological nature of cellular tumor infiltration stimulated by Multikine injection vs. SOC (at the time of planned surgery completion)

2.3 *Tertiary Efficacy Objectives*

The tertiary objectives of the study include:

(1) Compare overall survival for Multikine + CIZ + SOC vs. Multikine + SOC
(2) Compare tumor response for Multikine + CIZ + SOC vs. SOC (at the time of planned surgery completion)

2.4 *Other Efficacy Objectives*

Other efficacy objectives of the study include:

(1) Compare OS controlling for pre-defined histopathology thresholds, stage, location, and treatment interactions

(2) Compare LRC controlling for pre-defined histopathology thresholds, stage, location, and treatment interactions

(3) Compare PFS controlling for pre-defined histopathology thresholds, stage, location, and treatment interactions.

2.5 *Safety Objective*

The safety objectives include:

(1) Compare the incidence of related treatment-emergent adverse events (TEAEs) incidence excess for each of the Mulitikine treated arms vs. SOC, alone and in combination

(2) Compare the incidence of related serious adverse events (SAEs) incidence excess for the Multikine treated arms vs. SOC, alone and in combination.

3.0 Study Design

This is a Phase III open-label, multi-center, randomized study to determine the efficacy and safety of peri-tumoral and peri-lymphatic injection of Multikine (400 IU as IL-2 equivalent / daily dose for 3 weeks 5 times per week) given prior to SOC. The following treatment groups were enrolled:

(1) Multikine + CIZ + SOC

(2) Multikine + SOC

(3) SOC alone.

3.1 *Randomization*

An Interactive Web Response System (IWRS) was used for randomization and stratification. Subjects were randomized in a 3:3:1 allocation with Multikine + SOC (without CIZ) being the treatment group to be assigned least often in contrast to Multikine + CIZ + SOC and SOC alone. Subjects were stratified by **country**, by **tumor location** (tongue (oral portion only – base of the tongue excluded), floor of the mouth, cheek and soft palate), and **tumor stage** (Stage III and Stage IVa (T4N0-2, T1-3N2))

	Primary Tumor			
Nodal Involvement	**T1**	**T2**	**T3**	**T4**
N0	**NE**	**NE**	Stage III	Stage Iva
N1	Stage III	Stage III	Stage III	Stage Iva
N2	Stage Iva	Stage Iva	Stage Iva	Stage Iva

NOTE: Patients with distal metastasis were not allowed into the study.

A dynamic randomization was used to promote balancing across study sites within a country and globally.

3.2 *Sample Size Rationale*

3.2.1 Historical Basis

Primary Endpoint

Current SOC for SCCHN adopted by the oncology community [includes: post-operative concurrent chemoradiotherapy (CRTx) or radiotherapy (RTx)] was based on trials conducted by the Radiation Therapy Oncology Group (RTOG - NEJM 2004; 350(19): 1937), and the European Organization for the Research and Treatment of Cancer (EORTC – NEJM, 2004; 350(19): 1945). Both trials demonstrated improvements in the 3-year Overall Survival rates: an absolute increase of 12 % (48% vs. 60%) for the EORTC study and an absolute increase of 10% (47% vs. 57%) for the RTOG study.

This study assumed a 55% 3-year overall survival rate for SOC alone. This assumes 70-80% of the subjects were at lower risk at the time of surgery. As for the EORTC and the RTOG studies, a 10% absolute gain in overall survival (OS) is regarded as being clinically meaningful. The primary study goal to test Multikine + CIZ + SOC superiority vs. SOC alone was to reject the 55% 3-year overall survival rate for SOC alone against a 65% 3-year overall rate for Multikine + CIZ + SOC. The Phase 3 study is being conducted as an event-driven study where death is the event.

3.2.2 Assumptions and Calculations

The primary comparison was based on 80% power and a two-sided 5% Type I error to detect a 10% absolute survival advantage at 3 years (55% vs. 65%). Assuming exponential hazards, this yields a hazard ratio of 0.721. For this comparison, the log rank test requires a total of 298 deaths. The trial will be conducted as an event-driven

trial and will conclude once a total 298 deaths in Multikine + CIZ + SOC group and SOC alone group have been documented. The deaths in the Multikine + SOC arm are not included in the 298 deaths count.

Since death certificates or death registry information should be available for virtually all subjects, the calculations assume no losses to follow-up for overall survival. A 24-month (total) recruitment period and a 30-month follow-up period yields a sample size of 336 subjects in each of the Multikine + CIZ + SOC and SOC alone groups. Under a 3:3:1 randomization, this yields 112 subjects in the Multikine + SOC group for a total estimate of 784 evaluable subjects.

A total of 928 subjects were enrolled in the Phase 3 Study, globally.

4.0 Endpoint Timing Considerations

This study has the following time-critical intervals (TCIs) of interest:

A. Study entry to completion of study-planned surgery
B. Immediately after study-planned surgery including the completion of stage-specific protocol-directed therapy (RTx or CRTx) to and including Month 36
C. Immediately after study-planned surgery including the completion of stage-specific protocol-directed therapy (RTx or CRTx) to Study Exit
D. Immediately after Month 36 to Study Exit
E. Study Entry (date randomized) to Study Exit (date of last follow-up).

There will be efficacy and safety endpoint specific analyses reflecting these time-critical intervals.

5.0 Study Hypotheses

For the primary and secondary efficacy measures, a two-sided p-value of 0.05 or less will be considered to be statistically significant in comparing the Multikine treatments vs. SOC alone for superiority; for each pre-planned model involving OS, LRC, and PFS, contrast tests will be used make pairwise comparisons for all three treatment groups will be simultaneously compared without penalty. A Holm closed-sequential procedure (as described by Ye[1]) will be used to control the probability of Type I error for the secondary hypotheses. The primary efficacy hypothesis is tested first to be followed by the secondary efficacy hypotheses presented below.

5.1 *Primary Hypothesis*

OS for Multikine + CIZ+SOC vs. SOC controlling for pre-planned covariates

The primary null (H_0) and alternative hypotheses (H_a) are:

H_0: $h_{Multikine + CIZ + SOC} / h_{SOC} = 1$ vs.

H_a: $h_{Multikine + CIZ + SOC} / h_{SOC} < 1$,

where $h_{Multikine + CIZ + SOC}$ is the active (Multikine + CIZ + SOC) hazard rate and

h_{SOC} is the control (SOC) hazard rate for **overall survival time**.

TCI E (see Section 4.0) is the main interval of interest.

5.2 *Secondary Hypotheses*

The following secondary comparisons are also planned:

(1) OS for Multikine + SOC vs. SOC controlling for pre-planned covariates

H_0: $h_{Multikine + SOC} / h_{SOC} = 1$ vs.

H_a: $h_{Multikine + SOC} / h_{SOC} < 1$,

where $h_{Multikine + SOC}$ is the active (Multikine + SOC) hazard rate and

h_{SOC} is the control (SOC) hazard rate for **overall survival time**.

TCI E (see Section 4.0) is the main interval of interest.

(2) LRC for Multikine + CIZ + SOC vs. SOC controlling for pre-planned covariates

H_0: $h_{Multikine + CIZ+ SOC} / h_{SOC} = 1$ vs.

H_a: $h_{Multikine + SOC} / h_{SOC} < 1$,

where $h_{Multikine + CIZ+ SOC}$ is the active (Multikine + CIZ + SOC) hazard rate and

h_{SOC} is the control (SOC) hazard rate for **progression free survival time**.

TCI E (see Section 4.0) is the main interval of interest.

(3) PFS for Multikine + CIZ + SOC vs. SOC controlling for pre-planned covariates

H_0: $h_{Multikine + CIZ+ SOC} / h_{SOC} = 1$ vs.

H_a: $h_{Multikine + SOC} / h_{SOC} < 1$,

where $h_{Multikine + CIZ+ SOC}$ is the active (Multikine + CIZ + SOC) hazard rate and

h_{SOC} is the control (SOC) hazard rate for **LRC** (Loco-regional control) **failure time**.

TCI E (see Section 4.0) is the main interval of interest.

(4) QOL for Multikine + CIZ + SOC vs. SOC

H_0: $u_{Multikine + CIZ + SOC} = u_{SOC}$ vs.

H_a: $u_{Multikine + CIZ + SOC} > u_{SOC}$

where $u_{Multikine + CIZ + SOC}$ is the active (Multikine + CIZ + SOC) mean and

u_{SOC} is the control (SOC) mean for **QOL change from baseline for function scales and global score**.

H_0: $u_{Multikine + CIZ + SOC} = u_{SOC}$ vs.

H_a: $u_{Multikine + CIZ + SOC} < u_{SOC}$

where $u_{Multikine + CIZ + SOC}$ is the active (Multikine + CIZ + SOC) mean and

u_{SOC} is the control (SOC) mean for **QOL change from baseline for symptom scales**. Note a high score for a symptom scale/item represents a worse outcome.

TCIs A and B (see Section 4.0) will be of interest with TCI A being of primary interest.

(5) Histopathology differences following Multikine + CIZ + SOC vs. SOC

Separate hypotheses of distribution (D) shift will be tested for each histopathology measure.

H_0: $D_{Multikine + CIZ + SOC} = D_{SOC}$ vs.

H_a: $D_{Multikine + CIZ + SOC} \neq D_{SOC}$

where $D_{Multikine + CIZ + SOC}$ is the active (Multikine + CIZ + SOC) distribution and

D_{SOC} is the control (SOC) distribution for **each histopathology marker**.

The distribution shift is the noteworthy outcome.

TCI A (see Section 4.0) is the main interval of interest.

5.3 *Tertiary Hypotheses*

The following tertiary comparisons are planned:

(1) OS for Multikine + CIZ + SOC vs. Multikine + SOC controlling for pre-planned covariates

H_0: $h_{\text{Multikine + CIZ + SOC}}$ / $h_{\text{Multikine + SOC}}$ = 1 vs.

H_a: $h_{\text{Multikine + CIZ + SOC}}$ / $h_{\text{Multikine + SOC}}$ < 1,

where $h_{\text{Multikine + CIZ + SOC}}$ is the Multikine + CIZ + SOC hazard rate and

$h_{\text{Multikine + soc}}$ is the Multikine + SOC hazard rate for **overall survival time**.

TCI E (see Section 4.0) is the main interval of interest.

(2) Tumor response controlling for pre-planned covariates

Overall tumor response (OTR) is defined as the percent CR or PR per population; as determined by RECIST Criteria 1.0; overall response will be evaluated at study-directed surgery (following randomized therapy) and then after completion of subsequent RTx, or CRTx.

H_0: $OTR_{\text{Multikine + CIZ + SOC}}$ = OTR_{SOC} vs.

H_a: $OTR_{\text{Multikine + CIZ + SOC}}$ > OTR_{SOC}

where $OTR_{\text{Multikine + CIZ + SOC}}$ is the active (Multikine + CIZ + SOC) response rate and OTR_{SOC} is the control (SOC) response rate.

As per protocol Section 7.4, the determination of OTR per subject will reflect a blinded assessment of histopathology samples to determine patient-specific OTR given that the tumor may appear larger resulting from immune cell infiltration into the tumor at the interval prior to study planned surgery.

TCI A is of interest for study therapy and TCI B is of interest for disease-directed therapy (DDT).

5.4 *Other Hypotheses*

(1) OS for Multikine + SOC vs. SOC controlling for pre-planned covariates including histopathology markers

H_0: $h_{\text{Multikine + SOC}}$ / h_{SOC} = 1 vs.

H_a: $h_{\text{Multikine + SOC}}$ / h_{SOC} < 1,

where $h_{\text{Multikine + SOC}}$ is the active (Multikine + SOC) hazard rate and

h_{SOC} is the control (SOC) hazard rate for **OS**.

TCI E (see Section 4.0) is the main interval of interest.

(2) LRC for Multikine + CIZ + SOC vs. SOC controlling for pre-planned covariates including histopathology markers

H_0: $h_{Multikine + CIZ+ SOC} / h_{SOC} = 1$ vs.

H_a: $h_{Multikine + SOC} / h_{SOC} < 1$,

where $h_{Multikine + CIZ+ SOC}$ is the active (Multikine + CIZ + SOC) hazard rate and

h_{SOC} is the control (SOC) hazard rate for **PFS**.

TCI E (see Section 4.0) is the main interval of interest.

(3) PFS for Multikine + CIZ + SOC vs. SOC controlling for pre-planned covariates including histopathology markers

H_0: $h_{Multikine + CIZ+ SOC} / h_{SOC} = 1$ vs.

H_a: $h_{Multikine + SOC} / h_{SOC} < 1$,

where $h_{Multikine + CIZ+ SOC}$ is the active (Multikine + CIZ + SOC) hazard rate and

h_{SOC} is the control (SOC) hazard rate for **LRC**.

TCI E (see Section 4.0) is the main interval of interest.

5.5 *Safety Hypothesis*

No formal hypotheses are proposed, but the goal is to rule out a 10% higher related TEAE incidence for both Multikine arms vs. SOC, alone and in combination, and to rule out a 10% higher related SAE incidence for both Multikine arms vs. SOC, alone and in combination.

See Section 12.2 for the intervals of interest.

6.0 Multiplicity

All p-values will be generated for primary, secondary, and tertiary endpoints. However, a Holm closed-sequential procedure will be used to control the probability of Type I error for the secondary hypotheses to be tested after the primary hypothesis with the exception of

the individual OS, LRC, and PFS hypotheses which will simultaneously compare all three treatment groups simultaneously using contrast statements. To support registration claims, the Holm closed-sequential procedure will be used for the secondary efficacy endpoints only after the primary endpoint OS is tested statistically significant at 0.05 level. Thus, these secondary hypothesis tests will not require further Holm testing adjustment. With Holm-Bonferroni method, we will order the p-values from smallest to largest for the four comparisons of the remaining secondary endpoints, and compare the ordered p-values with 0.05/(2-k+1), where 2 is the number of secondary endpoints of interest (LRC, PFS) and k is the kth comparison in the sorted sequence, and k ranges from 1 to 2.

The secondary endpoints of interest will be tested with p-values considered as meaningful if the primary endpoint is statistically significant (two-sided $p \leq 0.05$).

The Holm procedure will not apply to the histopathology markers or to the QOL measures. Thus, for the hypothesis tests on QOL and tertiary endpoints, no Type I error is spent, so the tests will only be considered as descriptive.

7.0 Endpoints

7.1 *Primary Efficacy Endpoint*

Overall Survival Time

OS is defined as the number of months from randomization to the date of documented death or date of last follow-up. TCI E (see Section 4.0) is the main interval of interest, but TCIs C and D are also of interest.

7.2 *Secondary Efficacy Endpoints*

Secondary variables are as follows:

Progression-Free Survival

PFS is defined as the number of months from randomization to the date of first documented, progressive disease (any tumor recurrence, any new disease above clavicle or distant metastases) or the date of last follow-up or death. TCI E (see Section 4.0) is the main interval of interest, but TCIs C and D are also of interest.

Loco-Regional Control

LRC is defined as the number of months from randomization to the date of documented local or regional failure (recurrence or progression) or date of last follow-up or death. LRC failure includes the reappearance (recurrence) of disease (at the original tumor sites), progressive disease (but not distant metastases), or any new disease (including new disease in lymph nodes), above the clavicle, not

present at baseline. This is the “traditional” RTOG measure of local-regional control, also referred to as Freedom from Local Progression (FFLP-LRC). TCI E (see Section 4.0) is the main interval of interest, but TCIs C and D are also of interest.

For each of the time to event endpoints, the censoring rules for patients are described in the following table.

Condition	Date of Censoring
Randomized and not treated	Date randomized
Randomized and treated with no follow-up tumor assessments	Earlier of Date of last treatment or surgery
No tumor assessments at baseline or day before surgery	Earlier of Date of last treatment or surgery
No LRC or progression	Latest of EOS, LAFU, and Last Contact Date
New anti-cancer treatment started before LRC or progression	Date of last tumor assessment before new anti-cancer treatment
LRC, progression, or death after $\geq$2 missed follow-up visits	Date of last non-PD assessment

Quality of Life

Quality of life is measured by the EORTC QLQ-C30 scales/items and EORTC QLQ-H&N35 scales/items.

The EORTC QLQ-C30 incorporates 30 items and consists of 5 functional scales (physical, role, cognitive, emotional, and social functioning), 3 general symptom scales (fatigue, pain, and nausea/vomiting), a global QOL scale, and 6 specific symptom scales (dyspnea, insomnia, appetite, constipation, diarrhea, and financial impact). See the protocol for further QOL specifics. The disease-specific EORTC QLQ H&N35 is a supplement module to the QLQ-C30 and consists of 35 questions covering aspects of head and neck cancer. Quality of Life (QOL) data will be scored according to the algorithm described in the EORTC QLQ-C30 scoring manual. Missing data points for multi-item scales will be imputed using the method described in the scoring manual where at least half the items from a given scale were answered. TCIs A (study therapy) and B (during RTx or CRTx) are of primary interest.

Histologically Confirmed Complete Response Immediately Following Surgery

Narratives will be written for all documented complete responders (CRs) achieved following the protocol-specified surgery to indicate and describe their tumor pathology response (pCR as available from surgical specimens) and the subsequent impact on disease-free status following surgery through Month 36 and through the end of the study. Complete responders will be documented by pathology and imaging as available. The disease-free margins will be noted where available. The narrative will also include tabulated information on the cellular infiltrate tests for Multikine injected subjects (by group) as well as those treated by SOC alone. A table will be generated for pathology findings for each study treatment group. TCI A (see Section 4.0) is the main interval of interest.

Subsequent Complete or Partial Response Following RTx or CRTx

Narratives will also be written for all documented complete (CR) or partial (PR) responses achieved following completion of RTx or CRTx to indicate and describe their clinical response (documented by imaging [as available] in the absence of biopsies or surgical specimens) and the subsequent impact on disease-free status through Month 36 and through the end of the study. Complete and partial responders will be documented by imaging as available. The narratives will also include reference to the study tumor response including previous histology findings. A table will be generated for response outcomes for CRTx and RTx stratified by each study treatment group. TCI B (see Section 4.0) is the main interval of interest.

7.3 *Tertiary Efficacy Endpoints*

Tumor response will be derived based on RECIST criteria V1.0. There were no non-target lesions assessed at screening on the CRF page, so we will assume all lesions (i.e. the primary tumor and any clinically involved lymph node(s)) measured at screening were target lesions. The longest diameters (LD) of lesions from primary tumor and any clinically involved lymph node(s) are summed.

Below is the response criteria for target lesions (per RECIST criteria V1.0):

Complete Response (CR)	Disappearance of all target lesions
Partial Response (PR)	At least a 30% decrease in the sum of LD of target lesions, taking as reference the baseline sum LD
Progressive Disease (PD)	At least a 20% increase in the sum of LD of target lesions, taking as reference the smallest sum LD recorded (this includes the baseline sum if that is the smallest on study) or the appearance of one or more new lesions
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD taking as reference the smallest sum LD since the treatment started

The criteria used for evaluation of overall response at each visit are in the following table.

Target Lesions	New Lesions	Overall Response
CR	No	CR
PR	No	PR
SD	No	SD
PD	Yes or No	PD
Any	Yes	PD

We will use screening tumor measurement to assess TCI A (see Section 4.0) (i.e., change between baseline and day before/on surgery). Following surgery, all subjects would have their tumor baseline reset (new baseline) provided the clinical and pathology reports indicate a complete removal (or having residual disease present post-surgery) of the tumor and any clinically involved lymph nodes. i.e. if a subject has their tumor(s) completely removed at surgery, the sum of LD will be assumed to be 0 (zero). For Long-Term Follow-up, the focus will be on sustaining the complete response as well as avoiding progression of any existing residual tumor (i.e., not completely removed at surgery) or the appearance of new tumor[s] (and any clinically involved lymph nodes) recorded on the New Lesions CRF page. Given tumors are measured/imaged “only if clinically indicated” during long-term

follow-up, only a small portion of subjects may have data at long-term follow-up visits. TCI A (see Section 4.0) is the main interval of interest.

For TCI B, the following percents will be displayed by CRTx and RTx and stratified by study treatment group: (1) having no progression (primary, distal, local), (2) having primary progression, (3) having distal progression, (4) having primary recurrence above the clavicle (LRC failure), and (5) having primary or distal progression (PFS failure).

7.4 *Safety*

Safety will focus on the number and percentage of subjects with at least one TEAE, related TEAE, SAE, related SAE, serious TEAE, serious related TEAE, TEAE leading to discontinuation of study treatment and TEAE leading to death will be presented by treatment group (each treatment group and the combined Multikine treated groups). Events related to study treatment will be summarized overall (i.e. related to any study treatment) and for each individual study treatment.

The above tables will be rerun after removing deaths, recurrences, and progressions. The rationale is that deaths, recurrence, and progression are each components of the major study efficacy endpoints (overall survival, progression-free survival, and loco-regional control) in this study. Thus, these are expected outcomes on this trial. Therefore, collecting these events as either AEs or SAEs would be considered as “double counting”, and thus inappropriate for safety reporting purposes. FDA has published this position in the New England Journal of Medicine[2].

Thus, separate tables will be provided to show all reported AEs/SAEs that include deaths/recurrences/progressions as well as filtered tables that do not include these events.

See Section 12.2 for intervals of interest.

8.0 Analysis Populations

8.1 *All Patients Population*

All patients who have attended a screening visit and signed written informed consent, including replaced subjects if any, will be included in the All Patients population. This population will be used for patient accountability summaries and data listings.

8.2 *Intent to Treat (ITT) Population*

All study subjects who are randomized, regardless of treatment and trial group (Multikine treatment or SOC), including any replaced subjects, if any, will be included in the ITT analysis.

The ITT population will be used for a supportive efficacy analysis.

8.3 *ITT Evaluable (eITT) Population*

All ITT study subjects who are declared to be evaluable by the blinded Independent Evaluation Committee (IEC); see the Evaluability Protocol. The six exclusion criteria to remove cases from the ITT population are as follows:

- Informed consent violations (informed consent not signed)
- Critical eligibility violations (ineligible stage)
- Randomization violation (incorrect treatment administration)
- Multikine treatment violations (failure to administer sufficient dosages)
- Surgery violations (surgery not performed)
- Critical radiotherapy or chemotherapy violations (inconsistent with the risk group).

The eITT population will serve as the primary efficacy population.

8.4 *Per Protocol Evaluable (ePP) Population*

The following criteria will be used to define the ePP population by excluding cases from the eITT population:

(a) Multikine + CIZ + SOC Arm (**Group 1**): Eligible subjects receiving 12 Multikine injections (administrations) as randomized, having completed surgery (as defined in the protocol) and receiving at least; **cyclophosphamide IV** (as defined in the protocol), **2 courses of cisplatin** (if as a result of surgical findings the subject is slated for the concurrent chemoradiotherapy sub-group treatment) and receiving 75% of indomethacin, 75% of scheduled radiation, and at least 75% of all other protocol required treatments.

(b) SOC Arm (**Group 3**): Eligible subjects having completed surgery (as defined in the protocol) as randomized and receiving at least **two courses of cisplatin** (if as a result of surgical findings the subject will receive the concurrent chemoradiotherapy as a sub-group treatment) and receiving at least 75% of all other protocol required treatments (e.g., 75% of all scheduled radiation).

(c) Multikine Treatment + SOC Arm (**Group 2**): Eligible subjects receiving 12 Multikine injections (administrations) as randomized but without administration of any CIZ components, having completed surgery (as defined in the protocol) and receiving at least 2 courses of cisplatin (if as a result of surgical findings the subject will receive the concurrent chemoradiotherapy as a sub-group treatment) and receiving at least 75% of all other protocol required treatments (e.g., 75% of all scheduled radiation).

The ePP population will be used for a supportive efficacy analysis. The analysis will have limited value if there are fewer than 298 deaths in the ePP population.

8.5 *Safety Population*

All subjects signing informed consent and having undergone study procedures (e.g., tumor biopsy to assess meeting study entry criteria) and/or receiving any study therapy/treatment following randomization - such as one or more injections of Multikine, any CIZ components, surgery, or any components of the standard of care (i.e., surgery or any radiotherapy and/or chemotherapy) will be included in the safety population.

9.0 Analysis Timing

Given that the ITT Evaluable population is a subset of the ITT population, the primary analyses will be conducted once the ITT Evaluable population contains at least 298 documented deaths in the database (eCRF/EDC) in the combined Multikine + CIZ + SOC arm and SOC alone arm have been attained; the deaths in the Multikine + SOC arm are not included in the required 298 deaths count which has already been reached in the ITT population. The ITT Evaluable (primary) and the ITT (supportive) populations will be used for registration purposes so the plan is to conduct this final analysis once the ITT Evaluable population accumulates at least 298 deaths in the combined Multikine + CIZ + SOC arm and SOC alone arm. To provide follow-up closure and a cushion against losing cases during the final blinded IEC review, the primary analyses will be based on all available follow-up data inclusive of June 30, 2020 when the last follow-up request was issued.

In addition, the analysis will be performed concurrently for the ePP population to avoid potential analysis influence from the ITT and eITT population analyses.

10.0 Patient Disposition and Baseline Information

10.1 *Patient Disposition*

All subjects entered in the study will be accounted for in the disposition summarization report. The subject disposition per randomized treatment group will include data on randomized treatment, actual treatment, site enrolment, subject eligibility, subject compliance, progress through the study, follow-up, discontinuation, data on overall qualification status of all subjects, and an account of all identified protocol violations. The number of subjects will be displayed for those who discontinue before treatment begins, do not qualify for the per protocol analyses, who progress during the study, who exhibit loco-regional recurrence, who die during the study, or who go on to other treatments.

The progress of subjects through each treatment regimen (Multikine +/-CIZ + SOC) and SOC alone (control group) and through study completion will be shown in lifetable output and time to event graphs; the number of subjects will also be displayed per randomized treatment group for each of the study populations. Reasons for early withdrawal from study will be displayed. The timing of withdrawals will also be displayed using a Kaplan-Meier lifetable and compared using an unstratified log-rank test. If a subject has reason of “death” or “disease progression” for withdrawal, then the subject will be censored at the time of withdrawal, not being considered as an event. The pattern of being missing completely at random (MCAR) will also be assessed. A logistic regression with early withdrawal as the response, treatment group indicator, tumor location (tongue, floor of mouth, cheek, and soft palate), tumor stage (Stage III and Stage IVa), and region as the covariates, will be performed to examine the pattern for missing data. Baseline characteristics of the subjects who withdrew early will be compared to those that did not withdrew early.

10.2 *Baseline and Demographic Characteristics*

Demographic and baseline disease characteristic data will be listed and summarized using descriptive statistics for continuous variables and tabulated for categorical variables. All summaries will be presented by treatment arm as well as overall for the ITT, eITT, ePP, and Safety populations.

Demographic characteristics for evaluation will include age, gender, race, ethnicity, weight, height and BSA. For baseline disease characteristic: details and status of the primary tumor at baseline will be summarized by frequency counts of each component. This will include physical aspects, biopsy information, location and staging.

Categorical data will be compared between treatment arms using a two-sided Fisher exact test, ordinal data will be compared using a Wilcoxon rank sum test, and continuous data (e.g., age and BSA) will be compared using an unpaired t-test.

11.0 Efficacy Analysis

All efficacy data will be analyzed in the ITT, eITT, and ePP populations. The ITT Evaluable population is the primary population of interest.

11.1 *Primary Efficacy Analysis*

The time to event analysis will be the primary focus for OS; deaths beyond 3 years will be included in the analyses. In the analyses for the ITT and eITT populations, deaths due to COVID-19 will be censored. The primary efficacy analysis will be performed using the unstratified log rank test (as specified in the protocol) with stratified log rank and proportional hazard models being supportive given FDA preference for predefined multivariate models. The proportional hazard model will compare both Multikine arms vs SOC as the referent. TCIs C-E (see Section 4.0) are the intervals of interest, with TCI E being primary.

The stratification is categorized by tumor location (tongue, floor of mouth [referent], cheek, and soft palate), tumor stage (Stage III and Stage IVa [referent]) and geographical region, respectively.

Countries will be grouped by the following geographical regions and to best reflect medical practices:

- North America/Europe (EU): US, Canada, UK, Spain, France, Italy, Poland, Croatia, Hungary, Romania
- Europe (Non-EU): Serbia, Bosnia, Turkey
- Europe/Eurasia: Ukraine, Belarus, Russia
- Asia/West-Asia **[referent]**: Sri Lanka, India, Israel
- Asia-Pacific/Far-East: Malaysia, Philippines, Thailand, Taiwan

Four series of **proportional hazards models** will be used to evaluate study treatment; the first series will evaluate study treatment as well as baseline covariates (**tumor stage, tumor location, and geography**) while the second series will add on-study treatment interactions to the previously cited covariates. The respective model will include interactions between treatments and tumor stage, tumor site, and geography. Treatment will be modeled with SOC as the reference group to allow simultaneous comparisons to Multikine + CIZ + SOC (primary comparison) and to Multikine + SOC (secondary comparison). The third series adds negative surgical margin after surgery to the first series while the fourth series adds negative surgical margin after surgery to the second series; the four series will be performed for TCIs C-E with TCI E primary.

Four additional series of **proportional hazards models** will be used to evaluate study treatment and subsequent DDT (CRTx; RTx); the first series will evaluate study

treatment as well as baseline covariates (**tumor stage, tumor location, DDT, and geography**) while the second series will add on-study treatment interactions to the previously cited covariates. The respective model will include interactions between treatments and tumor stage, tumor site, DDT, and geography. Treatment will be modeled with SOC as the reference group to allow simultaneous comparisons to Multikine + CIZ + SOC (primary comparison) and to Multikine + SOC (secondary comparison). The third series adds negative surgical margin after surgery to the first series while the fourth series adds negative surgical margin after surgery to the second series; the four series will also be performed for TCIs C-E with TCI E primary.

Subgroup Analysis

Additional time to event analyses will be performed for **the risk group assignment reflecting concurrent disease-directed therapy** (DDT; radiotherapy or concurrent chemoradiotherapy) following surgery - as noted in the protocol; this analysis will exclude subjects without surgery or subjects receiving subsequent disease-directed therapy other than that prescribed by the protocol following (or instead of) surgery. A major prognostic factor for these subjects (in addition to tumor stage at study entry) is the respective “high risk” or “low risk” assigned protocol treatment made **after** surgery. Thus, stratification at study entry for this factor cannot be performed. As described above, both unstratified and stratified log rank tests and proportional hazards models will be performed by risk group; the proportional hazards models will include tumor stage, tumor location, geographic location, and treatment.

In addition, a display of the baseline stage (Stage III, Stage IVa) against subsequent DDT (radiotherapy, concurrent chemoradiotherapy) will be generated for each treatment group. The choice of adjuvant therapy will depend upon tumor characteristics post-surgery as assessed by NCCN Guidelines to determine high/low risk administration of radio/chemotherapy. Per NCCN Guidelines[3], “High Risk” are patients with certain disease characteristics following surgery who are deemed by the investigator to be at a “Higher Risk” for recurrence and thus will receive concurrent chemoradiotherapy [CRTx] (instead of radiotherapy [RTx] only recommended (by the NCCN Guidelines and in the protocol) for the “lower risk” for recurrence patients - following surgery). The percents able to “switch” from the recommended treatment for Stage IVa or Stage III (at “higher risk”, i.e., CRTx) to receiving only radiotherapy (RTx) recommendation based on the pathology and clinical assessment following surgery (but not due to the general medical status which would not allow the administration of Chemoradiotherapy due to expectation of excessive toxicity or patients inability to tolerate CRTx) will be of primary interest for this tabulation. The baseline tumor location against subsequent DDT, e.g. radiotherapy, concurrent chemoradiotherapy, will also be generated for each treatment group by stage (Stage III, Stage IVa). In addition, the percent per treatment

group that only received RTx after surgery will be compared to those that received CRTx after surgery for each study treatment group.

To assess the benefit of study-directed therapy and subsequent DDT, additional time to event analyses (Kaplan-Meier lifetables and unstratified log rank tests) will be performed for the following subgroups: (1) negative surgical margin after surgery, and (2) complete or partial response following subsequent DDT. TCIs C-E will be the main intervals of interest with TCI E primary.

Last, prospective stratification will yield subgroup analyses that are planned as an exploratory analysis of the primary efficacy endpoint, OS, using tumor location (tongue, floor of mouth, cheek, and soft palate), and tumor stage (Stage III, Stage IVa) and geographic region. TCIs C-E will be the intervals of interest with TCI E primary.

11.2 *Secondary Efficacy Analysis*

As for OS, the primary efficacy analysis will be performed using the unstratified log rank test (as specified in the protocol) with stratified log rank and proportional hazard models being supportive given FDA preference for predefined multivariate models.

Four series of proportional hazards models will be used to evaluate study treatment for PFS and LRC; the first series will evaluate study treatment as well as baseline covariates (tumor stage, tumor location, and geography) while the second series will add on-study treatment interactions to the previously cited covariates. The respective model will include interactions between treatments and tumor stage, tumor site, and geography. Treatment will be modeled with SOC as the reference group to allow simultaneous comparisons to Multikine + CIZ + SOC (primary comparison) and to Multikine + SOC (secondary comparison). The third and fourth series adds negative surgical margin after surgery to the first two series; the four series will be performed for TCIs C-E with TCI E primary.

Four additional series of **proportional hazards models** will be used to evaluate study treatment and subsequent DDT (CRTx; RTx); the first series will evaluate study treatment as well as baseline covariates (**tumor stage, tumor location, DDT, and geography**) while the second series will add on-study treatment interactions to the previously cited covariates. The respective model will include interactions between treatments and tumor stage, tumor site, DDT, and geography. Treatment will be modeled with SOC as the reference group to allow simultaneous comparisons to Multikine + CIZ + SOC (primary comparison) and to Multikine + SOC (secondary comparison). The third and fourth series add negative surgical margin after surgery to the first two series; the four series will also be performed for TCIs C-E with TCI E primary.

Last, to assess the benefit of study-directed therapy and subsequent DDT on PFS and LRC in subgroups, additional time to event analyses (Kaplan-Meier lifetable and stratified and unstratified log rank test) will be performed for the following subgroups: (1) negative surgical margin after surgery, and (2) complete or partial response following subsequent DDT. TCIs C-E will be the intervals of interest with TCI E primary.

11.3 *Other Secondary Efficacy Analyses*

The following secondary comparisons are also planned:

(1) OS in Multikine + CIZ + SOC vs. SOC

(2) LRC in Multikine + CIZ + SOC vs. SOC

(3) PFS in Multikine + SOC vs. SOC

(4) QOL in Multikine + CIZ + SOC vs. SOC

(5) Histopathology differences regarding cellular tumor infiltration following Multikine injection vs. SOC.

OS, LRC, PFS

The Holm closed-sequential procedure will be used to control the overall 5% Type I error for the secondary efficacy endpoints (items 1-3 above) corresponding to the primary TCI designations with TCI E primary for OS, LRC, and PFS. If the primary endpoint OS reaches two-sided statistical significance, we will look at secondary endpoints with Holm closed-testing procedure (see Section 6). In addition, the two Multikine treatment groups will be compared for these three endpoints as part of the previous comparisons of each Multikine group vs SOC.

LRC and PFS will be analyzed in the same way as OS as designated in Section 11.2.

Quality of Life

QOL will be assessed for all subjects prior to randomization (to establish Baseline), and then prior to and following (completion of) Multikine administration up to 1 day prior to surgery, after RTx and CRTx are completed, and at nominal 6, 12, 18, and 36 month visits. The change in QOL from baseline within and between treatment groups will be assessed. In addition, the change in QOL from prior to surgery at each visit during long term follow up period will also be analyzed. Of interest, is improvement of swallowing and oral function relative to baseline. The comparisons between treatment groups will be performed using ANOVA while the change in QOL from baseline within treatment group will be performed using a paired t-test. Two-sided mixed model repeated measures will be used to test for treatment differences with no adjustment for the Type I error. In addition, a 10-point improvement between

treatments at each long-term follow-up visit will be performed by an exact binomial test for each treatment group and using a two-sided Fisher Exact test for between group comparisons. TCIs A-B will be run.

Changes in QOL will be assessed using a longitudinal growth model including all post-baseline measurements. Study treatment, time point, treatment and time point interaction, as well as baseline covariates (baseline score, tumor stage, tumor location, and geography) and subject random effect will be included in the model. Compound symmetry will be assumed as the covariance structure of the longitudinal data. If the model does not converge with subject as random effect, the random statement will be dropped from the model. TCIs A-B will be run.

Histopathology

Immunohistopathology and histopathology marker parameters (see below with low and high thresholds noted for PH regression analyses [see Section 11.5.1]) will be assessed for all available and evaluable pathology specimens (tumor-specific averages per case to be determined by the Central Pathology at the end of the study) will be displayed per study treatment group using a cumulative distribution to be compared using a Kolmogorov-Smirnoff test for each of the following markers, ratios, and differences:

1. CD1a DC (10, 50)
2. CD3 pan-Tcell (100, 1000)
3. CD4 T helper (100, 1000)
4. CD8 cytotoxic T (50, 500)
5. CD4/CD8 (1, 2.5)
6. CD20 B cell (100, 1000)
7. CD25 IL2R (100, 1000)
8. CD68 MPH (100, 1000)
9. CD163 M2-MPH (50, 100)
10. CD208 mature DC (50, 100)
11. FOXP3 Treg (100, 500)
12. CD8/FOXP3 Treg (1, 2)
13. p46 NK (10%, 50%)
14. p16-HPV+ (10%, 50%)
15. MPOX neutrophil (50, 100)
16. pan-HLA-I (20, 50)
17. B2M (20, 50)
18. PD1 (10, 20)
19. PDL1 (10, 20)
20. CTLA4 (10, 20)
21. MR1 (20, 50)
22. SCC (squamous cell carcinoma: Yes/No).

Any missing cutoffs will be provided prior to database lock. All central pathology assessments will be performed on blinded samples by pathologists. TCI A will be of primary interest. A third-party statistical vendor (LCLLC - Dr. Lavin) will perform the pathology analysis after the study blind is broken.

- Narratives will be written for all histologically confirmed complete responses determined at surgery following study-directed therapy to indicate and describe their tumor pathology response (pCR as available from surgical specimens and imaging) and disease-free status following surgery through Month 36 and to the end of the study; surgical disease-free margins will be noted where available. The narrative will also include tabulated information on the type of cellular infiltrates present in tumors collected from Multikine injected subjects (by group) as well as those treated by SOC alone.
- Narratives will be written for all subsequent complete or partial response following RTx or CRTx to indicate and describe their clinical response (in the absence of biopsies or surgical specimens and imaging) and the subsequent impact on disease-free status through Month 36 and through the end of the study. The narrative will also include reference to the previous histology findings (as available).

11.4 *Tertiary Efficacy Analysis*

- Tumor Response

The number of subjects and percentage for each category (CR, PR, SD, PD) of overall tumor response (as determined by RECIST Criteria 1.0) will be presented by baseline tumor stage, tumor location, treatment group and visit, will be compared between the Multikine + CIZ + SOC vs. SOC groups. Results will also be presented for the following baseline covariates: tumor stage; tumor location, and geographic region. TCIs A (MK effect) and B (DDT effect) will be run.

- OS for Multikine + CIZ + SOC vs. Multikine + SOC

The hazard rate for overall survival time will be compared between Multikine + CIZ + SOC and Multikine + SOC using proportional hazards model. TCIs C-E are the intervals of interest.

11.5 *Other Analyses*

11.5.1 Disease-directed Surgery

The ability of Multikine to enable disease-directed surgery to be performed. The percentage of patients undergoing disease-directed surgeries per treatment group will be compared. The surgery rates will be compared using a two-sided Fisher Exact test. TCI A (MK effect) will be of primary interest.

11.5.2 Karnofsky Performance Status

The number and percentage of subjects under each Karnofsky Performance Status (KPS) score will be summarized per treatment group. The change in KPS from baseline within and between treatment groups will be assessed over time. The comparisons between treatment groups will be performed using repeated measures while the change from baseline within treatment group will be performed using a paired *t*-test. The repeated measures model will include study treatment, time point, treatment and time point interaction, baseline KPS, baseline covariates (tumor stage, tumor location, and geography). Two-sided test will be used with no adjustment for the Type I error. TCIs A (MK effect) and B (DDT effect) will be run.

11.5.3 Other Histopathology Analyses

For each biomarker (including the pre-defined ratio and differences), proportional hazard models for OS, LRC, and PFS will be run first for just stage, location, lower biomarker cutoff, higher biomarker cutoff, and treatment as covariates; the models will be repeated by adding treatment interactions with stage, location, and the biomarker cutoffs. The goals will be to assess: (1) if any biomarkers have favorable prognosis and (2) if any biomarkers downgrade or upgrade the stage. The analyses will also be repeated for separate tumor locations (within the oral cavity).

A third-party statistical vendor (LCLLC - Dr. Lavin) will perform the pathology analysis after the study blind is broken.

11.5.4 Other Disease-directed Therapies

The impact of other oncology therapies on study participants (not specifically prescribed by the protocol) such as: Keytruda, Opdivo, anti-EGFR, and other/or approved treatments for H&N or those used “off-label” and which are administered post-surgery (and/or given post-randomization) on OS will be examined in the ITT and eITT populations. The use of CheckPoint inhibitors, chemotherapies, monoclonal antibodies, anti-epidermal growth factor receptor (EGFR) and targeted small molecules before as well as after progression will be analyzed using time-dependent proportional hazard models to address subjects initiating additional disease-directed therapy prior to reaching the respective study endpoint.

11.5.5 Meta-analyses

Additional analyses are planned to compare the CS001P3 study treatments to pre-defined historical controls from the literature for both safety and efficacy. Safety will

be compared for TEAEs while efficacy will be compared for 2-year, 3-year, and 5-year OS and 2-year LRC rates.

A completed meta-analysis [reference available upon request] has identified 5 published surgical control studies with therapies qualifying as SOC. These studies provide 2-year, 3-year, and 5-year OS and 2-year LRC estimates as well as safety outcomes to compare to the three CS001P3 Study treatment groups. The analysis methodology will check if the 2-year, 3-year, and 5-year OS and 2-year LRC lie within or below the corresponding Kaplan-Meier lifetable two-sided 95% confidence intervals.

This is relevant since a favorable outcome for Multikine treatment safety may justify a lower efficacy standard given the absence of any new breakthrough therapies having an OS impact for the advanced primary SCCHN patient population over the past 30+ years. This will become a BLA review matter depending on study outcomes not yet known.

Furthermore, the meta-analysis will allow comparison of the 2-year LRC, 3-year OS, 5-year OS, and overall related TEAE SAE for the three CS001P3 study treatment groups to the five surgical controls studies (in the literature) representing SOC. For each outcome, results are to be expressed as percent to permit a chi square test to be performed to compare the observed vs the expected percent in support of the benefit for each (and both) of the two Multikine treatment groups relative to the historical control.

A third-party statistical vendor (LCLLC - Dr. Lavin) performed the meta-analysis for the five literature controls without needing to break the study blind.

11.5.6 SEER analyses

To further assist in such considerations, projected OS outcomes have been performed for the SEER database [reference available upon request] to correspond to the specific CS001P3 population. This allows a comparison of the overall survival curves to assess the performance of both Multikine treatments and the SOC treatment relative to the literature control should Multikine treatment safety be more favorable than the literature control. The analysis methodology will check if the 2-year, 3-year, and 5-year OS and 2-year LRC lie within or below the corresponding Kaplan-Meier lifetable two-sided 95% confidence intervals.

This is relevant since a favorable outcome for Multikine treatment safety may justify a lower efficacy standard given the absence of any new breakthrough therapies having an OS impact for the advanced primary SCCHN patient population over the past 30+ years. This will become a BLA review matter depending on study outcomes not yet known.

The review of the SEER database 18 for 2000-2016 (n=6,641), excluding SEER data from 2000 to 2010 (n=3,291), in order to be able to “match” the timing of

accrual/treatment of subjects in the CS001P3 Study (using SEER Stat Section 8.3.5), has informed us that, in the USA (SEER data) in the 'Oral Cavity and Soft-Palate' matching CS001P3 patients' tumor location (excluding the base of the tongue as in the CS001P3 Study) the survival outcome for the combined Stage III and IVa patients (n=3,350). This SEER-database set, designed to approximate the CS001P3 study patient population, yields 36.75% ('Observed') 5-year OS and 46.59% ('Observed') 3-year OS for SOC only. Of note is that the CS001P3 Study protocol estimated (circa 2007 - 2010) that the 3-Year OS for SOC, alone, was 55% and was designed to show a 10% absolute increase in OS using the investigational (Multikine) treatment regimen + SOC, over that which can be attained with SOC alone (in an event-driven study design).

Table 6: SEER Annual Survival Data: Oral Cavity (including: Tongue [but not base of Tongue], Floor of the Mouth + Cheek) + Soft Palate (as in study CS001P3).

Product-Limit Survival Estimates					
Time (Months)	**Survival**	**Failure**	**Survival Standard Error**	**Number Failed**	**Number Left**
12	0.7142	0.2858	0.00785	948	2330
24	0.5386	0.4614	0.00894	1474	1372
36	***0.4659***	0.5341	0.00939	1638	839
48	0.4198	0.5802	0.01000	1707	498
60	***0.3675***	0.6325	0.0116	1751	0

The above SEER analyses were already conducted independently by a third-party statistical vendor (LCLLC - Dr. Lavin) without needing to break the study blind.

12.0 Safety Analysis

Safety will be assessed by means of adverse events, laboratory data, and vital signs data. All safety data will be summarized by each treatment group for the safety population; separately, the two Multikine treatment arms will also be pooled in all safety displays. Safety analysis will also be analyzed by timing of the adverse event onset from randomization to surgery, from surgery to end of SOC and post SOC (also see Sections 2.5, 7.4 and 8.5). See Section 12.2 for the intervals of interest.

The major safety endpoints are the incidence of related TEAEs and related SAEs for each treatment group and for the combined Multikine treated groups.

Disease progression and death will not be considered to be adverse events as both are expected in this population and are endpoints in the study. Disease progression

and death adverse events will be separately tabulated, displayed, and compared between the three study treatment groups. Separate listings will be generated for all SAEs (including progression) as well as for another listing that removes progressions from the list of all SAEs.

12.1 *Study Drug Exposure and Follow-up*

The number of treatment visits at which Multikine is administered and the amount of each study drug delivered per visit and cumulatively will be summarized and tabulated. Subjects requiring delay and/or reduction in dose of either, study drug or radiotherapy and/or concurrent chemoradiotherapy will be displayed with reasons provided for dose reduction.

The mean follow-up time from randomization to study exit will also be presented for each treatment group.

12.2 *Adverse Events*

Results will be presented separately for each treatment group and for combined Multikine treated groups. All adverse events except disease progression will be reported regardless of causality. Severity of AE will be graded by the investigator according to the NCI CTCAE version 4.0. The adverse events will be reviewed and coded by sponsor or sponsor's representative using the latest MedDRA dictionary, Version 20.1, or higher. Coding will include system organ class (SOC) and preferred term (PT). AEs will be recorded in the electronic Case Report Form (eCRF) starting from the time that a subject signs the Informed Consent Form, through the three-year follow-up period, and to the end of the study.

Each AE will be classified as either a pre-treatment AE or a treatment-emergent AE (TEAE). Any AE which occurred before the date of randomization in the study will be considered as a pre-treatment AE. Any AE which started on or after the date of randomization in the study will be considered as a TEAE. AEs with partial or missing start dates will be imputed with the rules outlined in Section 14.0. TEAE classification will based on the imputed start dates.

TEAEs with relationship to study treatment recorded in the electronic CRF as "possibly related", "probably related", "related" or "definitely related" will be defined as related TEAEs.

Pre-treatment AEs will not be included in the summary tables but will be flagged in the data listings.

An overall summary of the number and percentage of subjects with at least one TEAE, related TEAE, serious TEAE, serious related TEAE, TEAE leading to discontinuation of study treatment and TEAE leading to death will be presented by treatment group (each treatment group and the combined Multikine treated groups).

Events related to study treatment will be summarized overall (i.e. related to any study treatment) and for each individual study treatment.

Summaries of the number and percentage of subjects with TEAEs by study treatment group will also be presented by system organ class and preferred term for:

- All TEAEs;
- Related TEAEs;
- Serious TEAEs;
- Serious related TEAEs;
- TEAEs leading to discontinuation of study treatment;
- TEAEs leading to death.

In addition, summary of the number and percentage of subjects with TEAEs as well as the number of events under each system organ class, and preferred term will be presented by maximum CTC grade and relationship to study treatment ("not related", "possibly related", "probably related", "related" or "definitely related") respectively.

Displays will also be provided for the timing of adverse events to provide insight into the Multikine-related adverse events distinct from adverse events related to surgery and SOC. The classification of timing will based on the imputed AE start dates (see Section 14.0 of this document for imputation rules). The four intervals of interest for the timing of the AEs will be:

(1) post-randomization AEs (start date of AE is on or after the date of randomization; these are the TEAEs in this study)

(2) post-randomization/pre-surgery AEs (start date of AE is on or after the date of randomization and before the date of surgery),

(3) post-surgery during SOC AEs (start date of AE is on or after the date of surgery and start date of AE is on or before 60 days post the last date of radiotherapy/chemoradiotherapy) and

(4) Post-SOC AEs (start date of AE is after 60 days post the last date of radiotherapy/ chemoradiotherapy).

The overall summary of the number and percentage of subjects with at least one TEAE, related TEAE, serious TEAE, serious related TEAE, TEAE leading to discontinuation of study treatment and TEAE leading to death will be presented by treatment group according to the timing of adverse events as detailed above. Events related to study treatment will be summarized overall (i.e. related to any study treatment) and for each individual study treatment.

Summaries of the number and percentage of subjects with TEAEs will be presented by system organ class and preferred term for each time-period as determined above.

In the summaries of AEs, if severity or relationship to study treatment is missing, the AE will be regarded as ‘severe’ or ‘related’ to the study treatment, respectively.

All verbatim descriptions and coded terms will be listed for all AEs. One listing will be produced for all AEs (pre-treatment and TEAEs) and a separate listing will also be produced for all serious AEs (SAEs).

A bar chart showing the percentage of patients with at least one TEAE for each treatment group (each treatment group and the combined Multikine treated groups) and for each time-period will be presented. Additional figures may be requested as appropriate.

12.3 *Clinical Laboratory Test Results*

Laboratory tests will be performed by local laboratories operating under full GLP standards. The laboratory test battery will include routine laboratory tests (hematology, chemistry, and urinalysis) and immunohistochemistry/pathology.

The laboratory evaluations performed for the safety analysis by treatment group will include:

Hematology:

White blood cell (WBC) count, red blood cell (RBC) count, Hemoglobin, Hematocrit, Platelet Count, Neutrophils, Bands, Lymphocytes, Monocytes, Eosinophils, Basophils and ESR.

Biochemistry:

TSH, sodium, potassium, chloride, bicarbonate, glucose, urea, urine, creatinine, urate, calcium, phosphate, total protein, urine, albumin, total bilirubin, material, alanine transaminase, aspartate transaminase, alkaline phosphatase, lactate dehydrogenase, cholesterol, triglycerides, glutamyl transferase.

The clinical laboratory parameters (hematology and biochemistry) will be summarized by treatment group and visit using descriptive statistics for continuous data only. In addition, the change from baseline to each visit will also be summarized.

The number and percentage of subjects, who have high, low, normal laboratory values compared to the reference ranges and clinically significant laboratory values, as determined by the investigators, will be summarized treatment group (each treatment group and the combined Multikine treated groups) and visit. Shift from baseline at each visit will be presented as well.

If more than one assessment occurred at any visit (i.e. repeat samples taken), the last valid (non-missing) value will be used in the summaries. Unscheduled laboratory data will be listed but will not be included in the summary tables.

12.4 *Vital Signs*

Summary statistics will be presented by treatment group (each treatment group and the combined Multikine treated groups) for each vital sign parameter: systolic and diastolic blood pressure, heart rate, and respiration rate and body temperature. In addition, the change from baseline to each visit will be presented.

The number and percentage of subjects, who have high, low, normal vital signs values compared to the reference ranges and significant change as specified in the following table, will be summarized by treatment group and visit.

		Clinical Importance Criteria			
Vital Signs	**Unit**	**Low**	**High**	**Decrease**	**Increase**
Systolic Blood Pressure	mmHg	< 85	> 160	≥20 and ≥40	≥20 and ≥40
Diastolic Blood Pressure	mmHg	< 45	> 100	≥10 and ≥20	≥10 and ≥20
Heart Rate	Bpm	< 50	> 110	≥15 and ≥30	≥15 and ≥30

12.5 *Concomitant Medication*

All medications will be coded using the latest World Health Organization Drug Dictionary (WHODRL DEC2010, or later versions) and Anatomical Therapeutic Chemical (ATC) classification system. Coding will include the drug class and drug name.

All medications will be classified as either prior or concomitant as follows:

- Prior medication will be defined as non-study medication with a stop date before the first dose of study treatment;
- Concomitant medication will be defined as non-study medication with:
    - Start or stop date after the date of randomization;
    - Start dates prior to randomization but which were ongoing after date of randomization (concomitant at baseline);
    - Partial start dates that indicate that the medication could be concomitant in relation to the date of randomization;
    - Completely missing start dates, unless their stop dates confirm otherwise (i.e. the stop date is before the date of randomization).

Prior and concomitant medication data will be presented in data listings.

13.0 Geographic Region Analysis

Geographic region as defined in Section 11 will also be assessed for regional consistency or advantage for each primary and secondary efficacy endpoint. a treatment-region interaction will be tested; in the event of significant treatment-region interactions ($p<0.05$), region imbalance will be addressed using proportional hazards analysis (for time to event endpoints) or longitudinal model analysis (QOL domains, Karnofsky status); see Section 11 for the specific model covariates including the tumor stage, tumor location, treatment, region, and treatment-region interactions. If there are no significant treatment-region interactions (two-sided $p>0.05$), the respective models will not include the region-treatment interactions in all proportional hazards models, time-dependent models, and longitudinal models; baseline covariates will include the tumor stage, tumor location, treatment, region. Last, the overall AE incidence rate will be analyzed by region with Cochran–Mantel–Haenszel (CMH) method[4,5].

14.0 Missing Data Conventions

Any unresolved missing time to event data will be imputed in advance prior to database lock at the Biostatistics Data Review Meeting (BDRM). If only the day is missing, then the 15th of the month will be used. If both the month and day are missing, then the midpoint of that year (July 1) will be used. If all are missing, then a decision will be made at the BDRM meeting based on the available facts; for example, if the a subject's death date is unknown, then the date that the site first learned of the death will be used as the surrogate date. Also see Section 7.2 for other censoring rules to be applied for all time to event endpoints. The same rules will be in place for the primary and secondary efficacy endpoints.

There will be no imputation for missing data other than the QOL scoring according to the QOL endpoint scoring methodology.

Longitudinal models for QOL and Karnofsky will use all available data without the need to impute any missing data.

Similarly, as for missing time to event data, any unresolved missing adverse event dates will be imputed in advance prior to database lock at the BDRM. Missing AE dates will be imputed relative to the informed consent date as well as study and SOC treatment dates. First principles of imputation will be applied; for an AE with only day missing, if the year and month of AE onset is the same as informed consent date, then the AE date will be imputed with the informed consent date; otherwise, set AE date to the 1st day of that month. For an AE with both month and day missing, if the year of AE onset is the same as informed consent date, the AE date will be imputed

with the informed consent date; if the year of AE onset is after informed consent year, set AE date to the January 1 of that year. If the AE onset date is completely missing, set the date to the informed consent date. The BDRM will be used to discuss the details of imputation.

Other considerations will also be taken into account for imputation. For example, the start date of any adverse event should not begin earlier than the treatment start date and, if that date is unknown, not earlier than the informed consent date. If the end date is missing, then the end date could be the next follow-up visit or contact time when the adverse event was known to have ended. Missing month and day could be imputed to be the earliest possible date after the later of the randomization date and the treatment start date.

The BDRM will utilize such adverse event timing logic relative to the informed consent date, the treatments start-dates (study therapy as well as SOC), and the following timing windows: when such data are available:

- Informed Consent to Surgery,
- Post-surgery to RTx or CRTx, and
- Post RTx or CRTx (i.e., after 30-60 days of last SOC treatment administration which is the first day of long term follow up).

For treatment for oral carcinoma recurrence, if only the start day is missing, then the 15th of the month will be used. If both day and month are missing, then July 1 of that year will be used. For complete missing date, a decision may be made at the BDRM meeting based on the available facts.

15.0 Interim Analyses

Interim analyses are to be performed through the study to assess Multikine safety, sample size assumptions, and futility as determined by the study IDMC (Independent Data [and Safety] Monitoring Committee).

The first interim analysis was performed (for the IDMC) when 40 Multikine treated subject's complete surgery, and an additional safety analysis performed when all subjects have completed the 30-day safety follow-up visit (Day 28-35).

Subsequent interim analyses are performed (for the IDMC) as mutually agreed upon with the IDMC; the schedule is to perform interim analyses of safety and efficacy approximately every 6 months based on the study meeting enrollment objectives.

The IDMC reviews these semi-annual safety and efficacy analyses and annual DSUR reports during the study to assess the feasibility of continuing the study and to assess the sample size according to the primary study endpoint to see if the sample size is adequate or requires modification (reduction or increase). This sample size re-estimation is conducted on blinded data, prior to any unblinding. The study

will not be stopped for superiority, but the study may be stopped for futility (without penalty). Thus, the overall significance level of the final analysis (two-sided p≤0.05) is maintained.

The blinded sample size adjustment follows the method discussed in Todd[6]. The method re-estimates the enrollment rate at the time of interim and the pooled survival rate. The survival rate at time that has not been observed in the interim data is adjusted from the assumed survival rate during the planning of trial. The adjustment is based on average difference between the observed survival rate and the assumed survival rate on the complementary log-log scale across different event time points or at the latest time point. Enrollment period and thus number of subjects needed can be estimated to reach the targeted number of deaths (d=298) in the combined comparator groups of the study.

The IDMC can recommend that the trial may stop for futility (non-binding) if the conditional power based on the current trend is lower than 20%. The conditional power is calculated by the following as per Cui[7]:

$C_p(\theta) = \Phi(\frac{\sqrt{t}*LR_1+\theta(1-t)-z_\alpha}{\sqrt{1-t}}),\ t = \frac{d_1}{d_2'}$

where Φ is the normal cumulative distribution function, and $\theta = -0.5 * \sqrt{d_2} * \log(HR)$, HR is the hazard ratio (Multikine + CIZ + SOC vs. SOC alone) future data follow, LR_1 is the standardized log-rank statistics from the accumulated data up to the interim analysis time point, d_1 is the cumulated number of deaths in both groups at the interim analysis. $d_2 = 298$ is the expected number of deaths in the end of trial in both groups.

Three conditional powers will be calculated assuming the future data follow,

1) The null hypothesis

$H_0: \theta = 0$

2) The alternative hypothesis

$H_\alpha: \theta = -0.5 * \sqrt{d_2} * \log(HR) = -0.5 * \sqrt{298} * \log(0.721) = 2.82$

3) The current trend:

$\theta = LR_{1\_obs}/\sqrt{t}$

The confidence interval for the conditional power under the current trend C_{p3} will be calculated assuming the observed interim data follow the current trend

Since $C_{p3} = \Phi(\frac{LR_1/\sqrt{t}-z_\alpha}{\sqrt{1-t}})$, $t = \frac{d_1}{d_2}$, and $\frac{LR_1/\sqrt{t}-z_\alpha}{\sqrt{1-t}} \sim N(\frac{\theta_1-z_\alpha}{\sqrt{1-t}}, \frac{t}{1-t})$, thus p percentile of the conditional power can be calculated by $\Phi(\sqrt{\frac{t}{1-t}}\Phi^{-1}(p) + \frac{\theta_1-z_\alpha}{\sqrt{1-t}})$,

and thus 95% CI can be produced by calculating its 2.5% and 97.5% percentile.

16.0 COVID-19 Study Risk Analysis and Mitigation Strategy

16.1 *Background*

FDA [8] and other regulatory agencies [9] have issued guidance for the conduct of clinical trials during the COVID-19 pandemic. Challenges have emerged regarding to the following factors for this study to be addressed as noted below:

- general or partial country quarantines
- site access restrictions or closures
- availability of site medical personnel and/or site staff for continuing study activities or study subjects becoming infected with/or die from COVID-19
- subject availability due to hardships or moving to safe environments
- study monitor travel restrictions.

The above items lead to the following difficulties for this study:

- acquiring subject follow-up
- resolving data querie
- verifying source data required to assure the integrity of trial information used for determining safety and efficacy.
- completing site monitoring.

16.2 *Overall Study Analysis Strategy*

In 2017, the ERC was established to review all matters regarding evaluability. Any potentially unevaluable cases will be referred to the ERC to determine possible eITT exclusion. All matters impacting evaluability will be referred to the ERC. The ERC will complete the same forms as was followed in 2017 when they first reviewed evaluability data. The BDRM will review all unresolved queries and will assess any possible major protocol deviations to determine possible ePP exclusions.

<u>BDRM Roles and Responsibilities</u>:

The BDRM will review all unresolved queries and review any possible major protocol deviations to identify ineligible cases. The BDRM will also review the Query Resolution Log. A Study Specific Procedure (SSP) will be issued by the responsible CRO.

Major protocol deviations will be identified and forwarded to the BDRM by the responsible CRO. Major protocol deviations include the following:

- Consent deviations
- Ineligibility
- SAE reporting delays
- Missed surgeries
- Study treatment variations.

The BDRM will review all remaining unresolved queries due to the following reasons:

- Site no longer operational
- Site does not answer the query prior to database lock
- Site has lost contact with the study subject.

The BDRM will not impute any missing, unknown, or inconsistent data in accordance with GCP. The BDRM will review all unresolved queries and confirm the corresponding Query Resolution Log (QRL) itemizing the impact of any unresolved query.

Unresolved Query Review Procedures:

The BDRM will assess all unresolved queries. They will consider data issues arising from open queries unable to be resolved by sites. The general classification process is described below.

The situations are as follows:

- Data are inconsistent. Data may or may not be present to correct the query.
- Data are missing.

In both situations:

- No data will be changed unless the site answers a specific query.
- No attempts will be made to record data, impute outcomes, or resolve inconsistencies.
- Do not enter NA in the absence of information.

- When an answer can be inferred, the most likely correct answers per query will be recorded in the QRL. If applicable, any possible impact will be noted on study endpoint conclusions, e.g. patient narrative, safety analysis, shift table, OS lifetable.

Examples for decision rules:

- A concomitant medication is administered but no adverse event is recorded: No imputation will be made.
- An adverse event is mislinked to a concomitant medication: Do not correct the error. Use existing eCRF information to complete the QRL.
- An adverse event end date is missing: NA will not be entered; data will be considered as “blank/unknown”. Record likely date range in the QRL.
- The cause of death is unknown: Do not try to guess cause of death if unavailable, leave as unknown cause of death.
- The result is unknown: Do not try to guess any missing data field.
- A Baseline CT scan indicates longest dimension >3 cm: Do not change TNM classification. Record correct TNM classification in the QRL.
- The lab units are incorrect: Review other lab data to make a reasoned judgement to avoid outliers / Standard ranges might be applied for MISSING ranges. Wrong ranges or wrong results may not be able to correct. Record likely correction in the QRL.
- A tumor response calculation error is found: Recompute the tumor arithmetic to determine tumor response using the data (measurements) in the eCRF entered by the site. Record likely correction in the QRL.
- An incorrect eligibility date is discovered: Correct a date inconsistency based on existing eCRF if information is available to make decision. Record likely correction in the QRL.
- The eCRF contains a later survival date than appears in the LAFU form: Scan the eCRF for the latest survival date. Record likely correction in the QRL.

<u>Imputation Strategies:</u>

These considerations impact GCP since data queries are required to finalize safety and efficacy determinations inclusive of events and dates. Prospective rules need

to be in place to define data convention and analysis approaches to address impact areas.

In the absence of site-specific availability, the following rules will apply using existing data:

- Efficacy events only known to occur between two possible dates can be viewed as occurring at the most conservative outcome, at the subject-level
- Efficacy analyses will be conducted once all queries are addressed.
- Safety events only known to start or finish between two possible dates can be viewed as occurring at the earliest possible date, which give worst case scenarios at the subject-level
- Safety analyses will be conducted once all queries are addressed.

In general, a worst-case assumption using existing data will be made per impacted endpoint.

Analysis Strategies:

The following analysis strategies will be followed:

- Eligibility: Any unconfirmed eligibility will result in inclusion in eITT and ePP populations
- Time to Event Outcomes: A worst case assumption will be made; the event must be documented through death confirmation (OS) or by biopsy, imaging, or clinical (PFS, LRC)
- Response Outcomes: A worst case assumption will be made regarding event timing; histopathology is required to establish pCR.

This strategy is intended to under-estimate efficacy while maintaining GCP compliance and minimizing risks to trial integrity.

Cause of death will also be obtained (as available) inclusive of COVID-19 death; subjects dying from COVID-19 would have otherwise lived so progression or death due to COVID-19 should be censored at the time of COVID-19 death.

COVID-19 Impact:

The COVID-19 impacted clinical study report (CSR) sections will further indicate:

1. A risk analysis as specified in the FDA Guidance [7] (specifically: Sections C.1-3 of the Guidance).
2. Contingency measures implemented to manage study conduct during disruption of the study as a result of COVID-19 control measures.

3. A listing of all subjects affected by the COVID-19 related study disruption by unique subject number identifier per investigational site, and a description of how the subject's participation was altered.
4. As described above, conventions, contingency analyses, and corresponding discussions that address the impact of implemented contingency measures (e.g., alternative procedures used to collect critical safety and/or efficacy data) on the safety and efficacy results reported for the study.

Risk Analysis Summary:

A risk analysis follows in response to FDA Guidance Section C.1-3:

A. Unresolved queries
    - Background:
        - All open queries are resolved prior to database lock or confirmed to be left as is
        - A Biostatistics Data Review Meeting (BDRM) is held to judge and classify any remaining open queries
    - COVID-19 Impact:
        - Sites may not allow staff to have access to study records to answer queries
        - Study coordinators may not be able to return to work either due to personal circumstances or limitation on the number of staff resources and availability at specific sites
    - Consequences:
        - New study coordinators or clinical staff may need to be trained
        - Site coordinators/staff may not have access to study records
        - Site coordinators/staff may not have access to study patients for survival updates
        - Sites may not be able to answer queries in a timely manner
    - Solutions:
        - Remote Monitoring
        - Remote query resolution (site dependent)
        - Source Data Verification via interview and video conference while maintaining patient confidentiality and adherence to HIPAA and

GDPR (or equivalent country-specific) rules at the same level as a site visit by CRAs for SDVs.

B. Unverified source data

- Background:
  - FDA expects sponsors to comply with preset source data verification (SDV) standards
  - CEL-SCI is committed to 100% SDV for all study data including queries but this may not be possible
- COVID-19 Impact:
  - Study monitors may not be able to travel to sites
  - Site coordinators may not be able to host study monitors to allow SDV to be completed
  - Site coordinators may not have access to patient medical records
- Consequences:
  - New study monitors may need to be trained
  - Delays are likely to be incurred until site coordinators resume full-time activities
  - Delays are likely until travel bans are lifted
  - Study monitors may be backed up to visit sites without adding resources
- Solutions:
  - Remote Monitoring
  - Remote query resolution (as possible – Site dependent)
  - Remove Source Data Verification via interview and video conference (see above)

C. Data analyses

- Background:
  - The SAP is signed prior to database lock
  - The SAP includes analysis methods and table, listing, and figure (TLF) shells
- COVID-19 Impact:
  - FDA recently issued guidelines for Sponsors to consider

        - Sponsor response must be study-specific, e.g., COVID-19 deaths are considered to be censored at the time of death and not to be counted towards n=298 required number of deaths for the study
        - Sponsor needs to set data rules for unresolved queries, e.g. which ones can be resolved, which ones cannot be resolved and do not have material impact on the study safety or efficacy (or must await resolution)
        - Sponsor strategy must be prospectively implemented before database lock
    - Consequences:
        - COVID-19 diagnoses and deaths need to be determined
        - COVID-19 data rules need to be defined and implemented
        - COVID-19 impacted subjects need to be queried
        - COVID-19 TLFs need to be prospectively defined, e.g. run for subset with no remaining queries, run using an applied data convention established prospectively
        - COVID-19 TLFs may need to be run in multiple ways
        - FDA may ask Sponsor to rerun analyses once all queries can be resolved.
    - Solutions
        - Subjects with COVID-19 diagnoses will experience adverse events which must be reported per 21 CFR 312 standards
        - Any deaths attributed to COVID-19 will not be counted towards the final death count; instead, such subjects will be censored at the time of COVID-19 death for all time to event endpoints
        - Efficacy events only known to occur between two possible dates can be viewed as occurring at the most conservative outcomes, at the subject-level
        - Safety events only known to start or finish between two possible dates can be viewed as occurring at the earliest possible date, which gives a worst-case scenario at the subject-level.

17.0 Changes to Planned Methodology

- ITT Evaluable (eITT) and PP Evaluable populations were added to make the population nesting more efficient. The PP population was dropped since it made sense to remove both unevaluable and ineligible patients at the same time. The ITT Evaluable (eITT) population will be used for the primary efficacy analysis.
- Designated TCIs were added to identify the study critical intervals of interest.
- Added the percent undergoing surgery following study treatment to see if study treatment led to a higher percent able to undergo surgery.
- Added analyses for Karnofsky Performance Status as another QOL dimension.
- Added OS, PFS, LRC analyses controlling for pre-planned covariates including histopathology markers
- Conservative imputation rules were imposed to accommodate missing data.
- Logistic and Poisson regression models have been dropped from safety analyses.
- Site pooling analyses was dropped in favor of analyses of pre-defined geographic regions.

18.0 Data Reporting Conventions

In the presentation of descriptive summary statistics, the minimum and maximum will be presented to the same number of decimal places as the variable being reported; the mean and median will be to +1 decimal place and the SD to +2 decimal places.

18.1 *Mock-up Template*

Refer to the TLF mockup document containing all proposed tables, listings, and figures.

19.0 References

1. Ye Y, Li, A, Liu L, Yao B. (2012) A group sequential Holm procedure with multiple primary endpoints Statist Med (wileyonlinelibrary.com) DOI: 10.1002/sim.5700.
2. Sherman RB, Woodcock J, Norden J, Grandinetti C, Temple RJ. New FDA Regulation to Improve Safety Reporting in Clinical Trials N Engl J Med 2011; 365:3-5.
3. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines): Head and Neck Cancers (2020). Version 1.2020. nccn.org
4. Cochran WG (1954) Some methods for strengthening the common χ^2 Tests". Biometrics. 10 (4): 417–451. doi:10.2307/3001616. JSTOR 3001616.
5. Mantel N, Haenszel W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute. 22 (4): 719–748. doi:10.1093/jnci/22.4.719. PMID 13655060.
6. Todd S, Valdés-Márquez E, West J. (2002). A practical comparison of blinded methods for sample size reviews in survival data clinical trials. Pharm Stat. 2012 Mar-Apr; 11(2):141-8.
7. Cui L, Hung HM, Wang SJ. (1999). Modification of Sample Size in Group Sequential Clinical trials. Biometrics 55:853-85.
8. FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic Guidance for Industry, Investigators, and Institutional Review Boards, March 2020.
9. EMA (EU/ICH) Guidance for medicine developers and companies on COVID-19. V.3, April/May 2020.

DocUUID : 6ad6803a-25bd-4f41-a70e-51e4a1baec22