STATISTICAL ANALYSIS PLAN

VERSION: 5.0

DATE OF PLAN:

01 November, 2021

STUDY DRUG:

Nelatimotide and Adegramotide (Ombipepimut-S)

(hereafter referred to as DSP-7888) Dosing Emulsion

PROTOCOL NUMBER:

Version 1.0 29MAR2017
Protocol Amendment 1.0 (Version 2.0) 30 Apr 2017,
Protocol Amendment 2.0 (Version 3.0) 18 Sep 2017,
Protocol Amendment 3.0 (Version 4.0 for South Korea Only) 06 Oct 2017,
Protocol Amendment 4.0 (Version 5.0 for South Korea Only) 13 Oct 2017,
Protocol Amendment 5.0 (Version 6.0) 03 Aug 2018,
Protocol Amendment 6.0 (Version 7.0) 24 Jun 2019,
Protocol Amendment 7.0 (Version 8.0) 05 Feb 2021

STUDY TITLE:

A Randomized, Multicenter, Adaptive Phase 3 Study of DSP-7888 Dosing Emulsion in Combination with Bevacizumab versus Bevacizumab Alone in Patients with Recurrent or Progressive Glioblastoma following Initial Therapy

SPONSOR:

Sumitomo Dainippon Pharma Oncology Inc. (SDP Oncology)

640 Memorial Drive

Cambridge, MA 02139, USA

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

SIGNATURE PAGE

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. REVISION HISTORY

Version #	Date	Revision Summary
0.1	31 Jul 2017	Initial Release Version. Based on Protocol Amendment 1
0.2	23 Oct 2017	Update with sponsor comments and with changes from Protocol Amendment 2.0
0.3	27 Aug 2018	Update with sponsor comments and with changes from Protocol Amendments 3.0, 4.0 and 5.0
0.4	30 Sep 2018	Update with sponsor comments
1.0	06 Dec 2018	Update with sponsor comments
2.0	09 Jan 2020	Add the IA/FA efficacy boundary adjustment Add CTL, WTL, HLA*A genotype correlation analysis Add tipping point sensitivity analysis Update the OS/PFS censoring rule
3.0	05 Feb 2021	Update the statistical design to an adaptive Phase 3 design Designate 12-month OS rate as the key secondary endpoint
4.0	29 September 2021	Clarification of conditional power (Section 4.3.4) Updated per protocol analysis set definition to include per protocol analysis set for PFS (Section 6.3) Clarify the baseline definitions for efficacy and safety parameters (Section 8.2) Update missing data imputation rules in Section 8.4 for missing or partial death date, post GBM treatment date and concomitant medication date. Update PFS definitions and censoring rules (Section 8.4.4) Update baseline height definition (Section 9.2) Update best overall response hierarchical order, reasons for NE, and definitions of ORR / DCR (Section 10.3.2) Clarify the biomarker analysis are exploratory endpoints and the analysis are exploratory in nature and removed biomarker analysis from SAP (Section 11) Update dose intensity definition (Section 12.1) Add the analysis for ISR (Section 12.3) Update lab CTCAE grading and summary (Section 12.4) Add the markedly abnormal ranges for vital signs (Section 12.5) Update the summary of ECG (Section 12.6)

Version #	Date	Revision Summary
5.0	01 November 2021	Update PFS Derivation Rule (Section 8.4.4) to included preliminary PD followed by confirmed PD is an event. Update the test statistic of 12 month OS rate difference in the case that the log-log transformation is not appropriate (Section 10.2.1) .

Table 1: List of Abbreviation

Abbreviation	Description
AE	Adverse Event
ATC	Anatomical Therapeutic Chemical
BBI	Boston Biomedical Inc.
Bev	Bevacizumab
BMI	Body Mass Index
BOR	Best Overall Response
CI	Confidence Interval
CISH	Chromogenic in situ Hybridization
CP	Conditional Power
CR	Complete Response
CRF	Case Report Form
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
CTL	Cytotoxic T Lymphocyte
DCR	Disease Control Rate
DDS	Dose-Determining Set
DEC	Dose Evaluation Committee
DLT	Dose Limiting Toxicity
DSP	Sumitomo Dainippon Pharma
ECG	Electrocardiogram
FA	Final Analysis
FDA	Food and Drug Administration
GBM	Glioblastoma Multiforme
GTR	Gross Total Resection
HLA	Human Leukocyte Antigen
HR	Hazard Ratio
IA	Interim Analysis
ICH	International Conference on Harmonization
i.d.	Intradermally
IDH	Isocitrate Dehydrogenase
IDMC	Independent Data Monitoring Committee

Abbreviation	Description
IHC	Immunohistochemistry
ISR	Injection Site Reaction
ITT	Intent-to-Treat
i.v.	intravenously
KPS	Karnofsky Performance Status
KM	Kaplan-Meier
LVEF	Left Ventricular Ejection Fraction
MedDRA	Medical Dictionary for Regulatory Activities
MGMT	O-methyl-guanyl-methyl-transferase
MTD	Maximum Tolerated Dose
MUGA	Multigated Acquisition Scan
N/A	Not Applicable
NANO	Neurological Assessment in Neuro-Oncology
NCI	National Cancer Institute
OS	Overall Survival
ORR	Objective Response Rate
PD	Progressive Disease
PFS	Progression-Free Survival
PMDA	Pharmaceuticals and Medical Devices Agency
PP	Per-Protocol
PR	Partial Response
PT	Preferred Term
QTc	Corrected QT Interval
QTcF	Fridericia's correction formula (msec)
RANO	Response Assessment in Neuro-Oncology
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis Software
SD	Standard Deviation
SD	Stable Disease
SDP Oncology	Sumitomo Dainippon Pharma Oncology, Inc
SOC	System Organ Class

Abbreviation	Description
SS	Safety Set
TEAE	Treatment Emergent Adverse Event
USAN	United States Adopted Name
WHO	World Health Organization
WT1	Wilm's Tumor 1

2. PURPOSE

The purpose of this statistical analysis plan (SAP) is to ensure that the data listings, summary tables and figures which will be produced to support the completion of the Clinical Study Report (CSR), and the statistical methodologies that will be used, are complete and appropriate to allow valid conclusions regarding the study objectives according to International Council for Harmonization (ICH) E9, ICH E3 and ICH E6. This analysis plan is based on the latest version of both the protocol and case report forms. Any ad hoc or unplanned analyses performed and not identified in this SAP will be documented separately.

2.1. Timing of Analysis

This study is divided into 2 parts. Part 1 is a safety lead-in phase. Part 2 is a randomized, active-controlled, multicenter, open-label, parallel-group, adaptive Phase 3 study of DSP-7888 Dosing Emulsion plus Bev versus Bev alone in patients with recurrent or progressive GBM following treatment with first line therapy consisting of surgery and radiation with or without chemotherapy. An adaptive, phase 3 design with two interim analyses (IAs) and one final analysis (FA) is employed in Part 2 of this trial. More detail regarding the timing of the respective analyses can be found in Section 4.3 . Unless otherwise specified, the analysis includes all data collected in the database through the time of the data snapshot or database lock. Additional data collected after the database lock for final analysis of the study will be prepared as an addendum to the study report according to regulatory or scientific need.

DSP7888 with United States Adopted Name (USAN) of Ombipepimut-S will be use in the SAP. Ombipepimut-S will be used in tables, figures, listings.

3. STUDY OBJECTIVES, HYPOTHESES AND DECISION RULES FOR PART 2

3.1. Primary Objective

- To compare the overall survival (OS) between treatment with DSP-7888 Dosing Emulsion plus bevacizumab (Bev) versus Bev alone in patients with recurrent or progressive glioblastoma (GBM) following initial therapy.

3.2. Key Secondary Objective

- To compare the 12-month survival rate between treatment with DSP-7888 Dosing Emulsion plus Bev versus Bev alone.

3.3. Other Secondary Objectives

- To compare the progression-free survival (PFS) of DSP-7888 Dosing Emulsion plus Bev to Bev alone;
- To compare the 6-month PFS of DSP-7888 Dosing Emulsion plus Bev to Bev alone;
- To compare the response rate of DSP-7888 Dosing Emulsion plus Bev to Bev alone;
- To compare the duration of response of DSP-7888 Dosing Emulsion plus Bev to Bev alone;
- To describe the adverse event (AE) profile of DSP-7888 Dosing Emulsion plus Bev

3.4. Exploratory Objectives

- To describe the relationship between pharmacodynamic parameters such as Wilms' Tumor 1 (WT1) peptide-specific cytotoxic T lymphocyte (CTL) activity and certain clinical findings (e.g., OS, response) in patients receiving DSP-7888 Dosing Emulsion plus Bev;
- To describe the relationship between expression of WT1 by immunohistochemistry (IHC) and chromogenic in situ hybridization (CISH) in tumor tissue and certain clinical findings (e.g., OS, response) in patients receiving DSP-7888 Dosing Emulsion plus Bev;
- To collect diffusion-weighted imaging for future analysis of imaging findings and correlation with disease progression.
- To characterize the NANO scale score profile post DSP-7888 Dosing Emulsion plus Bev or Bev alone treatment.

4. STUDY DESIGN

4.1. Summary of Study Design

This study will be divided into 2 parts. Part 1 is a safety lead-in phase. Part 2 is a randomized, active-controlled, multicenter, open-label, parallel group, adaptive Phase 3 study of DSP-7888 Dosing Emulsion plus Bev versus Bev alone in patients with recurrent or progressive GBM following treatment with first line therapy consisting of surgery and radiation with or without chemotherapy.

Patients may have undergone a second surgery at the time of recurrence or progression but may not have received any other anti-neoplastic therapy. Patients are randomized 1:1 to treatment with DSP-7888 Dosing Emulsion plus Bev or Bev alone.

Potential study subjects first undergo pre-screening to determine if they express the appropriate human leukocyte antigen (HLA) type. The DSP-7888 Dosing Emulsion is to induce a CTL response against cancer cells in subjects who express HLA-A*02:01, HLA-A*02:06, or HLA-A*24:02.

Following the HLA determination, a screening phase of up to 28 days determines eligibility. Following verification of eligibility and enrollment, DSP-7888 Dosing Emulsion is administered intradermally (i.d.), based on the following schedule:

- Induction Phase: once every 7 days (± 1 day) for (Doses 1 to 5);
- Consolidation Phase: once every 14 days (± 3 days) for (Doses 6 to 15) except for Day 183;
- Maintenance Phase: once every 28 days (± 7 days) until discontinuation (Dose 16 and thereafter).

Bev is administered every 14 ± 3 days at the labeled dose of 10 mg/kg intravenously (i.v.).

4.1.1. Study Part 1 (Safety Lead-In)

In Part 1, 3 patients will be treated with the combination of DSP-7888 Dosing Emulsion plus Bev in a single-arm manner. DSP-7888 Dosing Emulsion will be administered at a dose of 10.5 mg i.d. per the schedule described above and Bev at the dose and schedule described above.

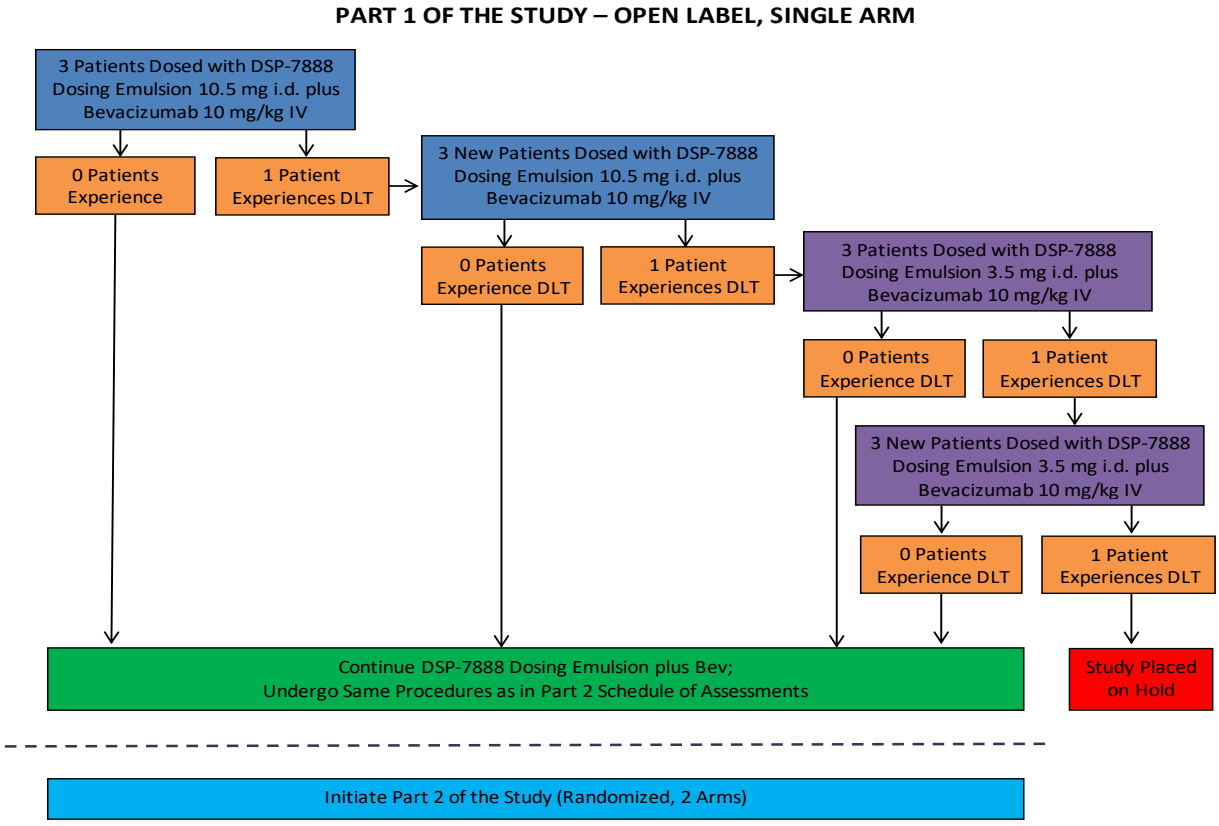
If 0 of these first 3 patients experiences dose-limiting toxicity (DLT), then the study will proceed to Part 2, which will randomize patients to the combination of DSP-7888 Dosing Emulsion plus Bev or to Bev alone.

If 1 of the first 3 patients experiences a DLT, then 3 more patients will be enrolled at the same treatment regimen. If no additional DLTs are observed at the same dose (with 1/6 patients experiencing a DLT), then the study will proceed to Part 2. However, if a total of 2 or more of the first 6 patients experience a DLT, then the dose of DSP-7888 Dosing Emulsion will be decreased to 3.5 mg i.d., and an additional 3 patients will be treated with the combination of DSP-7888 Dosing Emulsion plus Bev. If 0 of these 3 patients experiences DLT at this reduced dose of DSP-7888 Dosing Emulsion, then the study will proceed to Part 2 with DSP-7888 administered at 3.5 mg i.d.. If 1 of the 3 additional patients experiences a DLT at the reduced dose of DSP-7888 Dosing Emulsion, then 3 more patients will be enrolled at the same treatment

regimen. If no additional DLTs are observed at the reduced dose (with 1/6 patients experiencing a DLT), then the study will proceed to Part 2 with DSP-7888 administered at 3.5 mg i.d.. However, if a total of 2 patients or more experience a DLT at this reduced dose of DSP-7888 Dosing Emulsion plus Bev, then enrollment will be placed on hold until the sponsor decides how best to proceed.

In summary, if 0 of the first 3 patients or 1 of the first 6 patients experience a DLT, then the study will proceed to Part 2. The maximum tolerated dose (MTD) is determined only when 3, in the event of zero DLTs, or 6, in the event of 1 DLT, patients are treated at a dose level with ≤ 1/6 DLT. If at least 2/3 or 2/6 DLTs are observed, then the MTD has been exceeded. Dosing should stop, and the dose should be de-escalated to DSP-7888 3.5 mg. Then the process should be repeated at the reduced dose level of DSP-7888 3.5 mg. Part 1 of this study is presented in Figure 1 below.

Figure 1: Part 1 of the Study – Open Label, Single Arm



4.1.1.1. Dose Limiting Toxicity (DLT)

Dose-limiting toxicities will be ascertained during the DLT Evaluation Period for the patients who are enrolled in Part 1, which extends from the day of the first dose of study drug to completion of fifth dose at the assigned dose level (Days 1 through 29).

A dose limiting toxicity will be defined as any of the following AEs that are at least possibly related to the investigational agents per the Investigator and institutional guidelines:

- The occurrence of any Grade 3 or higher hematologic AE lasting more than 7 days;
- Any Grade 3 or higher non hematological AE that persists for more than 72 hours with the exceptions of transient nausea, vomiting, and diarrhea that are responding to supportive care;
- Any Grade 3 or higher allergic and/or hypersensitivity reactions;
- Any Grade 2 or higher autoimmune disease.

Patients who discontinue Part 1 of the study prior to the completion of the fourth dose for reasons other than DLT during the Induction Phase will be replaced.

The review of the study data to determine DLTs, increase the number of subjects in a given dose cohort, adjust the study dose due to safety concerns, as well as the final determination of the dose for Part 2 will be decided by a Dose Evaluation Committee (DEC). The DEC will consist of the study Medical Monitor, relevant clinical investigators of the study patients and the Syneos Health, LLC team and BBI (now SPDO).

4.1.2. Study Part 2

Based on the results of the DLT evaluation in Part 1, a 10.5-mg dose of DSP-7888 Dosing Emulsion was determined as the dose for Part 2.

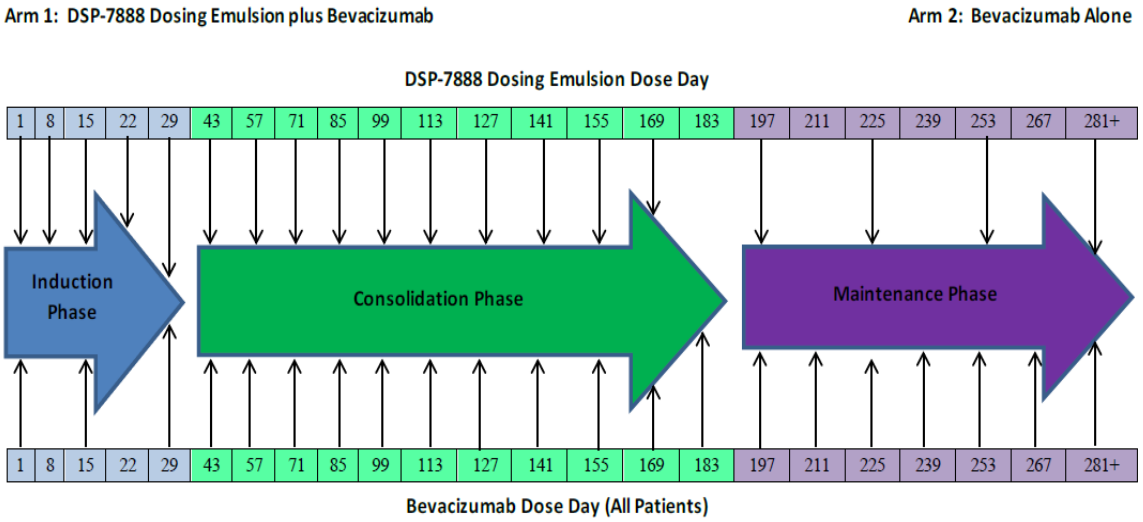
In Part 2, patients are stratified prior to randomization based on the extent of surgical resection in primary therapy (gross total resection [GTR] versus not GTR) and Karnofsky Performance Status (KPS) (See Protocol Appendix 16.1) scores (60 through 70 versus 80 through 100) and randomized to either:

a) DSP-7888 plus Bev, or
b) Bev alone

Patients randomized to the Bev arm of the study undergo evaluations on a similar schedule as patients randomized to treatment with the combination of DSP-7888 Dosing Emulsion plus Bev.

Part 2 of this study is presented in Figure 2 .

Figure 2: Part 2 of the Study – Randomized, Two Arms



4.1.3. Stopping Rules

Refer to protocol section 8.3.5.

4.2. Patient Selection and Study Duration

Between 3 and 12 patients were planned to be enrolled in Part 1 of the study (Section 4.1.1).

Part 2 of the study is an event-driven adaptive trial and its statistical power is determined by the target number of death events instead of the number of patients. For Part 2, if the study continues to the final analysis with approximately 260 death events (Section 4.3.3), approximately 338 patients with GBM will be enrolled from approximately 70 investigational sites in the United States (US), Canada, Japan, Taiwan, and South Korea. Part 2 of the study may be concluded sooner with early efficacy at IA2 (with approximately 185 death events from approximately 215 patients) or may be concluded at the final analysis with a re-estimated number of death events (between approximately 260 and 370 death events from approximately 338 and 480 patients). Note that Part 2 of the study is an event-driven adaptive trial and the statistical power of Part 2 is determined by the target number of death events instead of the number of patients. The actual number of randomized patients in Part 2 may be adjusted to observe the targeted number of events, depending on the outcome of the interim analyses and operational feasibility. The study may be terminated early for either efficacy or futility at IA1 and for efficacy only at IA2 or the study will continue until the target number of events is observed.

4.3. Hypotheses and Decision Rules in Part 2

4.3.1. Statistical Hypotheses

The primary study endpoint is OS in Part 2 of the study. The null and one-sided alternative statistical hypotheses in the terms of OS in Part 2 are:

H_0: Hazard Rate of DSP-7888 Dosing Emulsion plus Bevacizumab = Hazard Rate of Bevacizumab alone

H_a: Hazard Rate of DSP-7888 Dosing Emulsion plus Bevacizumab < Hazard Rate of Bevacizumab alone

The key secondary endpoint is 12-month OS rate in Part 2 of the study. The null and one-sided alternative statistical hypotheses are:

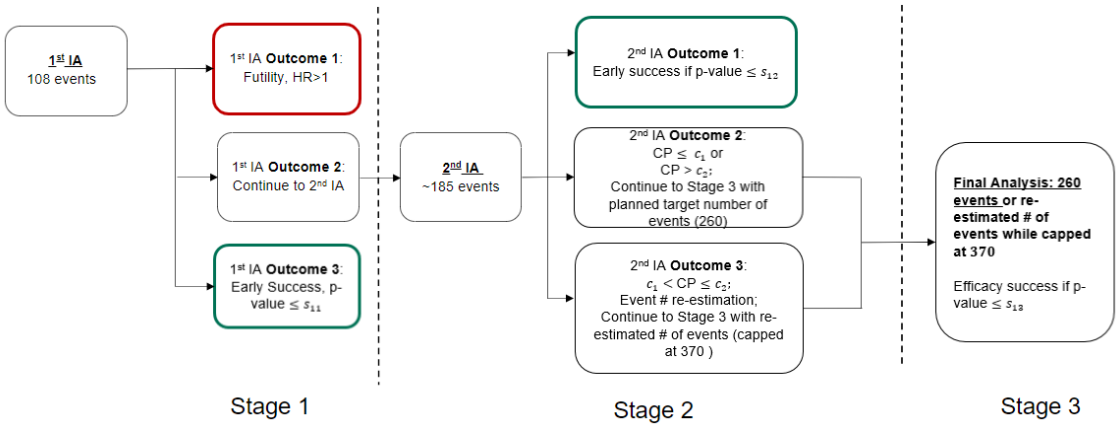
H_0: 12-month OS rate of DSP-7888 Dosing Emulsion plus Bevacizumab = 12-month OS rate of Bevacizumab alone

H_a: 12-month OS rate of DSP-7888 Dosing Emulsion plus Bevacizumab > 12-month OS rate of Bevacizumab alone

4.3.2. Adaptive Design

An adaptive design with two interim analyses and one final analysis is employed in Part 2 of this trial. Figure 3 presents the 3 stages of the adaptive design for Part 2 of the Study.

Figure 3: Adaptive Design for Part 2 of the Study



CP = conditional power; HR = hazard ratio; IA = interim analysis

4.3.3. Determination of Number of Events, Number of Patients and Power

Considering that the number of death events are 108 (observed at interim analysis 1 [IA1]), approximately 185 (IA2), and approximately 260 (FA), the study will have approximately 90% power to detect a 34% reduction in the HR of DSP-7888 Dosing Emulsion plus Bev versus Bev alone (i.e., HR = 0.663, which corresponds to a 1-year survival proportion of 45% in the DSP-7888 Dosing Emulsion plus Bev arm compared to 30% in the Bev alone arm). The number of death events at the FA could go up to 370 if event re-estimation is performed based on the conditional power calculated at IA2. Assuming an enrollment rate of 11 patients per month and patients are followed until the end of the study, with a 5% dropout rate per year, the study will require enrolling of approximately 338 - 480 patients.

Note that Part 2 is an event-driven adaptive trial and statistical power of Part 2 is determined by the target number of death events instead of the number of patients. The actual number of randomized patients in Part 2 may vary to observe the targeted number of events, depending on the outcome of the interim analyses, the duration of the study and operational feasibility. It is

envisioned that approximately 215 patients will be enrolled at IA2 with approximately 185 events, approximately 338 patients will be enrolled at the FA with 260 events if there is no event number re-estimation. If the number of events at the FA reaches the cap of 370 with event number re-estimation, approximately 480 patients will be enrolled in the study assuming that the ratio of the number of events to the number of patients at the FA is approximately 0.77 (assuming similar to the death event rate of 260 events/338 patients when the target event number is 260).

4.3.4. Decision Rules for the Interim and Final Analyses

As shown in Figure 3, IA1 was carried out after 108 death events; IA2 will be carried out after approximately 185 death events; lastly, FA will be carried out after approximately 260 death events or re-estimated number of death events while capped at 370. The target death event re-estimation will be determined by calculating the conditional power (CP) based on the IA2 results. The conditional power is defined as the probability of establishing a significant treatment effect on OS at the final analysis conditional upon the OS data available at IA2.

If the one-sided p-value at the IA1 is less than or equal to the prespecified significance level s_{11}, an early efficacy success would be claimed in terms of the primary endpoint of overall survival. If the HR at the IA1 is greater than 1, then the study may be stopped for futility. The futility boundary is non-binding and the totality of data including other efficacy data and safety data may weigh in the futility decision making besides the futility criteria for the HR.

Should the study neither stop for futility nor claim for early efficacy at the IA1, the study will continue to the IA2. The IA2 decision rules are described as follows:

[1] If the one-sided p-value at the IA2 is less than or equal to the pre-specified significance level s_{12}, an early efficacy success will be claimed in terms of the primary endpoint of overall survival. In this case, the primary endpoint is reached by the IA2, and the study may be concluded to be successful, and the clinical study report (CSR) may be written based on the IA2.

[2] If the early efficacy success cannot be claimed at the IA2, the event re-estimation will be performed using CP. This rule will be defined using two thresholds for CP, denoted by c_1 and c_2. In this study $c_1 = 0.4$ and $c_2 = 0.9$. These thresholds define the "underpowered zone" where it is most sensible to increase the number of OS events. The total number of events at FA may be modified after IA2 as follows:

1) If CP $\leq c_1$, retain the original number of events (ie, 260 events).

2) If CP $> c_1$ and CP $\leq c_2$, increase the number of events to achieve conditional power of 90% or up to a pre-defined cap. In this study the cap is set to be 370.

3) If CP $> c_2$, retain the original number of events (ie, 260 events).

[3] When the FA is conducted and the one-sided p-value is less than or equal to the pre-specified significance level s_{13}, then efficacy success in terms of the primary endpoint of overall survival will be claimed at the end of the study.

The CP will be computed at IA2 using the following approach. Let m_2 denote the actual number of events at IA2 and n_3 denote the planned number of events at FA prior to event count re-estimation. Also, let Z be the normally distributed interim test statistic (log-rank test) and

assume that a larger absolute value of this test statistic, |Z|, corresponds to a beneficial treatment effect. Conditional power at IA2 is given by

$$\mathrm{CP} = \Phi\left(a|Z| - bz_{1-s_{13}}\right),$$

where

$$a = \sqrt{\frac{n_3 - m_2}{m_2}} + \sqrt{\frac{m_2}{n_3 - m_2}},\ b = \sqrt{\frac{n_3}{n_3 - m_2}},$$

$z_{1-s_{13}}$ is the $100(1 - s_{13})$ percentile of the standard normal distribution and $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. The conditional power formula assumes that the future survival data are consistent with the observed data at IA2.

The CP function above is a monotone function of the interim hazard ratio. To prevent the potential back calculation of the interim HR, the event count adjustment rule may be modified to be a non-monotone function.

The details of the modified event count adjustment rule will not be described here in the SAP, will be provided to IDMC in a separate document, and will not be shared with the blinded study team, site and other blinded partners. This procedure is introduced to minimize the risk of the treatment effect being revealed to blinded partners by the event number re-estimation. The decision of event count adjustment will also consider other factors including potential clinical benefit and operational feasibility.

4.3.5. Multiplicity Adjustment

This adaptive design guarantees overall Type I error rate control in the strong sense with respect to all sources of multiplicity in this trial, ie, data-driven design adaptation and analysis of two endpoints. The combination function approach (Wassmer and Brannath, 2016) will be used to ensure strong Type I error rate control for the adaptive design. The hierarchical testing approach will be applied to testing of the primary endpoint and the key secondary endpoint using the method introduced in Maurer and Bretz (2013).

4.3.5.1. Efficacy Stopping Rules

The efficacy stopping rule (to claim efficacy success) for the primary and key secondary endpoints will be derived using the O'Brien-Fleming type alpha-spending function. This approach will be defined to be consistent with the efficacy stopping rule that was applied at IA1 as shown below.

The overall Type I error rate in the trial will be controlled at a one-sided $\alpha = 0.025$. The one-sided alpha levels at IA1, IA2 and FA will be computed using the Lan-DeMets error spending function approach with an O'Brien-Fleming stopping boundary (Jennison and Turnbull, 2000).

The results for one-sided alpha levels for the primary endpoint that are obtained by EAST 6.5 are shown in Table 2 . The same alpha-spending function with the same information fraction at IA1 and IA2 will be used for the key secondary endpoint, resulting the same boundaries for the key secondary endpoint, ie, $s_{21} = s_{11} = \alpha_1 = 0.0005$ and $s_{22} = s_{12}$.

Table 2: One-Sided Alpha Spending for the primary endpoint at Each Decision Point

Analysis	α spending at each planed analysis (α_i, i=1,2,3)	Cumulative α spent	P-value boundary for success (s_{1i}, i=1,2,3)
IA1: 108 events	0.0005	0.0005	0.0005
IA2: 185 events	0.0074	0.0079	0.0077
FA: 260 events	0.0171	0.025	0.0226

This flexible approach of alpha-spending function determines the rate or fraction at which the overall type I error (alpha) is to be spent during a trial. In case the actual number of events observed at IA2 deviates from 185 events, Table 3 provides the corresponding IA2 p-value boundaries for the primary endpoint that will be used for IA2 according to the actual events number at IA2 between 180 to 190 events.

Table 3: One-Sided Alpha Spending for the Primary Endpoint at IA2 Between 180 to 190 Events

IA2 Event number	α spending at IA2 (α_2)	Cumulative α spent	P-value boundary for success (s_{12})
180	0.0066	0.0071	0.0069
181	0.0067	0.0072	0.0071
182	0.0069	0.0074	0.0072
183	0.0070	0.0076	0.0074
184	0.0072	0.0077	0.0075
185	**0.0074**	**0.0079**	**0.0077**
186	0.0075	0.0081	0.0079
187	0.0077	0.0082	0.0081
188	0.0079	0.0084	0.0082
189	0.0081	0.0086	0.0084
190	0.0082	0.0087	0.0086

4.3.5.2. Adaptive Design and Combination P-Value Approach

To define the decision rules that will be applied at IA2 and FA, consider the log-rank test statistics for testing the primary hypothesis of no treatment effect at these decision points,

denoted by Z_{12} and Z_{13}, respectively (a one-sided version of the log-rank test statistic that follows a normal distribution is assumed).

The stagewise test statistics are defined as follows:

- Stage 2: $Z_{12}^* = Z_{12}$
- Stage 3: $Z_{13}^* = \frac{Z_{13}\sqrt{m_3} - Z_{12}\sqrt{m_2}}{\sqrt{m_3 - m_2}}$,

where m_2 is the actual number of events at IA2 and m_3 is the actual number of events at the FA. The stagewise test statistics are independent of each other under the null hypothesis of no effect. The stagewise p-values are computed from the null distributions of the stagewise test statistics, namely, the one-sided Stage 2 and 3 p-values are denoted by p_{12}^* and p_{13}^*, respectively.

In addition, let Z_{22}^* and Z_{23}^* denote the stagewise test statistics for the key secondary endpoint. The corresponding one-sided Stage 2 and 3 p-values will be denoted by p_{22}^* and p_{23}^*, respectively.

If the trial continues to FA, the significance of the treatment effect on the primary and key secondary endpoints will be evaluated at FA as follows. First, the combination p-values will be derived for the two endpoints using the weighted inverse-normal combination function, ie,

$$c(x, y) = 1 - \Phi\left(\sqrt{w}\Phi^{-1}(1 - x) + \sqrt{1 - w}\Phi^{-1}(1 - y)\right),$$

where $\Phi(x)$ denotes the cumulative distribution function of the standard normal distribution, and w and $1 - w$ are the pre-defined weights assigned to Stage 2 and 3 in the adaptive design, ie, 0.71 (185/260) and 0.29.

Using combination p-value, a significant effect will be established for the primary endpoint at FA if

$$c(p_{12}^*, p_{13}^*) \leq s_{13},$$

where s_{13} is the significance level for the primary endpoint at FA.

If this condition is met at FA and thus the primary hypothesis is rejected, a significant effect will be established for the key secondary endpoint at FA if

$$c(p_{22}^*, p_{23}^*) \leq s_{23}.$$

where s_{23} is the significance level for the key secondary endpoint at FA (**Table 4**).

The significance levels for the key secondary endpoint at FA will be derived as follows. First of all, the one-sided significance level at IA1 and IA2, denoted by s_{21} and s_{22}, are set to the same value as the significance level for the primary endpoint, i.e., $s_{21} = s_{11} = 0.0005$ and $s_{22} = s_{12}$

The significance level at FA, denoted by s_{23}, is computed from

$$P(P_2 > s_{22}, P_3 \leq s_{23}) = \alpha_3,$$

where P_2 and P_3 denote the random variables that have the same bivariate distribution as the one-sided p-values for the key secondary endpoint at IA2 and FA under the null hypothesis of no treatment effect. This significance level is easily computed from the joint distribution of Z_{22} and Z_{23} under the secondary hypothesis of no effect for this endpoint. This distribution will depend

on the estimated information level at IA2 and FA, i.e., I_2 and I_3, namely, Z_{22} and Z_{23} both follow standard normal distributions and the correlation between Z_{22} and Z_{23} is equal to the actual information fraction at the interim analysis, i.e., $\sqrt{I_2/I_3}$, where the information is the reciprocal of the variance .

$$I = \left[\sum_i \left(\left(\frac{\hat{\sigma}_{i1}}{log[\widehat{OS}_{i1}]}\right)^2 + \left(\frac{\hat{\sigma}_{i2}}{log[\widehat{OS}_{i2}]}\right)^2\right)\right]^{-1}$$

Note that, if the number of events is increased in the trial, the information value I_3 needs to be computed at the planned FA after 260 events.

For illustration, Table 4 defines the significance level for the key secondary endpoint at FA (s_{23}) for selected values of the actual information fraction (I_2/I_3) under the assumption that the information fraction at IA2 is set to be 0.71.

Table 4: One-sided significance levels for the key secondary endpoint at FA for various actual information fractions at IA2

Assumed information fraction at IA2	**Actual information fraction at IA2**	P-value boundary for success (significance level, s_{23})
0.71	0.60	0.0214
0.71	0.65	0.0219
0.71	0.70	0.0224
0.71	0.75	0.0229
0.71	0.80	0.0234

4.3.5.3. Hierarchical Testing Procedure

The hierarchical testing strategy for the primary and key secondary endpoints will be defined using the method introduced in Maurer and Bretz (2013). The key secondary endpoint will be tested sequentially beginning with the primary endpoint. It is important to point out that the common alpha spending function for the primary and key secondary endpoints, ie, the Lan-DeMets function with an O'Brien-Fleming stopping boundary, satisfies the condition (well-ordered condition) imposed in Maurer and Bretz (2013).

4.4. Treatment Assignment & Blinding

The study will be divided into 2 parts. Part 1 will be a non-randomized, open-label safety lead in Phase where all patients will be treated with the combination of DSP-7888 Dosing Emulsion plus Bev in a single-arm manner. Patients who discontinue Part 1 of the study for reasons other than DLT during the Induction Phase will be replaced.

Once the safety of the combination of DSP-7888 Dosing Emulsion plus Bev has been established, Part 2 (the randomized part of the study) will begin. In Part 2, patients will be stratified prior to randomization based on the extent of surgical resection in primary therapy

(GTR versus not GTR) and KPS class (60 through 70 versus 80 through 100) and will be randomized in a 1:1 ratio for DSP-7888 Dosing Emulsion plus Bev or Bev alone via the Endpoint Interactive Voice/Web Response System. The treatment assignment is to be obtained in Part 2 after the study patient has passed all inclusion and exclusion criteria and has been randomized in the study prior to the initiation of the study treatment or first dosing visit. Withdrawn patients will not be replaced in Part 2 of the study.

This is an open label study. The primary endpoint of Part 2 of this study is OS, the assessment of which is unlikely impacted by the open-label nature of the study.

To comply with ICH E6 (R2) Guideline for Good Clinical Practice (2016), which recommends the blinded review of planned analyses, and to minimize the risk of potential bias into interim and final analyses for the adaptive Phase 3 design in Part 2, blinding controls have been implemented on the aggregate data by treatment arm. Specifically, the IA results will be reviewed by the IDMC and will only be available to the IDMC and selected unblinded team members. The investigators, the patients, the study personnel who interact with the site and other blinded partners do not have the access to the IA results.

A blinding plan that details the blinding controls for the sponsor and its partners will be summarized in separate document.

4.5. Study Schedule of Assessments

Schedule of Assessments is provided in the protocol section 8.1.3.

5. ENDPOINTS

5.1. Primary Endpoint

The primary endpoint is OS and is defined as the interval between randomization and death from any cause. Patients who are lost to follow-up will be censored on the last day they were known to be alive.

5.2. Key Secondary Endpoint

The 12-month OS rate is defined as the proportion of patients alive 12 months after randomization.

5.3. Other Secondary Endpoints

- PFS is defined as the interval between randomization and progression, determined by central radiology review (the local investigator assessment will be used at IA) or death from any cause.
- 6-month PFS rate is defined as the proportion of patients alive at 6 months after randomization and without progressive neoplastic disease.
- The objective response rate is defined as the proportion of patients exhibiting a response (complete response [CR] plus partial response [PR]) based on the mRANO criteria as determined by the central radiology review (the local investigator assessment will be used at IA).
- The duration of response is defined as the interval between first documented oncological response and progression of disease or death from any cause, with response based on the mRANO criteria as determined by the central radiology review (the local investigator assessment will be used at IA).
- To describe the adverse event (AE) profile of DSP-7888 Dosing Emulsion plus Bev alone.

5.4. EXPLORATORY ENDPOINTS

In addition, exploratory endpoints include evaluating the amount of immune response by:

- Measurement of WT1 peptide-specific cytotoxic T lymphocyte (CTL) induction activity at each assessment time for patients receiving DSP-7888 Dosing Emulsion plus Bev.
- Determination of expression of WT1 protein via immunohistochemistry (IHC) and WT1 mRNA via Chromogenic in situ Hybridization (CISH) in tumor tissue from patients receiving DSP-7888 Dosing Emulsion plus Bev.
- Diffusion-weighted imaging results.
- NANO scale score.

6. ANALYSIS SETS

6.1. Safety Set

The Safety Set (SS, also called Safety Population) for Part 1 will include all patients who were administered at least one dose of study medication in Part 1. The Safety Set for Part 2 will include all randomized patients who were administered at least one dose of study medication in Part 2. Patients will be analyzed according to treatment received. The SS will be used for all analyses of safety endpoints and for the presentation of patients in all patient listings, unless otherwise specified

6.2. Intent-To-Treat Set

The Intent-to-Treat (ITT) Set (also called intent-to-treat population) will include all randomized patients in Part 2 of the study only, regardless of whether they received a dose of study drug or not. Patients will be analyzed according to their assigned treatment. The ITT Set will be used for all analyses of study disposition, baseline demographics, medical history, protocol deviations and efficacy endpoints et al. The presentation of patients in all efficacy listings will be based on the ITT Set.

6.3. Per-Protocol Set

The Per-Protocol (PP) Set (also called per-protocol population) will consist of all patients in Part 2 of the study who received at least one dose of study drug (DSP-7888 Dosing Emulsion or Bev), and who do not have major protocol deviations that may impact the primary analysis of efficacy. Per-Protocol Set may be different depending on the Efficacy Endpoints as follows:

- Per-Protocol Set for the Overall Survival Endpoint (PP-OS)
- Per-Protocol Set for the Progression Free Survival Endpoint (PP-PFS)
- Per-Protocol Set for the Overall Response Endpoints (PP-OR)

The criteria of protocol deviations and reasons for exclusion from the PP Population will be finalized prior to the database lock and unblinding (for IA2 and FA). Reasons for exclusion from PP Population may include, but not limited to, the following:

- Not meeting inclusion criterion
- Meeting exclusion criterion
- Meeting withdrawal criterion (e.g., a subject is subsequently found to be pregnant after inclusion in the study)

Protocol deviations will be identified and maintained in a separate document, and the list of patients with protocol deviations including those with major protocol deviations that may be excluded from efficacy analysis will be finalized before database lock and unblinding.

6.4. Dose-Determining Set

The Dose-Determining Set (DDS, also called DLT population) will consist of all patients in Part 1 of the study only who received the first 5 administrations of the combination of DSP-7888

Dosing Emulsion plus Bev or who are discontinued from the study during Part 1 due to the occurrence of a DLT.

7. PROTOCOL DEVIATIONS

The Investigator will notify the Sponsor or designee of any deviations from the protocol. Such contact with the Sponsor will be made as soon as possible to permit a decision as to whether the patient (for whom the deviation from the protocol was affected) is to continue in the study. The case records will describe both the details of and rationale for the protocol deviation.

Major Protocol Deviations (PD) in the ITT population will be summarized, and all PDs in the ITT population will be listed.

8. GENERAL ASPECTS FOR STATISTICAL ANALYSIS

8.1. General Methods

Statistical analysis of data from this study will be performed using SAS version 9.3 or more recent version.

Unless otherwise specified, study Part 1 summaries will be presented by dose level. Study Part 2 summaries will be presented by treatment groups. Summaries based on the ITT and PP populations will be presented by planned treatment groups, unless specifically stated otherwise. Safety analyses and other summaries based on the safety population will be presented by actual treatment received.

Statistical analyses will be descriptive in nature. Continuous variables will be summarized using the number of observations (n), mean, standard deviation (SD), median, minimum, and maximum. When needed, the use of other percentiles (e.g., 10%, 25%, 75%, and 90%) will be mentioned in the relevant section. Categorical variables will be summarized using number of observations (n), frequency and percentages of patients.

For categorical variables, a Fisher's exact test will be used for unstratified analyses of categorical variables and the Cochran-Mantel-Haenszel test will be used for stratified analyses. The Kaplan-Meier (KM) product limit method will be used to provide the survival estimates and the survival curves for the time to event variables for each treatment group and for subgroup of patients within each actual stratum [KPS (60 through 70, and 80 through 100), and extent of surgical resection in initial therapy (GTR versus not GTR)]. Treatment comparisons for time-to-event analyses will be performed using a stratified log-rank test. For patients who discontinue early or who are in the study without meeting criteria for the event, the time-to-event will be censored. More detailed censoring rules for OS and PFS are described in Section 7.3 of this document. Hazard ratios and their corresponding 95% confidence intervals will be estimated based on a stratified Cox's regression model using treatment group as the independent variable and the randomization stratification variables as strata.

Patient's tumor response will be assessed based on the mRANO (Neurotherapeutics. 2017 Apr; 14(2): 307–320) as assessed by both study investigators and by a central blinded review. See study Protocol Amendment 6, Sections 16.2 and 16.3 for details on tumor response determinations that are fully captured on the case report forms (CRFs) of the study.

One-sided p-values will be presented for treatment comparisons in the primary endpoint of OS and the key secondary endpoint of 12-month OS rate in Part 2. Confidence intervals (CIs) will be calculated at 95%, 2-sided.

All p-values will be rounded to 4 decimal places. P-values rounding to less than 0.0001 will be presented as "<0.0001".

Percentage values will be rounded to one decimal point (e.g., 52.3%, 8.9%). Dates will be printed as DDMMMYYYY. If only year and month are available, date will be displayed as --MMMYYYY. If only year is available, then just ----YYYY will be displayed.

All analyses will be performed using data pooled across sites.

All relevant patient data will be included in listings.

8.2. General Data Handling

Baseline measurements for Part 2:

For efficacy parameters: for example, baseline weight, KPS score, NANO, tumor scans, baseline will be defined as the last non-missing value within 35 days prior to or on the randomization date; when such a parameter within 35 days prior to or on the randomization date is missing, and patients receive treatment, then the first parameter recorded prior to or on first dose date (if on first dose date, visit = screen/prescreen visit) but after randomization can be used if it is available. If patients didn't receive any treatment, then the parameter recorded at the screening visit unless it occurs more than 35 days prior to randomization will be used.

For safety parameters, such as labs, ECG, physical exams and vital signs, the last non-missing record prior to first dose date or on first dose date (record from "screen/pre-screening visit" and the time is not after Bev administration time) will be used as baseline. Baseline window of 28 days prior to first dose date will be used.

Baseline for Part 1 will follow the Part 2 safety baseline definition.

For the safety parameters: unless otherwise specified, post baseline reporting period includes non-missing records collected from after first dose of study treatment until 30 days following last dose of study medication (DSP7888 and/or Bevacizumab).

Scheduled and unscheduled visits post baseline: unless otherwise specified, for the summary by visit post baseline, only scheduled visits will be included; for summary of the maximum or maximum changes post baseline, both scheduled visits and unscheduled visits will be included. All records (scheduled and unscheduled) regardless of post baseline reporting period will be listed in the data listings.

8.3. Key Definitions

The reference date for calculating Study day for Part 1 and safety measurements of Part 2 is the date of the first dose of study drug, which will be considered Day 1 in the study. Specifically, study days for post-baseline data will be calculated as follows:

Date of interest – Date of first dose + 1, if date of interest on or after the first dose date; or

Date of interest – Date of first dose, if date of interest prior to the first dose date.

For Part 2 efficacy measurements (such as PFS and OS), randomization date will be used as the reference date.

Duration of exposure for each patient is the length of time from the first dose date of study drug to the date of last dose of study drug, and will be calculated using the formula below:

Date of last dose of study treatment – Date of first dose of study treatment + 1.

8.4. Missing Data

8.4.1. Missing Partial Death Dates

It is recommended that the database be designed to mandate a complete death date. If there is a record for death, but the date is missing or is partial, it will be imputed based on the last contact date.

If the entire date is missing, the death date will be imputed as the day after the date of last contact.

If the day or both day and month are missing, the death date will be derived as below: 1) impute partial dates to be 1st day of the month if day of death is missing, or January 1st if both day and month of death are missing, 2) compare the imputed dates with one day after the last known to be alive date, select the latest date. For example, if a patient with partial death date is known to be March2020, and its last known to be alive date is March 16,2020, then considering the rules of both 1) and 2), its death date will be imputed as March 17 2020.

8.4.2. Missing Dates for Adverse Events (AEs) or Concomitant Medications

Partial or incomplete start/end dates in the adverse events (AE) or concomitant medications datasets will be imputed to help determine treatment-emergent adverse events and (prior or concomitant) medications.

For incomplete start dates for AE or concomitant medications, use the following imputation rules:

- Missing day - Impute the 1st of the month unless month is same as month of first dose of study drug then impute using the day of first dose date
- Missing day and month – impute 1st January unless year is the same as first dose date then impute using the first dose date
- Completely missing – impute using first dose date unless the end date suggests it must have started prior to first dose, in which case impute the missing date with 01JAN of the same year as the first dose date.

When imputing a start date, ensure that the new imputed date is sensible i.e. is prior to the end date of the AE or medication. If the imputed start date is after the end date, then the start date will be assumed equal to the end date.

For incomplete end dates for AE or concomitant medications, use the following imputation rules:

- Missing day - Impute the last day of the month unless month and year are the same as the month and year of the last dose of study drug, then impute last dose date
- Missing day and month – impute 31st December unless year is the same as last dose date then impute last dose date
- Completely Missing – need to look at whether the AE/medication is still ongoing before imputing a date and also when it started in relation to study drug. If the ongoing flag is not missing, then assume that AE is still present/medication is still being taken (i.e. do not impute a date). If the AE/medication is not ongoing and the start date is prior to first dose

date, then impute the end date using 1st dose date. If it started on or after first dose date, then impute a date that is after the last dose date.

When imputing an end date, ensure that the new imputed date is sensible i.e. is after the start date of the AE or medication. If the imputed end date is before the start date, then the end date will be assumed equal to the start date.

8.4.3. Missing Post GBM Treatment Start Date

For data collected from Post GBM Treatment CRF form start date:

For treated patients:

= 01JanYYYY, if only Year is available. After imputation if it is before last dose date, it will be set to last dose of study treatment + 1

= 01MMMYYYY, if only day is missing. After imputation if it is before last dose date, it will be set to last dose of study treatment + 1

= last dose date + 1 if complete missing.

For randomized but not treated patients:

= 01JanYYYY, if only Year is available. After imputation if it is before randomization, it will be set to randomization date.

= 01MMMYYYY, if only day is missing. After imputation if it is before randomization, it will be set to randomization date.

= randomization date if complete missing.

For concomitant procedure:

If the start date is completely or partially missing, then the imputed start date is:

= 01JanYYYY, if only Year is available. After imputation if it is before randomization, it will be set to randomization date.

= 01MMMYYYY, if only day is missing. After imputation if it is before randomization, it will be set to randomization date.

= randomization date if complete missing.

8.4.4. Censoring Rules for Time-to-Event Data

For time-to-event efficacy analyses, endpoints will be defined using the censoring rules presented in the following Table 5 Table 6 Table 7 for Part 2.

Table 5: Censoring Rules for Overall Survival for Part 2.

	Description	Situation	Date of Events or Censoring	Outcome
Overall Survival	The interval between randomization and death from any cause	Date of death	Date of death	Event
		Alive as of the last known follow-up,	Date of last known to be alive	Censored

	Description	Situation	Date of Events or Censoring	Outcome
		discontinued or lost to follow-up		

Table 6: PFS Censoring Rules by Confirmed PD for Part 2

Hierarchy Censoring	Situation	Date of Events or Censoring	Outcome
1	No baseline assessment of target lesion and non-target lesion within 35 days from randomization	Date of Randomization	Censored
	1st Confirmed PD within 2 assessment intervals [1] following previous adequate overall response assessment [2] or randomization if no previous adequate overall response assessment after randomization	Date of response assessment of 1st Confirmed PD	Event
2	New anti-cancer treatment started and no confirmed PD or death prior to that	Date of previous adequate overall response assessment prior to start of new therapy	Censor
3	Confirmed PD after 2 assessment intervals following previous adequate overall response assessment	Date of previous adequate overall response assessment [3]	Censor
	Preliminary PD following by death / confirmed PD within one assessment interval [4]	Date of Preliminary PD	Event
	For patient without confirmed PD, but death reported within 2 assessment intervals following previous adequate overall response assessment or randomization if no previous adequate overall response assessment after randomization	Date of death	Event
4	Have baseline target lesion or non-target lesion assessment performed within 35 days from randomization but no post-baseline overall response assessment (no confirmed PD by default) and no death within 2 assessment intervals following randomization	Date of Randomization	Censored
5	no confirmed PD and no death reported within 2 assessment intervals following previous adequate overall response assessment or randomization if no previous adequate overall response assessment after randomization	Date of latest adequate assessment or randomization date, whichever is later	Censored

Only mRANO assessments prior to new anti-cancer treatment will be used for deriving the PFS.

[1]As mRANO tumor assessments will be performed at Screening; after 8, 16, and 24 weeks on study; and then every 12 weeks thereafter, two assessment intervals are 126 days (16 weeks+ 14 days) if previous assessment is within 105 days (16 weeks – 7 days) from randomization; or 154 days (20 weeks + 14 days) if previous assessment is after 105 days but prior to 161 (24 weeks –

7 day) days from randomization, or 182 days (24 weeks + 14 days) if previous assessment is after 161 days from randomization.

[2]Adequate overall response assessments include the overall assessment of Confirmed CR, Confirmed PR, Preliminary CR, Preliminary PR, Pseudo Response, SD, Confirmed PsP, Preliminary PD, Confirmed PD (NE or NA is not included).

[3] Date of overall response assessment is the last date from tumor scan, NANO/KPS assessment from the same visit.

[4] One assessment interval is 70 days (8 weeks + 14 days) if previous assessment is prior to 161 days (24 weeks – 7 day) from randomization, or 98 days (12 weeks + 14 days) if previous assessment is after 161 days from randomization.

The censoring reason will be assigned according to the hierarchy order.

Preliminary Progression Free Survival is the interval between the date of randomization and the date of 1st progressive disease including Preliminary PD / Confirmed PD / Confirmed PsP or death, whichever is first reported.

Events and Sequential Censoring Rules for PD for preliminary PFS.

Table 7: Preliminary PFS Censoring Rules for Part 2

Hierarchy Censoring	Situation	Date of Events or Censoring	Outcome
1	No baseline assessment of target lesion and non-target lesion within 35 days from randomization	Date of Randomization	Censored
	1st PD [1] within 2 assessment intervals [2] following previous adequate overall response assessment [3] or randomization if no previous adequate overall response assessment after randomization	Date of response assessment of 1st PD [1]	Event
2	New anti-cancer treatment started and no PD [1]or death prior to that	Date of previous adequate overall response assessment prior to start of new therapy	Censor
3	PD [1] after 2 assessment intervals following previous adequate overall response assessment	Date of previous adequate overall response assessment [4]	Censor
	For patient without PD [1], but death reported within 2 assessment intervals following previous adequate overall response assessment or randomization if no previous adequate overall response assessment after randomization	Date of death	Event
4	Have baseline conventional MRI or CT scan performed within 35 days from randomization but no post-baseline overall response assessment (no PD [1] by default) and no death within 2 assessment intervals following randomization	Date of Randomization	Censored

5	no PD [1] and no death reported within 2 assessment intervals following previous adequate overall response assessment or randomization if no previous adequate overall response assessment after randomization	Date of latest adequate assessment or randomization date, whichever is later	Censored

PD includes Preliminary PD / Confirmed PD / Confirmed PsP.

[1]As mRANO tumor assessments will be performed at Screening; after 8, 16, and 24 weeks on study; and then every 12 weeks thereafter, two assessment intervals are 126 days (16 weeks+ 14 days) if previous assessment is within 105 days (16 weeks – 7 days) from randomization; or 154 days (20 weeks + 14 days) if previous assessment is after 105 days but prior to 161 (24 weeks – 7 day) days from randomization, or 182 days (24 weeks + 14 days) if previous assessment is after 161 days from randomization.

[3] Adequate overall response assessments include the overall assessment of Confirmed CR, Confirmed PR, Preliminary CR, Preliminary PR, Pseduo Response, SD, Confirmed PsP, Preliminary PD, Confirmed PD (NE or NA is not included).

[4] Date of overall response assessment is the last date from tumor scan, NANO/KPS assessment from the same visit.

The censoring reason will be assigned according to the hierarchy order.

For Part 1 patients, the first dose date will be used instead of randomization date in the above algorithm.

8.4.5. Missing Lab, Pharmacodynamic and Biomarker Data

Unless otherwise specified, missing lab, pharmacodynamic and biomarker data will be treated as such and no imputed values will be derived.

8.5. Analysis Visit Window

No analysis windows are planned for the study regarding data collected outside the protocol specified windows, see Section 3.9 for the detailed schedule of assessments and protocol windows. Analyses will not exclude patient data due to the patient's failure to comply with the visit scheduled, as data will be summarized based on the CRF study visit in which it was collected.

8.6. Pooling of Centers

No pooling of sites is planned for this study. Patients from all sites will be pooled together for data analysis.

8.7. Subgroups

Subgroup analysis will be presented for the primary endpoint by strata for randomization for the following 4 subgroups:

- KPS 60 to 70

- KPS 80 to 100
- Gross total resection
- Not gross total resection
- Country (United States, Canada, Japan, Korea, Taiwan),
- Geographic region (North America, Asia),
- Race (Caucasian, Asian, Black, Other),
- Sex (male, female, other) where "other" will includes patients with "unknow" or "other" or missing category.
- Age Group (<55, 55 through 64, 65 through 74, 75 through 84, ≥85),
- HLA (HLA-A*02 Positive Only, HLA-A*24 Positive Only, Both Positive),
- IDH Mutation (IDH1 Mutation: Yes, No, Missing; IDH2 Mutation: Yes, No, Missing, IDH1 or IDH2 Mutation: Yes, No, Missing),
- MGMT Methylation: Yes, No, Missing,
- Corticosteroids Use at Baseline: Yes, No
- Patients Who Had a Surgery After Most Recent Recurrence: Yes, No

If total number of patients in any subgroup is less than 10, no p-value or KM curve will be provided.

Subgroup analyses (e.g., Japan, Non-Japan) for efficacy and safety endpoint will be performed for Pharmaceuticals and Medical Devices Agency (PMDA) submission. Those analyses may not be included in the CSR.

9. DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS AND MEDICATION

9.1. Patient Disposition and Withdrawals

The number of patients pre-screened for HLA type, the number of patients formally screened for the study, the number of patients screen failed, and the number of patients randomized will be summarized. Additionally, frequency counts and percentages of all patients who are randomized but did not take study drug, and patients who are dosed but discontinue investigational product will be presented by dose level and overall in study Part 1; and by treatment group and overall in study Part 2 for patients in the ITT Set. All patients who discontinue investigational product will be identified, and their reason for ending the investigational product will be summarized.

Follow up time (months) defined by [last contact date (patients who still alive) or death date (patients who dead) - randomization date + 1]/30.4375 will be summarized.

A listing of all study patients, whether they discontinued the investigational product or not, and the reasons for investigational product discontinuations will be provided for the Safety Set .

The reasons for exclusion from analysis sets (including PP set) will be tabulated by study treatment and listed.

9.2. Demographic and Other Baseline Characteristics

Age (years) will be calculated as the number of years between the date of birth and the date of informed consent:

Age at screening = (screening informed consent date – date of birth +1)/365.25 and truncated to complete years

If height is recorded in inches (in), then height will be summarized in cm using the following conversion:

Height (cm) = Height (in) x 2.54

For Part 1, baseline height is defined as the last non-missing measurement prior to the first dose of study drug but within baseline window. In the event such a value is missing, the first measurement after the first dose of study drug is used.

For Part 2, baseline height is defined as the last non-missing measurement on or prior to the date of randomization but within baseline window. In the event such a value is missing, the first measurement after randomization is used.

If weight is recorded in pounds (lb), then weight will be summarized in kilograms (kg) using the following conversion:

Weight (kg) = Weight (lb) x 0.4536

Body Mass Index (BMI) in kg/m2 will be calculated as: Weight (kg)/[Height (m)2]

Demographic and baseline characteristic data (sex, race, ethnicity, age (years), height (cm), weight (kg), BMI (kg/m2), and HLA Class I results) will be summarized by dose level in Part 1 Safety Set, and by treatment group and overall in Part 2 ITT Set.

All demographic and baseline characteristic data will also be listed.

9.3. Medical / Surgery History

Medical/surgical history including diagnosis, signs & symptoms, procedures, the onset date, and status (ongoing or resolved) will be collected.

Medical/surgical history recorded at screening will be summarized by dose level in Part 1 Safety Set, and by treatment group and overall in Part 2 ITT Set via descriptive statistics. The number and percentage of patients reporting each System Organ Class (SOC) and/or Preferred Term (PT) will be presented. Patients with multiple occurrences of the same event will be counted only once within the given preferred term. Similarly, at the SOC level, a patient will be counted only once if the patient experienced more than one event within the given SOC.

The medical/surgical history that is specifically related to GBM as indicated by the "related to GBM" checkbox on the medical/surgical history CRF will be summarized separately from all other medical history.

Medical/Surgical history listings will be sorted by treatment group or dose level depending on the study part, patient number, onset date, end date, SOC, and PT.

9.4. GBM History and Other Baseline Characteristics

GBM history including date of initial diagnosis of primary site, date of most recent recurrence/relapse/progression and recurrent type will be obtained. The time in months since initial GBM diagnosis to randomization, the time in months since Initial GBM diagnosis to most recent recurrent/relapse/progression, the time in months since most recent recurrent/relapse/progression to Randomization, the frequency of most recent recurrent/type will be summarized by dose level in Part 1 Safety Set, and by treatment group and overall in Part 2 ITT Set. The date of initial diagnosis of primary site, date of most recent recurrence/relapse/progression and recurrent type will be listed by treatment group or dose level depending on the study part as captured on the CRF.

The type of primary surgery performed will also be summarized based on the data entered into the Gross Total Resection CRF (number of patients who had a gross total resection vs. Not gross total resection). Of those who did not have a gross total resection, the number who had a partial resection vs. Number who had no resection at all will be summarized.

In addition, MGMT promoter methylation, isocitrate dehydrogenase (IDH) 1 mutation and IDH 2 mutation data obtained during screening will be listed. The use of corticosteroids at baseline and the sum of NANO overall scale score at baseline will be summarized by dose level in Part 1 Safety Set, and by treatment group and overall in Part 2 ITT Set.

9.5. Medication

Medications will be coded using WHO Drug Dictionary Enhanced (WHO-DDE) -B2 201609.

Medications will be summarized using Anatomical Therapeutic Chemical (ATC) classification level 2 and Preferred base. All medications will be summarized by dose level in Part 1 Safety Set, and by treatment group and overall in Part 2 Safety Set. The analysis will be sorted by order

of decreasing frequency of ATC classification level 2 and preferred base in DSP-7888 Dosing Emulsion plus Bev arm.

Missing or partial start/end dates of medications will be imputed using the approach discussed in Section 8.4 .

Medication data collected on the study include systemic anti-cancer medications received prior to randomization, prior and concurrent medications that are not anti-cancer, and systemic anti-cancer medications received after discontinuing protocol treatment.

Prior and concomitant medication data for medications that are not anti-cancer will be listed and will include the primary indication for use, start date, end date, ATC classification level 2, preferred base, dose, dose unit, and route of administration. Medications will be flagged as being either prior medications, concomitant medications, or both. Within patients, medications will be ordered by start date, end date, ATC classification level 2 and preferred base.

9.5.1. Prior Medication

Prior medications are defined as medications that are not anti-cancer with at least one dose taken before the date of the first dose of the study drug.

9.5.2. Concomitant Medication

Concomitant medications are defined as medications that are not anti-cancer with at least one dose taken between the date of first dose (inclusive) and the date of last dose (inclusive) of study drug.

Concomitant medications will be summarized by dose level in Part 1 Safety Set, and by treatment group and overall in Part 2 Safety Set using the ATC classification level 2 terms. If a patient has taken a medication more than once, the patient will only be counted once in the total for the summary.

9.5.3. Prior Cancer Therapies

Prior cancer therapies (also known as prior systemic anti-cancer medications) will be summarized and listed separately from prior and concomitant medications that are not anti-cancer. The classification of the cancer therapies includes type of prior cancer therapy, regimen, start/end date, and best response captured on the Prior Cancer Therapy eCRF page. The categorization of prior systemic anti-cancer therapies was performed by the study sponsor.

For prior systemic anti-cancer medications, the number and percentage of patients that received each specific agent (as defined by preferred base) and best response will be summarized by dose level in Part 1 Safety Set, and by treatment group and overall in Part 2 ITT Set. Prior cancer therapies will also be listed and include the regimen as entered by the site, the type of prior cancer therapy entered, the categorization provided by the sponsor, start/stop dates, preferred base and best response, et al . The listing will be ordered by treatment group or dose level depending on the study part, patient, and the start date of the therapy.

9.5.4. Prior Radiation Therapies

Prior radiation therapies will be summarized and listed separately from other prior and concomitant medications. The summary of prior radiotherapy includes number of fractions,

fractional dose received (Gy), total dose received, best response, and reason for stopping captured on the Prior Radiation Therapy eCRF page. The specific type of radiotherapy received and the anatomic site will not be summarized but will be listed.

Time (months) since stop date of radiotherapy to first Dose date, number of fractions, total dose (Gy) and best response will be summarized by dose level in Part 1 Safety Set, and by treatment group and overall in Part 2 ITT Set.

The number and percentage of patients who had prior radiation therapy and best response will be summarized. Continuous variables such as the number of fractions, the time (months) since stop date of radiotherapy to first Dose date and total dose received will be summarized via descriptive statistics using the number of observations (n), mean, standard deviation (SD), median, minimum, and maximum.

Prior radiation therapies will also be listed and include the type of prior radiation therapy received, anatomic site, start date, stop date, number of fractions, fractional dose received, total dose received, best response, reason for stopping. The listing will be ordered by treatment group or dose level, depending on the study part, patient, and the start date of the therapy.

10. EFFICACY

All continuous variables will be summarized using the number of observations (n), mean, standard deviation (SD), median, minimum, and maximum. Categorical efficacy variables will be summarized using frequency counts and percentages. The analysis will be presented by treatment group and overall in Part 2 ITT Set. Analysis of the primary efficacy endpoint and selected secondary efficacy endpoints will be repeated using the PP Set.

10.1. Primary Efficacy Endpoint and Analysis

10.1.1. Primary Efficacy Analysis

The primary efficacy endpoint is OS. The primary statistical null hypothesis is that there is no difference between the 2 treatment groups (DSP-7888 Dosing Emulsion plus Bev and Bev alone) with regards to OS. This will be tested by the stratified log-rank test based on the actual stratification variables at baseline.

The KM method will be used to estimate and visualize the distribution of OS by treatment group. Additionally, KM curves and estimates by treatment group will be provided for the subgroups of patients (subgroups are defined in SAP Section 8.7). Descriptive summary statistics that are of interest, point estimates and 95% confidence intervals for overall survival will be provided for each treatment group at 3, 6, 9, 12, and additional months if estimable. These estimates will be computed from Kaplan-Meier survival curves for each treatment group. The primary treatment comparison will be based on a stratified log-rank test considering the 2 actual stratification factors of KPS class and extent of surgical resection in primary therapy (GTR versus not GTR). A stratified Cox's regression model will be performed with treatment group as the independent variable and the 2 stratification factors of KPS class (60 through 70 versus 80 through 100) and extent of surgical resection in primary therapy (GTR versus not GTR) as strata. The hazard ratio and respective 95% CI obtained from the stratified Cox model will be provided.

10.1.2. Sensitivity Analysis of Primary Endpoint

Sensitivity Analysis of OS includes:

The analyses of OS in the per-protocol population and safety set will be performed. An un-stratified log-rank test and Cox PH model may be used as sensitivity analyses for OS. A stratified log-rank test/stratified Cox PH model may be performed based on randomization stratification factors for ITT set.

Evaluation of the impact of new anti-cancer therapy on OS may be performed. The patients who receive subsequent (new) anti-cancer therapy may be censored at the time of new anti-cancer therapy. If applicable, other sensitivity analyses including the time-dependent Cox PH model may be used to adjust for the impact of new anti-cancer therapy.

Subgroup analysis will be performed considering the subgroup specified in Section 8.7 other subgroups where appropriate.

For each subgroup level, the median OS (or other quartiles) and a 2-sided 95% CI will be provided for each arm. The difference in OS between Arm 1 and Arm 2 may be analyzed by unstratified log-rank test. The nominal 2-sided p-value may be provided in an exploratory basis.

The estimated hazard ratio and 2-sided 95% CI may be provided based on un-stratified Cox PH model.

Forest plots of hazard ratios from the primary analysis or other subgroup analyses will be provided where appropriate.

A sensitivity analysis using stratified multivariate Cox proportional hazards model will be performed to account for potential imbalance for IDH/MGMT between two treatment arms by including the IDH mutation and/or MGMT methylation as covariates; potential treatment by covariate interaction effects may be explored as appropriate. Subgroup analysis for MGMT methylation and IDH mutation for the primary and key secondary efficacy endpoints will also be conducted.

At the final analysis, in the case there is significantly high dropout rate, e.g. (15% or higher overall) , when the primary endpoint is statistically significant, a tipping point sensitivity analysis may be carried out. See Section 15.1 for more detail.

In addition, the assumptions used in the primary analysis may be assessed using the method proposed in He, Fang and Su (2013) and He and Su (2015). The methodology is aimed at evaluating the significance of treatment effect in the presence of delayed effects as well as important deviations from the assumption of a constant hazard ratio function. The methodology will be applied as follows:

- Step 1. The survival data in the two treatment arms will be examined to determine whether one or more change points in the hazard ratio function exist. The changing point search will be limited to median survival from ITT set with pooling both arm togethers.
- Step 2. If at least one change point is identified, the changes points will be used to partition the trial period (up to the last event or censored observation) into multiple intervals. Since the hazard ratio function is expected to be close to a constant within each interval, interval-specific inferences will be performed, e.g., the log-rank test will be applied to compute the interval-specific treatment effect p-value and the Cox proportional hazards model will be used to estimate the interval-specific hazard ratio.

Additional sensitivity analysis may be performed using appropriate methods from the class of weighted log-rank tests (Zucker and Lakatos, 1990), including generalized piecewise weighted log-rank tests (Xu et al., 2018) with parameters that match the observed deviations from the assumed constant hazard ratio function.

10.2. Key Secondary Efficacy Endpoint and Analysis

10.2.1. 12-Month Overall Survival Rate

The 12-month OS rate is defined as the proportion of patients alive 12 months after randomization on the KM curve.

Treatment comparisons will be performed using the KM estimates and the respective variances according to the Greenwood's formula to construct a standard normally distributed Z-statistic. Let $\widehat{OS}_{i1}$ and $\widehat{OS}_{i2}$ denote the KM estimates of the 12-month OS rates for the two arms from Stratum i (Arm 1: DSP-7888+Bev; Arm 2: Bev only) and $\hat{\sigma}_{i1}$ and $\hat{\sigma}_{i2}$ denote the standard errors. The stratified test statistic T_{OS12} is obtained by

$$T_{OS12} = \frac{\sum_i \left(log(-log[\widehat{OS}_{i1}]) - log(-log[\widehat{OS}_{i2}]) \right)}{\sqrt{\sum_i \left(\left(\frac{\hat{\sigma}_{i1}}{log[\widehat{OS}_{i1}]} \right)^2 + \left(\frac{\hat{\sigma}_{i2}}{log[\widehat{OS}_{i2}]} \right)^2 \right)}}$$

which asymptotically follows a standard normal distribution N(0,1) under the null hypothesis.

$$\hat{\sigma}_{ik}^2 = \sum_{t_j \le t} \frac{d_{jik}}{Y_{jik}(Y_{jik} - d_{jik})}$$

where $k = 1$ indicates Arm 1, and $k = 2$ indicates Arm 2, d_{jik} is the number of events in strata i and Arm k and Y_{jik} is the number of patients at risk at event time t_j in strata i and Arm k (Klein 2007). Please note the log-log transformation is not appropriate when either of $\widehat{OS}_{i1}$ or $\widehat{OS}_{i2}$or both are 0 or 1. In this case the linear transformation will be used. The test statistic is defined as

$$T_{OS12} = \frac{\sum_i (\widehat{OS}_{i1} - \widehat{OS}_{i2})}{\sqrt{\sum_i \left((\widehat{OS}_{i1}\hat{\sigma}_{i1})^2 + (\widehat{OS}_{i2}\hat{\sigma}_{i2})^2 \right)}}$$

10.3. Other Secondary Efficacy Endpoint and Analyses

10.3.1. Progression-Free Survival (PFS)

This is defined as the interval between randomization and disease progression, determined by central review, or death from any cause, whichever comes first. The disease progression is based on mRANO criteria with confirmed PD. See Section 8.4.4 for the detailed events/censoring rules for PFS. The PFS based on the investigator assessment will be used as the primary analysis at IA and used as a sensitivity analysis to assess the robustness of the study conclusion on PFS at FA. Progression-free survival will be analyzed similar as OS.

The 6-Month Progression Free Survival

The six-month PFS rate is defined as the proportion of patients alive at 6 months after randomization and without progressive disease. Six-month PFS will be estimated considering only confirmed progressions by central review or death as events. Six-month PFS will be analyzed similarly to 12-month overall survival rate. The six-month PFS rate based on investigator assessment will be used as the primary analysis at IA and maybe used as a sensitivity analysis at FA.

10.3.2. Best Overall Response, Objective Response Rate (ORR) and Disease Control Rate

The best overall response (BOR) is the best response designation recorded from randomization until end of study or start of new anti-cancer treatment whichever occurs first. The best overall response will be derived based on following order: Confirmed CR > Confirmed PR > Preliminary CR > Preliminary PR > Pseudo Response > SD > confirmed Pseudo Progression > Preliminary PD > Confirmed PD. All other cases will be categorized as NE. The reasons for NE will be summarized below:

- No baseline assessment

- New anti-cancer therapy started before first post baseline assessment
- Death before mRANO assessment post baseline
- No post baseline assessment or all post-baseline assessment have overall response of NE/NA
- SD too early (< 28 days) since randomization date)
- PD* too late (> 119 days) since randomization date)

*PD too late (> 119 days) since randomization date will include preliminary PD and confirmed PD.

Pictorial presentations to examine changes in tumor burden such as spider and waterfall plots will also be provided.

The ORR is defined as the proportion of patients who achieve Confirmed CR or Confirmed PR based on the mRANO criteria as determined by the central review. Confirmed CR or Confirmed PR by the local investigator assessment will be used at IAs and may be included as supportive analysis at FA. The absolute and relative frequency of ORR will be presented by treatment group and within each randomization stratum (KPS 60 through 70, KPS 80 through 100, and extent of surgical resection [GTR versus not GTR]). Comparison of responses among the treatment groups will be done using stratified 1-sided Cochran-Mantel-Haenszel method, based on actual stratification factors. Treatment Difference of ORR and its 95% CI based on harmonic mean method adjusting actual stratification factors will be provided. For the analysis of ORR in the subgroups, differences between the two arms may be compared using a 1-sided Z-test via normal approximation. Treatment difference and its 95% CI based on normal approximation (unstratified) may be provided.

Disease Control Rate (DCR) is defined as the proportion of patients with BOR of Preliminary CR, Confirmed CR, Preliminary PR, Confirmed PR, Pseudo Response, Confirmed Pseudo Progression or SD.

BOR/ORR/DCR will be based on the central review at FA. Local investigator assessment will be used at IAs and may be included as supportive analysis at FA. DCR will be analyzed similarly as ORR.

10.3.3. Duration of Response

The duration of response (DOR) is defined as the interval between first documented response (Confirmed CR and Confirmed PR) based on the mRANO criteria as determined by the central radiology review and progression of disease or death from any cause. DOR will only be calculated for patients with ORR. Censoring rules is identical to the censoring rules presented for PFS. Central review assessment will be used at FA and local investigator assessment will be used at IAs and may be included as supportive analysis at FA. Descriptive analysis for DOR based on KM method may be included.

10.4. Exploratory Efficacy Endpoint and Analyses

The overall NANO score categories (Neurologic Response, Neurological Progression, Neurological Stability) will be summarized by visit and will be listed as appropriate. Overall NANO score changes over time may be evaluated graphically.

Corticosteroid systemic use (prior to treatment, on or after treatment, both prior and on or after treatment, or no prior or on/after treatment) will be summarized by arm for ITT set. Duration of Patients remaining off Corticosteroids for treatment of GBM will be summarized by arm. Corticosteroids use over time may be tabulated or presented graphically.

11. ANALYSIS OF BIOMARKER DATA

Biomarker data are exploratory endpoints and may include but are not limited to the following:

- Measurement of WT1 peptide-specific CTL induction activity
- Expression of WT1 via Chromogenic in situ Hybridization (CISH) and/or WT1 protein via immunohistochemistry (IHC) in tumor tissue

Analysis of biomarker data will be exploratory in nature.

12. SAFETY

The population used for safety analyses will be the Safety Set (SS). Safety will be assessed based on treatment exposure, adverse event (AE) or adverse drug reactions (ADR) reports, clinical laboratory data, ECG parameters, physical examinations, and vital signs. Analyses of the safety parameters will be presented by visits and/or time points where applicable. For continuous variables, descriptive statistics will include the number of patients (n), mean, standard deviation, median, minimum, and maximum. Frequencies and percentages will be displayed for categorical data. Percentages by categories will be based on the number of patients with no missing data, i.e., will add up to 100%. Categorical variables with a non-zero number of missing values will include an additional "Missing" category with no percentage provided (the category will be excluded from the distribution).

No inferential testing will be performed on safety data, which will be summarized by dose level in Part 1 and by treatment group in Part 2 Safety Set.

For patients who have a dose reductions/modification, all AEs (due to drug or otherwise) will be assigned to the associated initial dose combination/dose level.

Safety data will be presented graphically as deemed appropriate. This may include, but is not restricted to, presentation of parameters against time or shift plots. Appropriate scatter- or box-plots will also be considered to investigate trends in parameters compared to baseline.

12.1. Extent of Exposure

Study drug exposure including cumulative dose and duration of exposure will be summarized by study part and treatment group. Duration of exposure (i.e., weeks on study drug) will be calculated as the number of weeks from first to last dose date, (last exposure date - first exposure date + 1)/7. The number of patients with a dose reduction or dose interruption in each dose group or treatment group will also be summarized. The relative dose intensity defined as the actual dose intensity (ADI) delivered divided by the planned dose intensity (PDI) multiplied by 100 will also be presented.

12.1.1. Exposure of DSP7888

- Intended Treatment Duration (weeks) = (last dose date of DSP7888 + intended timing OR last dose of bevacizumab + intended timing which comes later– first dose date of DSP7888 or bevacizumab which comes first)/7. Intended timing is 7 days for induction phase, and 14 days for consolidation phase and 28 days for maintenance phase for DSP7888, and 14 days for Bev.
- Actual Cumulative Dose (mg) = Sum of all administered dose of DSP7888 (mg)
- Actual Dose Intensity (mg/week), ADI = Actual Cumulative Dose (mg) / Intended Treatment Duration (weeks)
- Cumulative Planned Dose = Sum of all protocol-defined planned doses (mg) during the intended treatment duration period
- Planned Dose Intensity (mg/week), PDI = Cumulative planned dose (mg) / Intended Treatment Duration (weeks)

- Relative Dose Intensity (%), RDI = 100 * ADI (mg/week) / PDI (mg/week)

Dose exposure of DSP7888 will also be summarized by phase: Induction Phase, Consolidation Phase, Maintenance Phase.

For the induction phase:

- Intended Treatment Duration (weeks) in the induction phase = (first dose date of DSP7888 in the consolidation phase - first dose date of DSP7888 in the induction phase)/7 or (last dose date of DSP7888 in the induction phase - first dose date of DSP7888 in the induction phase + 7) /7 if patients didn't enter the consolidation phase
- Actual Cumulative Dose (mg) in the induction phase = Sum of all administered doses of DSP7888 (mg) in the induction phase
- Actual Dose Intensity (mg/week), ADI in the induction phase = Actual Cumulative Dose (mg) / Intended Treatment Duration (weeks) in the induction phase
- Relative Dose Intensity (%), RDI in the induction phase = 100* ADI (mg/week) / 10.5 (mg/week)

For the consolidation phase:

- Intended Treatment Duration (weeks) in the consolidation phase = (first dose date of DSP7888 in the maintenance phase - first dose date of DSP7888 in the consolidation phase) /7 or (last dose date of DSP7888 in the consolidation phase - first dose date of DSP7888 in the consolidation phase + 14) /7 if patients didn't enter the maintenance phase
- Actual Cumulative Dose (mg) in the consolidation phase = Sum of all administered doses of DSP7888 (mg) in the consolidation phase
- Actual Dose Intensity (mg/week), ADI in the consolidation phase = Actual Cumulative Dose (mg) /Intended Treatment Duration (weeks) in the consolidation phase
- Relative Dose Intensity (%), RDI in the consolidation phase = 100* ADI (mg/week) / 5.25 (mg/week)

For the maintenance phase,

- Intended Treatment Duration (weeks) in the maintenance phase = (last dose date of DSP7888 in the maintenance phase - first dose date of DSP7888 in the maintenance phase + 28) /7
- Actual Cumulative Dose (mg) in the maintenance phase = Sum of all administered doses of DSP7888 (mg) in the maintenance phase
- Actual Dose Intensity (mg/week), ADI in the maintenance phase = Actual Cumulative Dose (mg) / Intended Treatment Duration (weeks) in the maintenance phase
- Relative Dose Intensity (%), RDI in the maintenance phase = 100* ADI (mg/week) / 2.625 (mg/week)

12.1.2. Exposure of Bevacizumab

- Intended Treatment Duration (weeks) = (last dose date of DSP7888 + intended timing OR last dose of bevacizumab +intended timing which comes later – first dose date of

DSP7888 or bevacizumab which comes first). Intended timing is 7 days for induction phase, and 14 days for consolidation phase and 28 days for maintenance phase for DSP7888, and 14 days for Bev.

- Actual Cumulative Dose (mg/kg) = Sum of all administered doses of bevacizumab (mg/kg) that the patient received. Actual dose received (mg) for each visit will be converted to dose in mg/kg by dividing the weight. The weight is derived on the last available weight prior to or on each date of Bevacizumab dosing.
- Actual Dose Intensity (mg/kg/week), ADI = Actual Cumulative Dose (mg/kg) / Intended Treatment Duration (weeks)
- Planned Dose Intensity (mg/kg/week), PDI = 5 mg/kg/week
- Relative Dose Intensity (%), RDI = 100* ADI (mg/kg/week) / PDI (mg/kg/week)

12.2. Treatment Compliance

N/A. All patients will be administered study drug in clinic

12.3. Adverse Events / Adverse Drug Reactions

Adverse events will be collected throughout the study, commencing from the time the informed consent form is obtained until 30 days after the last dose of study drug. When an SAE that is assessed as causally related (i.e., Definite, Probable, or Possible) occurs after completion of the clinical study, the Investigator will report the event to the Sponsor as a spontaneous report.

All adverse events that occur after administration of the first dose of study drug or that are present at baseline but worsen in severity after the first dose of study drug up to 30 days after the last dose of study treatment (or up to any time if a serious AE and considered related to study drug) will be considered as Treatment-Emergent Adverse Events (TEAEs). Any AEs that is not TEAE but occur after 30 days during the Post-Protocol Therapy will be considered as Post-AEs. The emphasis for AE analysis will be based on treatment-emergent AE (TEAE), however all AEs will be listed regardless of TEAE or not.

Adverse events will be summarized by dose level for Part 1 and by arm for Part 2.

Adverse events will be summarized by the SOC and PT based on the MedDRA dictionary version 19.1 or later.

For summaries of AEs by SOC and PT, a patient will be counted once at the SOC level and once at each PT within the SOC level. For summaries by SOC, PT, and maximum CTCAE grade, a patient will be counted once at the highest CTCAE grade level for which the event occurred at the SOC level and the highest CTCAE grade level for each unique PT within that SOC level. Therefore, patients may only contribute once to each PT and once to each SOC level summary.

The summaries presenting frequency of AEs by SOC and PT will be ordered by descending frequency of SOC and then, within an SOC, by overall descending frequency of PT in DSP-7888 Dosing Emulsion plus Bev arm.

The following AE tables will be provided:

- An overall summary of the number and percentage of patients reporting TEAEs, treatment-related TEAEs, serious TEAEs, Grade 3 (or worse) TEAEs, AEs considered as DLTs (Section 4.1.1.1 for definition of DLTs), TEAEs leading to dose withdrawn, and fatal AEs
- TEAEs by SOC and PT
- TEAEs by PT
- Serious TEAEs, by SOC and PT
- Serious TEAEs, by PT
- DLTs
- Study-treatment-related AEs, by SOC and PT
- TEAEs by CTCAE Grade, by SOC and PT
- TEAEs by Worst Grade
- TEAEs leading to death, by SOC and PT
- TEAEs of grade 3 or higher, by SOC and PT
- TEAEs of grade 3 or higher on worst grade and by PT
- TEAEs leading to withdrawal of study medication, by SOC and PT.
- TEAEs leading to dose modification or interruption of study medication by SOC and PT

For patients who have a dose modification, all AEs (due to drug or otherwise) will be assigned to the initial dose group.

Listings will be provided for AEs marked as DLTs. TEAEs will be flagged in listings.

In addition, Exposure Adjusted Incidence Rate (EAIR; He et al., 2015) of selected AEs may be summarized.

Serious Adverse Event (SAE) and Death

Treatment emergent SAEs and treatment emergent SAEs related to DSP-7888 will be summarized by SOC, PT and maximum CTCAE grade.

The number and percentages of patients who died on treatment, as well as the primary cause of death will be summarized. On-treatment death is defined as death within 30 days of last dose of study drug (DSP-7888 or Bev). A listing of death data will also be provided.

Injection Site Reaction (ISR)

The predominant adverse effect observed with DSP-7888 is injection site reaction. The Preferred Terms of TEAEs within the High Level Group Term of “Administration Site Reaction” (MedDRA version 19.1) will be used to identify events of injection site reaction occurring in the pivotal study and across the program.

Below analysis will be provided for ISR:

- ISR by SOC, PT, and maximum CTCAE Grade

- ISR by SOC, PT, and maximum CTCAE Grade by subgroup of Geographic Region, Race, Country, and HLA type.
- Onset time of first ISR for patients with at least one ISR: Time to onset is analyzed for patients with at least one ISR from the first dose of the DSP-7888 to the first occurrence of ISR regardless of CTCAE grade.
- Onset time of first ISR: Time to first ISR is defined as the time from the first dose of the DSP-7888 to the first occurrence of ISR . For patients with no ISR, the onset time will be censored as data cutoff date, end of study date, death date, or safety follow-up completion date (end of treatment date + 30 days) whichever is shorter.
- Onset time of the first Grade 3 or higher ISR for patients with at least one Grade 3 or higher ISR: Time to onset is analyzed for patients with at least one Grade 3 or higher ISR from the first dose of DSP-7888 to the first occurrence of the Grade 3 or higher ISR.
- Onset time of the first Grade 3 or higher ISR: Time to first Grade 3 or higher ISR is defined as the time from the first dose of DSP-7888 to the first occurrence of the Grade 3 or higher ISR. For patients with no Grade 3 or higher ISR, the onset time will be censored as data cutoff date, end of study date, death date, or safety follow-up completion date (end of treatment date + 30 days) whichever is shorter.
- Duration of Grade 3 or higher ISR and Duration of ISR may be analyzed in the CSR.

A listing of ISR will also be provided.

12.4. Laboratory Evaluation

All samples will be analyzed at a central laboratory and collected per Schedule of Assessments in Section 3.9. Local laboratories may be used as deemed necessary by the Investigator in order to make treatment related decisions.

The results of laboratory parameters will be graded according to NCI CTCAE v4.03 for below parameters (Table 8). For patients with both baseline and post baseline assessments, a summary of shift from baseline CTCAE grade to maximum postbaseline CTCAE grade will be provided. All records on or after the first dose (scheduled and unscheduled) until 30 days following the last dose of study drug will be considered as post baseline. Additional parameters may be added for the shift tables. Both scheduled and unscheduled laboratory test results will be included in the shift tables.

Table 8: CTCAE Lab Grading Parameters.

	↓	↑
HEMATOLOGY	Anemia	
HEMATOLOGY	White blood cell decreased	
HEMATOLOGY	Lymphocyte count decreased	
HEMATOLOGY	Neutrophil count decreased	
HEMATOLOGY	Platelet count decreased	
CHEMISTRY		Alanine aminotransferase increased

CHEMISTRY	Hypoalbuminemia	
CHEMISTRY		Alkaline phosphatase increased
CHEMISTRY		Aspartate aminotransferase increased
CHEMISTRY		Blood bilirubin increased
CHEMISTRY		Blood Urea Nitrogen Increased
CHEMISTRY	Hypocalcemia	Hypercalcemia
CHEMISTRY		Creatinine increased
CHEMISTRY	Hypoglycemia	Hyperglycemia
CHEMISTRY	Hypomagnesemia	Hypermagnesemia
CHEMISTRY	Hypophosphatemia	
CHEMISTRY	Hypokalemia	Hyperkalemia
CHEMISTRY	Hyponatremia	Hypernatremia

Blood Urea Nitrogen Increased is not in CTCAE grading, SDPO will define as use mg/dL as grade 1 is 23 - 26, grade 2 is 27-31, grade 3, >31

Listings of hematology and biochemistry, urinalysis and coagulation will be provided, including the test result, units, normal range, change from baseline, and CTCAE grade if available.

An eDISH (evaluation of Drug-Induced Serious Hepatotoxicity) plot for patients in Part 1 and Part 2 will be provided. A listing of liver function test results that corresponding to the eDISH plot will be provided by patients and potential Hy's law cases if any will be flagged in the listing.

12.5. Vital Signs

Body temperature will be summarized in °C. If body temperature is recorded as °F, then temperature will be converted to °C using:

Temperature (°C) = 5/9 (Temperature [°F]-32).

Summaries of post baseline markedly abnormal vital signs, including systolic blood pressure, diastolic blood pressure, pulse rate, weight change, and temperature will be presented. Markedly abnormal ranges for vital signs parameters are given in Table 9 . Post baseline includes non-missing records from both scheduled and unscheduled visits after the first dose of study drug until 30 days following the last dose of study treatment.

Table 9: Markedly Abnormal Ranges for Vital Signs

Parameter	Markedly Abnormal (Low)	Markedly Abnormal (High)
Systolic Blood Pressure	• Absolute value ≤ 90 mmHg for post baseline, or • a decrease from baseline ≥20 mmHg for change from baseline	• Absolute value ≥ 180 mmHg for post baseline, or • an increase from baseline ≥20 mmHg for change from baseline
Diastolic Blood Pressure	• Absolute value ≤ 50 mmHg for post baseline, or • a decrease from baseline ≥15 mmHg for change from baseline	• Absolute value ≥ 105 mmHg for post baseline, or • an increase from baseline ≥ 15 mmHg for change from baseline
Pulse	• Absolute value ≤ 50 bpm for post baseline, or • a decrease from baseline ≥ 15 bpm for change from baseline	• Absolute value ≥ 120 bpm for post baseline, or • an increase from baseline ≥ 15bpm for change from baseline
Weight Change	Weight loss from baseline ≥ 20%	Weight gain from baseline ≥ 20%
Temperature	≤ 35°C post baseline only	≥ 40°C post baseline only

A listing of vital signs results will also be presented using the Safety Set.

12.6. Electrocardiogram

A 12-lead Electrocardiogram (ECG) with categorical results (normal; abnormal, not clinically significant; abnormal, clinically significant) will be summarized in the shift summary from baseline to worst post baseline. Post baseline includes non-missing records collected (scheduled and unscheduled) after the first dose of study treatment until 30 days following the last dose of study treatment. QT measurements corrected by heart rate will be used for the data analysis and interpretation. QTc interval will be calculated using Fridericia's correction. The formula is:

Fridericia's correction: $QT_cF = QT / RR^{0.33}$

where unit of QT is milliseconds (msec) and unit of RR is second (sec). The number and percentage of patients in QT and QTcF (QT and QTcF interval: ≤ 450 msec, > 450 - ≤ 480 msec, > 480 - ≤ 500 msec, and > 500 msec) will be summarized at baseline and maximum post baseline. Categories of maximum changes from baseline (≤ 30 msec, > 30 - ≤ 60 msec, > 60 msec) will be summarized as well.

An ECG result listing that including QT, QTc, and Heart Rate will also be provided.

12.7. Karnofsky Performance Score

Summary tables by visit and change from baseline by visit for Karnofsky performance assessment (see Table below) will be provided.

All Karnofsky performance scores will be listed by treatment group and dose level.

12.8. Physical Examination

Physical examination, including documentation of body system, normal/abnormal and abnormality description and neurological examination will be presented in individual subject data listings.

12.9. Echocardiogram/Multigated Acquisition Scan

Echocardiogram and multigated acquisition scan (MUGA) will be assessed at baseline. Results including LVEF (%) will be listed.

12.10. Oxygen Saturation

Oxygen saturation will be assessed at baseline. Results will be listed.

13. CHANGE FROM ANLAYSIS PLANNED IN PROTOCOL

No major change from planned analyses was made.

14. REFERENCE

1. Protocol BBI-DSP7888-201G Original Protocol 29MAR2017

2. Protocol BBI-DSP7888-201G Protocol Amendment 1 Version 2.0 FINAL 30APR2017

3. Protocol BBI-DSP7888-201G Protocol Amendment 2 Version 3.0 FINAL 18SEP2017

4. Protocol BBI-DSP7888-201G Protocol Amendment 3 Version 4.0 FINAL 06OCT2017

5. Protocol BBI-DSP7888-201G Protocol Amendment 4 Version 5.0 FINAL 13OCT2017

6. Protocol BBI-DSP7888-201G Protocol Amendment 5 Version 6.0 FINAL 08JUL2018

7. Protocol BBI-DSP7888-201G Protocol Amendment 6 Version 7.0 FINAL 24JUN2019

8. Protocol BBI-DSP7888-201G Protocol Amendment 7 Version 8.0 FINAL 05FEB2021

9. Independent Radiology Review Imaging Charter Version 2 22JUL2020

10. Cui, L., Hung, H. M. J., Wang, S. (1999). Modification of sample size in group sequential clinical trial. Biometrics. 55, 853-857.

11. Shen, Y., Cai, J. (2003). Sample size re-estimation for clinical trials with censored survival data. Journal of the American Statistical Association. 98, 418-426.

12. Wassmer G, Brannath W. Group Sequential and Confirmatory Adaptive Designs in Clinical Trials. New York: Springer, 2016.

13. Willi Maurer & Frank Bretz (2013) Multiple Testing in Group Sequential Trials Using Graphical Approaches, Statistics in Biopharmaceutical Research, 5:4, 311-320, DOI: 10.1080/19466315.2013.807748

14. He X, Chen L, Lei L, Xia HA, Lee MLT (2015) A Simple Method for Estimating Confidence Intervals for Exposure Adjusted Incidence Rate and Its Applications to Clinical Trials. J Biom Biostat 6: 238.

15. He P, Fang L, Su Z. A sequential testing approach to detecting multiple change points in the proportional hazards model. Statistics in Medicine.2013; 32:1239-1245.

16. He P, Su Z. A novel design for randomized immuno-oncology clinical trials with potentially delayed treatment effects. Contemporary Clinical Trials Communications. 2015;1:28-31

17. Klein, J.P., Logan, B., Harhoff, M. and Andersen, P.K. (2007), Analyzing survival curves at a fixed point in time. Statist. Med., 26: 4505-4519.

18. Zucker DM, Lakatos E. Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. Biometrika. 1990; 77:853-864.

19. Xu Z, Park Y, Zhen B, Zhu B. Designing cancer immunotherapy trials with random treatment time-lag effect. Statistics in Medicine. 2018; 37: 4589-4609

15. APPENDIX

15.1. Tipping Point Sensitivity Analysis

To quantify the robustness of the primary study objective to departures from the censoring at random (CAR) assumption, sensitivity analyses under the censoring not at random (CNAR) assumption will be conducted for patients who discontinued from the study, e.g., lost to follow up.

Sensitivity analyses will be conducted using a multiple imputation (MI) framework with a Bayesian Piecewise Exponential Proportional Hazards model (PEPHM) as the imputation model and the primary analysis model (e.g., the Cox proportional hazards model) as the analysis model. Sensitivity analysis will proceed as follows.

1. *Fitting Bayesian PEPHM* with the number of intervals *k* from 3 to 10 to the analysis data set using PROC PHREG in SAS (BAYES statement), separately by treatment arm with the same covariates that are used in the primary analysis model. The cut-points are determined in a data-driven way to ensure an about equal number of events per interval within each treatment arm.
2. *Sampling from posterior distribution* of the model parameters (interval-specific hazards and regression coefficients, *m*=1,000 draws).
3. *Generating event times* for patients who discontinued early from the study (lost to follow up) using imputation from the posterior predictive distribution of missing data for the same arm, with covariate-specific hazard functions multiplied by a treatment-specific sensitivity parameter δ_i $(i = 0,1)$. For each draw from the posterior predictive distribution, simulate missing event times for each patient starting from the time this patient discontinued from the study (C_i) and retain this value as an imputed time to event or censor at the end of the follow-up period (T_{max}) if it exceeds T_{max}. This results in forming *m* imputed data sets.
4. *Analyzing* each imputed data set using the primary analysis model and computing the log-hazard ratios, log(HR), and associated standard errors for the treatment effect
5. *Combining* the m estimated log-hazard ratios and associated standard errors within single MI-inference, including the point estimate, confidence interval and p-value using Rubin's combination rules (using PROC MIANALYZE).

A tipping point analysis will be conducted by executing Steps 1-5 for different values of the sensitivity parameters δ_0 and δ_1, for the control and treatment arm, respectively. For each setting, a p-value is computed, and values of the sensitivity parameters are identified when the null hypothesis fails to be rejected (assume original test without imputation reject the null hypothesis). Most of time, the delta is varied only in the treatment arm ($\delta_1 > 1$), which implies an increased hazard for patients after their discontinuation from the treatment, as prior to discontinuation. both δ_0 and δ_1 can vary resulting in a two-dimensional map for the sensitivity analysis p-values. $\delta > 1$ implies an increased hazard for patients after their discontinuation from the treatment, while <1 implies a decreased hazard after discontinuation.

Another type of sensitivity analysis that can be easily implemented using the outlined framework is *control-based* imputation. For this type of sensitivity analysis, times to event in both arms (in Step 3) are computed from the survival distribution estimated from the control arm (Step 1).

The details on Steps 1-5 are provided as follows.

15.1.1. Fitting a Bayesian Piecewise Exponential Model to the Analysis Data Set

The imputation model will have a hazard function $h(t)$ incorporating p baseline covariates, $X_1, X_2, .., X_p$, i.e.,

$$h(t) = h_0(t)\exp(X^T\boldsymbol{\beta}),$$

Where

- *t* is the time since randomization.
- Piecewise baseline hazard $h_0(t) = \sum_{i=1}^{k} \lambda_{0i} I(\tau_{i-1} < t \leq \tau_i), i = 1, .., k$, the k intervals defined as $0 = \tau_0 < \tau_1 < \cdots < \tau_{k-1} < \tau_k = \infty$, λ_{0i} are constant baseline hazards for the *i*th interval and $I(\cdot)$ is the indicator function.
- Default diffuse priors for λ_{0i} are gamma G (10^{-4}, 10^{-4}), i.e., the prior mean is set to 1 and the prior variance is set to 10^4. Default priors for the regression coefficients $\boldsymbol{\beta}$ are independent normal $N(0,10^6)$. Sampling from the posterior distribution of the parameters λ_{0i} and $\boldsymbol{\beta}$ will be carried out via MCMC with default options for the number of burn-in iterations (NBI), number of iterations (NMC) and thinning (THINNING). If the diagnostics suggest the presence of autocorrelation, the thinning parameter should be increased accordingly.

15.1.2. Generating Event Times for Patients Lost to Follow up

Event times are generated from the piecewise exponential distribution using the inversion method that can be easily implemented with SAS custom code following the steps detailed below for each trial arm:

1. Set a sensitivity parameter $\delta \geq 1$. Note that setting $\delta = 1$ corresponds to imputation under CAR.
2. Draw $\tilde{\lambda}_{0j}, j = 1, .., k$ and $\tilde{\boldsymbol{\beta}}$ from the posterior distributions of $\boldsymbol{\lambda}_0$ and $\boldsymbol{\beta}$ (as explained in Section 1).
3. For patients to be imputed, evaluate $S_{\tilde{\boldsymbol{\lambda}}_0,\tilde{\boldsymbol{\beta}}}(c_i|\boldsymbol{x}_i)$, i.e., the survival function at c_i, the time when the *i*th patient was censored (early discontinued), with the parameters $\tilde{\boldsymbol{\lambda}}_0, \tilde{\boldsymbol{\beta}}$. Let $\tau_{j-1} < c_i \leq \tau_j$, $j = 1, .., k$ (τ_0=0 and $\tau_k = \infty$). Then
$$S_{\tilde{\boldsymbol{\lambda}}_0,\tilde{\boldsymbol{\beta}}}(c_i|\boldsymbol{x}_i) = \prod_{l=1}^{j-1} \exp\left(-(\tau_l - \tau_{l-1})\delta\tilde{\lambda}_{0l}\exp\left(\boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}\right)\right) \times \exp\left(-\left(c_i - \tau_{j-1}\right)\delta\tilde{\lambda}_{0j}\exp\left(\boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}\right)\right)$$
4. For the *i*th censored patient, simulate uniform values from $u_i \sim U(p_i, 1)$, where $p_i = 1 - S_{\tilde{\boldsymbol{\lambda}}_0,\tilde{\boldsymbol{\beta}}}(c_i|\boldsymbol{x}_i)$.
5. Compute the event time $\tilde{t}_i$ by using the inverse of the CDF for the time to event evaluated at $\tilde{\boldsymbol{\lambda}}_0, \tilde{\boldsymbol{\beta}}$

$$\tilde{t}_i = \begin{cases} -\dfrac{\log(1-u_i)}{\lambda_1(\boldsymbol{x}_i)}, 0 < u_i < 1 - S_1(\boldsymbol{x}_i) \\ \tau_{j-1} - \dfrac{\log(1-u_i) + \sum_{l=1}^{j-1}(\tau_l - \tau_{l-1})\lambda_l(\boldsymbol{x}_i)}{\lambda_j(\boldsymbol{x}_i)}, 1 - S_{j-1}(\boldsymbol{x}_i) \le u_i < 1 - S_j(\boldsymbol{x}_i), j = 2,..,k-1 \\ \tau_{k-1} - \dfrac{\log(1-u_i) + \sum_{l=1}^{k-1}(\tau_l - \tau_{l-1})\lambda_l(\boldsymbol{x}_i)}{\lambda_k(\boldsymbol{x}_i)}, 1 - S_{k-1}(\boldsymbol{x}_i) \le u_i \le 1 \end{cases}$$

where the piecewise constant covariate adjusted hazards are $\lambda_j(\boldsymbol{x}_i) = \delta\tilde{\lambda}_{0j}\exp(X^T\tilde{\boldsymbol{\beta}})$, and the associated survival function is $S_j(\boldsymbol{x}_i) = \prod_{l=1}^{j}\exp\big(-(\tau_l - \tau_{l-1})\lambda_l(\boldsymbol{x}_i)\big)$.
Note that the sensitivity parameter δ is incorporated into the hazard function $\lambda_j(\boldsymbol{x}_i)$. When performing imputation under the sensitivity analysis, δ can be treatment-specific, i.e., δ >1 in the treatment arm (assuming a higher hazard for patients after discontinuation compared to a patient with the same covariates who remained on treatment) and δ=1 in the control arm.

6 Compare the simulated time $\tilde{t}_i$with the maximum follow up time T_{max}. If the time exceeds T_{max}, censor the *i*th patient at T_{max}.

15.1.3. Analyzing Imputed Data

Each imputed data set is analyzed using the primary analysis model, e.g., the proportional hazards Cox model, for several values of δ. The resulting point estimates and standard errors are combined using PROC MIANALYZE to produce the final estimates of the treatment effect and p-values. The p-values are reported as a function of the sensitivity parameter δ and regions of δ's are found where the p-value becomes non-significant (tipping point analysis).