Statistical Analysis Plan

Protocol Title:	A Phase 1b Study Evaluating the Safety and Efficacy of AMG 757 in Combination with AMG 404 in Subjects with Small Cell Lung Cancer (SCLC)
Short Protocol Title:	AMG 757 and AMG 404 in Subjects with Small Cell Lung Cancer (SCLC)
Protocol Number:	20200439
NCT Number:	NCT04885998
Authors:	[redacted]
Sponsor:	Amgen, Inc. One Amgen Center Drive Thousand Oaks, CA, 91329, USA
SAP Date:	Original (v1.0) 6 September 2021 Amendment 1 (v2.0) 9 March 2022 Amendment 2 (v3.0) 27 March 2023 Amendment 3 (v4.0) 8 May 2023 Amendment 4 (v5.0) 14 August 2023 Amendment 5 (v6.0) 27 September 2024

Version Number	Date (DDMMMYYYY)	Summary of Changes, including rationale for changes
Original (v1.0)	06-September-2021	
[Amendment 1 (v2.0)]	09MAR2022	• Added section 7.1.1 as per the protocol amendment 2. • Replaced AMG 757 with tarlatamab as per the PA2. • Added the analysis sets as per PA2 except OR analysis set. • Updated table 7 to remove sensitivity analysis and added “Analysis set”. • Updated table 8 for the criteria 37.

[Amendment 2 (v3.0)]	27MAR2023	• Updated section 5 to add the definitions of Duration of disease control, modified the definitions of Duration of response, Progression-free survival, Overall survival, TEAE, treatment-related TEAE.
		• Updated section 6 to modify the definition of safety analysis set, interim efficacy analysis set.
		• Updated section 9.5 efficacy analysis table to add a column “Analysis Set”.
		• Updated 9.6.2 to include EOI related text.
		• Updated appendices
		  ○ A for the imputation of death dates.
		  ○ B for the text confirmed BOR by investigator, table 14-2.
		  ○ C to add PFS version.

[Amendment 3 (v4.0)]	08MAY2023	• Added ECG Analysis • Added definition for ECG baseline

		• Updated list of baseline characteristics parameters • Updated interim efficacy analysis set definition • Updated MedDRA version • Removed imputation rule for time to event end points from Appendix A
[Amendment 4 (v5.0)]	03AUGUST2023	• Updated ECG Analysis • Updated list of baseline characteristics parameters • Updated definition of TEAE for extended period • Added derivation logic for Last known alive date. • Added definition for TTR
[Amendment 5 (v6.0)]	27SEP2024	• Updated Study design (Section 3.1) for safety follow-up period • Updated TEAE definition (Section 5.0) to exclude the events directly related to lung cancer or disease progression • Updated the MedDRA version • Updated Adverse Events (Section 9.6.2) and Event of Interest list

Table of Contents

List of Table

List of Figures

List of Abbreviations

Abbreviation	Explanation
ACTH	Adrenocorticotropic hormone
ADA-IC	Anti-drug antibody related immune complexes
ADAs	Anti-drug antibodies
ADCC	Antibody-dependent cellular cytotoxicity
ADL	Activities of daily living
ALL	Acute lymphoblastic leukemia
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ASCO	American Society of Clinical Oncology
ASR	Annual Safety Report
AST	Aspartate aminotransferase
AUC	Area under the concentration-time curve
BID	Twice daily
BNP	B-Natriuretic Peptide
BP	Blood pressure
C	Cycle
CBC	Complete blood count
CI	Confidence intervals
C_{max}	Maximum serum concentration
C_{min}	Minimum serum concentration
CNS	Central nervous system
CRF	Case report form
CRP	C-reactive protein
CRS	Cytokine release syndrome
CSF	Cerebral spinal fluid
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
[redacted]	
DCR	Disease control rate
DLL3	Delta-like protein 3
DLRM	Dose level review meeting
DLRT	Dose level review team
DLT	Dose-limiting toxicity
DNA	Deoxyribonucleic acid
DOR	Duration of response

Abbreviation	Explanation
DSUR	Development Safety Update Report
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
EDC	Electronic data capture
EOI	End of infusion
EOT	End of treatment
eSAE	Electronic Serious Adverse Event
FFPE	Formalin-fixed paraffin-embedded
FIH	First-in-human
FSH	Follicle stimulating hormone
FT4	Free thyroxine
GCP	Good Clinical Practice
GI	Gastrointestinal
GSO	Global Safety Officer
GLP	Good Laboratory Practice
H	Hour
HBsAb	Anti-hepatitis B antibody
HBsAg	Hepatitis B antigen
HCV	Hepatitis C
HCVAb	Hepatitis C virus antibody
HIV	Human immunodeficiency virus
HLE	Half-life extended
HNSTD	Highest Non-Severely Toxic Dose
HR	Heart rate
HUVECs	Human endothelial cells
IB	Investigator Brochure
ICF	Informed consent form
ICH	International Council for Harmonisation
IEC	Independent Ethics Committee
IFN-γ	Interferon gamma
Ig	Immunoglobulin
IHC	Immunohistochemistry
IL-2	Interleukin-2
IM	Intramuscular

Abbreviation	Explanation
INR	International normalized ratio
IPIM	Investigational Product Instruction Manual
IRB	Institutional Review Board
IV	Intravenous
LDH	Lactate dehydrogenase
LH	Luteinizing hormone
LTFU	Long-term follow-up
LVEF	Left ventricular ejection fraction
MDRD	Modification of Diet in Renal Disease
MMSE	Mini Mental Status Examination
MRI	Magnetic resonance imaging
MTD	Maximum tolerated dose
mTPI	Modified toxicity probability interval
MUGA scan	Multigated acquisition scan
NCT	National Clinical Trials
NSAIDs	Non-steroidal anti-inflammatory drugs
OR	Objective response
OS	Overall survival
PAVA	Pool adjacent violators algorithm
PBL	Peripheral blood leukocytes
[redacted]	
PCR	Polymerase chain reaction
PD	Pharmacodynamics
PFS	Progression-free survival
PK	Pharmacokinetic(s)
PPP	Platelet Poor Plasma
PR	Partial Response
Pre	Pre-infusion
PT	Prothrombin time
PTT	Partial thromboplastic time
PTU	Propylthiouracil
RBC	Red blood cell
RECIST	Response Evaluation Criteria in Solid Tumors
RNA	Ribonucleic acid
RP2D	Recommended phase 2 dose
RR	Respiratory rate
SAE	Serious adverse event

Abbreviation	Explanation
SC	Subcutaneous
SCLC	Small Cell Lung Cancer
SCR	Screening
SFU	Safety follow-up
SGOT	Serum glutamic-oxaloacetic transaminase
SGPT	Serum glutamic-pyruvic transaminase
SJS	Stevens-Johnson syndrome
SOA	Schedule of Assessments
SSKI	Saturated solution of potassium iodide
T4	Thyroxine
TEN	Toxic epidermal necrolysis
TLS	Tumor lysis syndrome
TNF	Tumor necrosis factor
TSH	Thyroid stimulating hormone
ULN	Upper limit of normal
UPM	Unit probability mass
US	United States
USFDA	United States food and drug administration
VS	Vital signs
WBC	White blood cell

1. Introduction

The purpose of this Statistical Analysis Plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol amendment 5 for study **20200439**, dated 27 September 2024. The scope of this plan includes the interim analysis, the primary analysis and the final analysis that are planned and will be executed by the Amgen Global Biostatistical Science department unless otherwise specified. Pharmacokinetic, pharmacodynamic, exposure-response and biomarker analyses will be performed by Clinical Pharmacology Modeling and Simulation (CPMS) or biomarker group.

2. Objectives, Endpoints and Hypotheses

2.1 Objectives and Endpoints/Estimands

Objectives	Endpoints
Primary	
• To evaluate the safety, tolerability, and recommended phase 2 target dose of tarlatamab in combination with AMG 404	• Dose-limiting toxicities (DLTs), treatment-emergent and treatment-related adverse events, changes in vital signs, electrocardiograms (ECGs), and clinical laboratory tests
Secondary	
• To evaluate anti-tumor activity of tarlatamab in combination with AMG 404	• Objective response (OR) per modified response evaluation criteria in solid tumors (RECIST) v1.1, duration of response (DOR), disease control rate (DCR), progression-free survival (PFS), and overall survival (OS).
• To characterize the pharmacokinetics (PK) of tarlatamab in combination with AMG 404	• PK parameters including, but not limited to, maximum serum concentration (C_{max}), minimum serum concentration (C_{min}), and area under the concentration-time curve (AUC) over the dosing interval
Exploratory	
[redacted]	

Objectives	Endpoints
[redacted]	

2.2 Hypotheses and/or Estimations

No formal statistical hypothesis will be tested.

3. Study Overview

3.1 Study Design

This is a phase 1b, multicenter, open-label study evaluating the safety, tolerability, PK, pharmacodynamics (PD), and efficacy of tarlatamab in combination with AMG 404 in subjects with SCLC. The study will consist of dose exploration (Part 1) and dose expansion (Part 2). Study enrollment ended in [redacted]. A total of [redacted] subjects were enrolled including [redacted] subjects in Part 1, and [redacted] subjects in Part 2.

[redacted]

Part 1 (Dose Exploration)

The dose exploration part of the study will estimate the recommended phase 2 dose (RP2D) of tarlatamab in combination with AMG 404 using a modified toxicity probability interval (mTPI-2) design. A combination RP2D may be identified based on emerging safety, efficacy, and pharmacodynamic data prior to reaching an MTD.

AMG 404 will be administered as a short-term IV infusion (30 minutes) every [redacted] (± 3 days) (note if AMG 404 is initially administered on cycle 1 [redacted] there will be a [redacted] interval between the cycle 1 [redacted] and cycle 2 [redacted] dose). AMG 404 will be administered at the recommended phase 2 dose of [redacted] mg throughout the study. The starting dose of tarlatamab will be [redacted] mg with 1-step dosing ([redacted] mg on cycle 1 [redacted] followed by [redacted] mg on cycle 1 [redacted] [redacted] mg on cycle 1 [redacted] and [redacted] mg [redacted] thereafter). Thus far in the ongoing FIH study (Study 20160323), the highest target dose deemed safe and tolerable is [redacted] mg. Planned dose levels in Study 20160323 are

shown in **Table 1**. The highest planned target dose of tarlatamab will not exceed [redacted] mg in this study using the [redacted] regimen.

Table 1. Planned Target Doses per Dose Cohort Level

Dose Cohort Levels	Dose (mg) IV [redacted]
-1	
1	
2	
3	

IV = intravenous

Footnotes are referenced from first-in-human study (Study 20160323)

[redacted]

To mitigate the risk of CRS and to potentially optimize the PD activity of tarlatamab, a step dosing approach with 1-step dosing will be implemented as part of the dosing schedule. Based on emerging safety data and the recommendations of the DLRM, the dosing schedule may be adjusted to allow for AMG 404 to be administered initially on cycle 1 [redacted] or cycle 1 [redacted] Depending on which day AMG 404 is administered on in cycle 1, beginning in cycle 2, AMG 404 will be administered [redacted] beginning on cycle 2 [redacted] or cycle 2 [redacted] (refer the Table 1-6 and Table 1-7 of protocol section 1.3 for more details).

Part 1 may include one or more of the following planned dose levels of tarlatamab in combination with a fixed dose of AMG 404:

- Dose cohort Level 1: tarlatamab at a target dose of [redacted] mg administered IV [redacted] (with 1-step dosing) in combination with AMG 404 at [redacted] mg IV [redacted] beginning on cycle 1 [redacted]
- Dose cohort Level 2: tarlatamab at a target dose of [redacted] mg administered IV [redacted] (with 1-step dosing) in combination with AMG 404 at [redacted] mg IV [redacted] beginning on cycle 1 [redacted]

- Dose Cohort Level 3: tarlatamab at a target dose of [redacted] mg administered IV [redacted] (with 1-step dosing) in combination with AMG 404 at [redacted] mg IV [redacted] beginning on cycle 1 [redacted]
- Dose cohort Level -1: tarlatamab at a target dose of [redacted] mg administered IV [redacted] (with 1-step dosing) in combination with AMG 404 at [redacted] mg IV [redacted] (in case Dose Cohort Level 1 is not well tolerated) beginning on cycle 1 [redacted]
- Dose cohort Level -2: tarlatamab at a target dose of [redacted] mg administered IV [redacted] (with 1-step dosing) in combination with AMG 404 at [redacted] mg IV [redacted] (in case Dose Cohort Level -1 is not well tolerated) beginning on cycle 1 [redacted]
- Dose cohort Level -3: tarlatamab at a target dose of [redacted] mg administered IV [redacted] (with 1-step dosing) in combination with AMG 404 at [redacted] mg IV [redacted] (in case Dose Cohort Level -2 is not well tolerated) beginning on cycle 1 [redacted]

Based upon emerging PK, PD and safety data in the current study as well as the ongoing FIH study (20160323), alternative (intermediate) dose cohort levels, [redacted] as part of the de-escalation recommendations per the DLRM, or alternative dosing schedule(s), including additional step dosing strategies of tarlatamab, may be explored.

Step Dosing

Due to its known mechanism of action, subjects are at an increased risk for CRS during initiation of tarlatamab treatment. To mitigate this risk, tarlatamab will be administered with 1-step dosing. One-step dosing involves a first step dose on [redacted] mg), followed by a step dose on [redacted] (equal to the target dose) and the target dose on [redacted] then [redacted]

Dose exploration will begin with [redacted] subjects treated at Dose Cohort Level 1. The study DLT period is [redacted]. Once all subjects enrolled at a certain Dose Cohort Level are DLT evaluable, a DLRT meeting will be convened. Depending on observed safety data, the following may occur: 1) dose de-escalation to the next lowest Dose Cohort Level, 2) additional enrollment to the current Dose Cohort Level, or 3) dose escalation to the next highest Dose Cohort Level or initiation of enrollment in Part 2. Re-escalation

to the next higher Dose Level may be allowed, as appropriate. If re-escalation occurs, alternative (intermediate) dose levels may be explored per the recommendations of the DLRM, [redacted]

Dose escalation/de-escalation recommendations will be guided by a mTPI-2 model (Guo et al, 2017) with a target toxicity probability of 30%, equivalence toxicity interval of (25%, 33%) and probability of overdosing of 95%. Beta (1, 1) is used as a prior distribution.

The detailed information of escalation/de-escalation rules are defined in **Table 2**. The DLT decision starts from the [redacted] subject.

Table 2. Guideline for Escalation /De-escalation

Number of Evaluable Subjects	Number of DLTs	Decision
[redacted]	0	E
	1-2	D
	3	DU
[redacted]	0	E
	1	S
	2	D
	≥3	DU
[redacted]	0-1	E
	2-3	D
	≥4	DU
[redacted]	0-1	E
	2-3	D
	≥4	DU
[redacted]	0-1	E

	2	S
	3-4	D
	≥5	DU
[redacted]	≤2	E
	3-4	D
	≥5	DU
	≤2	E
	3	S
	4-5	D
	≥6	DU

D = de-escalate to the next lower dose level; DLT = dose-limiting toxicity; DU = the current dose is unacceptably toxic; E = escalate to the next higher dose; S = stay at the current dose

If late onset adverse events occur during a cohort, the DLRT may adaptively re-consider the doses evaluated within a cohort for subsequent dosing and/or possibly trigger a de-escalation or withholding of additional doses in subsequent cohorts. Additionally, if dose de-escalation occurs and based on emerging PK, PD and safety data, alternative (intermediate) dose levels may be explored per the recommendations of DLRM, [redacted]. The recommended phase 2 target dose will be estimated using isotonic regression (Ji et al, 2010) and recommended phase 2 target dose will be the dose level with estimated DLT rate closest to 0.30. In order to consider a certain dose level as the recommended phase 2 target dose at least [redacted] DLT evaluable subjects must be enrolled at the dose level.

Dose exploration phase will end once any of the following events occur:

- Highest planned dose level is determined to be safe and tolerable (minimum of [redacted] evaluable subjects overall).
- One of dose levels is determined to be safe and tolerable (minimum of [redacted] evaluable subjects) and the next higher dose level is determined to be unsafe and intolerable.
- All planned dose levels (including any intermediate doses or alternate dosing schedules) are determined to be unsafe and intolerable.

To minimize the number of subjects treated at potentially toxic dose levels, the first [redacted] subjects in each cohort will be enrolled sequentially with a minimum of [redacted] days between the first dose of these subjects.

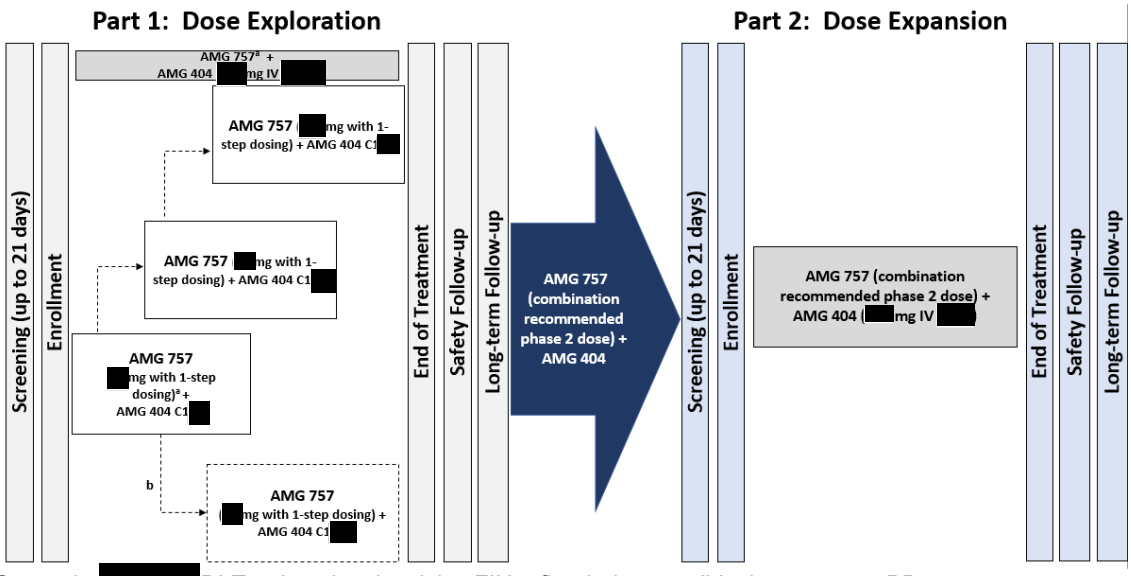
On the basis of a review of real-time safety data and available preliminary PK data, dose escalation may be halted or modified by the Sponsor as deemed appropriate.

Part 2 (Dose Expansion)

Upon completion of Part 1 of the study, enrollment will commence in Part 2 to confirm the safety and tolerability of the selected dose and to further evaluate the efficacy of tarlatamab in combination with AMG 404. Formal safety interim analyses will be performed for every [redacted] subjects who have had opportunity to have at least [redacted] of follow-up since first dose of tarlatamab. The details are provided in protocol Section 9.4.1.

The overall study design is described by the study schema in **Figure 1**.

Figure 1 Study Schema[c]



C = cycle; [redacted] DLT = dose level toxicity; FIH = first-in-human; IV = intravenous; PD = pharmac cs; Ph1 = phase 1; PK = pharmacokinetics; [redacted]

[a] The starting dose and associated dosing schedule of **tarlatamab** will be a target dose of [redacted] mg with 1-step dosing

[b] If the combination is not tolerated (based on target DLT rate > 30%), AMG 404 may be administered on cycle 1 [redacted] Based upon emerging PK, PD, and safety data in the current study as well as the ongoing FIH study (20160323), alternative (intermediate) dose cohort levels, [redacted] as part of the de-escalation recommendations may be explored.

[redacted]

Study Periods:

- Screening: up to [redacted] before enrollment
- Treatment: treatment continues until confirmed radiographic progression or disease progression
- End of treatment: should occur as soon as possible (within [redacted]) after the last dose of investigational product
- Safety follow-up: Upon permanent discontinuation from the study treatment for any reason, an SFU visit will be performed approximately [redacted] after the last dose (tarlatamab or AMG 404) or prior to initiation of other therapy, whichever occurs first. The second safety follow-up visit will be performed approximately [redacted] after the end of the last dose of AMG 404 or prior to initiation of other therapy, whichever occurs first.
- Subjects who discontinue AMG 404 but remain on treatment with AMG 757 will continue to have safety event collection at regularly scheduled study visits. The second safety follow-up visit need only occur if [redacted] has not elapsed from the end of AMG 404 treatment at the time of end of study treatment.
- Long-term follow-up: every 3 months (± 2 weeks) up to 1 year from the first dose of tarlatamab for all subjects who have not withdrawn consent by clinic visit, telephone, or chart review to assess for survival and/or the commencement of subsequent cancer therapy only

Subjects in this clinical investigation shall be referred to as “subjects”. For the sample size justification, see protocol section 9.2.

3.2 Sample Size

Approximately [redacted] subjects were anticipated to be enrolled in the study (up to [redacted] subjects in a dose exploration phase, and the remaining subjects enrolled in a dose expansion phase).

Study enrollment ended in [redacted]. A total of [redacted] subjects were enrolled in this study including [redacted] subjects in Part 1, and [redacted] subjects in Part 2.

Dose Exploration Phase:

Up to 6 planned dose cohorts will be examined during dose exploration. The planned dose levels for tarlatamab with a fixed dose of AMG 404 ([redacted] mg IV [redacted] are described in protocol section 4.1.

Dose exploration will begin with [redacted] subjects treated at Dose Cohort Level 1. The study DLT period is [redacted]. Once all subjects enrolled at a certain Dose Cohort Level are DLT evaluable, a Dose Level Review Team (DLRT) meeting will be convened. Depending on observed safety data, the following may occur: 1) dose de-escalation to the next lowest Dose Cohort Level, 2) additional enrollment to the current Dose Cohort Level, or 3) dose escalation to the next highest Dose Cohort Level or initiation of enrollment in Part 2. In order to consider a certain Dose Level as the recommended phase 2 target dose, at least [redacted] DLT-evaluable subjects must be enrolled at that Dose Level. In total, there are up to [redacted] subjects enrolled in dose exploration phase.

These sample sizes are based on practical considerations and are consistent with conventional oncology studies with the objective to estimate the recommended phase 2 target dose and to evaluate initial safety and tolerability. The probability of observing at least 1 DLT if the true DLT rate is 10 to 30% is provided in Table 3 for various number of subjects.

Table 3. Probability of Observing at least 1 DLT

Number of Subjects	10% DLT Rate		30% DLT Rate
[redacted]	19		55
	27		70
	34		80
	47		91
	57		96
	65		98

DLT = dose-limiting toxicity

Dose Expansion Phase:

If the sample size in the dose expansion phase of [redacted] There is 65% (88%) probability of observing at least 1 adverse event with 10% incidence rate, and the 95% exact confidence interval for 20% objective response (OR) is 3% to 56% (6% to 44%).

3.3 Adaptive Design

The guidelines described in protocol section 4.1 of the protocol for dose escalation or de-escalation to the next dose level are determined by using a mTPI-2 algorithm (Guo et al, 2017) with one practical modification. Consistent with conventional oncology phase 1 study designs (eg, 3+3 design) and given the imprecision with making decisions using as few as [redacted] subjects, in the instance of 1 DLT in the initial [redacted] subjects at a dose level then, as appropriate, the design allows expansion at the dose level beyond [redacted] subjects. All other dose escalation or de-escalation rules are determined strictly by the mTPI-2 algorithm as described below. The mTPI-2 algorithm employs a simple beta-binomial Bayesian model. Let p_T be the target toxicity level and ($p_T - \epsilon_1$, $p_T + \epsilon_2$) be the equivalence toxicity interval denoted as EI. The unit toxicity interval (0, 1) is divided into subintervals with equal length given by ($\epsilon_1 + \epsilon_2$). Let the under-dosing intervals (LI) denote for a set of intervals below EI, and the overdosing intervals (HI) for a set of intervals above EI. The 3 types of dosing intervals (EI, LI, HI) are associated with 3 different dose-escalation decisions. The LI correspond to a dose escalation (E), the HI correspond to a dose de-escalation (D), and proper dosing intervals (EI) correspond to staying at the current dose (S). This study design uses a target toxicity level, p_T of 30%, and EI of (25%, 33%).

Decision rules are based on the unit probability mass (UPM) calculated on these equal-length intervals. Given an interval and a probability distribution, the UPM of that interval is defined as the probability of the interval divided by the length of the interval. The mTPI-2 design calculates the UPMs for the each of equal-length dosing intervals. If the interval with the largest UPM is from LI, EI or HI, then the corresponding dose decision would be E, S, or D, respectively.

A dose level will be considered unsafe, with no additional subjects enrolled at that dose level, if it has an estimated 95% or more probability of exceeding the target DLT rate p_T (ie, P [DLT > p_T | data] > 95%). Based on this rule, the following instances would result in a dose level being considered unsafe

- [redacted] or more DLTs in ≤ [redacted] subjects
- [redacted] or more DLTs in ≤ [redacted] subjects

- [redacted] or more DLTs in ≤ [redacted] subjects
- [redacted] or more DLTs in [redacted] subjects

After the escalation phase is completed, final DLT rates at each dose level will be estimated by isotonic regression (Ji et al, 2010). The weighted least squares regression model will assume monotonic non-decreasing DLT rates with increasing dose and use the empirical (observed) DLT rates at each dose level as responses and dose level sample sizes as weights, along with the pool adjacent violators algorithm (PAVA) to estimate the DLT rate at each dose level. Given the DLT estimates for each dose level, the MTD will be selected from all tried dose levels that have not been previously declared to be unsafe with a DU decision according to the mTPI decision table. With this constraint, the MTD will be determined as the dose level with the DLT estimate closest to the target toxicity level of 30%.

4. Covariates [redacted]

4.1 Planned Covariates

The relationship between covariates and efficacy endpoints may be explored if appropriate.

The following baseline covariates may be summarized as appropriate for each study part: sex, age, race, ethnicity, weight, height, BMI, ECOG, smoking history, sum of diameters of target lesions at baseline, presence of metastatic disease related to current malignancy at initial diagnosis, presence of liver metastasis at initial diagnosis, presence of brain metastasis at initial diagnosis, limited or extensive stage disease at initial diagnosis, number of prior lines of anti-cancer therapy, prior PD-1 or PD-L1, platinum sensitivity to first line platinum based treatment, prior radiotherapy treatment, prior anti-cancer therapy, prior surgery for current malignancy, neurological examination, writing test, mini mental status score, and Echo/MUGA scan result.

[redacted]

5. Definitions

Adverse Event

An adverse event is defined as any untoward medical occurrence in a clinical trial subject. The event does not necessarily have a causal relationship with study treatment.

The definition of adverse event includes worsening of a preexisting medical condition. A preexisting condition that has not worsened during the study, or involves an intervention, is not considered an adverse event.

Age at Enrollment

Subject age at enrollment will be determined using the age in years reported in the clinical database.

Area under the concentration-time curve (AUC)

It is defined as actual body exposure to drug after administration of a dose of the drug.

Baseline

Unless otherwise specified, the baseline value for parameters/assessments scheduled to be performed on the same day as the earliest date of first administration of tarlatamab or AMG 404 is the last non missing value measured before the first administration of tarlatamab or AMG 404 on that day.

For parameters/assessments not scheduled to be performed (or scheduled but not performed) on the same day as the earliest date of first administration of tarlatamab or AMG 404 the baseline value is the value from the screening period measured closest to the day earliest date of first administration of tarlatamab or AMG 404. In the event that multiple assessments are done on the same day as the earliest date of first administration of tarlatamab or AMG 404 and there is no time associated with the assessments, the value associated with the last clinically planned event before the earliest date of first administration of tarlatamab or AMG 404 will be used as the baseline value.

Baseline ECG

Baseline ECG is the Cycle 1 [redacted] pre-dose assessment. If Cycle 1 [redacted] pre-dose assessment is missing, the last non-missing assessment(s) taken prior to the first administration of tarlatamab or AMG 404 treatment will be used as the baseline ECG.

Best Overall Response (BOR)

Best overall response (BOR) for a subject is the best observed post baseline disease response per modified response evaluation criteria in solid tumors (modified RECIST 1.1). Overall response assessments occurring after the start of the first subsequent anticancer therapy will not be included.

BOR is defined as the best response in the following order: CR, PR, SD, PD, or NE, where CR, PR and PD require confirmation by a repeat, consecutive scan at least 4 weeks after the first documentation of response. A best overall response of SD requires an on-study imaging of SD or better no earlier than [redacted] days after cycle 1 [redacted] otherwise the best overall response will be not evaluable (NE).

Please refer the Appendix B for more details.

Body Mass Index (BMI)

Body Mass Index should be calculated using the following formula

$$BMI\ (kg/m^2) = weight\ (kg) / [height\ (cm)/100]^2$$

Body Surface Area (BSA)

The Body Surface Area should be calculated using the following formula

$$BSA = 0.007184 \times weight(kg)^{0.425} \times height(cm)^{0.725}$$

Change from Baseline

Change from Baseline is the arithmetic difference between post-dose assessments and Baseline.

Change (absolute) from Baseline = (Post-baseline Value – Baseline Value)

C_{max}

Maximum observed serum concentration of tarlatamab.

C_{min}

Minimum observed serum concentration of tarlatamab.

Disease control rate (DCR)

DCR is defined as the incidence of a BOR of a confirmed response (CR/PR) or SD while on study as defined by modified RECIST 1.1.

Duration of Disease Control (DoDC)

The duration of disease control (in month) is defined as time from the earlier date of first dose of tarlatamab or AMG 404 to disease progression or death due to any cause, whichever occurs first. DoDC will be calculated only for subjects with a best overall response of CR, PR or SD.

[redacted]

Subjects will be censored following the censoring strategy described in the analyses of PFS (Appendix C).

Duration of Response (DOR)

The duration of response (in months) is defined as time from first evidence of CR or PR to disease progression or death due to any cause, whichever occurs first. DOR will be calculated only for subjects who achieve a best overall response of PR or CR.

DOR (Months)= (PD / death date - response start date + 1) ×12 / 365.25

Subjects will be censored following the censoring strategy described in the analyses of PFS (Appendix C).

Dose Limiting Toxicity (DLT)

Dose-limiting toxicity (DLT) is defined as tarlatamab-related toxicity with an onset within the first [redacted] following first dose with the criteria defined in section 6.2.1.4 of the Protocol.

End of Follow-Up

It is defined as when the last subject completes the last protocol-specified assessment in the study.

End of IP administration

End of IP Admin for each subject is defined as the date the decision was made to end IP as recorded on the End of Investigational Product Administration CRF page.

End of Study (Individual Subject)

End of study for each subject is defined as the date when the subject completed protocol-specified last visit shown in the schedule of activities. The date will be recorded on the End of Study CRF page.

End of Study (End of Trial)

It is defined as the date when the last subject across all sites is assessed or receives an intervention for evaluation in the study (i.e., last subject last visit), including any additional part in the study (eg, long-term follow-up, [redacted]), as applicable.

End of Treatment (EOT)

It is defined as the last assessment per the protocol-specified treatment phase of the study (i.e., end of last treatment cycle) for an individual subject. For subjects who prematurely discontinue investigational product treatment, the EOT should occur as soon as possible (within [redacted]) after the last dose of investigational product was administered.

Half-Life ($t_{1/2}$)

The time required for the observed concentration of a drug to be reduced by one-half.

Investigational Product

The term 'investigational product' is used in reference to tarlatamab and AMG 404.

Last Investigational Product Dose Date

The last IP date for each subject is defined as the latest date IP administered.

Last Known Alive Date:

Last known alive date is the latest date of the following dates before death date:

- Date of Enrollment on Subject Enrollment form
- Date First Taken, Date Last Taken on Concomitant Medications form, Other Protocol Required Therapy form, and Anti-Cancer Therapies (Subsequent) form
- Date Performed on ECOG Performance Status, Vital Signs, MUGA scan or Echocardiogram, Electrocardiogram and Mini Mental Status Examination forms
- Admission Date, Discharge Date on Hospitalizations form, Date of Examination on Physical Measurement form
- Date Collected on all lab forms (e.g., Pregnancy Test (Local Lab), Chemistry (Local Lab), Hematology (Local Lab), Coagulation (Local Lab), Urinalysis (Local Lab)) and in central lab data
- Start Date, Stop Date on Investigational Product Administration forms.
- Start date and Stop date on Cytokine Release Syndrome form

- Date subject ended on End of Investigational Product Administration (AMG 757), End of Investigational Product Administration (AMG 404), and End of Study, End of Safety Follow-up, and End of Radiographic Follow-Up forms where the primary reason is not death or lost to follow-up.
- Date Started and Date Ended but not Resulted in Death on Events form
- Subject Status Date if status is Alive on Survival Status form
- Date Performed on 'CT or MRI', 'Neurological Evaluation', 'Writing Test' forms
- Date of Imaging for Index Lesions, New Lesions and Non-Index Lesions
- Date Scan Performed for Tumor Response Assessment
- Date performed for Procedure, Transfusions, Liver Biopsy Results, Liver Imaging Results forms
- Date Consent Signed, Date Consent Withdrawn from Additional Consents/Withdrawal of Consent form

If death happened, then last known alive date is death date minus 1.

Death date sources eCRF forms:

- End of study, End of Radiographic Follow-Up and End of Safety Follow-up
- Events
- End of Investigational Product Administration (AMG 757), End of Investigational Product Administration (AMG 404)
- Survival Status

Long Term Follow-Up

Long term follow-up will be conducted every 3 months (±2 weeks) up to 1 year from the first dose of tarlatamab for all subjects who have not withdrawn consent by clinic visit, telephone, or chart review to assess for survival and/or the commencement of subsequent cancer therapy only.

For subjects who discontinued treatment for any reason other than confirmed progressive disease, every effort should be made to perform radiographic imaging (computed tomography [CT]/MRI) of the chest, abdomen, pelvis, and all other known sites of disease every 3 months until documentation of confirmed disease progression per modified RECIST 1.1, clinical progression, start of new anticancer therapy, or up to 12 months after the first dose of tarlatamab, whichever occurs first.

Modified Toxicity Probability Interval (mTPI)

The modified toxicity probability interval for target dose-limiting toxicity (DLT) probability of 30% is defined as equivalence toxicity of interval as (25%, 33%) and probability of overdosing of 95% with Beta (1,1) as the prior distribution.

Objective response (OR)

Objective Response is defined as a best overall response of either Complete Response (CR) or Partial Response (PR) per modified RECIST 1.1. Radiographic response

(Complete Response, Partial Response) requires confirmation by a repeat, consecutive scan at least 4 weeks after the first documentation of response.

Objective Response Rate (ORR)

ORR is defined as the incidence of a BOR of a confirmed CR or PR while on study as defined by modified RECIST 1.1. All subjects that do not meet the criteria for objective response by the analysis cut-off date will be considered as non-responders.

Objective Response Rate (ORR) based on Unconfirmed Response

ORR based on unconfirmed response is defined as the incidence of a BOR of CR or PR which do not require confirmation while on study as defined by modified RECIST 1.1. Only unconfirmed response awaiting confirmatory scan is considered in the ORR based on unconfirmed response.

Overall Survival (OS)

OS (in Months) is defined as the interval from the earlier date of first dose of tarlatamab or AMG 404 to the event of death due to any cause. Subjects still alive will be censored at the date last known to be alive. If the date last known to be alive is after the date that triggers the analysis (ie, the data cutoff date), the subject will be censored at the analysis trigger date.

[redacted]

Percent change from Baseline

Percent change from baseline is the arithmetic difference between post-baseline and baseline divided by baseline values times 100.

Change (percent) from Baseline = [(Post-baseline Value – Baseline Value) /

Baseline Value] x 100

Progression-free survival (PFS) per Modified RECIST 1.1

PFS (in Months) per modified RECIST 1.1 is defined as the interval from the earlier date of first dose of Tarlatamab or AMG 404 to the earlier of PD per modified RECIST 1.1 or death due to any cause.

[redacted]

The censoring rules for the progression-free survival analysis are detailed in Appendix C.

Relative Dose Intensity

Relative dose intensity is calculated as actual cumulative dose / planned cumulative dose.

- Actual cumulative dose [mg] is defined as the total dose given during the study treatment exposure up to the study visit. For subjects who did not take any drug the actual cumulative dose is 0 mg.
- Planned cumulative dose [mg] is defined as the per-protocol planned dose accumulated over the actual duration on study treatment, whether the actual visit is missing or not.

Safety Follow-Up (SFU)

The first SFU visit is defined as the subject visit up to [redacted] after last dose of tarlatamab or AMG 404, whichever occurs later, or prior to initiation of other anti-cancer therapy, whichever occurs first. The second SFU visit is defined as the subject visit up to [redacted] after the end of the last dose of AMG 404 or prior to initiation of other anti-cancer therapy, whichever occurs first.

Study Day 1

It is defined as the first day that tarlatamab or AMG 404 is administered to the subject.

Time to Response (TTR)

TTR is defined as the time from the earlier date of first dose of tarlatamab or AMG 404 to the date of the first documented CR or PR that is subsequently confirmed per Modified RECIST1.1. TTR will be calculated only for confirmed responders.

Treatment-Emergent Adverse Event (TEAE)

Events categorized as Adverse Events (AEs) starting on or after first dose of tarlatamab or AMG 404 as determined by "Did event start before first dose of investigational product" equal to "No" or missing on the Events eCRF and up to and including [redacted] after the last dose of tarlatamab or AMG 404 or the End of Study date, whichever is earlier. **Events that are directly related to lung cancer or disease progression (including, but not limited to, preferred terms "Small cell lung cancer", "Disease progression" etc.) will be excluded from TEAE analysis**.

The severity of each adverse event will be graded using the CTCAE version 4.0 **with the exceptions - CRS will be graded according to American Society for Transplantation and Cellular Therapy (ASTCT) (Lee et al, 2019)**. Adverse events will be coded using **latest version of** MedDRA.

Treatment-Emergent Adverse Event (TEAE) For Extended Period

An adverse event that occurs after the [redacted] since the last dose of tarlatamab or AMG 404 and up to [redacted] after last dose of tarlatamab or AMG 404 or the End of Study date, whichever occurs later. The severity of each adverse event will be graded using the CTCAE version 4.0. Adverse events will be coded using **latest version of** MedDRA.

Treatment-Related AE

A treatment-related AE is any treatment-emergent AE with the relationship flag on the Events eCRF indicating there is a reasonable possibility that the event may have been caused by tarlatamab or AMG 404. Event with missing relationship with protocol-defined treatment regimen in each cohort will be assumed to be treatment-related.

6. Analysis Sets

The following sub-sections describe the analysis sets to be used.

6.1 Full Analysis Set

Not applicable for this study.

6.2 Safety Analysis Set

The Safety Analysis Set will consist of all subjects who received at least 1 dose of any Investigational Products (tarlatamab/AMG 404). The analysis of all safety endpoints, unless noted otherwise, will be conducted on the Safety Analysis Set.

6.3 DLT Analysis Set

For the dose exploration part of the study, the analysis of DLT will be conducted on the DLT Analysis Set, defined as all subjects that are enrolled and received at least 1 dose of tarlatamab or AMG 404 with an evaluable DLT endpoint. The DLT endpoint is evaluable if either: 1) the subject experiences a DLT, or 2) the subject does not experience a DLT and receives tarlatamab or AMG 404 treatment as planned in cycle 1 and has been followed for safety events a minimum of [redacted] from date of first dose administration of tarlatamab or AMG 404.

6.4 Pharmacokinetic Analyses Set(s)

The PK Analysis Set will contain all subjects who have received at least 1 dose of the Investigational Product and have at least 1 PK sample collected. These subjects will be evaluated for PK analysis unless the number of data points required for analysis is not enough, or significant protocol deviations have affected the data, or if key dosing or sampling information is missing.

6.5 Interim Efficacy Analysis Set

The Interim Efficacy Analysis Set is a subset of the safety analysis set. The Interim efficacy analysis set includes subjects whose data cut-off date is at least 9 weeks after the first dose date of tarlatamab or AMG 404. Subjects who stopped disease assessments prior to 9 weeks will be included in this analysis set if the data cutoff is at least 9 weeks after their first dose date.

7. Planned Analyses

7.1 Interim Analysis and Early Stopping Guidelines

Safety data will be reviewed on an ongoing basis.

During dose exploration (Part 1) and formally during dose level review meetings (DLRMs), Amgen, in consultation with the site investigators, will review all available cumulative data by cohort prior to making dose escalation or dose de-escalation recommendation. Adverse events and DLTs observed in all subjects will be evaluated continually and fully integrated into all DLRMs and considered in all enrolment and dosing decisions.

During dose expansion (Part 2), Amgen will conduct evaluations of the ongoing grade 4 or higher treatment-related TEAEs to assess if the threshold for possible early trial termination has been reached. The safety interim will occur for every [redacted] subjects

enrolled who have had opportunity to have at least [redacted] of follow up since the first dose of tarlatamab. If this threshold is met, enrollment to dose expansion will be halted pending review of safety data by the DLRT. After receiving the DLRT recommendation, Amgen will choose to take one of the following actions.

- Terminate the trial
- Amend the protocol to potentially improve the benefit/risk for subjects (eg, increase safety monitoring, modify dose/schedule, mandate premedication)
- Continue dose expansion without any changes.

The stopping rules use a Bayesian approach proposed by Thall (Thall et al, 1995) to terminate the study if the posterior probability that the grade 4 or higher treatment related TEAE rate is greater than 20% is > 80%. The stopping boundaries assume a prior distribution of Beta (0.40, 1.60) are presented in **Table 4** and the operating characteristics with pre-specified batch size of [redacted] new subjects per batch are presented in **Table 5**.

Table 4. Stopping Boundary for Dose Expansion with Posterior Probability of 80% and Grade 4 or Higher Treatment-related Adverse Event Limit of 20%

Number of Subjects	Stop Subjects if Observing This Many Grade 4 or Higher Treatment-related Adverse Events
[redacted]	≥ 4
	≥ 6
	≥ 9
	Dose Expansion Complete

Table 5. Operating Characteristics with Batch Size of [redacted] Subjects

True Grade 4 or Higher Treatment-related Adverse Event Rate	Probability of Early Stopping of Dose Expansion	Average Dose Expansion Sample Size
0.10	2.0%	39.5
0.15	9.7%	37.6
0.20	25.8%	33.9
0.25	47.7%	28.8
0.30	69.2%	23.4

7.1.1 Criteria for Evaluating Treatment With Siltuximab

Amgen will conduct evaluations of the treatment and outcome of the CRS events treated with siltuximab on an ongoing basis to assess if the threshold for evaluating siltuximab treatment has been reached as outlined in the Table 6. If these criteria are met, an adhoc DLRM will be triggered to review safety data and available

pharmacokinetic, pharmacodynamics, and efficacy data. If recommended by DLRT, the use of siltuximab will resume. The criteria to trigger an adhoc DLRM to review siltuximab treatment use a Bayesian approach proposed by Thall et al, 1995; an adhoc DLRM will be triggered if the posterior probability that the CRS progression to Grade 3 rate of greater than 20% is > 80% or the posterior probability that the CRS progression to Grade 4 rate of greater than 7.5% is >80%; or observation of any grade 5 CRS after the event has been treated with siltuximab. The boundaries presented below assume a prior distribution of Beta (0.4, 1.6) for progression to grade 3 CRS and a prior distribution of Beta (0.15, 1.85) for progression to grade 4 CRS. The evaluations could occur more frequently if necessary to address emerging safety concerns. If the triggered ad hoc DLRM coincide with regular DLRM, they may be combined.

Table 6 . The Criteria for Evaluating the Use of Siltuximab

	Trigger DLRM if severity of any CRS event treated with siltuximab progresses to Grade 5	
Number of subjects treated with siltuximab	**Or this number of subjects with severity of CRS progressed to Grade 3 after being treated with siltuximab**	**Or this number of subjects with severity of CRS progressed to Grade 4 after being treated with siltuximab**
[redacted]	≥ 3	≥ 2
	≥ 4	≥ 2
	≥ 5	≥ 3
	≥ 6	≥ 3
	≥ 7	≥ 4
	≥ 9	≥ 4
	≥ 10	≥ 5
	≥ 11	≥ 5

CRS = cytokine release syndrome; DLRM = Dose Level Review Meeting

7.2 Primary Analysis

The primary analysis for this study will occur when target enrollment is complete and each subject either completes at least 6 months on study or withdraws from the study.

7.3 Final Analysis

The final analysis will be performed after the last subject has had an opportunity to complete the corresponding EOT visit/procedures.

8. Data Screening and Acceptance

8.1 General Principles

The objective of the data screening is to assess the quantity, quality, and statistical characteristics of the data relative to the requirements of the planned analyses.

8.2 Data Handling and Electronic Transfer of Data

The Amgen Global Study Operations-Data Management (GSO-DM) department will provide all data to be used in the planned analyses. This study will use the RAVE database. The database will be subject to edit checks outlined in the Clinical Data Management Plan (DMP). See details of this section in the DMP.

8.3 Handling of Missing and Incomplete Data

The following imputation for missing or incomplete data will be performed if required: Incomplete adverse event, concomitant medication, death and time to event dates missing data will be imputed as described in Appendix A.

8.4 Detection of Bias

Lack of protocol compliance and the potential for biased statistical analyses will be examined by assessing the incidence of important protocol deviations in each cohort. The clinical study team will identify and document the criteria for important protocol deviations.

8.5 Outliers

Pharmacokinetic (PK) concentration data will be evaluated for outliers by visual inspection, and decision to re-assay individual samples will be made in accordance with standard pharmacokinetic evaluation practice. Descriptive statistics will be used to identify potential outliers in key variables. Suspected outliers will be included in all analyses unless there is sufficient scientific justification to exclude them.

8.6 Distributional Characteristics

Where appropriate, the assumptions underlying the proposed statistical methodologies will be assessed. If required data transformations or alternative non-parametric methods of analyses will be utilized.

8.7 Validation of Statistical Analyses

Programs will be developed and maintained and output will be verified in accordance with current risk-based quality control procedures.

Tables, figures, and listings will be produced with validated standard macro programs where standard macros can produce the specified outputs.

The production environment for statistical analyses consists of Amgen-supported versions of statistical analysis software; for example, the SAS System version 9.4 or later.

9. Statistical Methods of Analysis

9.1 General Considerations

Descriptive statistics will be provided for selected demographics, safety, pharmacokinetic, efficacy data by dose, dose schedule, and time as appropriate. Descriptive statistics on continuous data will include means, medians, standard deviations and ranges, while categorical data will be summarized using frequency counts and percentages. Graphical summaries of the data may also be presented. Confidence intervals (CI) for proportions will be estimated using an exact method proposed by Clopper-Pearson (Clopper and Pearson, 1934). Kaplan-Meier methods will be used to estimate the median and percentiles for time to event endpoints with CI calculated using the Brookmeyer and Crowley (Brookmeyer and Crowley, 1982) method. Kaplan-Meier methods will be used to estimate landmarks for time to event endpoints (eg, 1-year OS) with the Greenwood formula (Kalbfleisch and Prentice, 1980) used to estimate the standard error used in CI calculation.

9.2 Subject Accountability

The number and percentage of subjects who were enrolled, received investigational product, discontinued from investigational product (including reasons for discontinuing, completed study), discontinued the study (including reason for discontinuing) will be summarized by dose/cohort, combining data from each part and overall. Key study dates for the first subject enrolled, last subject enrolled, and last subject's end of study will be presented.

A subject listing and summary noting inclusion in each analysis subset will be provided for all subjects enrolled. A subject listing noting duration of tarlatamab and AMG 404 administrations, reason for discontinuation of treatment, and reason for discontinuing the study will be provided.

9.3 Important Protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first subject's initial visit and updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, subcategory codes, and descriptions will be used during the course of the study. Eligibility deviations are defined in the protocol. The final IPD list is used to produce the Summary of IPDs table and the List of Subjects with IPDs. In addition, a separate listing of all inclusion and exclusion deviations will be provided.

9.4 Demographic and Baseline Characteristics

The following descriptive summaries of demographic and baseline characteristics will be produced by dose/cohort, combining data from each part (i.e., Part 1 – Dose exploration phase and Part 2 – Dose expansion phase) and overall.

The Demographic characteristics will include – age at enrollment in years (summary statistics), age groups (<65, ≥65), sex (Male, Female), race (White, Asian, Black or African American, vs other categories depending on frequency observed) and ethnicity (Hispanic, Not Hispanic). If multiple races have been reported for a subject, the subject will be categorized as multiple.

The baseline characteristics will include – Height, Weight, Body Mass Index, sum of diameters of target lesions at baseline, presence of metastatic disease, Eastern Cooperative Oncology Group (ECOG) performance status (0, 1), Time from initial diagnosis to enrollment, Disease stage at initial diagnosis (Limited, Extensive), Prior lines of therapy (1, 2, 3, >3), Prior anti-cancer therapy (Yes, No), Prior radiotherapy for current malignancy (Yes, No), prior PD-1 or PD-L1, platinum sensitivity to first line platinum based treatment, Prior surgery for current malignancy (Yes, No), neurological exams, writing test and mini-mental status exams, Brain metastases at initial diagnosis (No brain mets present, Brain mets present – treated, Brain mets present - not treated), Liver metastases at initial diagnosis (No liver mets present, Liver mets present), Smoking history (Never, Current, Former), ECHO or MUGA scan, if available.

A listing of the demographic and baseline characteristics will be provided if required. In additional listings of medical and surgical history, prior anti-cancer therapy and prior radiation therapy usage will be provided, if required.

9.5 Efficacy Analyses

Table 7. Efficacy Endpoint Summary Table

Endpoint	Primary Summary and Analysis method	Interim Analysis Set	Primary Analysis and Final Analysis Set
ORR	The percentage of subjects with an objective response (per modified RECIST criteria v1.1) will be summarized by planned dose level along with exact 2-sided 95% confidence interval using Clopper–Pearson method. Subjects without a post-baseline tumor assessment will be considered non-responders.	Interim Efficacy Analysis Set	Safety Analysis Set
DCR	The percentage of subjects with disease control (per modified RECIST criteria v1.1) will be summarized along with 95% exact 2-sided confidence interval using Clopper –Pearson method. Subjects without a post-baseline tumor assessment will be considered non-responders.	Interim Efficacy Analysis Set	Safety Analysis Set
DOR	The K-M quartiles along with 2-sided 95% CIs and rates for selected duration of response (eg, >3, >6, >9, >12 months) will be estimated by planned dose level using Kaplan-Meier method	Interim Efficacy Analysis Set	Safety Analysis Set
TTR	Time to response (TTR) will be summarized with the non-missing sample size (n), mean, standard deviation, median, Q1, Q3, minimum, and maximum for	N/A	Safety Analysis Set

Endpoint	Primary Summary and Analysis method	Interim Analysis Set	Primary Analysis and Final Analysis Set
	subjects who achieve best overall response of CR, or PR.		
PFS	KM curve for PFS, K-M quartiles along with 2-sided 95% CIs and rate for selected time points (eg, 6-months PFS, 1 year [12 months] PFS) will be estimated by planned dose level using Kaplan-Meier method.	Interim Efficacy Analysis Set	Safety Analysis Set
OS	KM curve for OS, K-M quartiles along with 2-sided 95% CIs and rate for selected time points (eg, 6-months PFS, 1 year [12 months] PFS) will be estimated by planned dose level using Kaplan-Meier method.	Interim Efficacy Analysis Set	Safety Analysis Set

The efficacy endpoints include ORR, DCR, DOR, TTR, PFS, and OS. For primary and final analyses, the analysis of efficacy endpoints will be done on the Safety Analysis Set. For interim analysis, the analysis of efficacy endpoints will be done on the Interim Efficacy Analysis Set. Overall response assessments occurring after the start of the first subsequent anticancer therapy will not be included in the efficacy analyses.

9.5.1 Analyses of Primary Efficacy Endpoint(s)/Estimand(s)

Not applicable for this study.

9.5.2 Analyses of Secondary Efficacy Endpoint(s)

The objective response rate and disease control rate (per modified RECIST criteria v1.1) will be summarized by cohort along with 95% exact 2-sided confidence interval using Clopper-Pearson method (Clopper and Pearson, 1934). The subjects without a post-baseline tumor assessment will be considered non-responders.

The duration of response (DOR) will be summarized only for those subjects who achieve a best overall response (either PR or CR). The distribution of DOR including median and quartiles will be presented by cohort using Kaplan-Meier (KM) method

along with 2-sided 95% CI using Brookmeyer and Crowley (Brookmeyer and Crowley, 1982) method. The KM rates of selected durations (eg, >3, >6, >9, >12 months) will also be reported. KM curve can be produced for DOR, provided at least [redacted] of subjects achieve either CR or PR.

The distribution of PFS including median and quartiles will be presented by cohort using Kaplan-Meier method along with 2-sided 95% CI using Brookmeyer and Crowley (Brookmeyer and Crowley, 1982) method. The standard error used in CI calculation will be estimated using the Greenwood formula (Kalbfleisch and Prentice, 1980). The KM rates of selected time points (eg, 6-month PFS, 1-year PFS) will also be reported.

The overall survival (OS) will be analyzed using the same approach as described for the PFS endpoints.

9.5.3 Analyses of Exploratory Efficacy Endpoint(s)

Not applicable.

9.6 Safety Analyses

9.6.1 Analyses of Dose Limiting Toxicities

The subject incidence of DLT will be tabulated by planned dose level. The analysis of DLTs will be conducted on DLT analysis set. The subject incidence of DLTs will be used to estimate final DLT rates for each dose level using isotonic regression (Ji et al, 2010). Refer the Section 3.3 of adaptive design for more details.

[redacted]

9.6.2 Adverse Events

The **most recent version of** Medical Dictionary for Regulatory Activities (MedDRA) will be used to code all events categorized as adverse events to a system organ class and a preferred term. AEs of interest (EOI) categories will be based on search strategies defined by Medical Coding.

The subject incidence of adverse events will be summarized for all treatment-emergent adverse events, treatment-related adverse events, serious adverse events, grade 3 and above, grade 4 and above and fatal adverse events, adverse events leading to discontinuation of investigational product, adverse events leading to interruption and/or reduction of investigational product.

Subject incidence of all treatment-emergent adverse events, treatment-related treatment-emergent adverse events, serious treatment-emergent adverse events, grade 3 or above treatment-emergent adverse events, fatal treatment-emergent adverse events, treatment-emergent/treatment-related adverse events leading to interruption and/or reduction of investigational product, and treatment-emergent/treatment-related adverse events leading to discontinuation of investigational product or other protocol-required therapies will be tabulated for each cohort by preferred term or by system organ class and preferred term in descending order of frequency of highest dose group. The tables will also be presented by worse grade, wherever appropriate or applicable.

The summary of the subject incidence of treatment-emergent adverse events, treatment-related adverse events, serious adverse events and fatal adverse events for the extended period will also be provided.

The severity of each adverse event will be graded using CTCAE 4.0 criteria **and CRS events will be graded using Lee et al, 2014 criteria**.

Adverse Events of Interest (EOI):

EOIs identified for tarlatamab:

Cytokine Release Syndrome (Broad), Cytokine Release Syndrome (Narrow), Neutropenia (Broad), Neutropenia (Narrow), Immune effector Cell Associated Neurotoxicity Syndrome (ICANS) and associated neurological events (Broad), Neurological Events [Nervous system disorders (SOC) + Psychiatric disorders (SOC)], and Hypersensitivity events [Hypersensitivity (SMQ Broad)/Anaphylactic reactions (SMQ Narrow)].

For the list of EOIs mentioned above, the subject incidence of treatment-emergent adverse events of interest (standardized MedDRA queries and/or Amgen Medical Queries), grade ≥ 2, grade ≥ 3, grade ≥ 4, serious, fatal, and events leading to interruption, reduction or discontinuation of IP will be summarized.

The subject incidence of above treatment-emergent EOIs will be tabulated by preferred term. CRS will be graded according to Lee et al, 2014.

The summaries of time to onset of EOI from first dose of IP (days), which comes earlier, time to onset of EOI after each dose from last prior dose of IP (days), which comes later(days), time to onset of CRS from first dose of tarlatamab (hours), time to onset of CRS from each last prior tarlatamab dose (hours), and duration of treatment-emergent

EOI will be provided. Additionally, summaries of time to onset and duration of EOI will be provided for grade 2 or higher and grade 3 or higher EOIs respectively.

9.6.3 Laboratory Test Results

The analyses of safety laboratory endpoints will include summary statistics at selected time points by cohort. Shifts in grades of safety laboratory values between the baseline and the worst on-study value will be tabulated by cohort.

Summaries of the absolute value and/or changes from baseline at each scheduled assessment will be provided for selected laboratory parameters of interest.

Shift tables indicating the change between the baseline and the maximum post dose CTCAE grades for an increased value, and the maximum post dose grade for a decreased value will be provided for selected laboratory parameters of interest. Unscheduled assessments will be included in the shift tables.

Potential Hy's law cases will be listed and may also be summarized by cohort. Hy's law cases will have the following three components: 1) ALT or AST > 3 X ULN, 2) ALP < 2 X ULN and 3) TB >2 X ULN, on any day. A listing of AST, ALT and Total Bilirubin values at each time point will be produced for the subjects suspected of Hy's law case.

9.6.4 Vital Signs

The analyses of vital signs will include summary statistics at selected time points by treatment group. Shifts in vital sign values between the baseline and the worst on-study value will be tabulated by treatment group.

The vital sign includes systolic/diastolic blood pressure, heart rate, respiratory rate, temperature and pulse oximetry and will be summarized using descriptive statistics. The summary statistics will be presented for baseline and scheduled post-baseline visits. In addition, summaries for changes from baseline over time may be summarized, provided the scope and size of changes is feasible.

Shifts in scores for ECOG performance status scores between the baseline and each scheduled / assessed time point will be tabulated. ECOG performance status scores will be summarized for each cohort at the relevant time points.

9.6.5 Physical Measurements

The analyses of physical measurements will include summary statistics at baseline by treatment group. Depending on the size and scope of the changes, the summaries of

changes from baseline over time may be provided. These summaries will also include the unscheduled visits.

9.6.6 Electrocardiogram

Summaries over time and/or changes from baseline over time will be provided for available ECG parameters.

[redacted]

9.6.8 Exposure to Investigational Product

Subject exposure to investigational product and combination therapy will be summarized using descriptive statistics. The number of doses per subject, cumulative dose (mg), relative dose intensity (%), treatment duration (weeks) will be summarized. The number and percent of subjects with dose modifications (eg, dose reductions, dose interruptions) and reason for modification will also be summarized. Subject-level data may be provided instead of the summary if the subject incidence is low or single dose is given.

9.6.9 Exposure to Non-investigational Product

Not applicable for this study.

9.6.10 Exposure to Other Protocol-required Therapy

Descriptive statistics will be produced to describe the exposure to other protocol specified therapy (i.e., corticosteroids such as dexamethasone and tocilizumab).

Descriptive statistics of the number of doses per subject, cumulative dose (mg), relative dose intensity (%), treatment duration (weeks).

9.6.11 Exposure to Concomitant Medication

Number and proportion of subjects receiving therapies of interest may be summarized by preferred term for each cohort as coded by the World Health Organization, 2018. A subject listing of all prior and concomitant medications may be provided.

9.7 Other Analyses

9.7.1 Analyses of Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Endpoints

Pharmacokinetic parameters will be determined from the time concentration profile using standard non-compartmental approaches and considering the profile over the complete sampling interval. [redacted].

9.7.2 Analyses of Clinical Outcome Assessments

Not applicable for this study.

9.7.3 Analyses of Health Economic Endpoints

Not applicable for this study.

[redacted]

10. Changes From Protocol-specified Analyses

There are no changes to the protocol-specified analyses.

11. Literature Citations / References

Brookmeyer R, Crowley J. A confidence interval for the median survival time. *Biometrics*. 1982;38:29-41.

Clopper CJ and Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*. 1934;26:404-413.

Guo W, Wang SJ, Yang S, Lynn H, Ji Y. A Bayesian interval dose-finding design addressing Ockham's razor: mTPI-2. *Contemp Clin Trials*. 2017;58:23-33.

Ji Y, Liu P, Li Y, Bekele BN. A modified toxicity probability interval method for dosefinding trials. *Clin Trials*. 2010;7:653-663.

Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. John Wiley & Sons, Inc., New York, 1980. xi + 321 pp

Thall PF, Simon RM, Estey EH. Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Statistics in Medicine*. 1995;14:357-379.

12. Prioritization of Analyses

The definitive list of TFLs to be produced for this study will be documented in the Program Index.

13. Data Not Covered by This Plan

Exploratory data not included in this plan and may be analyzed at later date or by different Amgen Department.

14. Appendices

Appendix A. Date Imputation Rules

Imputation rules for partial or Missing Dates for Adverse Events and concomitant Medication are detailed below. **Date of on-study anti-cancer therapy will be imputed using the same rule when only the day is missing (no imputation when month or year is missing)**.

Imputation Rules for Partial or Missing Start Dates

Start Date		Stop Date						
		Complete: *yyyymmdd*		Partial: *yyyymm*		Partial: *yyyy*		
		< 1st dose	**≥ 1st dose**	**< 1st dose *yyyymm***	**≥ 1st dose *yyyymm***	**< 1st dose *yyyy***	**≥ 1st dose *yyyy***	missing
Partial: *yyyymm*	= **1st dose** *yyyymm*	2	1	2	**1**	n/a	1	1
	≠ **1st dose** *yyyymm*		2		2	2	2	2
Partial: *yyyy*	= **1st dose *yyyy***	3	1	3	1	n/a	1	1
	≠ **1st dose *yyyy***		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1=Impute the date of first dose; 2=Impute the first of the month; 3=Impute January 1 of the year; 4=Impute January 1 of the stop year
Note: For subjects who were never treated (first dose date is missing), partial start dates will be set to the first day of the partial month or first day of year if month is also missing.

Note: If the start date imputation leads to a start date that is after the stop date, then do not impute the start date.

1. **Imputation rules for partial or missing stop dates:**

Initial imputation

- If the month and year are present, impute the last day of that month.
- If only the year is present, impute December 31 of that year.
- If the stop date is entirely missing, assume the event or medication is ongoing.

If the imputed stop date is before the start date, set stop date to missing.

If the imputed stop date is after the death date, impute as death date.

2. Imputation rules for partial or missing death dates:

- If death year and month are available but day is missing:
  - If mmyyyy for last contact date = mmyyyy for death date, set death date to the day after the last contact date.
  - If mmyyyy for last contact date < mmyyyy for death date, set death date to the first day of the death month.
  - If mmyyyy for last contact date > mmyyyy for death date, data error and do not impute.
- If both month and day are missing for death date or a death date is totally missing, do not impute.

Note that the last contact date refers to the last contact (i.e. a visit or an assessment) with patient instead of family members. Last contact date would be derived from the latest patient visit/assessment date.

Appendix B. BOR Calculation Algorithm

Confirmed BOR by Investigator

At each time point, BOR will be derived based upon the evaluated time points up to and including the current assessment. The following rules will apply to BOR:

- CR is better than PR is better than SD is better than PD is better than NE
- For a BOR of SD, a duration of ≥ [redacted] since the cycle 1 [redacted] is required
- For a BOR of CR, PR or PD, confirmation is required in a consecutive assessment ≥ [redacted] after the initially observed assessment of CR, PR or PD except the following cases
    - An unlimited number of intermittent assessments of NE or CR < [redacted] can occur between the initial response and the confirmation of CR. For example, BL, CR, NE, NE, NE, CR – the CR at post-baseline 1 is confirmed at post-baseline 5. SD is not allowed between CR and subsequent confirmation CR.
    - An unlimited number of intermittent assessments of NE, SD, PR or CR < [redacted] can occur between the initial response and the confirmation of PR. For example, BL, PR, SD, SD, SD, PR – the PR at post-baseline 1 is confirmed by post-baseline 5.
    - An unlimited number of intermittent assessments of NE can occur between the initial PD and the confirmation. (No reference though)
    - PD doesn't need confirmation if PD is the last evaluable assessment PD and determines BOR or PD time of PFS.

Table 8 provides the BOR determination per modified RECIST 1.1 for trials where confirmation for response and progressive disease are required. A BOR determined by Table 8 is considered a Confirmed_BOR. Table 9 outlines the steps to derive Confirmed_BOR (step 1) given investigator's time point assessments.

Interim BOR by Investigator

At interim futility analyses, due to lack of sufficient follow-up time for confirmation san, study team may choose to report both confirmed responders and unconfirmed responders (subjects had an initial PR or CR and still has potential for future confirmative scans). Interim BOR is defined to include unconfirmed responders in addition to those who achieve Confirmed BOR. Table 9 step 2 is the step to derive these unconfirmed responders.

Confirmed and Interim BOR by Central Review

Confirmed and Interim BOR by Central Review will be derived from Time point response by Central. The steps are same as derivation for Confirmed and Interim BOR by Investigator (Table 8 and Table 9)

Table 8. BOR per Modified RECIST 1.1

Criterion	Timepoint T1 Response	T1 ≥ (X-1)* [redacted] after C1[redacted]?	Timepoint T2 Response	T2 ≥ (X-1)* [redacted] after C1[redacted]?	T2 ≥ [redacted] after T1?	BOR_temp
C1	CR	Yes	CR	-	Yes	CR
C2			CR	-	No	SD
C3			PR, SD	-	-	Query data*
C4			PD	-	-	SD
C5			NE, No further evaluations			SD
C6		No	CR	-	Yes	CR
C7			CR	Yes	No	SD
C8			CR	No	No	NE
C9			PR, SD	-	-	Query data*
C10			PD	-	-	NE**
C11			NE, No further evaluations			NE
C12	PR	Yes	CR, PR	-	Yes	PR
C13			CR, PR	-	No	SD
C14			SD	-	-	SD
C15			PD	-	-	SD
C16			NE, No further evaluations			SD
C17		No	CR, PR	-	Yes	PR
C18			CR, PR	Yes	No	SD
C19			CR, PR	No	No	NE
C20			SD	Yes	-	SD
C21			SD	No		NE
C22			PD	-	-	NE**
C23			NE, No further evaluations			NE
C24	SD	Yes	CR, PR, SD, PD, NE, no more evaluation			SD
C25		No	CR, PR, SD	Yes	-	SD
C26			CR, PR, SD	No	-	NE
C27			PD	-	-	NE**
C28			NE, No further evaluations			NE
C29	PD	-	CR, PR, SD	Yes		SD
C30			CR, PR, SD	No		NE
C31			PD		Yes	PD
C32			PD		No	NE**
C33			NE, No further evaluations			NE**
C34	NE	-	NE, No further evaluations			NE
C35		-	CR, PR, SD	Yes	-	SD
C36		-	CR, PR, SD	No	-	NE
C37		-	PD	-	-	NE**

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, and NE = not evaluable.

X is the scan schedule of every X weeks per protocol. X = 8

*If a CR is truly met at first time point, then any disease seen at a subsequent time point, even disease meeting PR criteria relative to baseline, makes the disease PD at that point (since disease must have reappeared after CR). Best response would depend on whether minimum duration for SD was met. However, sometimes 'CR' may be claimed when subsequent scans suggest small lesions were likely still present and in fact the patient had PR, not CR at the first time point. Under these circumstances, the original CR should be changed to PR and the best response is PR.

** For rows C10, C22, C27, C32, C37, if timepoint T2 response of PD is the last evaluable assessment, then BOR_temp is PD, otherwise the BOR_temp is NE. For row C33, if timepoint T1 response of PD is the last evaluable assessment, then BOR_temp is PD, otherwise the BOR_temp is NE.

Table 9. BOR Derivation Step

Step 1. Derive Confirmed BOR at each visit

Derive *Confirmed BOR* at each visit:

i) At Visit 1: Using the Visit 1 scan result and refer to **Table 8** to derive *Confirmed BOR*.

ii) At Visit 2 onward:

(a) If current visit is CR, PR or PD: Find last scan that is at least [redacted] before current visit. Derive *BOR_temp* as below:

Last scan	Current	*BOR_temp*
CR	CR	CR
PR	CR	PR
PR	PR	PR
PD	PD	PD

(b) If none of above fits, find last scan reference **Table 8** to derive *BOR_temp*.

iii) Current visit Confirmed BOR = best of (BOR_temp, last visit confirmed BOR). Use rule CR > PR > SD > PD > NE

Step 2. Derive Interim BOR at last visit prior to DCO

For subjects who discontinue tumor assessment (i.e. had confirmed PD, next therapy, or end of study):	For subjects with potential for more assessment (i.e. no confirmed PD, next therapy, or end of study), unconfirmed PR/CR can be considered responders.

Assign *Interim BOR* = *Confirmed BOR*	• If any scan is CR and haven't get the opportunity to receive next scan, then *Interim BOR* = CR. • If any scan is PR and haven't get the opportunity to receive next scan, then *Interim BOR* = PR. • Else, *Interim BOR* = *Confirmed BOR* at the latest scan.

Appendix C. PFS Calculation and Censoring Rule

Table 10 defines PFS censoring rule per modified RECIST 1.1 PD from radiographic scans.

Table 10. PFS Censoring Rule

Situation up to DCO/EOS	Date of Event or Censor	Outcome
No evaluable post-baseline tumor assessments, no death recorded	Date of Randomization date (or first dose date of IP in non-randomized trials)	Censor
PD	Date of first detection of PD	Event
No PD, but death recorded	Date of death	Event
Start of new anti-cancer therapy prior to any PD or death	Date of last evaluable tumor assessment before or on start of new anti-cancer therapy	Censor
No PD, no death recorded, no start of new anti-cancer therapy	Date of last evaluable tumor assessment	Censor
Death or PD immediately after 2 or more missed tumor assessment (**18 weeks**)	Date of last evaluable tumor assessment prior to missing assessment(s) [a]	Censor

DCO = Data Cutoff; EOS = End of Study; PD = Progressive disease
[a] This supersedes the previous rules that result in PFS event at date of PD or death.

Note: PD's does not require confirmation if based on RECIST 1.1.