Statistical Analysis Plan

Protocol Title:	A Phase 1b Study Evaluating the Safety, Tolerability, Pharmacokinetics and Efficacy of Delta-like Protein 3 Half-life Extended Bispecific T-cell Engager Tarlatamab in Subjects with De Novo or Treatment Emergent Neuroendocrine Prostate Cancer (DeLLpro-300)
Short Protocol Title:	A Phase 1b Study of Tarlatamab in Subjects with Neuroendocrine Prostate Cancer
Protocol Number:	20200040
NCT Number:	NCT04702737
Authors:	[redacted] Biostatistician, IQVIA [redacted] Sr Associate Biostatistics, Amgen Inc.
Sponsor:	Amgen Inc. One Amgen Center Drive Thousand Oaks, CA 91320 Phone: +1 805-447-1000
SAP Date:	**Document Version** — **Date** **Amendment 3 (v4.0)** — 24 March 2023

Version Number	Date (DDMMMYYYY)	Summary of Changes, including rationale for changes
Original (v1.0)	01OCT2021	NA
Amendment 1 (v2.0)	29JUN2022	Changes per protocol amendment 5 (PA5). Update text 'AMG 757' to 'tarlatamab'. Update Objective Response Rate Analysis Set to Interim Efficacy Analysis Set. Update to endpoint definitions, CTCAE version update to 5.0, MedDRA version update to 25.0. Added Body Mass Index as baseline characteristics. Addition of exposure duration in days and cycles. Update to Dose Limiting Toxicity criteria per PA5.
Amendment 2 (v3.0)	31OCT2022	Changes per protocol amendment 6. Update to use RECIST 1.1 instead of

		PCWG3 modified RECIST 1.1 for objective response, duration of response and disease control rate. CTCAE version changed from 5.0 to 4.0 for consistency with protocol and DES guidance (In SAP version 2.0, CTCAE version was updated from 4.0 to 5.0, however protocol was not updated).
Amendment 3 (v4.0)	24MAR2023	Section 4.1: Updated text from 'Sum of diameters target lesions at baseline (mm)' to 'Sum of diameters of target lesions at baseline based on investigator assessment (mm)'. Section 5.1: Updated text from 'Duration of Exposure' to 'Treatment Duration'. Removed formula for exposure duration in days calculation. Added definition for relative dose intensity. Section 5.2: Updated treatment-emergent adverse event duration from 42 days to 47 days. Updated text 'the IP' by 'tarlatamab' for treatment-related adverse event definition. Section 5.3: Updated definition language for Best Overall Response, Disease Control Rate, Duration of Response, Objective Response Rate, Radiographic Progression and Radiographic Progression-free survival. Section 5.4: Updated definition language for [redacted] Section 6.2: Updated language for Safety Analysis Set definition. Section 6.6: Updated text in Interim Efficacy Analysis Set definition from 'day 1' to 'study day 1'. Section 6.7: Updated language for DLT Analysis Set definition. Added following analysis sets: RECIST 1.1 Evaluable by Central Reviewer Analysis Set, RECIST 1.1 Evaluable by Investigator Assessment Analysis Set, [redacted] Evaluable Based on Central Laboratory Data Analysis Set, [redacted] Evaluable Based on Local Laboratory Data Analysis Set, [redacted] Section 9.4: Added clarification that ethnicity will be reported only for subjects

		with race as 'white'. Removed 'Tobacco use'. Updated text from 'Sum of diameters of target lesions at baseline (mm)' to 'Sum of diameters of target lesions at baseline based on investigator assessment (mm)'. Section 9.5.2: Added details for use of central data and/or local data for primary analysis. Table 9-1: Updated text 'quartiles' to 'percentiles'. Updated analysis set for the following endpoints: Objective Response, Duration of Response, Disease Control Rate. Section 9.5.3: Added details for use of central data and/or local data for primary analysis. Table 9-2: Updated analysis set for the following endpoints: [redacted] Section 9.6.8: Updated section to present only number of doses, cumulative dose (mg), relative dose intensity (%), treatment duration (weeks) for exposure summary. Section 10: Added details for changes from protocol-specified analysis. Appendix B: Updated appendix title from 'Censoring Rules for Progression Free Survival' to 'Censoring Rules for Radiographic Progression-free Survival'. Table 14-2: Updated table title from 'Censoring Rules for Progression Free Survival' to 'Censoring Rules for Radiographic Progression-free Survival'. Updated language for censoring rules. Appendix C: Added options for the SAS procedure. Table 14-3: Updated table footnote. Appendix E: Updated appendix title from 'Bone Progression Status as per PCWG3' to 'Radiographic Progression per RECIST 1.1 with PCWG3 Modifications'. Added heading 'Overall Bone Progression Status (Bone Scan)' for existing Table 14-5. Updated definitions in Table 14-5. Added new heading 'Radiographic Progression per RECIST 1.1 with PCWG3 Modifications' and Table 14-6.

		Radiographic Progression by RECIST 1.1 with PCWG3 Modifications under the new heading.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Explanation
AE	Adverse event
[redacted]	
BMI	Body mass index
BOR	Best overall response
CI	Confidence intervals
COVID-19	Novel coronavirus disease
CR	Complete response
CRF	Case report form
CRS	Cytokine release syndrome
CT	Computed tomography
[redacted]	
CTCAE	Common terminology criteria for adverse events
DCO	Data cut-off
DCR	Disease control rate
DOR	Duration of response
DLL3	Delta-like protein 3
DLRM	Dose level review meeting
DLRT	Dose level review team
DLT	Dose limiting toxicity
ECG	Electrocardiogram
ECOG	Eastern cooperative oncology group
EI	Equivalence toxicity interval
[redacted]	
EOS	End of study
EOT	End of treatment
[redacted]	
FIH	First-in-human
[redacted]	
GSO-DM	Global study operations-data management
HI	Overdosing intervals
IPD	Important protocol deviation
IV	Intravenous

IP	Investigational product
KM	Kaplan-Meier
[redacted]	
LI	Under-dosing intervals
LTFU	Long term follow-up
MedDRA	The medical dictionary for regulatory activities
MRI	Magnetic resonance imaging
MTD	Maximum tolerated dose
mTPI-2	Modified toxicity probability interval design
NE	Not evaluable
NEB	No evidence of bone lesion
NEPC	Neuroendocrine prostate cancer
NP	Not performed
OR	Objective response
ORR	Objective response rate
OS	Overall survival
PAVA	Pool adjacent violators algorithm
PCWG3	Prostate cancer working group 3
PD	Pharmacodynamic
PD	Progressive disease
PDc	Confirmed PD
PDu	Unconfirmed PD
PD-1 (L1)	Programmed cell death protein 1 (Ligand-1)
[redacted]	
PFS	Progression-free survival
PK	Pharmacokinetic
PR	Partial response
[redacted]	
Q2W	Every 2 weeks
RECIST	Response evaluation criteria in solid tumors
RP2D	Recommended phase 2 dose
RR	Relapsed/refractory
SAP	Statistical analysis plan
SD	Stable disease

SFU	Safety follow-up
SUV	Standardized uptake value
TEAE	Treatment-emergent adverse event
TLS	Tumor lysis syndrome
TNM	Tumor, node, metastatic
TRAE	Treatment-related adverse event
UPM	Unit probability mass
WHO	World health organization

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 6 for study 20200040, **Tarlatamab (AMG 757)** dated 29 September 2022. The scope of this plan includes the interim analyses, the primary analysis and the final analysis that are planned and will be executed by the Amgen Global Biostatistical Science department unless otherwise specified.

2. Objectives, Endpoints/Estimands and Hypotheses

2.1 Objectives and Endpoints/Estimands

Objectives	Endpoints
Primary	
• Evaluate the safety and tolerability of tarlatamab	• Treatment-emergent adverse events, treatment-related adverse events, and changes in vital signs, electrocardiogram (ECG), and clinical laboratory tests
• Determine the maximum tolerated dose (MTD) or recommended phase 2 dose (RP2D)	• Dose limiting toxicities (DLTs)
Secondary	
• Evaluate anti-tumor activity of tarlatamab as assessed by additional measures	• Objective response (OR) per Response Evaluation Criteria in Solid Tumors (RECIST 1.1) • Duration of response (DOR) per RECIST 1.1 • Radiographic progression-free survival (PFS) per RECIST 1.1 with Prostate Cancer Working Group 3 (PCWG3) modifications • Overall Survival (OS) • Disease Control Rate (DCR) per RECIST 1.1
• Characterize the pharmacokinetics (PK) of tarlatamab	• PK parameters for tarlatamab following intravenous (IV) administration including, but not limited to, maximum serum concentration (Cmax), minimum serum concentration (Cmin), area under the concentration-time curve

	(AUC) over the dosing interval, accumulation ratio, and half-life (t1/2)

Estimand(s) for Primary Objective(s)

Not applicable for this study.

Estimand(s) for Secondary Objective(s)

Not applicable for this study.

Exploratory
[redacted]

2.2 Hypotheses and/or Estimations

No statistical hypotheses will be tested.

3. Study Overview

3.1 Study Design

This is an open label phase 1b study evaluating tarlatamab monotherapy. Tarlatamab will be administered as a short-term intravenous (IV) infusion every 2 weeks (Q2W) (with step dosing) in a 28-day cycle as monotherapy in subjects with de novo or treatment-emergent Neuroendocrine Prostate Cancer (NEPC).

The study will consist of 2 parts: dose exploration (Part 1), dose expansion (Part 2, [redacted]

Dose Exploration (Part 1)

The dose exploration part of the study will enroll up to 20 subjects with relapsed/refractory NEPC. At the time of initiation of the study, the starting dose will be the highest dose deemed safe and tolerable in the ongoing phase 1 trial of tarlatamab in subjects with Small Cell Lung Cancer (SCLC) (Note: based on emerging safety data from Study 20160323, the starting dose may be lower than the highest dose deemed safe and tolerable). The planned highest dose in this study is not anticipated to exceed the maximum tolerated dose (MTD) as established in the ongoing Study 20160323.

Dose exploration will begin with treating 3 to 4 subjects at the Dose Level 1. The study dose limiting toxicity (DLT) period is 28 days. Once all subjects enrolled at a certain dose level are DLT evaluable, a dose level review team (DLRT) will convene. For any dose level, and depending on observed safety data, the following recommendation may occur:

- dose de-escalation to the next lowest dose level
- additional enrollment to the current dose level
- dose escalation to the next highest dose level

Dose escalation/de-escalation recommendations will be guided by a modified toxicity probability interval design (mTPI-2) model (Guo et al, 2017) with a target toxicity probability of 0.3 (protocol Table 4-2) and equivalence toxicity interval of (25%, 33%). The operating characteristics in protocol Table 4-3 provide the probability of selecting each dose level and the expected number of treated patients at each dose level for given hypothetical true DLT rates.

If late onset of 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. The MTD will be estimated using isotonic regression (Ji et al, 2010) and the MTD will be the dose level with the estimated DLT rate closest to 0.30. In order to consider a certain Dose Level as the MTD at least 6 DLT evaluable subjects must be enrolled at that dose level. No more than 10 DLT evaluable subjects will be enrolled at any specific dose level in dose exploration phase.

At the time of enrollment, the dose-escalation phase (Part 1) will be eliminated if the recommended phase 2 dose (RP2D) or MTD has been selected in the ongoing phase 1 trial of tarlatamab in small cell lung cancer. Tarlatamab will then be dosed based upon the RP2D/MTD selected in small cell lung cancer trials. Note: if Part 1 is initially eliminated based upon the RP2D/MTD selection in the phase 1 trial in small cell lung cancer, dose escalation may be opened in the future to explore additional dose(s) and dosing schedule(s) as described above.

Dose Expansion (Part 2)

Once the RP2D/MTD has been determined, enrollment will commence in the dose expansion phase to confirm the safety and tolerability of the selected dose and to further evaluate anti-tumor activity.

One or more dosing schedules as described in Part 1 may be pursued in parallel during the expansion phase. Dosing schedules involving [redacted] as noted in protocol Table 1-3 may be pursued in parallel or instead of the step dosing approach based on recommendations from the emerging safety, efficacy, and PK/PD data from the ongoing phase 1 trial in SCLC.

Based on emerging data from the tarlatamab monotherapy in this indication and emerging data from other trials with tarlatamab, the protocol may be amended to include a combination of tarlatamab with anti-PD-1 (L1) agent.

The overall study design is described by a study schema in protocol section 1.2.

3.2 Sample Size

Approximately 100 subjects will be enrolled in the study, with up to 20 subjects to be enrolled in the dose exploration phase (Part 1) and approximately 80 subjects to be enrolled in the dose expansion phase (Part 2).

The sample size in the dose exploration phase (Part 1) is based on practical consideration and is consistent with conventional oncology studies with the objective to estimate the MTD. With 3 subjects in a dose level, there is a 27% to 70% probability of observing at least 1 DLT if the true DLT rate is 10% to 33% and with 4 subjects in a dose level, there

is 34% to 80% probability. With 6 subjects in a dose level, there is a 47% to 91% probability of observing at least 1 DLT if the true DLT rate is 10% to 33% and with 10 subjects in a dose level, there is a 65% to 98% probability.

In the dose expansion phase (Part 2), up to [redacted] subjects will be enrolled. Forty subjects will be enrolled in Part 2 [redacted]

3.3 Adaptive Design

The guidelines described in protocol Section 4.1 for dose escalation or de-escalation to the next dose level are determined by using a mTPI-2 algorithm (Guo et al, 2017). The mTPI-2 algorithm employs a simple beta-binomial Bayesian model. Let p_T be the target toxicity level and (p_T − ϵ1, p_T + ϵ2) be the equivalence toxicity interval denoted as EI. The unit toxicity interval (0, 1) is divided into subintervals with equal length given by (ϵ1 + ϵ2). 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 (i.e., P [DLT > p_T | data] > 95%). Based on this rule, the following instances would result in a dose level being considered unsafe.

- 3 or more DLTs in ≤ 4 subjects
- 4 or more DLTs in ≤ 6 subjects
- 5 or more DLTs in ≤ 9 subjects
- 6 or more DLTs in 10 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 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 and Subgroups

4.1 Planned Covariates

The relationship between relevant covariates and efficacy endpoints may be explored if appropriate. The following baseline covariates may be summarized as appropriate for each study part:

- Race
- Ethnicity
- Sex
- Age at enrollment in years
- ECOG performance status at baseline
- Prior lines of therapy
- Metastatic
- Sum of diameters **of** target lesions at baseline **based on investigator assessment** (mm)
- [redacted]

4.2 Subgroups

[redacted]

5. Definitions

5.1 General Definitions

Baseline

For any variable, unless otherwise specified, the baseline is the last non-missing assessment taken prior to the first administration of tarlatamab. Where baseline measurements are taken on the same day as the tarlatamab administered and no times are present, it will be assumed that these measurements are taken prior to the study specified treatment being administered.

Body Mass Index (BMI)

BMI is calculated using the following formula:

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

Change from Baseline

Change from **b**aseline is calculated using the following formula:
Change from **b**aseline = **p**ost-baseline **v**alue – **b**aseline **v**alue

End of Investigational Product (IP)

End of IP 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 (End of Trial)

The end of study date 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); following any additional parts in the study (e.g., long-term follow-up, [redacted], as applicable.

End of Study (Individual Subject)

End of study for each subject is defined as the date the subject last completed a protocol-specified procedure. The date will be recorded on the End of Study CRF page.

End of Study (Primary Completion)

The primary completion is defined as the date when the last subject is assessed or receives an intervention for the final collection of data for the primary endpoint(s).

If the study concludes prior to the primary completion date originally planned in the protocol (i.e., early termination of the study), then the primary completion will be the date

when the last subject is assessed or receives an intervention for evaluation in the study (i.e., last subject last visit).

Investigational Product (IP)

Investigational product refers to tarlatamab.

Long Term Follow-up (LTFU)

LTFU will be conducted every 3 months up to 3 years from the last 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 progressive disease, every effort should be made to perform radiographic imaging (CT/MRI/bone scintigraphy) of the chest, abdomen, pelvis, bone, and all other known sites of disease every 3 months until documentation of disease progression per PCWG3 guidelines, clinical progression, start of new anticancer therapy, or up to 3 years after the last dose of tarlatamab, whichever occurs first.

Treatment-related serious adverse events and fatal serious adverse events regardless of the cause should be collected during long-term follow-up.

Percent Change from Baseline

Percent change from baseline is calculated using the following formula:
Percent **c**hange from **b**aseline = [(**p**ost-baseline **v**alue – **b**aseline **v**alue) / **b**aseline **v**alue] x 100

Relative Dose Intensity

Relative dose intensity is calculated using the following formula:

Relative dose intensity (%) = (Actual cumulative dose / Planned cumulative dose) x 100; where actual cumulative dose = sum of received doses, planned cumulative dose = planned dose accumulated over the actual treatment duration, i.e., 100 mg x (number of records in exposure dataset - 1) + 1 mg. Actual and planned cumulative dose up to the study visit are considered.

Safety Follow-up (SFU)

Upon permanent discontinuation from the study treatment for any reason to further assess the risk of delayed adverse events, a SFU visit will be performed approximately 42 (+ 5) days after the end of the last dose of tarlatamab.

Study Day

If the date is after the date of study day 1, then study day is calculated as following:

Study day = (date - date of **s**tudy **d**ay 1) + 1

If the date is before the date of study day 1, then study day is calculated as following:

Study day = (date – date of **s**tudy **d**ay 1)

Study Day 1

Study Day 1 is defined as the date of the first dose of tarlatamab administration to the subject.

Treatment Duration

Treatment duration is calculated using the following formula:

Treatment duration (weeks) = (last dose **date** – first dose **date** + 1) / 7.

5.2 Primary Endpoint-related Definitions

Dose-Limiting Toxicity (DLT)

DLT is defined as any qualifying toxicity that is at least possibly related to tarlatamab with an onset within the first 28 days following first dose with either of the following criteria:

- Grade 3 adverse event lasting more than 3 days (with the exception of fatigue and grade 3 non-febrile neutropenia that improves to ≤ grade 1 within 3 weeks including the use of growth factor support per neutropenia management guidelines [protocol Section 6.8.6]);

- ≥ Grade 4 adverse event regardless of duration (with the exception of grade 4 non-febrile neutropenia lasting less than or equal to 7 days including the use of growth factor support per neutropenia management guidelines [protocol Section 6.8.6]).

- Grade 5 adverse event

Laboratory parameters of Grade 3 or 4 of any duration, not considered clinically relevant, will not be considered DLTs.

Grade 4 electrolyte abnormalities that last < 72 hours, are not clinically complicated, and resolve spontaneously or respond to conventional medical interventions will not be considered DLTs.

The following adverse events will NOT be adjudicated as DLT regardless of the grade or duration:

- Lymphopenia
- Fever
- Tumor lysis syndrome (TLS) including associated manifestations attributable to TLS (e.g., electrolyte abnormalities, renal dysfunction, hyperuricemia)

DLT Evaluable

The DLT endpoint is evaluable if either: 1) the subject experiences a DLT, or 2) the subject does not experience a DLT and receives IP as planned in cycle 1 and has been followed for safety events a minimum of 28 days from start of treatment.

Maximum Tolerated Dose (MTD)

The MTD will be selected from all tried dose levels that have not been previously declared to be unsafe with a decision according to the mTPI decision table. The MTD will be estimated using isotonic regression (Ji et al, 2010) and the MTD will be the dose level with the estimated DLT rate closest to 0.30.

Treatment-Emergent Adverse Event (TEAE)

TEAE is defined as adverse events starting on or after first dose of tarlatamab as determined by "Did event start before first dose of investigational product" equal to "No" or missing on the Events CRF, and up to and including 4**7** days after the last dose of tarlatamab excluding the events reported after End of Study date.

Treatment-related Adverse Event (TRAE)

TRAE is defined as treatment-emergent AE that per investigator review has a reasonable possibility of being caused by **tarlatamab** as indicated on the Events CRF page.

5.3 Secondary Endpoint-related Definitions

Best Overall Response (BOR) per RECIST 1.1

The subject's best response assignment will depend on the findings of both target and non-target disease and will also take into consideration the appearance of new lesions and confirmation of response. Confirmation of progressive disease is not required. Best overall response (BOR) will be based on all post-baseline disease assessments that occur prior to the initiation of subsequent anticancer treatment. At least 7 weeks from the first dose of IP must elapse without radiological disease progression to meet the minimum criteria for SD duration in order to assign a best overall response of SD. In general, subjects not classifiable under the RECIST 1.1

response categories due to inadequate data or early death will be classified as non-evaluable (NE) for BOR but will be counted in the denominator of all response rate calculations. The BOR will depend on the findings from MRI/CT scans. BOR per RECIST 1.1 has the following possible categories: complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD) or not evaluable (NE). The rules for deriving the BOR are detailed in Appendix D.

Disease control rate (DCR)

DCR is defined as the **proportion** of subjects with a best overall response of confirmed response (CR/PR) or SD as per RECIST 1.1.

Duration of Response (DOR) per RECIST 1.1

DOR is defined as the time from the date of an initial objective response per RECIST 1.1 to the earlier of progressive disease per RECIST 1.1 or death for subjects with an objective response. **DOR** will be censored following the same censoring strategy as radiographic progression-free survival as detailed in Appendix B.

Objective Response (OR) per RECIST 1.1

OR is defined as best overall response of PR or CR per RECIST 1.1, confirmed by an assessment at least 4 weeks later. Subjects who do not experience a PR/CR or do not have any follow-up tumor assessments will be regarded as non-responders.

Objective Response Rate (ORR)

ORR is defined as the **proportion** of subjects with a best overall response of **confirmed** CR or PR per RECIST 1.1.

Overall Survival (OS)

OS is defined as the time from the start of treatment until 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 (i.e., the data cut-off date), the subject will be censored at the data cut-off date.

Radiographic Progression:

A radiographic progression occurs if either of the following occurs:

- Progressive disease **in soft tissue** per RECIST 1.1.
- Bone **disease** progression **as** per PCWG3. **Bone disease progression for the first post-treatment bone scan (e.g. week 8) showing two or more additional lesions relative to screening/baseline must be confirmed by the next bone**

scan (e.g. week 16) showing two or more additional lesions relative to the first post-treatment bone scan (e.g. week 8). Bone disease progression for scans after the first post-treatment scan (e.g. week 16) showing two or more additional lesions relative to first post-treatment bone scan (e.g. week 8) must be confirmed by a later scan showing two or more additional lesions relative to the first post- treatment scan (e.g. week 8).

Rules to derive timepoint radiographic progression by RECIST 1.1 with PCWG3 modifications are detailed in Table 14-6.

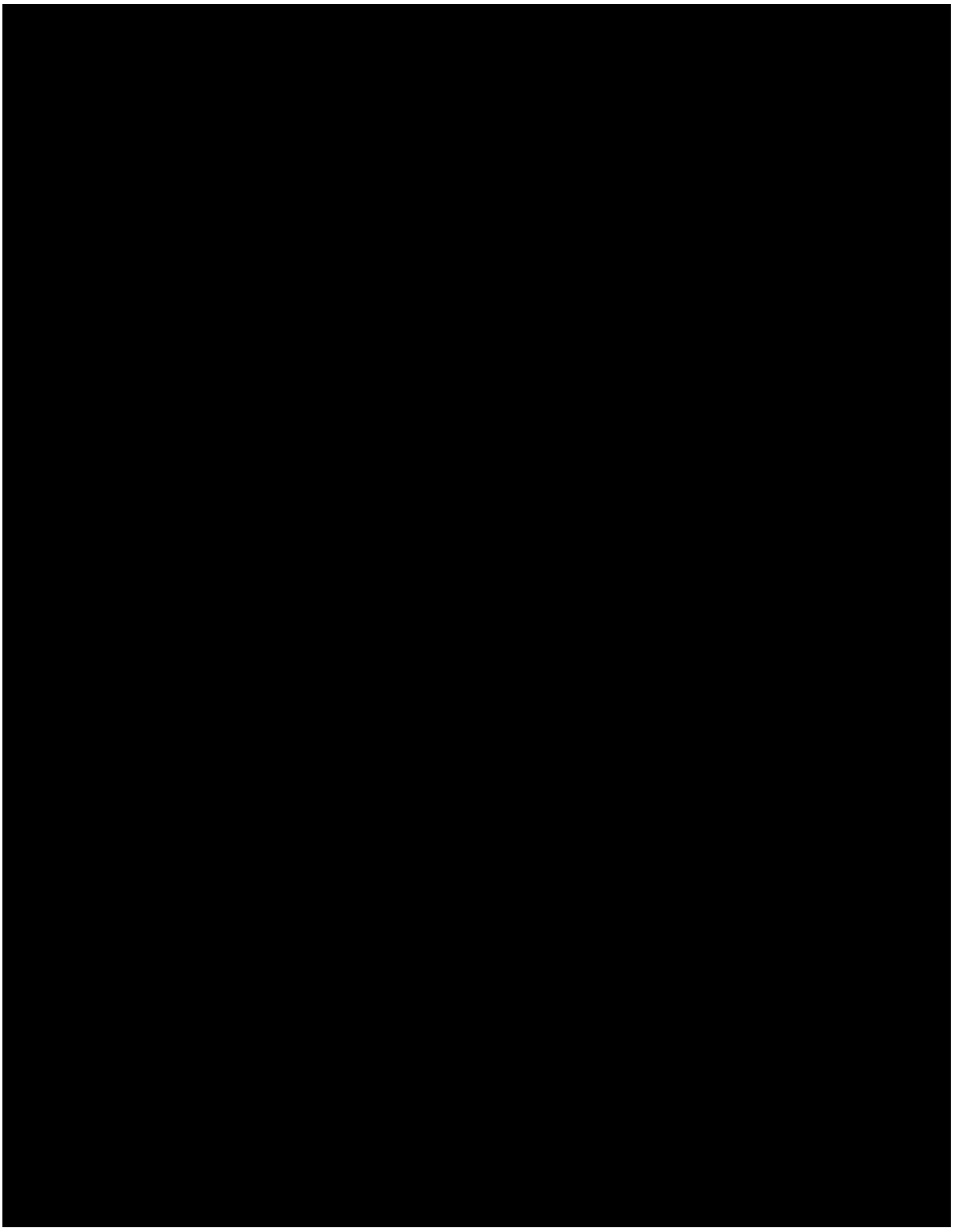
Radiographic Progression-free Survival (rPFS)

rPFS is defined as the interval from the first dose of tarlatamab to the earlier of a radiographic progression or death from any cause. **Censoring is at the last evaluable disease assessment date (tumor assessment or bone scan) based on the** rules detailed in Appendix B.

PFS = (PD or death date - start of treatment date + 1) x 12 / 365.25

5.4 Exploratory Endpoint-related Definitions

[redacted]



6. Analysis Sets

6.1 Full Analysis Set

The Full Analysis Set is the same as the Safety Analysis Set as described in Section 6.2.

6.1.1 Primary Analysis Set

Not applicable to the study.

6.2 Safety Analysis Set

The Safety Analysis Set **is defined as** all subjects who received at least 1 dose of tarlatamab. The analysis of all safety and efficacy endpoints, unless noted otherwise, will be conducted on the Safety Analysis Set.

6.3 Per Protocol Set(s)

Not applicable to the study.

6.4 Health-related Quality-of-Life or Health Economics Analyses Set(s)

Not applicable to the study.

6.5 Pharmacokinetic/Pharmacodynamic Analyses Set(s)

The PK Analysis Set contains all subjects who received at least 1 dose of tarlatamab 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. For all study parts, PK Analysis Set will be used to conduct the analysis of PK data, unless otherwise specified.

6.6 Interim Analyses Set(s)

At interim analysis with cut-off determined, the DLT data will be analyzed based on DLT Analysis Set, PK parameters will be analyzed based on PK Analysis Set, interim efficacy analysis will be performed on the Interim Efficacy Analysis Set. All other data will be reported using Safety Analysis Set, unless otherwise specified.

The Interim Efficacy Analysis Set consists of all subjects in the Safety Analysis Set who have had the opportunity to be followed for at least 9 weeks starting from **study** day 1. Subjects who stopped disease assessment 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.

6.7 Study-specific Analysis Set(s)

6.7.1 DLT Analysis Set

DLT Analysis Set is defined as all subjects from the Safety Analysis Set who are DLT evaluable **as defined in section** 5.2. For each part of the study, the analysis of DLT will be conducted on the DLT Analysis Set.

6.7.2 RECIST 1.1 Evaluable by Central Reviewer Analysis Set (/by Investigator Assessment)

RECIST 1.1 Evaluable by Central Reviewer Analysis Set is defined as all subjects that are enrolled, received at least 1 dose of tarlatamab, have measurable baseline disease per RECIST 1.1 with PCWG3 modifications as assessed by central reviewer and had the opportunity to be followed for at least 9 weeks starting from study day 1. Subjects who stopped disease assessment 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. This analysis set will be used for RECIST 1.1 related endpoints assessed by central reviewer for interim, primary and final analyses.

RECIST 1.1 Evaluable by Investigator Assessment Analysis Set is similarly defined using RECIST 1.1 with PCWG3 modifications as assessed by an investigator. This analysis set will be used for RECIST 1.1 related endpoints assessed by investigator for interim, primary and final analyses.

[redacted]

7. Planned Analyses

The following interim, primary and final analysis are planned.

7.1 Interim Analysis and Early Stopping Guidelines

Safety data will be reviewed on an ongoing basis. During dose exploration and formally during DLRMs, Amgen, in consultation with the site investigators, will review all available cumulative data by dose level prior to making dose escalation or dose de-escalation recommendations. Adverse events and DLTs observed in all subjects will be evaluated continually and fully integrated into all DLRMs and considered in all enrollment and dosing decisions.

For DLRMs, data will be subject to ongoing checks for integrity, completeness and accuracy; however, there is no requirement to resolve outstanding data issues ahead of the snapshot. All available data up to and including the data cut-off date will be included in the analysis based on an "as-is" snapshot of the database without data locking.

During dose expansion, Amgen will conduct evaluations of the ongoing grade 4 or higher treatment-related adverse event rate to assess if the threshold for possible early trial termination has been reached. If this threshold is met in either group in Part 2, enrollment to both dose expansion groups 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.

1) Terminate the trial

2) Resume Dose Exploration with a schedule as outlined in the protocol and

recommended by the DLRT.

3) Amend the protocol to potentially improve the benefit/risk for subjects (e.g., increase safety monitoring, modify dose/schedule, mandate premedication)

4) Continue dose expansion without any changes.

The stopping rules use a Bayesian approach proposed by Thall and colleagues (Thall et al, 1995) to terminate the study if the posterior probability that the grade 4 or higher treatment-related adverse event rate is greater than 20% is > 80%. The stopping boundaries assume a prior distribution of Beta (0.40, 1.60) as presented in Table 7-1 and the operating characteristics with pre-specified batch size of 10 new subjects per batch are presented in Table 7-2. The operating characteristics in Table 7-2 provide the probability of stopping the trial early for given hypothetical true rate of grade 4 or higher treatment-related adverse events, whereas the stopping criteria in Table 7-1 are based on situations where the empirical evidence would result in a posterior probability of ≥ 80% that the true grade 4 or higher treatment-related adverse event rate is ≥ 20%.

Table 7-1. Stopping Boundary 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
10	≥ 4
20	≥ 6
30	≥ 9
40	≥ 11
50	≥ 13
60	≥ 15
70	≥ 18
80	Dose Expansion Complete

Table 7-2. Operating Characteristics with Batch Size of 10 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%	78.6
0.15	11.9%	73.1
0.20	37.2%	60.3
0.25	69.8%	43.3

0.30	91.0%	29.1

Interim futility analyses (non-binding) will be performed [redacted] in a continuous manner using Bayesian predictive probability (Lee & Liu, 2008). Interim futility analyses will begin in each group after approximately 20 subjects have been treated and have a minimum potential follow-up of 9 weeks. Interim futility analyses will be based on the Interim Efficacy Analysis Set. Following this initial interim futility analysis, subsequent interim futility analyses will be performed after every 10 subjects in each group have reached the data cutoff criteria, until all subjects are enrolled and evaluated.

An additional interim analysis will be performed [redacted] once each subject in this group either completes at least 6 months on study or withdraws from the study. For interim analyses a response is defined as either a confirmed or unconfirmed CR or PR as per RECIST 1.1. For the final analysis, a response is defined as a confirmed CR or PR as per RECIST 1.1.

For the interim analysis, data will be subject to ongoing checks for integrity, completeness and accuracy with the expectation that outstanding data issues for selected CRF pages as deemed critical by the study team are resolved ahead of the snapshot to the extent possible. All available data up to and including the data cut-off date will be included in the analysis based on an “as-is” snapshot of the database with targeted data cleaning without data locking.

The Target Value (TV) is set to demonstrate a 18% increase in ORR (i.e., TV = 30%) over the benchmark 12% ORR. For Part 2, at final analysis:

- NoGo criteria will be met if the probability that the true ORR exceeds the pre-specified TV of 30% is less than 5% (i.e., NoGo: P[ORR > TV] < 5%).

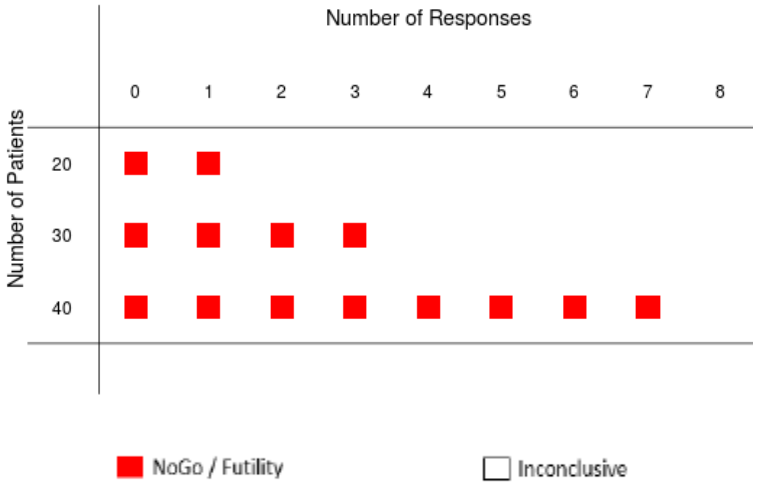
Given the existing observed data during the continuous monitoring stage, the Bayesian predictive probability is obtained by calculating the probability of reaching a NoGo decision should the treatment group be enrolled and evaluated to the maximum planned final sample size of 40 in each group in Part 2. Further enrollment to the respective group in Part 2 may be terminated if futility criteria are met. Notice the NoGo criteria is for interim futility analysis purpose only.

- Futility is met if it is predicted that there is a high probability of reaching a NoGo decision upon full enrollment of 40 subjects in the respective group in Part 2

given the existing observed data (ie, predictive probability of a NoGo decision > 95%).

The decision matrix for continuous monitoring of ORR is provided in Figure 7-1 assuming a prior distribution of Beta (1, 1). For example, given TV of 30%, 3 or fewer observed responders after 30 subjects have reached the data cutoff criteria would trigger stopping enrollment claiming futility due to a high probability of reaching a NoGo decision upon full enrollment to 40 subjects in the respective group, given the existing observed data in that group.

Figure 7-1. Decision Matrix for Continuous Monitoring of ORR



Operating characteristics of this continuous monitoring method are presented in Table 7-3 and are based on 100 000 simulations using the prespecified TV value, and NoGo and Futility definitions.

A false NoGo is defined as observing either futility criteria during the continuous monitoring stage or NoGo criteria at final analysis when the true ORR is greater than or equal to the TV (ie, False NoGo: P[Futility or NoGo Decision | True ORR ≥ TV]).

Table 7-3. Guidelines for Futility Analysis

True ORR (BM ORR = 12%)	Probability of Futility	Expected N	% Decision Prior to Final
10%	96%	29	67%
15%	76%	35	35%
20%	44%	38	14%
25%	19%	39	5%
30%	6%	40	1%
35%	1%	40	0%

40%	0%	40	0%

BM ORR = benchmark objective response rate; TV = target value.
Criteria: TV = 30%; NoGo: Pr(ORR > TV) < 5%; Futility: PP(NoGo) > 95%;
Frequency: 10 subjects.

7.2 Primary Analysis

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

For the primary analysis, data will be subject to ongoing checks for integrity, completeness and accuracy with the expectation that outstanding data issues are resolved ahead of the lock. The data supporting the interim analysis will be locked to prevent further changes.

7.3 Final Analysis

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

For the final analysis, data will be subject to ongoing checks for integrity, completeness and accuracy with the expectation that all outstanding data issues are resolved ahead of the final lock.

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.

8.3 Handling of Missing and Incomplete Data

Partially missing adverse event, concomitant medication and death dates will be imputed as outlined 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 study part. The clinical study team will identify and document the criteria for important protocol deviations.

8.5 Outliers

Descriptive statistics will be used to identify potential outliers in key variables and/or parameters. Suspected outliers will be included in all analyses unless there is sufficient scientific justification to exclude them.

8.6 Distributional Characteristics

The assumptions underlying the proposed statistical methodologies will be assessed as appropriate. Data transformations or alternative non-parametric methods of analyses will be utilized should there be a requirement.

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

Analyses will be split by study part and group (for part 2 only) unless otherwise specified

[redacted]

Descriptive statistics will be provided for selected demographics, exposure, safety, PK, efficacy data. 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 (e.g., 1-year overall survival [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 percent of subjects who were enrolled, received investigational product, completed investigational product, discontinued from investigational product (including

reasons for discontinuation), completed study, discontinued the study (including reasons for discontinuation) will be summarized. Key study dates for the first subject enrolled, last subject enrolled, and last subject's end of study will be presented.

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 the demographic and baseline characteristics will be produced:

- Race (White, Asian, Black or African American, and other categories depending on frequency observed). If multiple races have been reported for a subject, the subject will be categorized as 'Multiple'.
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino) **for subjects reporting 'White' race**
- Sex (Female, Male)
- Age at enrollment in years
- ECOG performance status at baseline (0, 1, 2, 3, 4)
- Weight (kg)
- Height (cm)
- BMI (kg/m^2)
- Time from initial diagnosis to enrollment (months)
- Histopathology **t**ype (Adenocarcinoma, Small cell neuroendocrine prostate cancer, Prostate cancer with neuroendocrine differentiation, Other)
- Prior lines of therapy (0, 1, 2, 3, >=4)
- Prior radiotherapy for current malignancy (Yes, No)
- Prior surgery for current malignancy (Yes, No)

- Metastatic (Yes, No)
- Number of metastatic sites (0,1, 2, >=3)
- Brain metastases (Yes, No)
- Liver metastases (Yes, No)
- Bone metastases (Yes, No)
- Sum of diameters of target lesions at baseline **based on investigator assessment** (mm)
- [redacted]

9.5 Efficacy Analyses

The following efficacy analysis will be performed.

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)/Estimand(s)

For the primary analysis, central reviewer data when available will be used for the following secondary efficacy endpoints: objective response per RECIST 1.1, duration of response per RECIST 1.1, disease control rate per RECIST 1.1 and radiographic progression-free survival per RECIST 1.1 with PCWG3 modifications. Local investigator data may also be used for primary analysis if deemed necessary.

Analyses of **s**econdary **e**fficacy **e**ndpoints are defined in the Table 9-1.

Table 9-1. Secondary Efficacy Endpoint Summary Table

Endpoint	Statistical Analysis	Analysis Set
Objective response (OR) per RECIST 1.1	The proportion of subjects with an OR will be summarized along with the Clopper-Pearson (Clopper and Pearson, 1934) exact 95% CI.	**Local data: RECIST 1.1 Evaluable by Investigator Assessment Analysis Set; Central data: RECIST 1.1 Evaluable by Central Reviewer Analysis Set**
Duration of response (DOR) per RECIST 1.1	The distribution of DOR, including the median and **percentiles** will be summarized using the Kaplan-Meier (KM) method with CI calculated using the Brookmeyer and Crowley (Brookmeyer and Crowley, 1982) method. Landmarks at the selected time points (e.g. 3, 6, 9 and 12 months) will be reported along with	**Local data: RECIST 1.1 Evaluable by Investigator Assessment Analysis Set; Central data:**

Endpoint	Statistical Analysis	Analysis Set
	95% CI using the Greenwood formula (Kalbfleisch and Prentice, 1980) to estimate the standard error in CI calculation.	**RECIST 1.1 Evaluable by Central Reviewer Analysis Set**
Radiographic progression-free survival (rPFS) per RECIST 1.1 with PCWG3 modifications	The distribution of rPFS, including the median and **percentiles** will be summarized using the Kaplan-Meier (KM) method with CI calculated using the Brookmeyer and Crowley (Brookmeyer and Crowley, 1982) method. rPFS rate at the selected time points (e.g. 3, 6, 9 and 12 months) will be reported along with 95% CI using the Greenwood formula (Kalbfleisch and Prentice, 1980) to estimate the standard error in CI calculation.	Interim: Interim Efficacy Analysis Set Primary/Final: Safety Analysis Set
Overall survival (OS)	The distribution of OS, including the median and **percentiles** will be summarized using the Kaplan-Meier (KM) method with CI calculated using the Brookmeyer and Crowley (Brookmeyer and Crowley, 1982) method. OS rate at the selected time points (e.g. 3, 6, 9 and 12 months) will be reported along with 95% CI using the Greenwood formula (Kalbfleisch and Prentice, 1980) to estimate the standard error in CI calculation.	Interim: Interim Efficacy Analysis Set Primary/Final: Safety Analysis Set
Disease **c**ontrol **r**ate (DCR) per RECIST 1.1	The proportion of subjects with disease control (i.e. a BOR of CR, PR or SD**)** with corresponding exact 95% CI will be calculated using the Clopper-Pearson (Clopper and Pearson, 1934) method.	**Local data: RECIST 1.1 Evaluable by Investigator Assessment Analysis Set; Central data:**

Endpoint	Statistical Analysis	Analysis Set
		RECIST 1.1 Evaluable by Central Reviewer Analysis Set

9.5.3 Analyses of Exploratory Efficacy Endpoint(s)/Estimand(s)

[redacted] Table 9-2.

Table 9-2. Exploratory Efficacy Endpoint Summary Table

Endpoint	Statistical Analysis	Analysis Set
[redacted]		

Endpoint	Statistical Analysis	Analysis Set
[redacted]		

Endpoint	Statistical Analysis	Analysis Set
[redacted]		

9.6 Safety Analyses

9.6.1 Analyses of Primary Safety Endpoint(s)

Analyses of Primary Safety Endpoints are defined in the Table 9-3.

Table 9-3. Safety Endpoint Summary Table

Endpoint	Statistical Analysis	Analysis Set
DLT	A listing and summary of the subject incidence of DLT will be provided.	DLT Analysis Set
TEAE	The subject incidence of adverse events will be summarized for all TEAEs, TRAEs, serious TEAEs, grade 3 or above TEAEs, fatal TEAEs, TEAEs leading to withdrawal of investigational product, TEAEs leading to interruption of investigational product.	Safety Analysis Set
Vital Signs	Summary of non-missing sample size (n), mean, standard deviation, median, minimum, and maximum for continuous data will be estimated for absolute and change from baseline data.	Safety Analysis Set
ECG	No statistical analysis is planned.	
Lab	Summary of non-missing sample size (n), mean, standard deviation, median, minimum, and maximum for continuous data will be estimated for absolute and change from baseline data. Frequencies and percentages for categorical data will be reported. Shifts in grades of safety laboratory values between the baseline and the worst on-study value will be tabulated. Potential Hy's Law cases will be summarized.	Safety Analysis Set

9.6.2 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 25.0 or later will be used to code all events categorized as adverse events to a system organ class and a preferred term.

The subject incidence of adverse events will be summarized for all TEAEs, TRAEs, serious TEAEs, grade 3 or above TEAEs, fatal TEAEs, TEAEs leading to withdrawal of investigational product, TEAEs leading to interruption of investigational product.

Subject incidence of all TEAEs, TRAEs, serious TEAEs, grade 3 or above TEAEs, fatal TEAEs, and TEAEs leading to withdrawal and interruption of investigational product or other protocol-required therapies will be tabulated by system organ class and preferred term in descending order of frequency. Where appropriate the tables will also be presented by worst grade. The severity of each adverse event will be graded using CTCAE version 4.0. The above AE tables will not be created if two or fewer subjects experience the AE.

Subject incidence of events of interest (standardized MedDRA queries and/or Amgen Medical Queries) will also be summarized according to their categories and preferred term.

9.6.3 Laboratory Test Results

Laboratory data will be summarized using standard descriptive statistics at each scheduled time point in the study. 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.

The subject incidence of potential cases of Hy's Law will be summarized. Hy's law cases 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 may be produced for the subjects suspected of fulfilling the criteria for Hy's law.

9.6.4 Vital Signs

Vital signs data will be reviewed for each subject. Depending on the size and scope of the changes, the analyses of vital signs may include summary statistics over time and/or changes from baseline over time.

Shifts in scores for ECOG performance status between the baseline and each assessed post-baseline timepoints will be tabulated. ECOG performance status scores will be summarized at relevant time points.

9.6.5 Physical Measurements

The analysis of physical measurements will include summary statistics at baseline.

9.6.6 Electrocardiogram

The ECG measurements are performed as per standard of care for routine safety monitoring, rather than for assessment of potential QTc effect. No statistical analyses of ECG measurements are planned.

9.6.8 Exposure to Investigational Product

Subject exposure to tarlatamab will be summarized using descriptive statistics. Descriptive summary will be provided for **number of doses**, cumulative dose (mg), relative dose intensity (%) and **treatment** duration (weeks).

Listing of the unique manufacturing lot numbers, and a listing of the subjects administered each manufacturing lot number will be provided.

9.6.9 Exposure to Non-investigational Product

Not applicable for this study.

9.6.10 Exposure to Other Protocol-required Therapy

Descriptive statistics may be produced to summarize the exposure to dexamethasone 8 mg IV (or equivalent dose of other corticosteroids) and/or Prophylaxis with IV hydration 1 L normal saline.

9.6.11 Exposure to Concomitant Medication

The number and proportion of subjects receiving concomitant medication will be summarized by preferred term or category as coded by the World Health Organization Drug (WHO DRUG) dictionary.

9.7 Other Analyses

The following additional analyses will be performed.

9.7.1 Analyses of Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Endpoints

For tarlatamab, PK parameters will be determined from the time concentration profile using standard non-compartmental approaches and considering the profile over the complete sampling interval. Based on the review of the data, exploratory analyses to describe the relationship between tarlatamab exposure and either PD effect and/or clinical outcome may also be performed. These exploratory analyses may not be part of the clinical study report.

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.

9.7.4 Analyses of Biomarker Endpoints

[redacted]

9.7.5 Analyses for COVID-19 Impact

Impact of COVID-19 on study conduct and patient safety may be analyzed as per the guidelines in GDE-409220 if applicable.

10. Changes From Protocol-specified Analyses

Protocol section 9.3.1 mentions interim efficacy analysis to be performed using Interim Efficacy Analysis Set. But the interim efficacy analysis will be performed using RECIST 1.1 Evaluable by Investigator Assessment Analysis Set and RECIST 1.1 Evaluable by Central Reviewer Analysis Set for local data and central data respectively for the following endpoints: objective response per RECIST 1.1, duration of response per RECIST 1.1 and disease control rate per RECIST 1.1.

11. Literature Citations / References

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

Clopper CJ, Person ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometricka.* 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 dose finding 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.

Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clinical Trials. 5(2):93-106. 2008

Scher HI, Morris MJ, Stadler WM, et al. Trial design and objectives for castration-resistant prostate cancer: Updated recommendations from the prostate cancer clinical trials working group 3. *J Clin Oncol*. 2016;34:1402-1418.

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

There is no prioritization of analyses.

13. Data Not Covered by This Plan

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

14. Appendices

Appendix A. Imputation Rules for Partial or Missing Dates for Adverse Events, Concomitant Medications and Death

The start and stop dates for adverse events, concomitant medications and death will be imputed using the following algorithm:

Table 14-1. Imputation Rules for Partial or Missing Start Dates

Start Date		Stop Date: Complete: *yyyymmdd* $< 1^{st}$ dose	Stop Date: Complete: *yyyymmdd* $\geq 1^{st}$ dose	Stop Date: Partial: *yyyymm* $< 1^{st}$ dose *yyyymm*	Stop Date: Partial: *yyyymm* $\geq 1^{st}$ dose *yyyymm*	Stop Date: Partial: *yyyy* $< 1^{st}$ dose *yyyy*	Stop Date: Partial: *yyyy* $\geq 1^{st}$ dose *yyyy*	Stop Date: missing
Partial: *yyyymm*	$= 1^{st}$ dose *yyyymm*	2	1	2	1	n/a	1	1
	$\neq 1^{st}$ dose *yyyymm*		2		2	2	2	2
Partial: *yyyy*	$= 1^{st}$ dose *yyyy*	3	1	3	1	n/a	1	1
	$\neq 1^{st}$ 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.

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

Imputation rules for partial or missing stop dates:

- For partial stop date mmyyyy, impute the last of the month.
- For partial stop date yyyy, impute December 31 of the year.
- For completely missing stop date, do not impute.
- If the stop date imputation leads to a stop date that is after the death date, then impute the stop date as the death date.

- If the stop date imputation leads to a stop date that is before the start date, then there is a data error and do not impute the stop date. (i.e. set the stop date as missing).

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. Censoring Rules for Radiographic Progression-free Survival

Table 14-2. Censoring Rules for Radiographic Progression-free Survival

Situation up to DCO/EOS	Date of Event or Censor	Outcome
PD	Date of first detection of PD	Event
No PD, but death recorded	Date of death	Event
Subject has no post-baseline radiographic tumor assessment and a vital status of alive or unknown	First dose date of IP	Censor
Subject does not show progression according to RECIST 1.1 or bone scan using PCWG3 criteria	Date of the last scan	Censor
Subject remains on study treatment and prior scans do not show radiographic progression	Date of the last scan showing no disease progression	Censor
Subject receives subsequent therapy known or intended for prostate cancer during the study prior to the radiographic progression	Date of the last scan showing no disease progression	Censor
Subject misses >= 2 consecutive planned radiographic scans (18 weeks) or has >= 2 consecutive unreadable scans (18 weeks)	Date of the last scan showing no disease progression	Censor

Appendix C. Code Fragments

95% Confidence Interval by Clopper Pearson Method

proc freq data= <data>;

tables <variable>/ binomial(exact) ;

run;

Kaplan Meier Method

proc lifetest data=<data> **<CONFTYPE=LINEAR for Brookmeyer and Crowley method> <CONFTYPE=LOGLOG for Kalbfleisch and Prentice method>**;

time T * Status(0);

run;

where

T is time to event;

Status is censoring variable;

and 0 is flagged for censored observation.

Appendix D. BOR per RECIST 1.1

BOR by investigator will be derived based on the timepoint response collected in the RECIST 1.1 CRF. BOR by central review will be derived from timepoint response shared by the central imaging vendor. At each time point, BOR will be derived based upon the evaluated timepoints up to and including the current MRI/CT 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 ≥ 49 days since the cycle 1 day 1 is required
- For a BOR of CR or PR confirmation is required. PR can be confirmed by a PR or CR. The confirmation assessment must be in a consecutive assessment ≥ 28 days after the initially observed assessment of CR or PR except the following cases:
  - An unlimited number of intermittent assessments of NE or CR <28 days 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 <28 days 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 at post-baseline 5.

Table 14-3 provides the BOR determination per RECIST 1.1 where confirmation of response (CR/PR) is required. Table 14-4 outlines the steps to derive Confirmed_BOR (step 1) given timepoint assessments.

At interim futility analyses, due to lack of sufficient follow-up time for confirmation scan, 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 14-4 step 2 provides details to derive the unconfirmed responders.

Table 14-3. BOR per RECIST 1.1 where Confirmation of CR/PR is Required

Criterion	Timepoint T1 Response	T1 ≥ 49 days after C1D1?	Timepoint T2 Response	T2 ≥ 49 days after C1D1?	T2 ≥ 28 days 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	-	-	PD
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	-	-	PD
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	-	-	PD
C28			NE, No further evaluations			NE
C29	PD		-			PD
C30	NE	-	NE, No further evaluations			NE
C31		-	CR, PR, SD	Yes	-	SD
C32		-	CR, PR, SD	No	-	NE
C33		-	PD	-	-	PD

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

*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. A CR may also be followed by a PR when the target lesion assessment

is CR and **the non-target lesion assessment is NE. In both of these situations,** the best response is PR.

Table 14-4. BOR Derivation Step

<table>
<tr><th colspan="2">Step 1. Derive Confirmed_BOR at each visit</th></tr>
<tr><td colspan="2">Derive Confirmed_BOR at each visit:
i) At Visit 1: Use the Visit 1 scan result and refer to Table 14-3 to derive Confirmed_BOR.
ii) At Visit 2 onward:
(a) If current visit is CR: Find last scan that is not NE and at least 28 days before current visit. If current visit is PR: Find last scan that is not NE or SD and at least 28 days before current visit. Derive BOR_temp as below:
<table>
<tr><th>Last scan</th><th>Current</th><th>BOR_temp</th></tr>
<tr><td>CR</td><td>CR</td><td>CR</td></tr>
<tr><td>PR</td><td>CR</td><td>PR</td></tr>
<tr><td>PR</td><td>PR</td><td>PR</td></tr>
</table>
(b) If none of above fits, find last scan reference Table 14-3 to derive BOR_temp.
iii) Current visit Confirmed_BOR = best of (BOR_temp, last visit confirmed_BOR). Use rule CR > PR > SD > PD > NE</td></tr>
<tr><th colspan="2">Step 2. Derive Interim_BOR at last visit prior to DCO</th></tr>
<tr><td>For subjects who discontinue tumor assessment (i.e. had PD, next therapy, or end of study):

Assign Interim_BOR = Confirmed_BOR</td><td>For subjects with potential for more assessment (i.e. no PD, next therapy, or end of study), unconfirmed PR/CR can be considered responders.
• If any scan is CR and haven’t got the opportunity to receive next scan, then Interim_BOR = CR.
• If any scan is PR and haven’t got the opportunity to receive next scan, then Interim_BOR = PR.
• Else, Interim_BOR = Confirmed_BOR at the latest scan.</td></tr>
</table>

Appendix E. Radiographic Progression per RECIST 1.1 with PCWG3 Modifications

Overall Bone Progression Status (Bone Scan)

Table 14-5. Bone Progression Status per PCWG3

Bone Progression Status	Definition
Non-PD	**Persistence of one or more bone lesions.**
PD	**Bone lesions fulfilling the requirements for new lesions and confirmation of radiographic progression.**
Not performed (NP)	**Only relevant if a bone scan is not performed at that visit.**

Radiographic Progression per RECIST 1.1 with PCWG3 Modifications

Combine RECIST 1.1 visit response collected in the data set and bone progression status (PCWG3) derived from Table 14-5 to get radiographic progression status by RECIST 1.1 with PCWG3 modifications as outlined in Table 14-6.

Table 14-6. Radiographic Progression by RECIST 1.1 with PCWG3 Modifications

RECIST 1.1 Visit Response (soft tissue/visceral)	Bone Progression Status	Bone Lesion (Present/Absent)	Timepoint Radiographic Progression by RECIST 1.1 with PCWG3 Modifications
PD	**Any**	**Any[b]**	**PD**
Any	**PD**	**Any[b]**	**PD[e]**
SD, Non-CR/Non-PD[h] , PR, CR	**Non-PD[f] , NEB[g] , or NP[a]**	**Any[b]**	**Non-PD**
NED[c]	**Non-PD[f]**	**Any[b]**	**Non-PD**
NED[c]	**NP[a]**	**Any[b]**	**NE[d]**

NE[d]	Non-PD[f] , or NP[a]	Any[b]	NE[d]

a. Bone scan was not done or issue with the quality of the image.

b. Any result whether measured or missing.

c. No evidence of soft tissue disease at all visits.

d. Not evaluable.

e. Following the confirmation of PD at the subsequent bone scan, the overall radiological visit response of progression for the visit will be programmatically derived as PD using the progression date (previous bone scan date where new lesions first seen) captured on the eCRF.

f. See Table 14-5: Bone Progression Status per PCWG3.

g. No evidence of bone lesion at baseline.

h. Valid RECIST 1.1 response option only for subjects with non-measurable soft tissue disease only at baseline.