Statistical Analysis Plan

Protocol Title:	A Phase 1 Study to Evaluate Safety, Tolerability, Pharmacokinetics and Pharmacodynamics of AMG 404, a Programmed Death-1 (PD-1) Antibody, [redacted] in Patients With Advanced Solid Tumors
Short Protocol Title:	AMG 404 in Patients with Advanced Solid Tumors
Protocol Number:	20180143
Authors:	[redacted]
Sponsor:	Amgen Inc. One Amgen Center Drive Thousand Oaks, CA 91320 USA
SAP Date:	**Document Version** — **Date** Original (v1.0) — 25 April 2019 Amendment 1 (v2.0) — 25 July 2022 **Amendment 2 (v3.0) — 19 August 2022**

Version Number	**Date (DDMMMYYYY)**	**Summary of Changes, including rationale for changes**
Original (v1.0)	25APR2019	NA
Amendment 1 (v2.0)	25JUL2022	Aligned with changes made in protocol amendment 6 dated 20 May 2022.
Amendment 2 (v3.0)	**19AUG2022**	**Updated BOR, DOR, DoSD, PFS definition** **Updated Appendix C**

NCT Number: NCT03853109
This NCT number has been applied to the document
for purposes of posting on Clinicaltrials.gov

Table of Contents

[redacted]

List of Tables

List of Appendices

List of Abbreviations and Definition of Terms

Abbreviation or Term	Definition/Explanation
AE	Adverse event
AUC	Area under the serum concentration-time curve
BOR	Best overall response
CI	Confidence Interval
C_{max}	Maximum observed serum concentration
CPMS	Clinical Pharmacology Modeling and Simulation
CR	Complete response
CTCAE	Common Terminology Criteria for Adverse Events
CT/MRI	Computed tomography/ Magnetic resonance imaging
DCR	Disease control rate
DLRM	Dose level review meeting
DLRT	Dose level review team
DLT	Dose Limiting Toxicity
dMMR	Mismatch repair deficient
DNA	Deoxyribonucleic acid
DOR	Duration of response
DoSD	Duration of stable disease
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
Electronic Source Data (eSource)	Source data captured initially into a permanent electronic record used for the reconstruction and evaluation of a trial.
Enrollment	Date of enrollment confirmation letter from Amgen
FIH	First in Human
ICH	International Conference on Harmonisation
IP	Investigational product
IPD	Important protocol deviation
IV	Intravenous
LTFU	Long term follow-up
MedDRA	Medical Dictionary for Regulatory Activities
mg	milligram
MSI-H	Microsatellite instability-high
MTD	Maximum tolerated dose
(m)TPI	(modified) Toxicity Probability Interval
NCI	National Cancer Institute
NE	Not evaluable

Abbreviation or Term	Definition/Explanation
NMPA	National Medical Products Administration
NSCLC	Non-small cell lung cancer
ORR	Objective response rate
PD	Progression of disease
PD	Pharmacodynamics
PD-1/PD-L1	Programmed death-1
PET	Positron Emission Tomography
PFS	Progression free survival
PK	Pharmacokinetics
PR	Partial response
QTc	QT interval corrected for heart rate using accepted methodology
QTcB	Bazett-corrected QT Interval
QTcF	Fridericia-corrected QT Interval
RECIST	Response evaluation criteria in solid tumors
RNA	Ribonucleic acid
RP2D	Recommended phase 2 dose
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Stable disease
TPS	Tumor Proportion Score
t_{max}	Time to achieve C_{max}
WBC	White blood cells
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 20180143, AMG 404 dated 20 May 2022. The scope of this plan includes the primary analyses and the final analyses that are planned and will be executed by the Amgen Global Biostatistical Science department unless otherwise specified.

2. Objectives, Endpoints and Hypotheses

2.1 Objectives and Endpoints

Objectives	Endpoints
Primary	
• To evaluate the safety and tolerability and determine the recommended phase 2 dose (RP2D) of AMG 404 in patients with advanced solid tumors	• Dose limiting toxicities (DLTs), treatment-emergent adverse events, treatment-related adverse events, and changes in vital signs, and clinical laboratory tests
[redacted]	
Secondary	
• To evaluate the pharmacokinetics (PK) of AMG 404, as a monotherapy [redacted]	• PK parameters of AMG 404 including, but not limited to, maximum observed serum concentration (C_{max}), time to achieve C_{max} (t_{max}), and area under the serum concentration-time curve (AUC)
• To assess the immunogenicity of AMG 404, as a monotherapy [redacted]	• Subject incidence of anti-AMG 404 antibodies
• To evaluate the preliminary antitumor activity of AMG 404, as a monotherapy [redacted]	• Objective tumor response, duration of overall response, progression-free survival, disease control rate (DCR), and duration of stable disease measured by CT/MRI and assessed per modified RECIST 1.1

Exploratory

[redacted]

2.2 Hypotheses and/or Estimations

AMG 404 as a monotherapy [redacted] is safe and well tolerated in patients with advanced solid tumors.

3. Study Overview

3.1 Study Design

This is a first-in-human (FIH), multicenter, non-randomized, open-label, phase 1 study to evaluate safety, tolerability, pharmacokinetics (PK) and pharmacodynamics of AMG 404 as a monotherapy [redacted], in subjects with advanced solid tumors. The study may be conducted in Australia, United States, France, Poland, Spain, Turkey, United Kingdom, Belgium, South Korea, Taiwan, Japan, China, Singapore, Canada, Brazil, and Mexico. Other countries or regions may participate. The study will consist of up to 13 cohorts. Cohorts 1-9 will evaluate the safety, tolerability, PK and pharmacodynamics of AMG 404 administered every 4 weeks (Q4W) with the following doses: Cohort 1 – 240 mg, Cohort 2 – 480 mg, Cohort 3 – Expansion at the dose of 480 mg (subjects enrolled outside of China), Cohort 4 – Exploratory Dose of 1050 mg, Cohort 5 – China and National Medical Products Administration (NMPA) certified sites in Taiwan and/or Hong Kong Specific Expansion at the recommended phase 2 dose (RP2D) with a safety lead-in, Cohort 6 – Expansion at the RP2D , Cohort 7 – Expansion at the RP2D in specific indications as specified in Section 3.1 of protocol , Cohort 8 – Expansion at the RP2D in subjects with

tumors that are microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR), and Cohort 9 – Expansion at the RP2D in subjects with non-small cell lung cancer (NSCLC) that is PD-L1 positive, Tumor Proportion Score (TPS) ≥ 50% (Cohorts 6-9 will enroll subjects outside of China).

In Cohorts 1-9, AMG 404 will be administered intravenously (IV) over approximately 30 minutes, every 4 weeks in subjects with advanced solid tumors. Cohorts 3 and 4 can be opened once Cohort 2 has been completed and determined to be tolerable. Cohorts 5-9 can be opened once the first 6-9 subjects in Cohort 4 have completed the DLT evaluation period and data have been reviewed to determine the RP2D. The DLT evaluation period will be 28 days. The Dose Level Review Meeting (DLRM) may recommend escalation to the next planned dose, escalation to an intermediate dose (a dose lower than the next planned dose), continuation or delay in dosing, repetition or expansion of a cohort, de-escalation to a lower dose, or termination of the study. Administration of AMG 404 (in all cohorts) may continue until evidence of disease progression, intolerance to study medication, withdrawal of consent, or in the absence of the above up to 12 months if subject achieves a complete response, and up to

24 months if partial response or stable disease**.** Subjects with complete response may continue treatment beyond 12 months, up to a maximum of 24 months of total therapy, with permission from the medical monitor. If a subject completes or discontinues treatment for reasons other than progression, tumor evaluations should continue according to the Schedule of Activities (for Cohorts 1-9**,** refer to Table 2-1**,** [redacted] from the protocol) until progression or start of a new treatment regimen.

Cohorts 1, 2 and Cohort 4

The doses of 240 mg, 480 mg and 1050 mg will be examined for safety, tolerability, PK, and pharmacodynamics of AMG 404 in subjects with advanced solid tumors. Rules for dose determination, and dose expansion are derived using a modified Toxicity Probability Interval (mTPI) model (Ji et al. 2010) with a target toxicity probability interval (TPI) of 0.20 to 0.33 with a TPI > 0.33 defined as excessive toxicity.

- Subjects will not be dosed at a level above the dose level indicated by the mTPI model. In particular, subjects will not be dosed at the current dose level or higher when the mTPI model indicates stopping enrollment at the current dose level and patients will not be dosed at an escalated dose level when the mTPI indicates staying at the current dose level.
- After reviewing a broad range of safety information including vital signs, laboratory values and detailed adverse events profiles, the team may choose to dose subjects below the dose level indicated by the mTPI model. For example, the team may choose to stop enrollment at the current dose level even if the mTPI model indicates staying at the current dose level.

The Dose Level Review Team (DLRT) will review data, monitor safety, and make recommendations on dose modifications based on all subjects that have been enrolled. A Dose Level Review Meeting will be conducted after Cohorts 1, 2 and 4 and may convene ad hoc to review safety data if deemed necessary.

Three dose levels are planned to be tested:

- Cohort 1: 240 mg every 4 weeks (N = 2-4)
- Cohorts 2 and 3: 480 mg every 4 weeks (Cohort 2: N = 6-9; Cohort 3: N = 20)
- Cohort 4: 1050 mg every 4 weeks (N = 20)

Dose escalation will begin with a run-in dose level of 240 mg (Cohort 1). If no DLTs are observed in Cohort 1, dose escalation will continue to the next planned dose level in Cohort 2. If DLTs are observed in Cohort 1, subsequent dosing will be determined by

Amgen after consulting with the Dose Level Review Team (DLRT) and constrained by the mTPI model.

The table below shows rules for dose determination and dose expansion per the mTPI model.

Table 3-1. Guideline for Dose Level Decisions for Cohort 2 (480 mg) and Cohort 4 (1050 mg)

No. of Subjects	Number of DLT observed at current dose		
	ESCALATE[a]	STAY AT CURRENT DOSE	STOP ENROLLMENT AT CURRENT DOSE
3	N/A	≤ 1	≥ 2
4	N/A	≤ 1	≥ 2
5	N/A	≤ 2	≥ 3
6	0	1 – 2	≥ 3
7	≤ 1	2 – 3	≥ 4
8	≤ 1	2 – 3	≥ 4
9	≤ 1	Maximum sample size reached	Maximum sample size reached

[a] For Cohort 4 (1050 mg), if the rules for dose escalation are met, enrollment may be stopped as there is sufficient evidence that 1050 mg dose is safe and tolerable; and this is the highest planned dose level being investigated on study.

Cohort 3

Upon completion of Cohort 2 of the study and depending on data obtained, dose expansion may proceed with the planned dose of 480 mg in Cohort 3 (N = 20) outside of China.

Cohorts 5-9

Upon completion of the first 6-9 subjects in Cohort 4 of the study and depending on data obtained, dose expansion may proceed with the RP2D in Cohort 5 with a safety lead-in (from China or NMPA certified sites in Taiwan and/or Hong Kong only) and Cohorts 6-9 (from all countries participating on the study, excluding China). The RP2D will be determined following review of available data obtained in Cohorts 1 through 4 including safety, tolerability, PK and pharmacodynamics data.

Cohort 5 is a China/Taiwan/Hong Kong-specific expansion cohort with a safety lead-in at one dose below the RP2D followed by evaluation at RP2D, as appropriate. Up to 12 subjects will be enrolled in China, and/or in Taiwan, and/or Hong Kong (If Taiwan or Hong Kong, only NMPA certified sites will participate). [redacted]

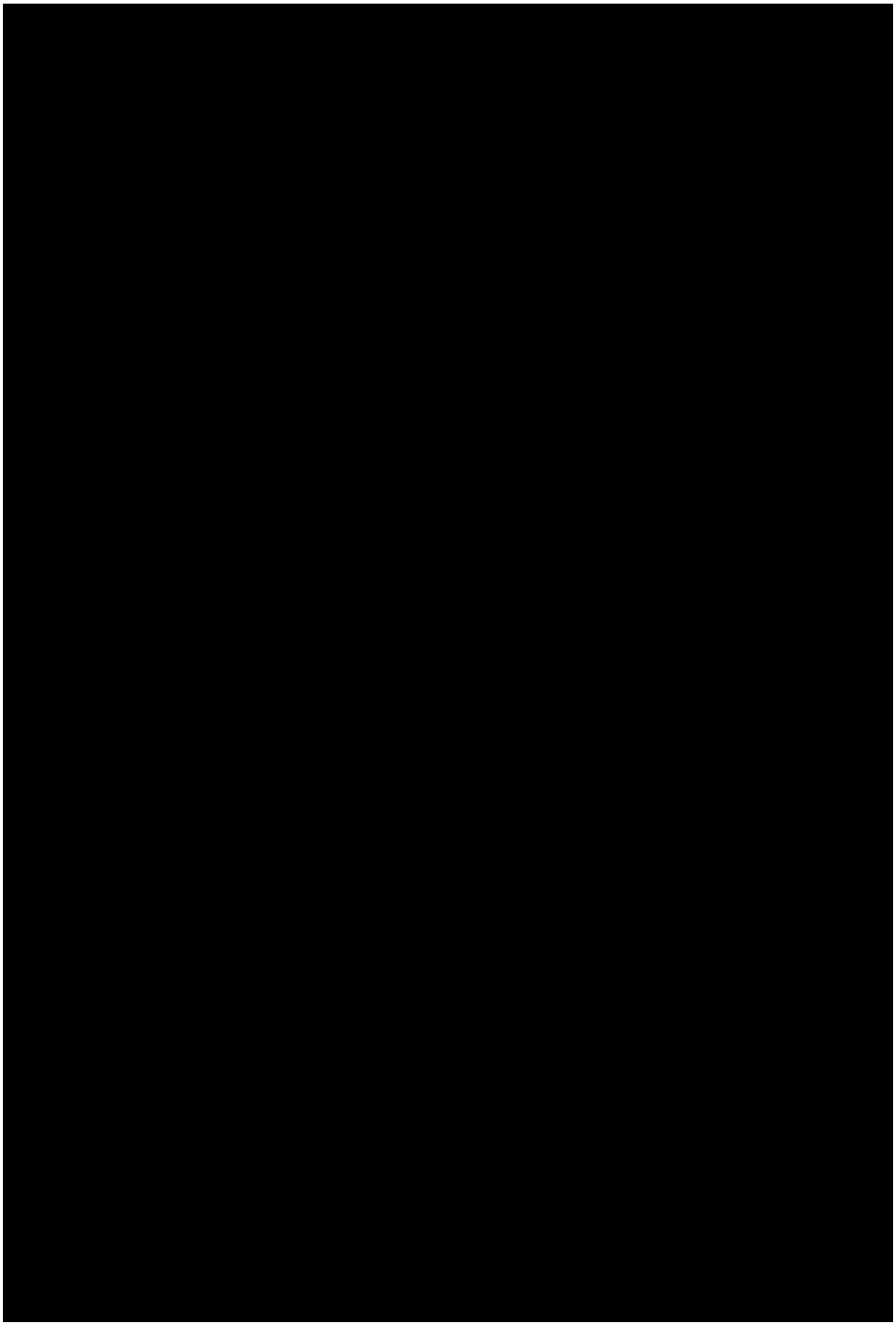
Intra-subject dose escalation for Cohorts 1-9

Intra-Subject dose escalations are allowed on this study for subjects enrolled, where applicable. Subjects enrolled to Cohort 1 who complete the DLT period may proceed to 480 mg once Cohort 2 has been deemed safe by the DLRT; subjects receiving 480 mg (in Cohorts 1-3) who complete the DLT period may proceed to the RP2D once determined if a dose higher than 480 mg is selected as the RP2D. Subjects enrolled to the safety lead-in dose in Cohort 5, may escalate to the RP2D dose level once the RP2D is determined safe following the China/Taiwan/Hong Kong specific DLRT recommendation.

All intra-subject dose escalations may occur after consultation with the sponsor if:

- No DLT has been reported for this subject during or after completion of the DLT period

Subjects who do not proceed to the higher dose may receive additional cycles at the original dose.



[redacted]

3.2 Sample Size

Up to 275 evaluable subjects will be enrolled in this study. Approximately 2 to 4 DLT-evaluable subjects will be enrolled in Cohort 1. Approximately 6 to 9 DLT-evaluable subjects will be enrolled in Cohort 2, Approximately 20 subjects will be enrolled each in Cohorts 3, 4 and 6, all outside of China. Approximately 12 subjects will be enrolled to Cohort 5 in China/Taiwan/Hong Kong. Approximately 40 subjects, respectively, will be enrolled to Cohort 7, Cohort 8 and Cohort 9. [redacted].

With 2 subjects in Cohort 1, there is a 36 – 55% probability of observing at least 1 DLT. With 4 subjects in Cohort 1, there is a 59 - 80% probability of observing at least 1 DLT if the true DLT rate is 20 – 33%.

Similarly, with 6 subjects per cohort, there is a 74 - 91% probability of observing at least 1 DLT. With 9 subjects per cohort, there is an 87 - 97% probability of observing at least 1 DLT if the true DLT rate is 20-33%.

In Cohorts 3, 4, 6, [redacted], a subject number of 20 will provide an 88% probability of observing at least one adverse event with 10% incident rate. An exact 80% binomial

confidence interval (CI) will be provided for overall response rate. With the 20 subjects and 25% overall response rate, the expected 80% CI would be 13% to 42%.

In Cohort 5, a subject number of 12 will provide a 72% probability of observing at least one adverse event with 10% incident rate. An exact 80% binomial confidence interval (CI) will be provided for overall response rate. With the 12 subjects and 25% overall response rate, the expected 80% CI would be 10% to 48%.

In Cohort 7, Cohort 8, and Cohort 9, the sample size is N~40 subjects. A subject number of N~40 subjects will provide an 87% probability of observing at least one adverse event with 5% incident rate. An exact 80% binomial confidence interval (CI) will be provided for overall response rate. With the 40 subjects and 25% overall response rate, the expected 80% CI would be 16% to 36%.

3.3 Adaptive Design

Dose level decisions are based on a modified toxicity probability interval design (mTPI). The mTPI was developed by Ji et al. (2010). The mTPI models the probability of toxicity for each dose level using a Bayesian model where each dose level has the same prior on the probability of toxicity, a Beta (1,1). When subjects are treated at the current dose level, the posterior probability of toxicity is updated using the observed data from this level. Dose level recommendation are made based on this posterior probability of toxicity, using three toxicity probability intervals (TPI).

- Under-dosing TPI: DLT rate from 0 to < 20%
- Target TPI: DLT rate from 20% to 33%
- Over-dosing: DLT rate > 33%

For the current dose level and after adjusting for the width of the under-dosing TPI, if the DLT rate is most likely in the under-dosing TPI then the recommendation is to dose escalate. If the DLT rate is most likely in the target TPI then the recommendation is to stay at the current level. If the DLT rate is most likely in the over-dosing TPI then the recommendation is to stop the enrollment. If DLTs are observed, the maximum tolerated dose (MTD) is the dose level with a DLT toxicity rate closest to 26.7% (Ji et al. 2010).

During enrollment and treatment at the RP2D using data from cohorts 3 and 6 and additionally using data from cohorts 7-9, objective response rate will be monitored with futility stopping rules described in Section 7.1. These futility stopping rules are selected using practical considerations.

During enrollment and treatment at the RP2D using data from cohorts 3 and 6 and cohorts 7-9, subject incidence of grade 4 treatment related adverse events will be monitored with safety stopping rules described in Section 7.1. The safety stopping rules use a Bayesian approach proposed by Thall, Simon, and Estey (1995) to terminate a cohort if the posterior probability that the subject incidence of grade 4 or higher treatment-related adverse events is greater than 20% is > 80%. The stopping boundaries assume a prior distribution of Beta (0.40, 1.60).

4. Covariates [redacted]

4.1 Planned Covariates

The relationship of covariates to efficacy endpoints may be explored if appropriate.

[redacted]

5. Definitions

Adverse Event (AE)

An adverse event is any untoward medical occurrence in a clinical study subject irrespective of a causal relationship with the study treatment. An adverse event can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease (new or exacerbated) temporally associated with the use of a treatment, combination product, medical device or procedure.

Age at enrollment

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

AUC

Area under the concentration-time curve reflects the actual body exposure to drug after administration of a dose of the drug (M. Hidalgo et al).

Baseline

For any variable, unless otherwise specified the baseline is the last non-missing assessment taken prior to the first administration of AMG 404. Where baseline measurements are taken on the same day as the study specified treatment and no times are present, it will be assumed that these measurements are taken prior to the study specified treatment being administered.

For parameters/assessments not scheduled to be performed (or scheduled but not performed) on the same day as the first administration of AMG 404, the baseline value is the value from the screening period measured closest to the day of first administration of AMG 404.

Baseline ECG Values in Triplicate

The baseline ECG is defined as the mean of the sets of triplicate screening ECGs and 1 set of triplicate ECGs obtained pre-dose on day 1; the mean of values in a triplicate should be calculated before taking the mean of the triplicate averages. When an ECG is missing within a triplicate, all available data will be averaged.

Bazett-corrected QT Interval (QTcB)

The Bazett correction will be calculated from the investigator reported QT (msec) and RR interval (msec), as follows: $QTcB=QT/(RR/1000)^{1/2}$.

Best Overall Response (BOR)

In general, the subject's best response assignment will depend on the findings of both target and non-target lesions and will also take into consideration the appearance of new lesions and confirmation of response.

Best overall response (BOR) will be based on all post-baseline disease assessments that occur prior to the initiation of subsequent anticancer treatment.

BOR for a subject is the best observed disease response per RECIST 1.1 in the following order: CR, PR, SD, PD, or NE. A best overall response of CR and PR require confirmation by a repeat, consecutive scan at least 4 weeks after the first documentation of response or progression.

A best overall response of SD requires an on-study imaging of SD or better no earlier than 7 weeks (49 days) after cycle 1 day 1.

Please refer Appendix C for deriving BOR algorithm as per RECIST 1.1.

Change from Baseline

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

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

Change (percent) from Baseline = [(Post-baseline Value – Baseline Value)/Baseline Value] x 100

C_{max}

The maximum concentration of drug obtained in the serum.

Disease control rate

DCR is defined as the proportion of subjects in whom objective response (CR or PR) or stable disease (SD) is determined as per RECIST v1.1.

Dose Limiting Toxicity (DLT)

Defined as the severe toxicities during the first cycle of cancer therapy. A Dose Limiting Toxicity (DLT) is defined as any AE (meeting the criteria listed in section 8.4.2 of protocol) occurring during the first treatment cycle of AMG 404 (day 1 through day 28) where relationship to AMG 404 cannot be ruled out.

The DLT window (ie, DLT-evaluable period) will be the first 28 days of AMG 404 treatment (starting cycle 1, day 1).

Duration of Overall Response (DOR)

Duration of overall response is defined as the time from the first documentation of objective response, until the first documentation of disease progression, or death due to any cause, whichever occurs first. Only subjects who have achieved objective response will be evaluated for DOR. Duration of response will be censored as per PFS censoring rule.

Duration of Stable Disease (DoSD)

Duration of stable disease will be measured from the start of treatment until the first documentation of disease progression or death due to any cause. DoSD will be calculated only in subjects with best overall response SD.

Duration of stable disease will be censored as per PFS censoring rule.

End of treatment

Defined as the last assessment for the protocol specified treatment phase of the study for an individual subject.

End of Study

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 (ie, last subject last visit), following any additional parts in the study (eg, long-term follow-up), as applicable.

End of Study for a particular subject

The end of study for a particular study subject will be the day of the last scheduled long term follow up (LTFU) call.

Enrollment

A subject is considered enrolled when the investigator decides that the subject has met all eligibility criteria (date eligibility worksheet been signed by investigator).

Fridericia-corrected QT Interval (QTcF)

The Fridericia correction will be calculated from the investigator reported QT (msec) and RR interval (msec), as follows:

$$QTcF=QT/(RR/1000)^{0.33}.$$

Investigational Product

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

Last IP Dose Date

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

[redacted]

[redacted]

Long term follow-up visit

Long-term follow-up visit will be conducted 6 months (± 1 week) after the last dose of AMG 404.

Measurable baseline disease

Measurable baseline disease is defined as “Lesions must be accurately measured in at least one dimension (longest diameter in the plane of measurement is to be recorded) with a minimum size of 10 mm by CT scan or MRI. To be measurable a lymph node must be $\geq$ 15mm in short axis when assessed by CT scan or MRI. All tumor measurements must be recorded in millimeters. Note: Previously irradiated lesions are non-measurable except in cases of documented progression of the lesion since the completion of radiation therapy.”

Objective Tumor Response

Objective response is defined as a tumor response assessment of either complete response (CR) or partial response (PR) measured by PET/CT, CT or MRI and assessed per RECIST v1.1. Subjects who do not experience PR/CR or do not have any follow-up tumor assessments will be regarded as non-responders. This endpoint will be determined only for subjects with measurable disease at baseline.

Overall Survival

Overall survival is defined as the time from the first dose of IP until death due to any cause. OS is censored at the last date known to be alive.

Primary Completion

The primary completion date 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), for the purposes of conducting the primary analysis, whether the study concluded as planned in the protocol or was terminated early.

Progression Free Survival (PFS)

Progression-free survival is defined as the time from the first dose of IP until first documentation of radiological disease progression or death due to any cause, whichever occurs first. PFS will be censored at the last evaluable post-baseline tumor assessment date; otherwise, at first dose of IP. Progression will be based on RECIST v1.1.

Safety Follow-up visit

Upon permanent discontinuation from the study treatment for any reason, two safety follow-up visits are required. A safety follow-up visit will be performed approximately 30 (+3) days after the last administration of study drug, and also 140 days (±7 days) after the last administration of AMG 404.

Serious Adverse Event

A serious adverse event (SAE) is defined as an adverse event that

- is fatal
- is life threatening (places the subject at immediate risk of death)
- requires in-patient hospitalization or prolongation of existing hospitalization
- results in persistent or significant disability/incapacity
- is a congenital anomaly/birth defect

- other medically important serious event

Study Day

Post study day: study day= (date - date of Study Day 1) + 1

Pre-study day: study day= (date – date of Study Day 1)

Study Day 1

It is defined as the first day that investigational product is administered to the subject. The day before Day 1 is referenced as Day -1.

t_{max}

The time takes to reach a maximum serum concentration of administered drug.

Treatment emergent AE

A treatment-emergent adverse event is any adverse event starting on or after the first administration of investigational product, as determined by the flag indicating if the AE started prior to first dose on the event eCRF and up to and including 140 days after the last IP dose date or end of the study, whichever occurs earlier.

Treatment related AE

A treatment-related adverse event is any treatment-emergent adverse event that per investigator review has a reasonable possibility of being caused by the investigational product. In the unlikely event that the relationship is missing, the treatment-emergent event will be considered treatment-related and documented in a footnote of the treatment-related summary.

6. Analysis Sets

The analysis of all endpoints, unless noted otherwise, will be conducted on the Safety Analysis Set.

6.1 Safety Analysis Set

The safety analysis set will consist of all subjects who received at least one dose of AMG 404.

6.2 DLT Evaluable Set

The Dose Limiting Toxicity Analysis Set will contain DLT-evaluable subjects. The DLT window (ie, DLT-evaluable period) will be the first 28 days of AMG 404 treatment (starting cycle 1, day 1) for Cohorts 1-9. [redacted]

[redacted] The subject will be DLT evaluable if the subject has completed the DLT window as described above or experienced a DLT any time during the DLT window or has received at least 90% of the planned dose of investigational product(s) and is followed for at least 1 cycle. A subject will not be DLT evaluable if he/she drops out before completion of the DLT-evaluable period for reasons other than a DLT.

The primary assessment of DLT will include DLT-evaluable subjects from dose exploration cohorts. If data is available, assessment of DLT including DLT-evaluable subjects from dose expansion cohort may be considered.

6.3 Pharmacokinetic / Pharmacodynamic Analyses Set(s)

Pharmacokinetic Analyses Set

The PK Analysis Set will contain all subjects who have received at least 1 dose of AMG 404 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.

7. Planned Analyses

7.1 Interim Analysis and Early Stopping Guidelines

Safety data will be reviewed on an ongoing basis. In the dose level review meetings, Amgen, in consultation with the site investigators, will review all available cumulative data by cohort prior to making dose determination decisions. Adverse events and DLTs observed in all subjects will be evaluated continually and fully integrated into all dose level review meetings and considered in all enrollment and dosing decisions.

Additional details regarding the dose review meeting are provided in sections 8.4.1 and 13.3 of the protocol.

There is a planned treatment of N~40 subjects at the RP2D with enrollment to cohorts 3 and 6. Additionally, there is a planned treatment of N~40 subjects at the RP2D with enrollment to Cohort 7, N~40 subjects enrolled to Cohort 8 and N~40 subjects enrolled to Cohort 9. For each of these cohorts (Cohort 3/6, Cohort 7, Cohort 8, Cohort 9), futility will be assessed after treating 15 and 25 subjects for at least 3 months. If the observed rate of responses is consistent with a lower than 15% response rate, enrollment and treatment at the RP2D may be terminated due to futility. For purposes of assessing

futility, a response is defined as an objective response per RECIST 1.1. The guidelines for early termination due to futility are as follows:

Number of Treated Subjects	Futility Termination Guideline
15	1 or fewer responders
25	4 or fewer responders
40	Enrollment to dose expansion cohort complete

If the true response rate is 15% then these termination guidelines result in a 69.4% probability of terminating dose expansion early with an expected sample size of 26.4 subjects. If the true response rate is 30% then there is a 90% probability of continuing enrollment to N~40 total subjects.

During this enrollment and treatment at the RP2D combining data from cohorts 3, 6, 7, 8 and 9 (up to 160 total subjects enrolled) and separately for each cohort including cohorts 10-13, Amgen will conduct evaluations of the ongoing subject incidence grade 4 or higher treatment-related adverse event rate to assess if the threshold for possible early termination of a cohort has been reached. If the threshold is met in the combined monotherapy data of expansion cohorts 3, 6, 7, 8, and 9 enrollment to all ongoing cohorts in the study, [redacted] will be halted pending review of safety data by the DLRT. If the threshold is met in a specific cohort, then enrollment to the specific cohort 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.

When overall safety threshold exceeded in combined monotherapy data of expansion cohorts 3, 6, 7, 8, and 9:

1) Terminate the trial
2) Amend the protocol to potentially improve the benefit/risk for subjects (eg, increase safety monitoring, modify dose/schedule, mandate premedication)
3) Continue monotherapy dose expansion cohorts and other cohorts, as appropriate, without any changes

When cohort-specific safety threshold exceeded:

1) Terminate enrollment to the cohort
2) Amend the protocol to potentially improve the benefit/risk for subjects in the cohort of concern (eg, increase safety monitoring, modify dose/schedule, mandate premedication)
3) Continue enrollment to the specific cohort (as appropriate) without any changes

The methods for deriving the stopping boundaries overall for the combined monotherapy data of expansion cohorts 3, 6, 7, 8 and 9, or for within a cohort are described in Section 11.3 of the protocol with the stopping boundaries overall for the combined monotherapy data of expansion cohorts 3, 6, 7, 8 and 9 presented in Table 7-1 and stopping boundaries for within a cohort presented in Table 7-2. Operating characteristics with pre-specified batch size of 10 new subjects per batch are presented in Table 7-3 (overall the combined monotherapy data of expansion cohorts 3, 6, 7, 8 and 9) and in Table 7-4 (cohort with planned sample size of 40 subjects). The operating characteristics in Table 7-2 and Table 7-4 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 and Table 7-2 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 Overall for Combined Monotherapy Data of Cohorts 3, 6, 7, 8, and 9 With Posterior Probability of 80% and Grade 4 or Higher Treatment-Related Adverse Event Limit of 20%

Number of subjects	Stop study if observing this many subjects with grade 4 or higher treatment-related adverse events
10	≥ 4
20	≥ 6
30	≥ 9
40	≥ 11
50	≥ 13
60	≥ 15
70	≥ 18
80	≥ 20
90	≥ 22
100	≥ 24
110	≥ 26
120	≥ 28
130	≥ 31
140	≥ 33
150	≥ 35
160	Combined Monotherapy Data for Cohorts 3, 6, 7, 8 and 9 Complete

Table 7-2. Stopping Boundary for Enrollment within Cohorts 3, 6-[redacted] Individually With Posterior Probability of 80% and Subject Incidence of Grade 4 or Higher Treatment-Related Adverse Event Limit of 20%

Number of subjects treated	Stop cohort enrollment if observing this many subjects with grade 4 or higher treatment-related adverse events
5	≥ 3
10	≥ 4
20	≥ 6
30	≥ 9
40	Cohort Enrollment Complete*

* Enrollment for cohorts 7, 8, and 9 completes after n=40 subjects enrolled. Enrollment for cohorts 3, 6, [redacted] completes after n=20 subjects enrolled. [redacted] Early stopping rules (eg, at 5 or 10 treated subjects) are the same regardless of planned maximum sample size in a cohort.

Table 7-3 Operating Characteristics With Batch Size of 10 Subjects (Combining All Monotherapy Data of Expansion Cohorts 3, 6, 7, 8 and 9)

True grade 4 or higher treatment-related adverse event rate	Probability of early stopping of dose expansion	Average sample size of overall combined monotherapy data
0.10	2.1 %	156.9
0.15	12.7 %	143.1
0.20	46.4 %	105.8
0.25	86.4 %	58.7
0.30	98.9 %	31.5

Table 7-4. Operating Characteristics with Batch Size of 10 Subjects (Cohort-Specific Enrollment With Maximum Sample Cohort Size of n=40 Subjects)

True grade 4 or higher treatment-related adverse event rate	Probability of early stopping of cohort enrollment	Average cohort sample size
0.10	2.0%	39.5
0.15	9.7%	37.6
0.20	25.8%	33.9
0.25	47.8%	28.8
0.30	69.2%	23.4

7.2 Primary Analysis

The primary analysis will be conducted when target enrollment is complete, and each subject had the opportunity to complete 6 months on study or withdraws from study.

7.3 Final Analysis

The final analysis will be conducted after all cohorts and dose-expansion subjects have ended the study. Primary and final analysis may be combined in case all subjects have ended study close to the time point of the primary analysis.

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.

[redacted]

[redacted]

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

Outliers will not be excluded unless scientifically justified. Pharmacokinetic (PK) serum concentration data will be evaluated for outliers by visual inspection, and decisions to re-assay individual samples will be made in accordance with standard PK evaluation practice.

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, PK, pharmacodynamics [redacted] data by dose cohort or time as appropriate. Unless otherwise stated, the data analysis will be conducted using subjects in the Safety analysis set. Subjects in the dose escalation part of the analysis will be summarized by each AMG 404 dose level.

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. When data are summarized by time, the values recorded against the scheduled time points listed in the protocol will be used. When assessing minimum/maximum increases or decreases over the study, all assessments, including unscheduled assessments will be used. Data listings will include all available data from all enrolled subjects unless specified otherwise.

All descriptive statistics and graphical summaries may be presented separately for Cohort 5, [redacted] versus all the other cohorts. Similarly, all the efficacy and the safety analyses may be conducted separately for Cohort 5, [redacted] versus all the other cohorts.

Efficacy endpoints will be summarized separately for Cohort 7, Cohort 8, Cohort 9, [redacted].

Confidence intervals (CI) for proportions will be estimated using an exact method proposed by Clopper-Person (Clopper & 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 method (Brookmeyer & Crowley, 1982). Kaplan-Meier methods will be used to estimate landmarks for time to event endpoints (eg, 1-year OS) with the Greenwood formula (Kalbfleisch & Prentice, 1980) used to estimate the standard error used in CI calculation.

9.2 Subject Accountability

A summary of subject disposition and investigational product disposition will be provided for all subjects enrolled in the study ie, the number and percent of subjects who were enrolled, received investigational product, discontinued from investigational product (including reasons for discontinuing), completed study, discontinued the study (including reasons for discontinuing) will be summarized by cohort.

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. 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.

A protocol deviation of listing of subjects impacted due to COVID-19 will also be provided.

9.4 Demographic and Baseline Characteristics

Demographic characteristics including age, age groups (18 to 65, 66 and above), sex, race, ethnicity and baseline characteristics, which may include height, weight, Eastern Cooperative Oncology Group (ECOG) performance status, baseline disease stage, number of prior anti-cancer therapies, cancer-related surgery and primary tumor type will be summarized using descriptive statistics.

If multiple races have been reported for a subject, the subject will be categorized as multiple.

9.5 Efficacy Analyses

In general, the analysis of efficacy endpoints will be based on the Safety Analysis Set unless otherwise specified.

9.5.1 Analyses of Primary Efficacy Endpoint(s)

No efficacy parameter is considered in primary endpoints.

9.5.2 Analyses of Secondary Efficacy Endpoint(s)

For all subjects treated at the RP2D and separately by cohort, the following analyses will be done. The proportion of subjects with confirmed objective tumor response (partial and complete response) and disease control rate (DCR) (partial or complete control) with corresponding exact 80% CI will be calculated and tabulated. Similarly, the proportion of subjects and 80% CI will be tabulated for 1-year duration of overall response, 1-year PFS, 1-year duration of stable disease and 1-year OS. Kaplan-Meier curve will be presented for duration of overall response, PFS, OS and duration of stable disease with estimates for rate and 80% CI at selected weeks based on the Brookmeyer and Crowley method (Brookmeyer & Crowley, 1982) if data allows.

9.5.3 Analyses of Exploratory Efficacy Endpoint(s)

No efficacy parameter is considered in exploratory endpoints.

9.6 Safety Analyses

9.6.1 Analyses of Primary Safety Endpoint(s)

Unless otherwise specified, statistical analyses on safety endpoints will be done using subjects from the safety analysis set, which includes subjects that are enrolled and received at least 1 dose of AMG 404.

The analysis of the subject incidence of dose limiting toxicities (DLT) will be based on the DLT evaluable set defined in Section 6.2. A summary of the subject incidence of dose limiting toxicities (DLT) with corresponding 95% CI will be provided.

9.6.2 Adverse Events

The most current version of the Medical Dictionary for Regulatory Activities (MedDRA) will be used to code all adverse events to a system organ class and a preferred term. The identification of adverse events of special interest is a continuous process. Events may be identified and documented as the safety profile of the drug is characterized. The severity of each adverse event will be graded using CTCAE version 5.0.

Tables will be provided for Subject incidence of all treatment-emergent, serious, treatment-related, those leading to withdrawal of investigational product, events of special interests and fatal adverse events by system organ class and preferred term in descending order of frequency. Where appropriate the tables will also be presented by worst grade.

Details of each treatment-emergent adverse event will be listed. Listings and/or narratives of any on-study deaths, serious treatment-emergent adverse, and grade 3 or higher treatment-related adverse, including adverse events leading to discontinuation of AMG 404, also will be provided should they occur.

9.6.3 Laboratory Test Results

Laboratory data will be summarized using standard descriptive statistics at each scheduled time point in the study. For continuous parameters, a summary of the changes from baseline to each post dose laboratory assessment will also be produced. Shifts in selected laboratory parameters between baseline and the worst on-study value will be summarized according to the NCI CTCAE toxicity grades. Safety Laboratory collection includes chemistry, hematology, pregnancy test (for women of childbearing potential) either urine or serum. The parameters described in Table 13-1 of the protocol will be collected, converted to Amgen standard units.

Selected individual analytes will be summarized by cohort using standard descriptive statistics at each of the scheduled time points through the study up to the safety follow-up visit. For continuous parameters, a summary of the changes from baseline to each

post dose laboratory assessment will be produced for each cohort. Tables of shifts from baseline to the worst-case on-study increased and decreased values (graded according to the NCI Common Toxicity Grading Criteria) will be provided for selected laboratory parameters with available NCI-CTCAE grades. Unscheduled assessments will be included in the shift tables.

Subject incidence of suspected Hy's law cases (Hy's law predicts potential for drug-related hepatotoxicity) will be summarized by cohort. A listing of alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, and total bilirubin values at each time point may be produced for the subjects suspected of Hy's law case.

Boxplots may be plotted by time point and cohort for selected analytes.

Urine analysis

Blood, protein and glucose will be graded in the following manner: 0='0 or Trace', 1='1+', 2='2+', 3='3+', 4='4+'. Microscopic parameters (WBC, RBC, epithelial cells, bacteria, casts, crystals) will be graded in the following manner: 0='0-4 none, rare, occasional', 1='5-50 moderate, few', 2='>50 many, heavy, too numerous to count'. Data recorded as "Present" will not be included in the summary.

The number and percent of subjects with a worst post-dose presence of blood in the urine of '0 or Trace', '1+','2+', '3+', or '4+' will be presented (also for protein and glucose in the urine).

Depending on the size and scope of changes in urinalysis data, summaries over time and/or changes from baseline over time may also be provided. Unscheduled assessments will be included in the analysis.

9.6.4 Vital Signs

Depending on the size and scope of changes, summaries of vital signs data (blood pressure (BP), respiratory rate, heart rate and temperature) over time and/or changes from baseline overtime may be provided. Shifts in vital sign values from baseline over time may also be tabulated if required.

9.6.5 Physical Measurements

The analysis of body weight may include summary statistics at selected time points for each treatment cohort.

9.6.6 Electrocardiogram

Summaries over time and/or changes from baseline over time will be provided for all ECG parameters. Subjects' maximum change from baseline in (QTcF, QTcB) will be categorized and the number and percentage of subjects in each cohort will be summarized. Subjects' maximum post baseline values will also be categorized and the number and percentage of subjects in each group will be summarized.

Subjects will be categorized into the following groups per their maximum change from baseline in QTcF and QTcB. Unscheduled assessments will be included in the determination of the maximum change.

- <=30 msec
- >30 – 60 msec
- >60 msec

The number and percentage of subjects in each group will be summarized.

Subjects will also be categorized into the following groups per their maximum post baseline QTcF. Unscheduled assessments will be included in the determination of the maximum post baseline value.

- <=450 msec
- >450 – 480 msec
- >480 – 500 msec
- >500 msec

The number of subjects in each group will be summarized for each cohort.

9.6.7 Antibody Formation

A listing of subjects who develop anti-AMG 404 antibodies (binding) may be provided. Blood sample(s) for antibody testing are to be collected according to the time points specified in the Schedule of Activities for the measurement of anti-AMG 404 antibodies. Samples testing positive for binding antibodies may be further characterized. Additional blood samples may be obtained to rule out anti-AMG 404 antibodies during the study. Subjects who test positive for anti-AMG 404 antibodies at the final scheduled antibody timepoint and have clinical sequelae that are considered potentially related to the anti-AMG 404 antibody response will be asked to return for additional follow-up testing. Sample collection and testing will occur approximately every 3 months from the safety follow up visit until: (1) anti-AMG 404 antibodies are no longer detectable; or (2) the subject has been followed for a period of at least 1-year (± 4 weeks) post administration

of AMG 404. All follow-up results, both positive and negative, will be communicated to the sites. More frequent testing (eg, every month) or testing for a longer period of time may be requested in the event of safety-related concerns.

The incidence and percentage of subjects who develop anti-AMG 404 antibodies, will be tabulated overall, and by cohort.

9.6.8 Exposure to Investigational Product

Descriptive statistics will be produced to describe the exposure to investigational product by treatment dose. Number of cycles started, Number of doses of investigational product, the cumulative dose by unit and average dose delivered per day and duration of exposure will be summarized.

Summaries of the number and percentage of subjects with dose change/withheld and reason for change/withheld will be provided. Details for each AMG 404 administration will be listed for every subject. In addition, a 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 Other Protocol-specified Treatment

Not applicable.

9.6.10 Exposure to Concomitant Medication

All medication will be coded using the WHO drug dictionary.

The number and proportion of subjects receiving concomitant medications will be summarized by preferred term or category.

9.7 Other Analyses

9.7.1 Analyses of Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Endpoints

The PK parameters of AMG 404 including, but not limited to, C_{max}, t_{max}, and AUC for serum AMG 404 will be estimated using non-compartmental methods for subjects where intense PK sampling was collected (ie, subjects in Cohorts 1-6 [redacted], at least 10 subjects in Cohorts 7-9). [redacted] The parameter estimates will be summarized by dose level using means, geometric means, standard deviations, coefficients of variation, medians, minimums, and maximums. Individual concentration-time profiles will be summarized by dose level. [redacted]. Summary statistics will be computed for each sampling time and parameter as appropriate. The relationship between AMG 404 exposure and efficacy/safety may be conducted.

Nominal sampling times will be used for individual concentration-time plots and tables. Actual dose administered and actual sampling times will be used for the calculation of PK parameters for each subject. The reasons for excluding any sample from the analyses will be provided. Above analyses will be conducted by Amgen Clinical Pharmacology Modeling and Simulation (CPMS).

The relationship between AMG 404 exposure and QT/QTc interval changes may be inspected graphically and model-based PK/pharmacodynamic analyses may be conducted to examine the relationship further.

[redacted]

11. Literature Citations / References

AMG 404 Investigator's Brochure. Thousand Oaks, CA. Amgen Inc.

Brookmeyer, R. and Crowley, J. (1982) A Confidence Interval for the Median Survival Time, Biometrics, 38, 29 - 41.

Clopper CJ, Person ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 1934; 26:404-413.

Ji Y, Liu P, Li Yisheng, Bekele BN. A modified toxicity probability interval method for dose-finding trials. Clinical Trials 2010; 7:653-663.

Kalbfleisch, J.D. and Prentice, R.L. (1980) The statistical analysis of failure time data. John Wiley and Sons, New York.

M. Hidalgo et al. (eds.), Principles of Anticancer Drug Development, Cancer Drug Discovery and Development, DOI 10.1007/978-1-4419-7358-0_2.

Thall, PF, Simon, RM, and Estey, RM. Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. Statistics in Medicine. 1995; 14(4):357-79.

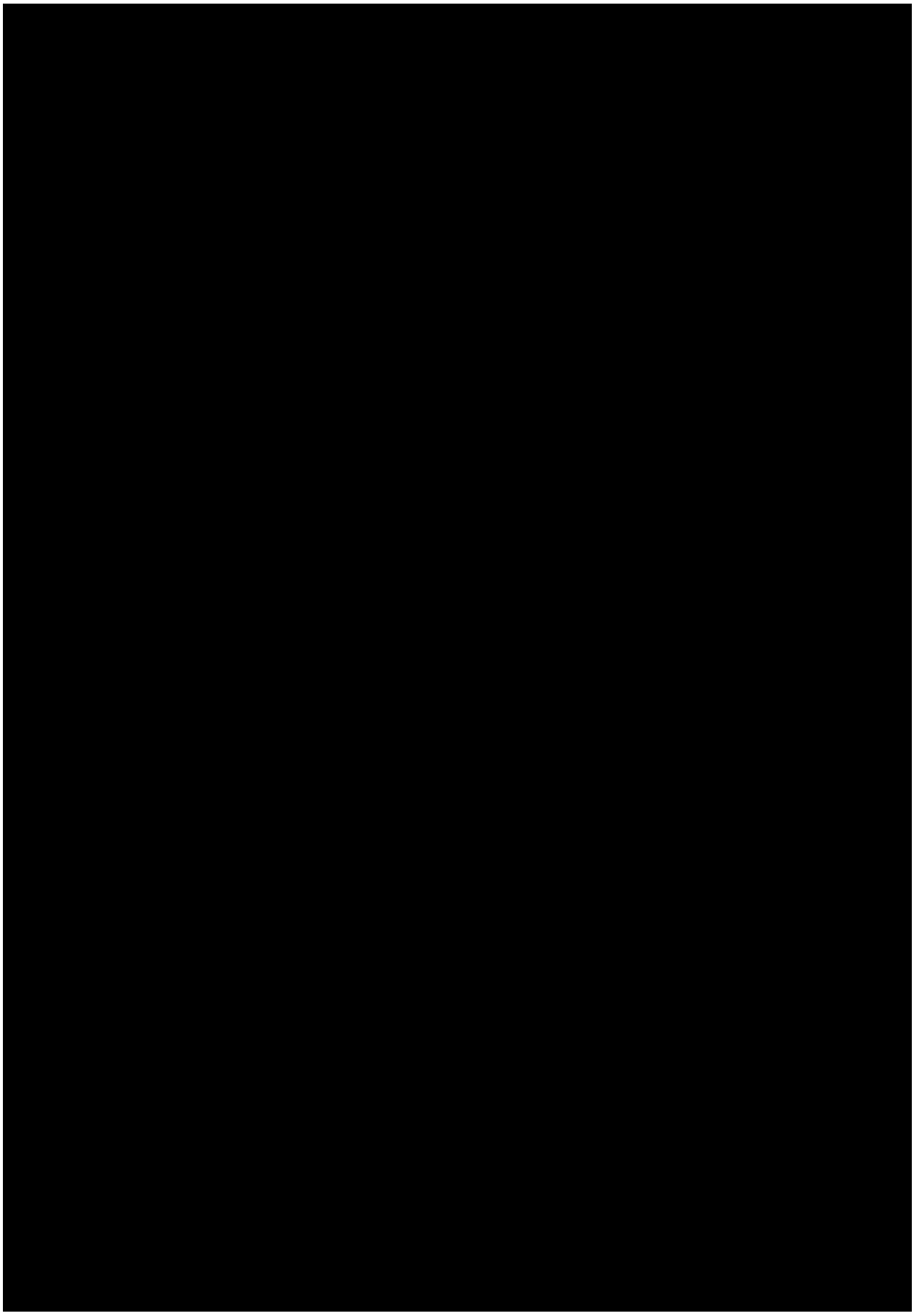
12. Prioritization of Analyses

There is no prioritization of analyses.

13. Data Not Covered by This Plan

Not applicable.

14. Appendices



Appendix B. Code Fragments

Example:

95% Confidence Interval by Clopper Pearson Method:

```
proc freq data=<data> ;
tables <variable>*<variable> / binomial(exact) ;
run;
```

[redacted]

- **Code fragment for K-M method:**

```
proc lifetest data=data1 method=km conftype=loflog alphaqt=0.05
by cohortn;
  time avail*cnsr (1);
run;
```

Appendix C. BOR per RECIST (v1.1) Criteria

The protocol requires that a PD must be confirmed. The analysis of modified RECIST 1.1 BOR endpoints will follow standard RECIST 1.1 convention which does not require confirmation of PD.

Best Overall Response (BOR) determination per RECIST 1.1 is given in below in Table 14-2, the BOR considered is a Confirmed BOR.

Table 14-2. BOR per RECIST 1.1

Criterion	Time point T1 Response	T1 ≥ 49 days after C1D1?	Time point T2 Response	T2 ≥ 49 days after C1D1?	T2 ≥ 28 days after T1?	BOR
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			PR, SD	-	-	Query data*
C9			PD	-	-	PD
C10			NE, No further evaluations			NE
C11	PR	Yes	CR, PR	-	Yes	PR
C12			CR, PR	-	No	SD
C13			SD	-	-	SD
C14			PD	-	-	SD
C15			NE, No further evaluations			SD
C16		No	CR, PR	-	Yes	PR
C17			CR, PR	Yes	No	SD
C18			SD	Yes	-	SD
C19			PD	-	-	PD
C20			NE, No further evaluations			NE
C21	SD	Yes	CR, PR, SD, PD, NE, no more evaluation			SD
C22		No	CR, PR, SD	Yes	-	SD
C23			CR, PR, SD	No	-	NE
C24			PD	-	-	PD
C25			NE, No further Evaluations			NE
C26	**PD**	**-**	**-**	**-**	**-**	**PD**
C27	NE	-	NE, No further evaluations			NE
C28		-	CR, PR, SD	Yes	-	SD
C29		-	CR, PR, SD	No	-	NE
C30		-	PD	-	-	PD

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, and 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 and the best response is PR.

A CR must be confirmed by CR, a PR can be confirmed by a PR or CR.

An unlimited number of intermittent assessments of NE can occur between the initial response and the confirmation. For example, BL, CR, NE, NE, NE, CR – the CR at post baseline 1 is confirmed at post baseline 5.