STATISTICAL ANALYSIS PLAN

A Phase 1 Study Evaluating the Safety, Tolerability, Pharmacokinetics, Pharmacodynamics, and Efficacy of AMG 211 Administered as Continuous Intravenous Infusion in Subjects with Relapsed/Refractory Gastrointestinal Adenocarcinoma

Protocol Number 20130354

Version: Version 1.0

Date: 19 May 2014

Authors: PPD

NCT Number: 02291614
This NCT number has been applied to the document
for purposes of posting on clinicaltrials.gov

TABLE OF CONTENTS

TABLE OF ABBREVIATIONS

Abbreviation/Acronym	Definition
AE	Adverse event
BQL	Lower limit of quantifications
CI	Confidence interval
cIV	Continuous intravenous
CL	Systemic clearance
C_{ss}	Steady-state drug concentration in plasma during constant-rate infusion
CSR	Clinical study report
CT	Computed tomography
CTCAE	Common terminology criteria for adverse events
DLT	Dose limiting toxicity
DMP	Data management plan
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EOS	End of study
irCR	Immune-related Complete Response
irPD	Immune-related Progressive Disease
irRC	Immune-related Response Criteria
irSD	Immune-related Stable Disease
PK	Pharmacokinetic(s)
PKDM	Pharmacokinetics and drug metabolism
MRI	Magnetic resonance imaging
MTD	Maximum tolerated dose
PD	Progressive disease
PR	Partial response
RECIST	Response evaluation criteria in solid tumors
SAE	Serious adverse event
SAP	Statistical analysis plan
SLD	Sum of longest diameters

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 for AMG 211 Study 20130354 dated 24 March 2014. The scope of this plan includes the interim analysis and the final analysis that are planned and will be executed by the Biostatistics department or designee unless otherwise specified eg, standard PK tables may be provided by PKDM.

2. OBJECTIVES

2.1 Primary Objectives

- Evaluate the safety and tolerability of AMG 211 in adult subjects with relapsed/refractory GI adenocarcinomas
- Determine the maximum tolerated dose (MTD) and/or biologically active dose (eg, recommended phase 2 dose)

2.2 Secondary Objectives

- Describe the pharmacokinetics (PK) of AMG 211
- Determine the formation of anti-AMG 211 antibodies
- Evaluate time to progression and to evaluate the proportion of subjects progression-free at 6 months
- Evaluate the anti-tumor activity of AMG 211 by evaluating
  - The number and proportion of subjects with objective response according to the modified Immune-related Response Criteria (irRC)
  - The duration of response and time to response

2.3 Exploratory Objectives

- Evaluate the protein, nucleic acid, and cellular biomarkers in blood (eg, cytokines, lymphocyte subsets) before and following treatment with AMG 211
- Evaluate the effects of genetic variations in cancer genes and multiple other aspects of the tumor and tumor microenvironment before and following treatment with AMG 211 which may include CEA expression, markers of necrosis or apoptosis, and potential changes in the nature and number of tumor infiltrating lymphocytes

3. STUDY OVERVIEW

3.1 Study Design

This is a multi-center, open-label, sequential dose-escalation study evaluating AMG 211 as a cIV infusion in adult subjects who have relapsed/refractory GI adenocarcinoma. Table 1 presents an overview of the study design. Additional details follow below.

Table 1. Study Design Overview

Part	Cohort	Dose Level	Number of Subjects
Dose Escalation	1	200 µg/d for 7 days in cycle 1 200 µg/d for 14 days in cycle 2 and all subsequent cycles	3-6
	2	200 µg/d for 14 days in all cycles	3-6
	3	400 µg/d for 14 days in all cycles	3-6
	4	800 µg/d for 14 days in all cycles	3-6
	5	1600 µg/d for 14 days in all cycles	3-6
	6	3200 µg/d for 14 days in all cycles	3-6
	7	6400 µg/d for 14 days in all cycles	3-6
28-day Schedule	X-1	Dose Level -1 of 14-day Schedule × 28 days	3-6
	X-2	MTD × 28 days	3-6
Dose Expansion	A	MTD or Highest Tested Dose	10

Abbreviations: d = day; DLT = dose-limiting toxicity; DRLM = dose level review meeting; EC50 = 50% of the maximal effective concentration level; MTD = maximum tolerated dose; PD = pharmacodynamic; PK = pharmacokinetic.
[a] Alternative dosing regimens (eg, 28-day infusions) may be explored after a DLRM decision based on 1) occurrence of DLT(s) or other emerging safety data; 2) PK/PD data demonstrating drug exposure that is in the range or above the mean EC50 concentration of 29 ng/mL; 3) absence of objective responses in the 14-day infusion schedule.

In the dose-escalation cohorts, subjects will be assigned sequentially into cohorts as they open. Three subjects will initially be enrolled into each dose-escalation cohort with at least a 48-hour interval between treatment of subject 1 and subject 2 of each cohort. In the event that a DLT is observed, additional subjects may be enrolled for a total of up to 6 evaluable subjects.

The dose-expansion cohort will be initiated after completion of the dose-escalation cohorts. Up to 10 subjects will be enrolled and treated at the dose level determined in the dose-escalation cohorts.

Subjects will be treated with AMG 211 until confirmed disease progression per modified irRC (Appendix D in the protocol), or intolerance of study treatment, or clinically significant deterioration of health status requiring discontinuation, whichever occurs first. Due to the mechanism of action of immune-enhancing therapies, subjects may experience an apparent enlargement of existing lesions or the appearance of new lesions prior to maximal clinical benefit of AMG 211.

For the schedule of assessment, please refer to section 7.1 in the protocol.

3.1.1 Dose Escalation Phase

The dose-escalation part of the study is aimed at determining the MTD, safety, tolerability, PK, and PD of AMG 211. Planned dose levels for the dose-escalation cohorts are as follows: 200, 400, 800, 1600, 3200, and 6400 µg/d. Alternative dose levels may be explored based on emerging data.

Subjects in cohort 1 of the dose-escalation cohorts will receive AMG 211 for 7 days followed by 21 days off treatment (cycle 1). After completing cycle 1, if no DLTs are observed, all subjects in cohort 1 will be allowed to continue treatment with the same dose of AMG 211, but the infusion duration will be 14 days followed by 14 days off treatment.

In the subsequent dose-escalation cohorts, AMG 211 will be administered at escalating doses in a 2 weeks on and 2 weeks off schedule. Based on the results, the Dose Level Review Team (DLRT) may choose to extend the infusion duration from 14 to 28 days (a 4 weeks on and 2 weeks off schedule) if the 14-day administration schedule has been well tolerated and the safety profile allows for a longer infusion duration, and no anti-tumor response has been observed in the 14-day administration schedule.

Moreover, the DLRT may decide to enroll in parallel into a cohort with a 28-day treatment duration or to prioritize one infusion schedule.

Dose escalation decisions will consider the incidence of DLTs among DLT-evaluable subjects that occur during the first 28 days of treatment. For definition of DLT-evaluable subjects, please refer to section 3.4 in the protocol.

A safety review of dose escalation is planned after all subjects in dose escalation have had the opportunity to complete 6 months of the treatment.

3.1.1.1 Dose Escalation and Dose Stopping Rules

Escalation to a higher dose cohort will only proceed when the previous dose(s) has (have) been found to be reasonably tolerated based on available study data through the DLT period for all DLT-evaluable subjects enrolled within a cohort and upon unanimous decision of the DLRM members. Dosing changes will be made on a treatment cohort basis and a case-by-case basis. Dose schedule changes will be made on a case-by-case basis. Dose decisions are planned after every 3 to 6 DLT-evaluable subjects are enrolled throughout the dose-escalation cohorts. Decisions made during the meeting will be documented in a dose level review memo.

The DLRT considers the recommendation for subsequent doses based on the following rules:

- Based on the modified TPI Bayesian model, the next dose is the one with the highest probability of the target TPI (0.20, 0.35), but with a less than 0.25 probability of an excessive or unacceptable TPI.
- If ≥ 1 subject has a DLT at a dose level, then dose escalation cannot occur unless there are 6 or more DLT-evaluable subjects at that dose level. However, if a transient cytokine-release syndrome is reported as the DLT, but resolves within 48 hours and can be managed with prophylactic corticosteroids, the DLRT may decide to escalate the dose with the institution of prophylactic corticosteroid administration.
- The maximum allowed dose increase will be 1 dose level above the maximum of previous evaluated doses.
- Dose escalation can be achieved by evaluating a higher dose level with the same infusion duration (eg, 800 µg/day for 14 days to 1600 µg/day for 14 days), or by evaluating the same dose level with a prolonged infusion duration (eg, from 800 µg/day for 14 days to 800 µg/day for 28 days). The DLRT may also decide to explore in parallel both options mentioned above after the dose-escalation decision is made, or to prioritize one schedule. In this case, 2 TPI models may be run in parallel to obtain dose recommendations for these 2 schedules separately.
- Intermediate dose levels not pre-specified may be considered based on the optimal dose recommended by the TPI model.

The DLRT is responsible for making the decision to end dose escalation. The DLRT may consider the modeling of dose escalation complete if 1 of the following rules is met:

- The highest planned dose level is evaluated and no DLTs occur at any dose level. In this case the maximum administered dose will be used in a dose-expansion cohort.
- The TPI Bayesian model recommends the same dose 3 times (not necessarily sequentially).

- A total of approximately 24 DLT-evaluable subjects have been enrolled.

3.1.2 Dose Expansion Phase

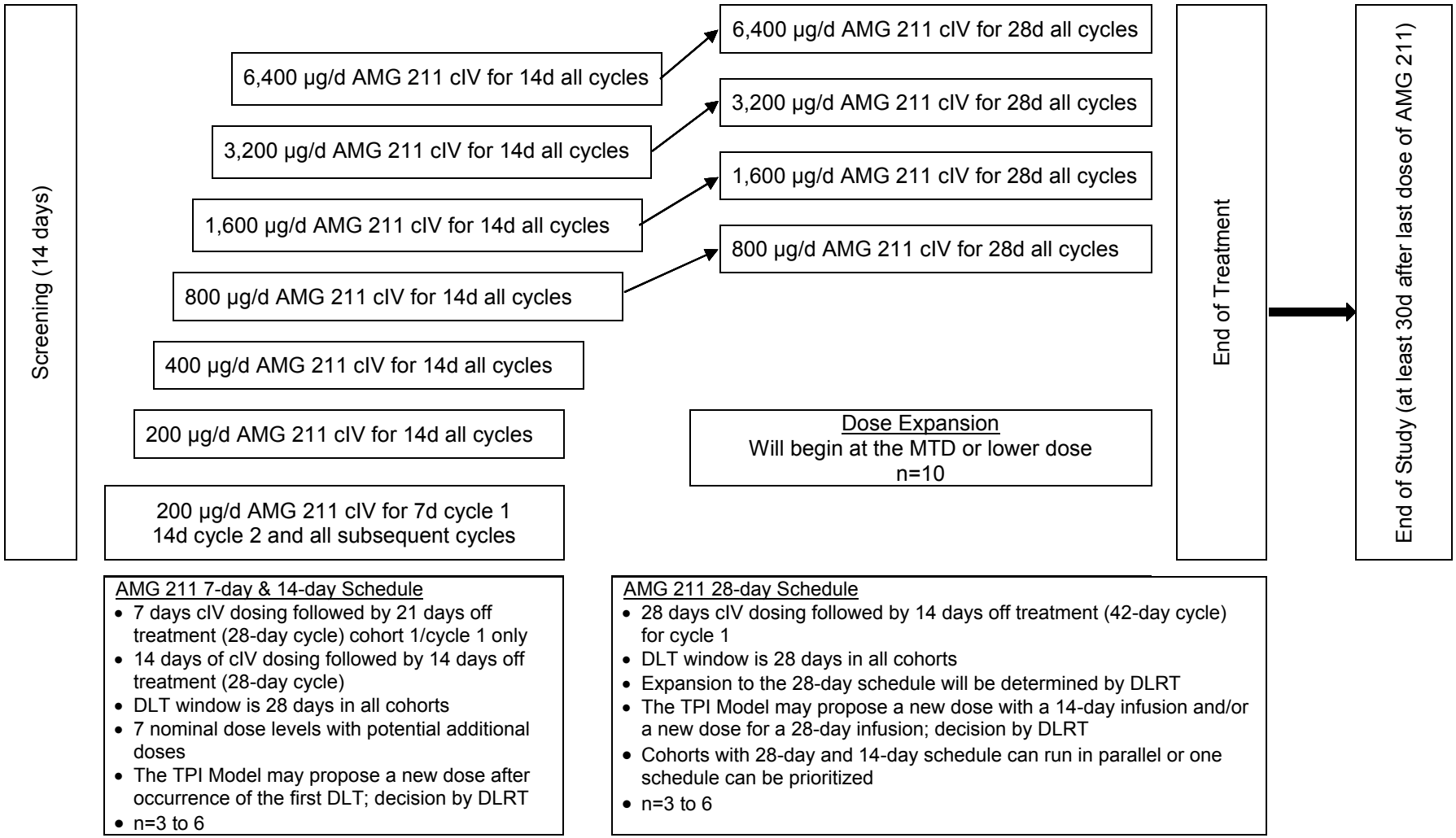
At the completion of the dose-escalation cohorts, additional subjects (up to 10) will be enrolled into a dose-expansion cohort to gain further clinical experience with AMG 211.

The dose to be evaluated will be at or below the MTD from the dose-escalation cohorts.

A final estimate of the recommended phase 2 dose will be based on the toxicity probability interval (TPI) Bayesian model using all DLT-evaluable subjects.

See Section 3 of the protocol for further details.

3.1.3 Study Schema



cIV = continuous intravenous; d = day; DLRT = dose level review team; DLT = dose-limiting toxicity; MTD = maximum tolerated dose; TPI = toxicity probability interval

3.2 Sample Size

It is anticipated that up to 34 subjects will be enrolled in this study. Approximately 24 subjects will be enrolled in the dose-escalation cohorts and up to 10 subjects will be enrolled in the dose-expansion cohort.

The sample size in dose-escalation cohort is determined empirically and is consistent with this type of study using a modified TPI Bayesian model design (Bailey et al, 2009; Neuenschwander et al, 2008).

Refer to Appendix E in the protocol for details of the modified TPI Bayesian model, the average sample size, and other design operating characteristics.

In the dose-escalation cohorts, with 3 DLT-evaluable subjects per dose level, there is a 49% to 73% probability of observing at least 1 DLT if the true DLT rate is 20% to 35%.

In the dose-expansion cohort, an exact 80% binomial confidence interval (CI) will be provided for the ORR. With 10 subjects and 20% ORR, the expected 80% CI and its half-width would be 5.4% to 45.0% and 19.8%, respectively. In this case, the probability of observing at least 1 ORR among 10 subjects is 89.2% and the probability of observing at least 2 ORRs is 62.4%.

4. STUDY ENDPOINTS

4.1 Primary Endpoints

- Safety: subject incidence of adverse events, DLTs, and clinically significant changes in vital signs, ECGs, physical examination findings, and clinical laboratory tests

4.2 Secondary Endpoints

- PK of AMG 211 after cIV infusion across 2 cycles
- Incidence of anti-AMG 211 antibody formation
- Time to progression
- 6-month progression-free rate
- Efficacy parameters: overall response rate (ORR; per modified irRC), duration of response, time to response

4.3 Exploratory Endpoints

- Changes in methylation and mutations in circulating free DNA present in plasma
- Lymphocyte counts, T-cell activation and immune checkpoint regulator status

- Changes in serum cytokine levels
- Biomarkers and mutations in tumor cells
- Changes in CTCs
- sCEA serum levels

5. HYPOTHESES AND/OR ESTIMATION

The hypothesis for this phase 1 study is that objective responses according to modified irRC will be observed at dose levels that achieve acceptable safety and tolerability.

6. DEFINITIONS

Adverse Event (AE)

An adverse event is defined as any untoward medical occurrence in a clinical trial subject. The event does not necessarily have a causal relationship with study treatment. The definition of adverse events includes worsening of a preexisting medical condition. Worsening indicates that the preexisting medical condition (eg, diabetes, migraine headaches, gout) has increased in severity, frequency, and/or duration, and/or has an association with a significantly worse outcome. A preexisting condition that has not worsened during the study, or involves an intervention such as elective cosmetic surgery or a medical procedure while on study, is not considered an adverse event.

Age at Enrollment

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

AUC_{inf}

Area under the concentration-time curve from time zero to infinity.

Baseline

Unless otherwise specified, the baseline value for parameters/assessments scheduled to be performed on the same day as the first administration of AMG 211, is the last value measured before the first administration of AMG 211 on that day. For parameters/assessments not scheduled to be performed (or scheduled but not performed) on the same day as the first administration of AMG 211, the baseline value is the value from the screening period measured

closest to the day of first administration of AMG 211. In the event that multiple assessments are done on the same day as the first administration of AMG 211 and there is no time associated with the assessments, the value associated with the last clinically planned event before the first administration of AMG 211 will be used as the baseline value.

Best Overall Response

Best overall response for a subject is the best observed post baseline disease response per modified Immune-Related Response Criteria (irRC). Any irCR or irPR must be confirmed by consecutive repeat assessments performed at least 4 weeks after the criteria for response are first met. The confirmatory scan can also be performed at the next scheduled imaging visit. An irPD must be confirmed by consecutive repeat assessments performed at least 4 weeks apart in the absence of rapid clinical deterioration. The confirmation scan may also be performed at the next imaging assessment as defined by the schedule of assessment after the first observation if the subject has stable or improved clinical status. Subjects with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at the time should have the reason for treatment discontinuation classified as non-confirmed disease progression. The investigator reported response information collected on the CRF and/or Independent review committee assessments will be used to determine the response.

Change From Baseline

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

C_{ss}

Steady-state drug concentration in serum calculated as average concentration between achievement of plateau and EOI during constant-rate infusion.

Complete Response

Complete response is defined using the modified irRC as outlined in Appendix D of the protocol.

Dose Limiting Toxicity (DLT)

Please refer to section 6.2.1.3 in the protocol for definition of DLT.

Duration of Response

Duration of response is defined as the number of days between the date of the first tumor assessment indicating an objective response through to the subsequent date of progression as classified by modified irRC or death due to any cause, or where applicable date of censoring [date of first progressive disease assessment or death or date of censoring – date of the first objective response result +1]. Subjects who respond and have not progressed while on study or died for reasons other than disease progression will be censored at the date of assessment of the last evaluable radiological assessment. Subjects who do not achieve an objective response will be excluded from the analysis of duration of response.

End of Investigational Product Administration

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

End-of-Study

The end of study date for an individual subject is recorded on the end of study CRF page. End of study (EOS) for an individual subject is defined as the date of the final study visit (EOS visit) when safety assessments and procedures are performed. The EOS visit should occur approximately 4 weeks (+ 1 week) after the last dose of AMG 211 or at the initiation of other therapy. Subjects who complete the EOS visit will be considered to have completed the study. Primary Completion: the time when the last subject is assessed or receives an intervention for the purposes of final collection of data for the primary analysis; the primary analysis for this study will occur after all dose-escalation cohort subjects and dose-expansion cohort subjects have completed 6 months of treatment or have discontinued treatment with AMG 211.

End of Trial: the time when the last subject is assessed or receives an intervention for evaluation in the study; the final analysis will occur at this time.

Enrollment

A subject is considered enrolled on day 1 (cycle 1 day 1) when the cIV infusion with investigational product is first administered. See protocol section 5.0 for further details.

Fridericia-corrected QT Interval (QTcF)

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

$QTcF=QT/(RR/1000)^{1/3}$

Investigational Product

The term 'investigational product' is used in reference to AMG 211.

Last Investigational Product Dose Date

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

Maximum Tolerated Dose (MTD)

Given the toxicity probability interval (TPI) specified, after each cohort, the model's recommended MTD dose for evaluation is the dose with the highest probability of the target TPI, but with a less than 0.25 probability of an excessive or unacceptable TPI. The actual dose selected at each dose decision may be at or below the model's recommended dose as determined at a DLRM by the dose level review team (DLRT) after considering all information.

A final model estimate of the MTD will be obtained utilizing all the dose-escalation cohort and dose-expansion cohort DLT-evaluable subjects.

Percent Change From Baseline

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

Progression-Free Survival Time

Progression-free survival time is calculated as the number of days from the first administration of AMG 211 to the first assessment of disease progression as per modified irRC or clinical progression, whichever occurs first or death due to any cause, or if applicable date of censoring (date of progressive disease or death or censoring – date of study day 1 +1). Progression-free Survival Time will incorporate the independent radiologic review and/or the Investigator-reported assessment. The progression-free rate at 6 months is defined as the proportion of subjects with no disease or clinical progression at 6 months from the first administration of AMG 211.

The following censoring strategies for missing assessment dates will be used for the progression free survival analysis:

If a subject's disease has not progressed and the subject is alive, progression-free survival time will be censored at the last date they are known to be progression-free (ie, the last evaluable radiological assessment date).

If a subject has no tumor evaluation in the study, progression-free survival time will be censored at the date of the first administration of AMG 211.

Subjects who withdraw consent to participate in the study prior to disease progression will be censored at their last evaluable radiological assessment date.

Screened

A subject is considered in screening once a consent form has been signed.

Study Day

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

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

Study Day 1

Study day 1 is defined as the first day of administration of the investigational product after enrollment. The day prior to Study Day 1 is considered Day -1.

Treatment-Emergent Adverse Event (TEAE)

A treatment-emergent adverse event is any adverse event starting on or after the first dose of investigational product, as determined by the flag indicating if the adverse event started prior to the first dose on the Adverse Events Summary CRF, and up to and including 30 days after the end of investigational product.

Treatment-Related AE

A treatment-related AE is any treatment-emergent AE that per investigator review has a reasonable possibility of being caused by the investigational product.

Time to Progression (TTP)

Time to progression is calculated as the number of days from the first administration of AMG 211 to the first objective assessment of disease progression as per modified irRC or if applicable date of censoring (date of progressive disease or censoring – date of study day 1 +1). Time to progression will incorporate the independent radiologic review and/or the Investigator-reported assessment.

The following censoring strategies for missing assessment dates will be used for the time to progression analysis:

If a subject's disease has not progressed and the subject is alive, time to progression will be censored at the last date they are known to be progression-free (ie, the last evaluable radiological assessment date).

If a subject has died, time to progression will be censored at the date of death.

If a subject has no tumor evaluation in the study, progression-free survival time will be censored at the date of the first administration of AMG 211.

Subjects who withdraw consent to participate in the study prior to disease progression will be censored at their last evaluable radiological assessment date.

Time to Response

Time to response is calculated as the number of days from the first administration of AMG 211 to the first objective assessment of response as per modified irRC or if applicable date of censoring (date of first response or censoring – date of study day 1 +1).

The following censoring strategies for missing assessment dates will be used for the time to response analysis:

If a subject did not respond, time to response will be censored at the latest evaluable radiological assessment date.

If a subject has no tumor evaluation in the study, time of response will be censored at the date of the first administration of AMG 211.

Tumor Burden

Tumor burden is derived as the sum of the products of the 2 longest perpendicular diameters (SPD) of all index lesions (5 lesions per organ and up to 10 total lesions) and new measurable lesions (≥5 × 5 mm; up to 5 new lesions per organ and 10 total lesions) if any, per irRC guidelines.

- Tumor burden = $SPD_{index\ lesions} + SPD_{new,\ measurable\ lesions}$

7. ANALYSIS SUBSETS

The analysis of all endpoints, unless noted otherwise, will be conducted on the Safety Analysis Set defined in Section 7.1.

7.1 Safety Analysis Set

The Safety Analysis Set will consist of all subjects that are enrolled and receive at least 1 dose of AMG 211. For the dose escalation part of the study subjects analyzed according to the initial AMG 211 dose received. For the dose expansion, subjects will be analyzed according to the dose selected for this part of the study.

7.2 Pharmacokinetic Concentration Analysis Set

The pharmacokinetic (PK) concentration analysis set will contain all subjects who received investigational product and have at least one PK sample collected.

7.3 Dose Escalation Analysis Set

The analysis of DLT will be restricted to DLT-evaluable subjects. The DLT window is 28 days for all cohorts and is independent of the length of the infusion. A subject is not DLT-evaluable if he/she drops out before completion of cycle 1 (defined as completed 90% of the planned dose in cycle 1) for reasons other than an adverse event related to the investigational product. This analysis Set will consist of all available data from all DLT-evaluable subjects, with subjects analyzed according to the treatment received initially as described for the Safety Analysis Set. Ineligible subjects may be replaced.

8. INTERIM ANALYSIS AND EARLY STOPPING GUIDELINES

Dose decision analyses are planned after every 3 to 6 DLT-evaluable subjects are enrolled throughout the dose-escalation cohorts. The DLRT considers the recommendation for subsequent doses based on the following rules:

- Based on the modified TPI Bayesian model (Bailey et al, 2009; Neienschwander et al, 2008), the next dose is the one with the highest probability of the target TPI (0.20, 0.35), but with a less than 0.25 probability of an excessive or unacceptable TPI.
- If ≥ 1 subject has a DLT at a dose level, then dose escalation cannot occur unless there are 6 or more DLT-evaluable subjects at that dose level. However, if a transient cytokine-release syndrome is reported as the DLT, but resolves within 48 hours and can be managed with prophylactic corticosteroids, the DLRT may decide to escalate the dose with the institution of prophylactic corticosteroid administration.
- The maximum allowed dose increase will be 1 dose level above the maximum of previous evaluated doses.
- Dose escalation can be achieved by evaluating a higher dose level with the same infusion duration (eg, 800 µg/day for 14 days to 1600 µg/day for 14 days), or by evaluating the same dose level with a prolonged infusion duration (eg, from 800 µg/day for 14 days to 800 µg/day for 28 days). The DLRT may also decide to explore in parallel both options mentioned above after the dose-escalation decision is made, or to prioritize one schedule. In this case, 2 TPI models may be run in parallel to obtain dose recommendations for these 2 schedules separately.
- Intermediate dose levels not pre-specified may be considered based on the optimal dose recommended by the TPI model.

The DLRT is responsible for making the decision to end dose escalation. Generally, the DLRT may consider the modeling of the dose escalation complete if 1 of the following rules is met:

- The highest planned dose level is evaluated and no DLTs occur at any dose level. In this case the maximum administered dose will be used in a dose-expansion cohort.
- The TPI Bayesian model recommends the same dose 3 times (not necessarily sequentially).
- A total of approximately 24 DLT-evaluable subjects have been enrolled.

The Amgen Medical Monitor and Amgen Global Safety Officer can make dose decisions without convening the DRLT in the following limited circumstances:

- If a cohort has at least 3 DLT-evaluable subjects and 1 subject experiences a DLT, then the Amgen Medical Monitor and Amgen Global Safety Officer may decide to expand the number of subjects treated in this cohort.
- The Amgen Medical Monitor and Amgen Global Safety Officer may decide to open the next cohort using a de-escalated dose (de-escalated from the dose level of the most recent cohort), if this de-escalated dose had been previously evaluated by the DLRT.

A safety review of dose escalation is planned after all subjects in dose escalation have had the opportunity to complete 6 months of the treatment.

9. DATA SCREENING AND ACCEPTANCE

9.1 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.

9.2 Data Handling and Electronic Transfer of Data

Amgen's Clinical Data Management department will provide all data to be used in the planned analyses. The database will be subject to edit checks outlined in the Clinical Data Management Plan (DMP). See details of this section in the DMP.

9.3 Handling of Missing and Incomplete Data

The following imputation for missing or incomplete data will be performed if required:

Incomplete adverse event and concomitant medication dates missing data will be imputed as described in Appendix A.

Non-pharmacokinetic measurements (eg. biomarker data) that are below the lower limit of quantification will be considered equal to half of the lower limit of quantification for all analyses unless specified otherwise.

Please refer to section 10.8 for details of handling missing or incomplete PK concentration data.

9.4 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.

9.5 Outliers

Outlier data will not be excluded unless scientifically justified.

PK plasma concentration data will be evaluated for outliers by visual inspection and decisions to re-assay individual samples will be made in accordance with standard PKDM practices**.**

9.6 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.

9.7 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 and S-plus.

10. STATISTICAL METHODS OF ANALYSIS

10.1 General Principles

Descriptive statistics will be provided for selected demographic, safety, PK, PD, and biomarker data by dose schedule and time as appropriate. Descriptive statistics on continuous data will include means, medians, SDs, 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. In general, data listings will be sorted by dose, subject, and time.

The primary analysis is planned after all dose-escalation cohorts and dose-expansion subjects have had the opportunity to complete 6 months of treatment.

A final analysis is planned after all dose-escalation cohorts and dose-expansion subjects have ended the study.

10.2 Subject Accountability

The number and percent of subjects who were screened, enrolled, received investigational product, completed investigational product, discontinued from investigational product (including reasons for discontinuing),completed study, discontinued the study (including reasons for discontinuing) will be summarized by dose schedule and time.

Key study dates for the first subject enrolled, last subject enrolled, last subject's end of study will be presented.

A subject listing and summary noting inclusion in each analysis subset will be provided for all subjects enrolled. A subject listing noting duration of AMG 211 administration, reason for discontinuation of treatment, and reason for discontinuing the study will be provided. A list of subjects screened but not enrolled (screen failures) will be provided.

10.3 Important Protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first

patient visit and updated during the IPD reviews throughout the study prior to database lock. If a snapshot is being taken during the study rather than a database lock at the end of the study, categories should be updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, sub-category 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.

10.4 Demographic and Baseline Characteristics

Demographic (ie, age, age groups [< 65, >= 65 and >= 75], sex, race, ethnicity, and baseline characteristics (height, weight, and Karnofsky performance status) will be summarized by dosing schedule and overall using descriptive statistics. If multiple races have been reported for a subject, the subject will be categorized as multiple races as well as by the combination of races.

A listing of the demographic and baseline characteristics will be provided. In addition listings of medical history (including antecedent hematologic or oncologic disease, other diseases/symptoms such as fatigue, bleeding, and infection whether resolved and ongoing), surgical history, prior anti-cancer usage and prior radiotherapy usage will be provided.

10.5 Safety Analyses

10.5.1 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 16.0 or later will be used to code all adverse events to a system organ class and a preferred term. The subject incidence of adverse events will be summarized for all treatment-emergent, serious treatment emergent, treatment-related, serious treatment-related, those leading to withdrawal of investigational product or other protocol-required therapies, fatal adverse events, and significant treatment emergent adverse events. The identification of significant adverse events 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 4.03 criteria.

Subject incidence of all treatment-emergent, serious, treatment-related, serious treatment-related, those leading to withdrawal of investigational product or other protocol-required therapies, significant treatment-emergent adverse events, and fatal adverse events 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 above adverse event tables will not be created if two or fewer subjects experience the adverse event.

Details of each adverse event will be listed. Listings and/or narratives of any on-study deaths, serious and significant treatment-emergent adverse events, including early withdrawals due to adverse events, also will be provided should they occur.

10.5.1.1 Dose Limiting Toxicities

The analysis of the probability of dose limiting toxicities (DLT) will be based on the Dose Escalation Analysis Set defined in Section 7.3. A listing and summary of the subject incidence of dose limiting toxicities (DLT) will be provided should they occur.

The planned primary nominal doses are 200, 400, 800, 1600, 3200 and 6400 µg/d. Potential intermediate doses with a 2- fold increment and a 50% reduction from the starting dose will be considered after DLT is observed. The dose-escalation cohorts will use a TPI Bayesian design to guide dose escalation. The MTD target TPI for DLT is (0.20, 0.35), and TPIs of (0.35, 0.60) and (0.60, 1.00) are defined as excessive and unacceptable, respectively. The design seeks to identify a dose most likely to have a DLT rate in the target TPI, but with overdose control that limits the possibility that the dose has an excessive or unacceptable DLT rate (Babb et al, 1998). The probability of a DLT at dose level d_i is assumed to follow a Bernoulli distribution with probability p_i where the logit of p_i increases linearly with the log of the standardized dose in the following 2-parameter logistic model:

$$\log [p_i / (1-p_i)] = \text{logit}(p_i) = \log[a] + \exp(\log[b]) \log (d_i / d_{max})$$

where a and b are random variables and d_{max} is the maximum planned dose.

The probability of each TPI and of a DLT will be summarized by dose along with the estimated dose-toxicity curve. A final estimate of the MTD will be estimated from the TPI Bayesian model utilizing all DLT-evaluable subjects. Please refer to Appendix C for more details on the TPI model and its operation characteristics.

10.5.2 Laboratory Test Results

10.5.2.1 Chemistry, Hematology and Coagulation

Individual chemistry, hematology and coagulation laboratory data will be listed and select parameters of interest plotted. Values outside the normal laboratory reference ranges will be flagged as high or low on the listings. CTCAE grades will also be highlighted where appropriate. Unscheduled assessments will be incorporated in the laboratory analyses where possible.

The number and percentage of subjects experiencing treatment emergent laboratory toxicities with worst post dose CTCAE grades of ≥1, ≥2, ≥3 and 4 will be presented. The direction of the laboratory worsening will be denoted. The summary will be presented for all laboratory

parameters for which at least one subject experienced a treatment emergent toxicity with a worst grade ≥ 3.

Additionally, the number and percentage of subjects experiencing 1, 2, 3 and 4 worsening CTCAE grade shifts from baseline will be presented. The direction of the laboratory worsening will again be denoted.

Shifts 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.

A listing of CTCAE grade 3 or higher laboratory toxicities will be provided. This listing will include all laboratory data for the subject and laboratory parameter of interest in order to provide proper context. A flag will indicate the grade 3 or higher toxicity.

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

A summary of the change from baseline to the post dose maximum, time to post-dose maximum, change from baseline to the post dose minimum, and the time to the post dose minimum may also be provided for selected parameters of interest.

Potential Hy's law cases will be listed and may also be summarized.

10.5.2.2 Urinalysis

Individual urinalysis data will be listed and select parameters of interest plotted.

Blood, protein, glucose and ketones will be graded in the following manner: 0='0 or Trace', 1='1+', 2='2+', 3='3+', 4='4+'. Microscopic parameters (WBC, 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'.

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

Summaries of the absolute value and/or changes from baseline at each scheduled assessment will also be provided.

10.5.3 Vital Signs

Vital signs will be listed for each subject. The analysis of vital signs will include summary statistics at selected time points for each subject. Shifts in vital sign values from baseline over time will be tabulated for each subject.

10.5.4 Electrocardiogram (ECG)

All on-study electrocardiogram (ECG) data will be listed and select parameters of interest may be plotted.

Summaries over time and/or changes from baseline over time will be provided for all 12-lead ECG parameters.

The analysis of Fridericia's (QTcF) QT correction will use the results recorded on the CRF. Subjects' maximum change from baseline in QT interval corrected by Fridericia's formula will be categorized into the following groups per their maximum change from baseline in QTcF. 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 dosing group.

In addition, the relationship between plasma concentration of AMG 211 and change from baseline in QTcF will be explored graphically.

10.5.5 Karnofsky Performance Status

Karnofsky performance status scores will be summarized at each assessed time point. The change in scores from baseline to each assessed time point and from baseline to safety follow-up will also be summarized.

10.5.6 Physical Measurements

The change in weight from baseline to each scheduled assessment time point will be summarized.

10.5.7 Antibody Formation

Positive anti-AMG 211 anti-body data will be listed and reviewed for each subject. Summaries of positive anti-AMG 211 anti-body test results over time may be provided.

10.5.8 Exposure to Investigational Product

Descriptive statistics will be produced to describe the exposure to investigational product by dosing schedule. The number of cycles, number of doses of investigational product and the total dose in µg will be summarized.

Details for each AMG 211 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.

10.5.9 Exposure to Concomitant Medication

All medication will be coded using the WHO drug dictionary. A subject listing of all prior and concomitant medications will be presented.

10.6 Efficacy Analyses

Statistical analyses of efficacy endpoints will be considered exploratory. Efficacy data will be conducted on the Safety Analysis Set. Dose/exposure-response relationship may be explored graphically.

10.6.1 Disease Response Analyses

All disease response data will be reported descriptively in the form of listings for all subjects. Listings will be produced for all subjects in the dose-escalation cohorts and the dose-expansion cohorts indicating the time to progression, objective response, time to response, and duration of response.

Unscheduled tumor assessments will be included in the response classifications for any analysis incorporating disease response data.

10.6.2 Best Overall Response

The number and percentage of subjects with a best overall disease/tumor response of complete response (irCR), partial response (irPR), stable disease (irSD) or progressive disease (irPD) will be presented for the dose escalation and/or expansion parts of the study. Subjects with best overall response of irSD should meet a minimum duration of 8 weeks at the time of scan.

In addition, the proportion of subjects treated at MTD with an overall response (ORR) and corresponding exact 80% confidence interval calculated using the Clopper-Pearson method (Clopper and Pearson, 1934) will be tabulated. Unevaluable subjects treated at MTD will be treated as non-responders and included in the calculation of ORR.

10.6.3 Time to Progression

Kaplan Meier curve will be presented for time to tumor progression with estimates for rates and 80% CI at selected weeks. Subjects treated at MTD will be included for this analysis.

10.6.4 Progression Free Survival

The proportion of subjects that are progression free at 6 months with corresponding exact 80% CI will be calculated using the Clopper-Pearson method. Subjects treated at MTD will be included for this analysis.

Kaplan Meier curve will also be presented as a sensitivity analysis with estimates for rates and 80% CI if subjects come off study prior to 6 months for reasons besides disease progression.

10.6.5 Tumor Burden

At each on-study radiological assessment, tumor burden will be derived as the sum of the products of the longest perpendicular diameters of the all index and new measurable lesions will be recorded. Baseline tumor burden will be the sum of the products of the 2 longest perpendicular diameters (SPD) of all the index lesions. The percentage change from baseline in tumor burden and percentage change from nadir will be recorded on the eCRF:

Individual subject time profiles of the percentage change from baseline will be generated for each part of the study. The initial dose group and primary tumor type will be highlighted.

For the dose escalation, the maximum percentage improvement (maximum decrease or minimum increase over the whole study) in the tumor burden will be plotted for each subject (waterfall plot). The initial dose group, primary tumor type and best overall response will be highlighted. This analysis will be repeated for subjects in the dose expansion part of the study.

In addition, summary statistics on percentage change from baseline by time point will be generated for subject treated at MTD.

10.6.6 Tumor Biopsy

Tumor biopsy data will be listed.

10.6.7 Duration of Response

Duration of response will be analyzed using the Kaplan-Meier method. Kaplan-Meier curves and median duration along with 80% CI will be presented. Subjects without an objective response are excluded from the analyses of this endpoint.

10.6.8 Time to Response

Time to response will be analyzed using the Kaplan-Meier method. Kaplan-Meier survival curves and median time along with 80% CI will be presented.

10.7 Analyses of Exploratory Endpoints

Graphical summaries of changes from baseline over time in biomarker levels may be provided. Association between baseline levels or presence of tumor biomarkers will be explored graphically for objective response and PD effects. If evident, correlation analyses for changes in biomarkers and tumor response or progression free survival will be performed.

10.8 Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Endpoints

AMG 211 serum concentration-time data will be used to determine the PK parameters using non-compartmental methods and will be listed for individual subjects. The following parameter will be estimated: C_{ss}, calculated as average concentration between achievement of plateau and EOI. Only in cycle 1 the following parameters will be estimated: $t_{1/2}$ and AUC_{inf}. Other parameters such as CL and V_{ss} may be estimated. Post dose serum concentrations below the lower limit of quantifications (BLQ) will be set to missing for non-compartmental analysis or 1/2 BQL for potential population PK analysis while those at predose will be set to zero. Actual dosing and sampling time will be used for PK analysis. The reasons for excluding any sample from the analyses will be provided. Other types of analysis such as compartmental modeling and population PK may be performed if necessary dependent on the data. The result of additional analysis may be reported separately from the clinical study report (CSR).

Individual concentration-time data will be tabulated and presented graphically. Mean concentration-time profiles for each dose will be provided using scheduled time. PK parameters will be summarized for each dose level using descriptive statistics, including subject numbers, means, standard deviations, medians, minimums, and maximums.

For dose proportionality analyses, C_{ss} will be dose normalized and log transformed prior to analysis. A linear regression may be performed using $\ln(C_{ss})$ as the dependent variable and ln(dose) as the independent variable. The scatter plots will show these estimated curves as well as the 90% confidence region around these curves. Tables will be generated to show the parameter estimates as well as the 90% confidence interval for the slope of the model. Given the broad dose range included in this study, the dose proportionality analyses may be restricted to the dose range being considered for the next stage of development.

PK/PD modeling may be performed if data are adequate to explore the relationship between AMG 211 exposure and various PD endpoints.

10.9 Changes from Protocol-Specified Analyses

There are no changes to the protocol-specified analyses.

11. LIST OF PLANNED TABLES, FIGURES AND LISTINGS [TFLs]

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

12. LITERATURE CITATIONS / REFERENCES

Babb, J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Stat Med* 1998;17:1103-1120.

Bailey S, Neuenschwander B, Laird G, Branson M. A Bayesian case study in oncology phase I combination dose-finding using logistic regression with covariates. *J Biopharm Stat*. 2009;19:469-484.

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

Neuenschwander B, Branson M, Gosponer T. Critical aspects of the Bayesian approach to phase I cancer trials. *Stat Med*. 2008;27:2420-2439.

13. PRIORITIZATION OF ANALYSES

There is no prioritization of analyses.

14. DATA NOT COVERED BY THIS PLAN

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

15. APPENDICES

Appendix A. Technical Detail and Supplemental Information Regarding Statistical Procedures and Programs

Imputation Rules for Partial or Missing Stop Dates

If the month and year are present, impute the last day of the month. If only the year is present, impute December 31 of that year. If the stop date is entirely missing, assume the event or medication is ongoing. If a partial or complete stop date is present and the 'ongoing' or 'continuing' box is checked, then it will be assumed that the AE or conmed stopped and the stop date will be imputed, if partial.

Imputation Rules for Partial or Missing Start Dates

		Stop Date						
		Complete: yyyymmdd		Partial: yyyymm		Partial: yyyy		Missing
Start Date		<1st Dose	≥1st Dose	<1st Dose yyyymm	≥1st Dose yyyymm	<1st Dose yyyy	≥1st Dose yyyy	
Partial: yyyymm	=1st Dose yyyymm	2	1	2	1	N/A	1	1
	≠ 1st Dose yyyymm		2		2	2	2	2
Partial: yyyy	=1st Dose yyyy	3	1	3	1	N/A	1	1
	≠ 1st Dose yyyy		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1 = Impute the date of first dose
2 = Impute the first of the month
3 = Impute January 1 of the year
4 = Impute January 1 of the stop year

Note: For subjects who were never treated (first dose date is missing), partial start dates will be set to the first day of the partial month.

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

CCI

Appendix C. Toxicity Probability Interval (TPI) Bayesian Design

The dose-escalation cohorts will use a TPI Bayesian design to guide dose escalation. The MTD target TPI for DLT is (0.20, 0.35), and TPIs of (0.35, 0.60) and (0.60, 1.00) are defined as excessive and unacceptable, respectively. The design seeks to identify a dose most likely to have a DLT rate in the target TPI, but with overdose control that limits the possibility that the dose has an excessive or unacceptable DLT rate (Babb et al, 1998). The probability of a DLT at dose level d_i is assumed to follow a Bernoulli distribution with probability p_i where the logit of p_i increases linearly with the log of the standardized dose in the following 2-parameter logistic model:

$$\log [p_i / (1-p_i)] = \text{logit}(p_i) = \log[a] + \exp(\log[b]) \log (d_i / d_{max})$$

where a and b are random variables and d_{max} is the maximum planned dose.

A minimally informative prior distribution (Neuenschwander et al, 2008) was selected for theta = (log a, log b) where the probability that the true DLT rate is ≤ 0.60 at the lowest dose (200 µg/day for 7 days) is 0.90 and the probability the true DLT rate is ≤ 0.20 at the maximum dose (6400 µg/day for 14 days) is 0.21. Median values for p_i were interpolated per the logistic model. For each d_i, 2 quantiles for p_i were selected from a Beta distribution with the target median and a precision < 2. This set of quantiles fully specified a target prior for theta. A bivariate normal distribution for theta was assumed where (log a) has a normal distribution with mean ma and standard deviation sa, and (log b) has a normal distribution with mean mb and standard deviation sb, and r is the correlation between (log a) and (log b). Numerical integration was used to calculate $Pr[p_i \leq q(d_i)]$, where $q(d_i)$ is a quantile for dose d_i. An optimal bivariate normal distribution was estimated that achieved the minimum sum of squared difference between achieved and specified quantiles across all doses. The bivariate normal distribution prior solution has ma = 0.0536, sa = 1.685, mb = -0.1469, sb = 1.001, and r = -0.216.

The operating characteristics of the TPI Bayesian design were evaluated via simulation. The cohort size was fixed to be 3 subjects. All simulated studies started with an initial dose of 200 µg/day for 7 days and subsequent doses were selected based on the rules specified in Section 8.

The design was evaluated for two possible dose-response scenarios consistent with the design's model: "Low" and "High" MTDs. The target MTD of these 2 scenarios are set to be 1/3, and this target was set at a dose of 400 µg/day for 14 days, and 3200 µg/day for 14 days for the Low High scenarios, respectively. An "acceptable" final dose was defined as one with the

highest or second highest target TPI probability and a less than 0.25 excessive or unacceptable TPI probability. The acceptable doses by scenario were 200, 400 for Low and 1600, 3200 for High. The bivariate normal distribution for each scenario was estimated as per the prior. A Bernoulli probability for each subject in the simulations was selected from a random sample for the bivariate normal distribution for a given scenario, and a random DLT outcome for the subject was generated based on this DLT probability.

The simulation results are summarized in the table below. The average sample size increased from 15.7 to 20.8 in Low to High MTD scenarios. Early DLTs resulted in close to 5% chance of stopping with no dose selection in both cases due to the over-dose control criteria. The average Low and High final dose was about 360 and 1600 µg/d. The rate of acceptable dose selection is higher in Low scenario (83.1%), comparing with High scenario (51.4%). The average number of DLTs and DLT rate is higher in Low scenario than High scenario due to early occurrence of DLTs.

Simulation Results of TPI Model

	Low **Acceptable Dose: 200, 400**	**High** **Acceptable Dose: 1600, 3200**
Number of Subjects (IQR)	15.7 (15, 18)	20.8 (18, 24)
Acceptable Rate (95% CI)	83.1% (80.6%, 85.4%)	51.4% (48.3%, 54.5%)
No Dose Rate (95% CI)	4.5% (3.3%, 6.0%)	3.7% (2.6%, 5.1%)
Dose (IQR)	358.0 (200, 400)	1583.8 (800, 1600)
Number of DLTs (IQR)	3.8 (2, 4)	3.1 (2, 4)
DLT Rate	24.1%	15.0%

CI = confidence interval; DLT = dose-limiting toxicity; IQR = interquartile range

Appendix D. DLRM Outputs List

Listings:

1. subject demographics
2. date, time, dose of investigational product administration
3. electrocardiogram (ECG) results, change from baseline
4. vital signs
5. reported adverse events (AEs)
6. safety laboratory data (chemistry, hematology, Urinalysis, coagulation)
7. medical history
8. concomitant medications
9. prior chemotherapy data

Figures

1. line plot of vital signs by time course
2. line plot of all safety labs by time course
3. line plot of ECGs results by time course
4. DLT probability plot at each dose [TPI design if DLT is observed]