Statistical Analysis Plan

Protocol Title:	A Phase 3 Multicenter, Randomized, Open Label, Active-controlled, Study of AMG 510 Versus Docetaxel for the Treatment of Previously Treated Locally Advanced and Unresectable or Metastatic NSCLC Subjects With Mutated KRAS p.G12C
Short Protocol Title:	A Phase 3 Study to Compare AMG 510 With Docetaxel in NSCLC Subjects With *KRAS p.G12C* Mutation (CodeBreak 200)
Protocol Number:	20190009
NCT Number:	Not available
Authors:	[redacted] Biostatistics Senior Manager, GBS, Amgen Inc. [redacted] Study Statistician, IQVIA.
Sponsor:	Amgen Inc. One Amgen Center Drive Thousand Oaks, CA, 91320
SAP Date:	**Document Version** — **Date** **Amendment 3 (v5.0)** — 21 July 2022

NCT Number: NCT04303780
This NCT number has been applied to the document for purposes of posting on Clinicaltrials.gov

Version Number	**Date (DDMMMYYYY)**	**Summary of Changes, including rationale for changes**
Original (v1.0)	17Dec2019	Original SAP
Amendment 1 (v2.0)	07May2021	Update to align with protocol amendment 3 Update definitions (section 5) Add pre-specified subgroup variables (section 4.2) Add sensitivity analyses for OS (9.5.2.1) Add COVID-19 analyses (section 9.7.5) Add Appendix A, B, C, D and E.
Amendment 2 (v4.0)	13Oct2021	Remove ECG analysis section to align with PA3 Update language for ORR summary measure and OS primary analysis timing to align with PA3, and OS data review after the data cutoff date (Section 9.5) Add imputation method for missing treatment relationship in AE reporting (Section 9.6.2). Update the imputation method for the date of the post-treatment anti-cancer therapy Appendices A. Update the MedDRA version to 24.0 or later and add safety summaries by SOC, HLT and PT (Section 9.6.2).
Amendment 3 (v5.0)	**21Jul2022**	**Remove the unintended word "consecutive" regarding the confirmation window of CR/PR in the definitions of "Best Overall Response (BOR)" and "Objective Response as per RECIST 1.1 (OR)", to be consistent with RECIST 1.1 in section 5.** **Update the language for the PFS 2 analyses for clarity (Section 9.5.3)**

Table of Contents

List of Appendices

List of Abbreviations

Abbreviation or Term	Definition/Explanation
AE	Adverse Event
AFT	Accelerated Failure Time
ANC	Absolute Neutrophil Count
AUC	Area Under the Concentration-Time Curve
BICR	Blinded Independent Central Review
BOR	Best Overall Response
C_{max}	Maximum Observed Concentration
CNS	Central Nervous System
COP	Confirmation Of Progression
CPMS	Clinical Pharmacology Modeling and Simulation
CR	Complete Response
CRF	Case Report Form
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DAP	Data Access Plan
DCR	Disease Control Rate
DMC	Data Monitoring Committee
DOR	Duration Of Response
ECG	Electrocardiogram
EQ-5D-5L	Euroqol-5 Dimension
EOI	Events Of Interest
EORTC QLQ-C30	European Organization For Research And Treatment Of Cancer Quality-Of-Life Questionnaire Core 30
EOS	End Of Study
EOT	End Of Treatment
FDA	Food and Drug Administration
GEE	Generalized Estimating Equations
GSO-DM	Global Study Operations-Data Management
HR	Hazard Ratio
IA	Interim Analysis
ICF	Informed Consent Form
IP	Investigational Product
IPCW	Inverse Probability of Censoring Weighting
IPD	Important Protocol Deviation
IRB/EC	Institutional Review Board/Ethics Committee
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog (Protein)
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog (Dna)
KRASG12C	Kras Protein With A G12c Mutation At The Protein Level
KRAS p.G12C	Kras Dna With A Mutation Resulting In A G12c Mutation At The Protein Level
LSE	Last Subject Enrolled
LTFU	Long Term Follow Up
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model with Repeated Measures
MRI	Magnetic Resonance Imaging
NCI	National Cancer Institute
NE	Not Evaluable
NSCLC	Non-Small-Cell Lung Carcinoma

OR	Objective Response
ORR	Objective Response Rate
OS	Overall Survival
PA	Primary Analysis
PD	Progressive Disease
PK	Pharmacokinetic(s)
PFS	Progression-Free Survival
PR	Partial Response
PRO	Patient-Reported Outcomes
PRO-CTCAE	Patient-Reported Outcomes Version of The Common Terminology Criteria for Adverse Events
QLQ-LC13	Quality-Of-Life Questionnaire Lung Cancer Module
QOL	Quality Of Life
RAS	Rat Sarcoma Viral Oncogene Homolog
RDI	Relative Dose Intensity
RECIST	Response Evaluation Criteria In Solid Tumors
RPSFT	Rank Preserving Structural Failure Time
SD	Stable Disease
SFU	Safety Follow Up
SOA	Schedule Of Activities
t_{max}	Time to Reach Maximum Concentration
TEAE	Treatment Emergent Adverse Event
TTR	Time To Response
VAS	Visual Analogue Scale
WHO DRUG	World Health Organization Drug

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 3 for study 20190009, AMG 510 dated 15 February 2021. The scope of this plan includes the interim analysis, the primary analysis and the final analysis that are planned except for patient-reported outcomes (PRO) related endpoints and will be executed by the Amgen Global Biostatistical Science department unless otherwise specified. Details of the analysis plan for PRO related endpoints are described in a separate PRO Supplemental Statistical Analysis Plan. Pharmacokinetic, pharmacodynamic, exposure-response and biomarker analyses will be performed by Clinical Pharmacology Modeling and Simulation (CPMS) or biomarker group.

2. Objectives, Endpoints and Hypotheses

2.1 Objectives and Endpoints/Estimands

Objectives	**Endpoints**
Primary	
• To compare the efficacy of AMG 510 versus docetaxel as assessed by progression-free survival (PFS) in previously treated subjects with *KRAS p.G12C* mutated non-small cell lung cancer (NSCLC)	• PFS - defined as time from randomization until disease progression or death from any cause, whichever occurs first for all subjects. Progression will be based on blinded independent central review (BICR) of disease response per Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1).
Attributes	Primary Estimand
• Target Population	• Subjects with previously treated locally advanced and unresectable or metastatic NSCLC with *KRAS p.G12C* mutation
• Primary Endpoint	• PFS
• Summary Measures	• Hazard Ratio (HR)
• Intercurrent Events and Strategies	• Start of new anti-cancer therapy prior to PFS event • PFS is censored at the date of last evaluable assessment before or on start of new anti-cancer therapy.
Primary Estimand Description	
HR of PFS between AMG 510 and docetaxel, for subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS p.G12C mutation, before or on start of new anti-cancer therapy	
Key Secondary	
• To compare the efficacy of AMG 510 versus docetaxel as assessed by: • Overall Survival (OS)	• Overall survival - defined as time from randomization until death from any cause. • Objective response (complete response [CR] + partial response [PR]), assessed per RECIST v1.1. Response will be assessed by BICR. Complete

• Objective response rate (ORR)	response and PR require confirmatory repeat radiologic assessment at no less than 4 weeks after the original response. The normal subsequent assessment is acceptable to confirm response.

• To compare patient-reported outcomes (PRO) as assessed by: European Organization for Research and Treatment of Cancer Quality-of-life Questionnaire Core 13 (EORTC QLQ-LC13) and European Organization for Research and Treatment of Cancer Quality-of-life Questionnaire Core 30 (EORTC QLQ-C30)	• Change from baseline (cycle 1 day 1) over time to week 12 in disease related symptoms of:
	  • Dyspnea as measured by a 4-item dyspnea domain from QLQ-C30 and QLQ-LC13
	  • Cough as measured by QLQ-LC13
	  • Chest Pain as measured by QLQ-LC13
	• Change from baseline over time to week 12 in
	  • Physical functioning as measured by QLQ-C30
	  • Global health status as measured by QLQ-C30

Attributes	Secondary Estimand - OS
• Target Population	• Subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS p.G12C mutation
• Key Secondary Endpoint	• OS
• Summary Measures	• Hazard Ratio (HR)

• Intercurrent Events and Strategies	• Start of new anti-cancer therapy
	  • OS will be estimated regardless of subsequent anti-cancer therapy.
	• Crossover from control group to treatment group
	  • OS will be estimated regardless of crossover.

Secondary Estimand Description - OS	
HR of OS between AMG 510 and docetaxel, for subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS p.G12C mutation, regardless of subsequent anti-cancer therapy and/or crossover.	
Attributes	Secondary Estimand – Objective Response
• Target Population	• Subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS p.G12C mutation
• Key Secondary Endpoint	• Objective Response
• Summary Measures	• Odds Ratio
• Intercurrent Events and Strategies	• Discontinuation of treatment prior to achieving an objective response (PR or CR) • Subjects will be considered as non-responders.
Secondary Estimand Description – Objective Response	
Odds ratio of objective response between AMG 510 and docetaxel, for subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS	

p.G12C mutation. Subjects who discontinued treatment prior to achieving an objective response are considered as non-responders.	
Attributes	Secondary Estimand - PRO
• Target Population	• Subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS p.G12C mutation
• Key Secondary Endpoint	• PRO
• Summary Measures	• Change from baseline to Week 12
• Intercurrent Events and Strategies	• Start of new anti-cancer therapy including crossover before Week 12 • PRO measurements before or on start of new anti-cancer therapy including crossover will be used to estimate treatment effect.
Secondary Estimand Description - PRO	
Change from baseline to Week 12 in PRO endpoints between AMG 510 and docetaxel, for subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS p.G12C mutation. PRO measurements before or on start of new anti-cancer therapy including crossover will be used to estimate treatment effect.	
Secondary	
• To compare efficacy of AMG 510 versus docetaxel as assessed by: duration of response (DOR), time to response (TTR), and disease control rate (DCR)	• Duration of response - defined as time from first evidence of PR or CR to disease progression or death due to any cause, whichever occurs first. Progression will be based on an BICR assessment of disease response per RECIST v1.1. • Time to response - defined as time from randomization to first evidence of PR or CR. • Disease control rate - defined as rate of confirmed objective response (CR or PR) + stable disease per RECIST v1.1 of at least 6 weeks measured
• To compare the safety and tolerability of AMG 510 versus docetaxel	• Subject incidence of treatment-emergent adverse events, treatment-related adverse events, changes in vital signs, and clinical laboratory tests.
• To compare the effect of treatment with AMG 510 on other treatment and disease related symptoms, and health related quality of life relative to docetaxel	• Change from baseline over time to week 12 for the remaining subscales for QLQ-LC13 and QLQ-C30 • Time to deterioration for the subscales for QLQ-LC13 and QLQ-C30 • Summary scores at each assessment and changes from baseline of visual analogue scale (VAS) scores as measured by EuroQol-5 Dimension (EQ5D5L)
• To characterize the pharmacokinetics (PK) of AMG 510 and its major metabolites	• Pharmacokinetic (PK) parameters of AMG 510 including, but not limited to, maximum plasma concentration (C_{max}), area under the plasma concentration-time curve (AUC) on Days 1 and 8, and pre-dose (trough) concentrations through cycle 4

Exploratory

[redacted]

Objectives	Endpoints
Exploratory (Continued)	
[redacted]	

2.2 Hypotheses and/or Estimations

The following 2-sided hypotheses will be tested in this trial.

Primary endpoint:

H_{01}: PFS survival distribution of AMG 510 group is the same as for docetaxel group versus H_{11}: the 2 PFS survival distributions are different.

Key secondary endpoints:

H_{02}: odds ratio of objective response between AMG 510 group and docetaxel group = 1 versus H_{12}: odds ratio of objective response between AMG 510 group and docetaxel group ≠ 1

H_{03}: OS survival distribution of AMG 510 group is the same as for docetaxel group versus H_{13}: the 2 OS survival distributions are different.

H_{04}: difference in mean change from baseline over time to week 12 in symptom of dyspnea as measured by a 4-item dyspnea domain from QLQ-C30 and QLQ-LC13 = 0 versus H_{14}: difference in mean change from baseline over time to week 12 in symptom of dyspnea as measured by a 4-item dyspnea domain from QLQ-C30 and QLQ-LC13 ≠ 0

H_{05}: difference in mean change from baseline over time to week 12 in symptom of cough as measured by QLQ-LC13 = 0 versus H_{15}: difference in mean change from baseline over time to week 12 in symptom of cough as measured by QLQ-LC13 ≠ 0

H_{06}: difference in mean change from baseline over time to week 12 in symptom of chest pain as measured by QLQ-LC13 = 0 versus H_{16}: difference in mean change from baseline over time to week 12 in symptom of chest pain as measured by QLQ-LC13 ≠ 0

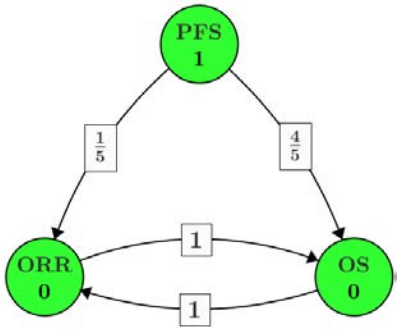
H_{07}: difference in mean change from baseline over time to week 12 in physical functioning as measured by QLQ-C30 = 0 versus H_{17}: difference in mean change from baseline over time to week 12 in physical functioning as measured by QLQ-C30 ≠ 0

H_{08}: difference in mean change from baseline over time to week 12 in global health status as measured by QLQ-C30 = 0 versus H_{18}: difference in mean change from baseline over time to week 12 in global health status as measured by QLQ-C30 ≠ 0

The hypotheses of the primary efficacy endpoint and key secondary efficacy endpoints will be tested using the following graphical multiple testing procedure (Maurer and Bretz, 2013) to control the study-level overall type I error rate below 1-sided 0.025 levels. A hypothesis can be re-tested repeatedly with a different nominal level that is propagated from rejecting other hypothesis test(s).

Figure 2-1 illustrates the Maurer-Bretz multiple testing procedure among PFS, OS and ORR. The fractions on the directed arrows indicate the proportion of α propagated from one hypothesis to the other when its hypothesis is rejected. Starting with PFS, if the hypothesis of PFS is rejected, ORR will be tested using 1-sided α/5 (0.005) level. With the rejection of ORR hypothesis, OS will be tested using 1-sided full α (0.025) level. If ORR hypothesis is failed to be rejected, OS will be tested using 1-sided 0.0001 level if at PFS IA, otherwise using 4α/5 (0.02) level. With the rejection of OS hypothesis, ORR can be retested using 1-sided full α (0.025) level.

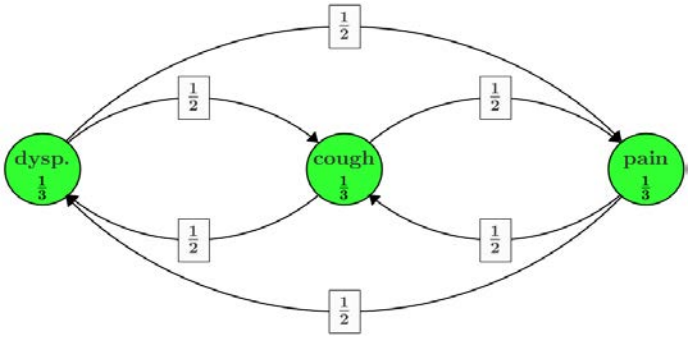
Figure 2-1. Initial Graph of Maurer-Bretz Multiple Testing Procedure for PFS, OS, and ORR.



ORR = objective response rate; OS = overall survival; PFS = progression-free survival

If all 3 hypotheses of PFS, OS, ORR are rejected, the next 3 endpoints of change from baseline over time to week 12 in 3 lung cancer symptoms will be tested using Holm's procedure, illustrated in Figure 2-2, including change from baseline over time to week 12 for the symptom of dyspnea as measured by a 4 item dyspnea domain from QLQ-C30 and QLQ-LC13 (dyspnea), change from baseline over time to week 12 for the symptom of cough as measured by QLQ-LC13 (cough), and change from baseline over time to week 12 for the symptom of chest pain as measured by QLQ-LC13 (pain). Hypotheses are rejected sequentially based on the smallest p-value.

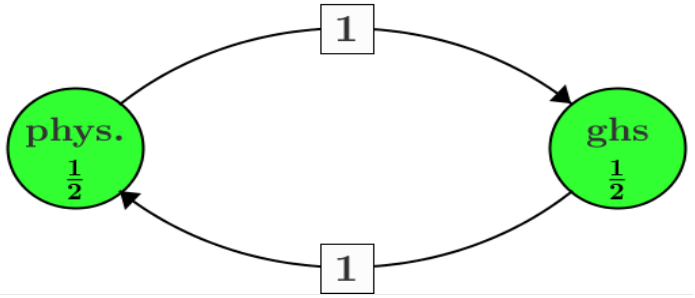
Figure 2-2. Holm's Procedure for Change From Baseline Over Time to Week 12 in 3 Lung Cancer Symptoms



Dysp = dyspnea

If all 6 hypotheses listed above are rejected, the next 2 PRO endpoints will be tested using the Holm's procedure, illustrated in Figure 2-3, including change from baseline over time to week 12 in physical functioning as measured by QLQ-C30 (phys.), change from baseline over time to week 12 in global health status as measured by QLQ-C30 (ghs). Hypotheses are rejected sequentially based on the smallest p-value.

Figure 2-3. Holm's Procedure for Change From Baseline Over Time to Week 12 on Physical Functioning and Global Health Status



phys = physical; ghs = global health status

3. Study Overview

3.1 Study Design

This is a phase 3, multicenter, randomized, open label, active-controlled, study to evaluate the efficacy, safety, and tolerability of AMG 510 versus docetaxel in subjects with previously treated locally advanced and unresectable or metastatic NSCLC with *KRAS p.G12C* mutation. The study will be conducted at approximately 290 sites globally. The study will consist of a screening period, a treatment period, a safety follow-up (SFU) period, and long-term follow-up period. Approximately 330 previously treated subjects with locally advanced and unresectable or metastatic NSCLC with centrally confirmed *KRAS p.G12C* mutation will be enrolled and randomized 1:1 to receive either AMG 510 or docetaxel.

Subjects will be stratified by number of prior lines of therapy in advanced disease (1 versus 2 versus > 2), race (Asian versus non-Asian), and history of central nervous system (CNS) involvement (present or absent).

Cycle 1 day 1 will be defined as the first day subject receives study medication. A cycle is 21 days in length ± 3 days, unless a delay is medically necessary. A ± 3-day window is allowed for protocol required assessments, unless otherwise specified. Tumor assessment will be conducted (by magnetic resonance imaging [MRI] and/or contrast enhanced computerized tomography [CT]) during screening (within 28 days of day 1); every 6 weeks from cycle 1 day 1; (± 7 days) (at weeks 7, 13, 19, 25, 31, 37, 43, and 49), and then at 9 week intervals (± 7 days) thereafter until independent central confirmation of progression, start of another anti-cancer therapy, withdrawal of consent, lost to follow up, or death, whichever occurs earliest. Subjects' scans will undergo independent central confirmation of progression (COP) at the time of first progressive disease (PD). Subjects that undergo treatment beyond progression or crossover, from docetaxel to AMG 510, will continue to receive scans after confirmation of first PD.

Once subjects on the docetaxel group are determined to have radiological progression according to RECIST v1.1 by the investigator and progressive disease undergoes independent central confirmation, they will be given the opportunity to crossover and receive AMG 510. Alternatively, should an early efficacy of the study be noted by the data monitoring committee (DMC), crossover will be considered for all subjects who randomized into the docetaxel group (so that they will be able to immediately receive AMG 510). All subjects remaining on study will be followed for survival, as specified in the SOA, until the pre-specified OS events are reached, regardless of the OS analysis result at the PFS primary analysis.

Tumor assessment and response will be confirmed by BICR who will evaluate disease progressions and responses without the knowledge of randomization assignments, in accordance with the RECIST v1.1.

In all subjects, after end of investigational product, information regarding date of progression, the type and duration of subsequent therapies, response to subsequent therapy, date of progression on subsequent therapy, and survival data will be collected. Subjects who discontinue treatment prior to RECIST v1.1 disease progression (eg, due to unacceptable toxicity) will also continue to be followed with tumor assessments until independent central confirmation of disease progression, withdrawal of consent, or start of another anti-cancer therapy, then followed for subsequent anti-cancer therapy and survival.

Subjects that consent to treatment beyond progression or consent to crossover, from docetaxel to AMG 510, will continue to receive investigational product after independent central confirmation of progression at the time of first progressive disease (PD).

In select cases, subjects may continue on treatment following radiologic progression if they are continuing to demonstrate clinical benefit (see protocol Section 8.1.6 and 8.1.7).

An independent (external to Amgen) data monitoring committee (DMC) will review safety and efficacy data as per DMC charter. Interim safety analysis will be conducted after approximately 50, 100, and 200 subjects have been enrolled and have had the opportunity to complete at least 6 weeks of study treatment, and then at approximately 6-month intervals until the primary analysis (PA) of PFS. Interim analysis (IA) data will be reviewed by the DMC. There will be 2 planned PFS efficacy analyses. The PA of PFS will occur when approximately 230 PFS events have been observed. The PFS PA may be delayed to ensure that the enrollment is finished and the delayed PA will be

triggered when the last randomized subject has had the opportunity to have at least 6 weeks of follow up. An IA of PFS for superiority is planned when approximately 70% (160 events) of the total PFS events have been observed from both groups, or when the enrollment is finished and the last randomized subject have had the opportunity to have 6 weeks of follow up, whichever occurs later.

Early efficacy at the proposed PFS interim analysis will be claimed if the observed PFS difference meets the pre-specified statistical significance as well as being considered clinically meaningful. If early efficacy is claimed at the PFS interim analysis, PFS interim analysis will be considered as PFS primary analysis. More details are included in the DMC charter.

To ensure trial integrity, a Data Access Plan (DAP) is established to govern the restricted data access to Amgen internal but product-independent personnel to make a decision on the study conduct, following DMC's recommendation. More details are included in the DAP.

If the primary endpoint PFS achieves statistical significance, all subjects remaining on study will still be followed for their survival data until the OS primary analysis.

One administrative interim summary for OS will be performed at the PFS interim analysis if PFS interim result achieves statistical significance.

There will be two planned OS analyses testing for superiority of AMG 510 over docetaxel, one OS interim analysis and one OS primary analysis. The OS interim analysis will be performed either at the time when PFS achieves statistical significance or after 175 OS events (~53% maturity) have been observed, whichever is later.

The OS primary analysis when at least 198 OS events (~60% maturity) have been observed which is expected to be at approximately 3 months after PFS primary analysis.

The analysis of ORR will be performed at the time when PFS is claimed statistically significant and the last randomized subject has had the opportunity to have at least 12 weeks of follow up. The final analysis will be performed after the last subject completes long-term follow-up (LTFU). In the case that PFS results are not statistically significant at the primary analysis, the sponsor may stop the study and subjects will not be followed for OS any further.

3.2 Sample Size

The PA of PFS will occur when approximately 230 PFS events have been observed. With 230 PFS events, the study will have ~90% power to show a statistically significant PFS at the 2.5% 1-sided significance level if the true treatment effect hazard ratio (HR) is assumed 0.65 for the AMG 510 group versus the control group. The sample size is chosen to have 70% maturity realizing PFS events, which is approximately 330 subjects with 1:1 randomization ratio into the AMG 510 group and the docetaxel group. The PFS PA may be delayed to ensure that the enrollment is finished and the delayed PA will be triggered when the last randomized subject have had the opportunity to have at least 6 weeks of follow up.

An IA is planned when approximately 70% (160) of the target PFS events have been observed from both groups, or when the enrollment is finished and the last randomized subject have had the opportunity to have 6 weeks of follow up, whichever occurs later. The monitoring boundary for early stopping for efficacy will be based on an O'Brien-Fleming type alpha spending function for multiplicity adjustment with 0.007 one-sided alpha spent at the PFS IA. The actual information fraction will be calculated based on the number of observed events at the time of the analysis. Under exponential distribution, the minimum detectable difference for success in this design is an HR of 0.68 between AMG 510 group and the docetaxel group with 160 PFS events, 70% of the target PFS events, and an HR of 0.769 at the PA with 230 PFS events. Assuming an enrollment rate of 40 subjects per month after a 3-month ramp-up period, with a total sample size of 330, it is estimated that approximately 19 months will be required to reach 230 PFS events, and 13 months will be required to reach 70% (160) of the target PFS events. This estimation is based on a median PFS of 5 months for the control group (Charpidou et al, 2019) and 7.7 months for the AMG 510 group, and a 10% dropout rate.

If the primary endpoint PFS achieves statistical significance, all subjects remaining on study will still be followed for their survival data until the OS primary analysis, to enable OS analyses and a robust description of the totality of the data.

An administrative interim summary for OS will be performed at the PFS interim analysis with approximately 107 OS events (~32% maturity) observed, if PFS achieves statistical significance at the interim analysis. A nominal alpha of 0.01% (negligible impact on overall type I error rate) will be spent on this OS interim summary.

There will be two planned OS analyses testing for superiority of AMG 510 over docetaxel, one OS interim analysis and one OS primary analysis. The OS interim analysis will be performed either at the time when PFS achieves statistical significance or after 175 OS events (~53% maturity) have been observed, whichever is later. Assuming the actual crossover rate of approximately 30% at the time of PFS primary analysis for subjects on the control group who have disease progression, then with 175 OS events, the study has ~96% probability to observe a HR < 1 when the true OS HR is 0.75.

The OS primary analysis will occur when at least 198 OS events (~60% maturity) have been observed, which is expected to be at approximately 3 months after PFS primary analysis. The estimation is based on a median OS of 9 months for the control group and 12 months for the AMG 510 group (OS HR=0.75). The multiplicity will be adjusted as necessary based on O'Brien-Fleming type alpha spending function.

3.3 Adaptive Design

Not applicable

4. Covariates and Subgroups

4.1 Planned Covariates

Covariates may be incorporated in selected models of efficacy endpoints. In addition to the stratification factors for randomization, number of prior lines of therapy in advanced disease (1 versus 2 versus > 2), race (Asian versus non-Asian), history of CNS involvement (yes versus no), the following additional covariates may be included:

- region
- best response on immediate prior therapy
- age
- sex
- race
- ECOG status
- liver metastasis at baseline
- disease stage
- smoking history
- histology
- brain metastasis at baseline
- bone metastasis at baseline
- PD-L1 protein expression

- [redacted]
- [redacted]
- Presence of other co-mutations at baseline
  - [redacted]

4.2 Subgroups

In addition to the stratification factors for randomization, number of prior lines of therapy in advanced disease (1 versus 2 versus > 2), race (Asian versus non-Asian), history of CNS involvement (yes versus no), primary and selected secondary endpoints will be examined in the following subgroups to investigate the consistency of treatment effects:

- region (North America and Europe vs rest of world)
- best response on immediate prior therapy; primary refractory (progression on first scan) **vs** suboptimal response (stable disease) vs recurrent (initial response with subsequent growth)
- age (< 65 vs ≥ 65)
- sex (male vs female)
- race (white vs black vs Asian vs other)
- ECOG status (0 vs 1)
- liver metastasis at baseline (yes vs no)
- stage (locally advanced vs unresectable vs metastatic)
- smoking history (never vs current vs former)
- histology (squamous vs non-squamous)
- brain metastasis at baseline (yes vs no)
- bone metastasis at baseline (yes vs no)
- PD-L1 protein expression (< 1% vs ≥ 1% and < 50% vs ≥ 50%)
- [redacted]
- presence of other co-mutation at baseline (yes vs no)
  - [redacted]

In the event that there are insufficient number of subjects in the subgroup (ie, less than 10% of the whole population), relevant subgroups may be combined.

Country specific subgroup analysis may be conducted for regulatory requirement in each country.

5. Definitions

Baseline

For any variable, unless otherwise specified, the baseline is the last non-missing assessment taken prior to the first administration of any investigational products. 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.

Best Overall Response (BOR):

Best overall response (BOR) for a subject is the best response recorded from the start of the treatment until disease progression/recurrence per RECIST v1.1. Overall response assessments occurring after the start of the first subsequent anticancer therapy will not be included. No additional derivations will be required to classify the response.

BOR is defined as the best response in the following order: CR, PR, SD, PD, or NE, where CR and PR require confirmation by a repeat scan at least 4 weeks after the first documentation of response. A best overall response of SD requires an on-study imaging of SD or better no earlier than 6 weeks after cycle 1 day 1; otherwise the best overall response will be not evaluable (NE)

Change from Baseline

Change from baseline is the arithmetic difference between a post-baseline value 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

[redacted]

Disease Control Rate (DCR)

DCR is defined as the proportion of subjects with the best overall response of CR or PR or of stable disease per RECIST v1.1 >= 6 weeks.

Duration of Response (DOR)

DOR is defined as time from the first evidence of PR or CR to subsequently confirmed disease progression by BICR or death due to any cause (whichever comes first):

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

DOR will be calculated only for subjects who achieve a confirmed best overall response of PR or CR. Subjects will be censored using the same censoring rule for PFS as listed in Appendix D if applicable.

Last Known Alive Date

Last known alive date is the latest date of the following dates before death date. Other Case Report Form (CRF) data may be explored to get the last known alive date.

- Date of Randomization on Subject Enrollment CRF
- Date First Taken, Date Last Taken on Concomitant Medications CRF
- Date Performed on ECOG Performance Status, Vital Signs, Echocardiogram, Electrocardiogram, Transfusions, Surgery, Procedure CRFs
- Admission Date, Discharge Date on Hospitalizations, CRF Date of Examination on Physical Measurement CRF
- Date Collected on Reproductive Status and Pregnancy Test (Local Lab), Chemistry (Local Lab), Hematology (Local Lab), Coagulation (Local Lab) CRFs, Urinalysis (Local Lab) and in central lab data
- Start Date, Stop Date on Investigational Product Administration CRF
- Start Date, Stop Date on Other Protocol Required Therapy CRF
- Date Started and Date Ended or Resulted in Death on Events CRF
- Start date, Stop date on Concurrent Radiotherapy, Anti-Cancer Therapies CRF
- Subject Status Date if status is Alive on Survival Status CRF
- Assessment Date of CT or MRI and Date of Tumor Response Assessment
- Date of Clinical Outcome Assessment

Investigational Product (IP)

The term ‘investigational product’ is used in reference to AMG 510 and docetaxel.

Last Investigational Product Dose Date

The last IP date for each subject is defined as the latest date non-missing administration of AMG 510 and docetaxel.

Objective Response as per RECIST 1.1 (OR)

Objective response is defined as a complete response or partial response, as defined by RECIST 1.1. Confirmation is determined by a repeat assessment no less than 4 weeks from the date of first documented assessment. Response will be assessed by BICR.

Objective response rate (ORR)

ORR is defined as the proportion of subjects with a best overall response of confirmed CR or confirmed PR per RECIST v1.1.

Overall Survival (OS)

OS is defined as the time from the date of randomization until event of death due to any cause.

OS = (Date of death– date of randomization + 1) ×12 / 365.25

Subjects still alive will be censored at the date last known to be alive. If the date last known to be alive is after the date that triggers the analysis (ie, the data cutoff date), the subject will be censored at the date of last contact through the analysis trigger date.

Progression-Free Survival (PFS) by BICR

PFS by BICR is defined as the time from the date of randomization to disease progression by BICR or death due to any cause (whichever comes first).

PFS = (PD / death date - randomization date + 1) ×12 / 365.25

The censoring rules are available in Appendix D.

Progression-Free Survival (PFS) by Investigator

PFS by investigator is defined as the time from the date of randomization to disease progression by investigator or death due to any cause (whichever comes first).

PFS = (PD / death date - randomization date + 1) ×12 / 365.25

The censoring rules are available in Appendix D.

[redacted]

[redacted]

Relative Dose Intensity (RDI) of AMG 510:

RDI of AMG 510 is calculated as actual dose intensity / planned dose intensity, where

- Cumulative actual dose (mg) is defined as the total dose given during the study treatment exposure. For subjects who did not take any drug the cumulative actual dose by definition is 0 mg.
- Actual dose intensity for subjects with non-zero duration of exposure is defined as: cumulative actual dose (mg) / actual duration of exposure (days), where the actual duration of exposure = date of last dose – date of first dose + 1. For subjects who did not take any drug, the actual dose intensity is 0 mg/day
- Cumulative planned dose is the per-protocol planned dose accumulated over the planned duration on study treatment.
- Planned dose intensity for subjects with non-zero duration of exposure is defined as: cumulative planned dose (mg) / planned duration of exposure (days). For this study, the planned dose intensity is 960 mg/day.
    - If subject is on the IP then the planned duration of exposure = date of last dose – date of first dose + 1.
    - If subject has end of treatment (EOT) then the planned duration of exposure = max (EOT, date of last dose) – date of first dose + 1.

Relative Dose Intensity (RDI) of Docetaxel:

RDI of docetaxel is calculated as actual dose intensity / planned dose intensity, where

- Cumulative actual dose (mg/m^2) is defined as the total dose given during the study treatment exposure divided by BSA (m^2). For subjects who did not take any drug the cumulative actual dose by definition is 0 mg.
- Actual dose intensity for subjects with non-zero duration of exposure is defined as: cumulative actual dose (mg/m^2) / actual duration of exposure (days), where actual

duration of exposure = date of last dose – date of first dose + 21. For subjects who did not take any drug, the actual dose intensity is 0 mg/m^2/day.

- Cumulative planned dose is the per-protocol planned dose accumulated over the planned duration on study treatment.
- Planned dose intensity for subjects with non-zero duration of exposure is defined as: cumulative planned dose (mg/m^2) / planned duration of exposure (days). For this study, the planned dose intensity is 3.57 mg/m^2/day.
    - If subject is on the IP then the planned duration of exposure = (date of last dose + 20) – date of first dose + 1.
    - If subject has EOT then the planned duration of exposure = max (EOT, date of last dose + 20) – date of first dose + 1.

Study Day

For subjects who have administered at least one dose of IP,

Post study day 1: study day = (date – date of first dose) + 1.

Pre study day 1: study day = (date – date of first dose).

For subjects who have not administered any dose of IP,

Post study day 1: study day = (date – date of randomization) + 1.

Pre study day 1: study day = (date – date of randomization).

Time to Response (TTR)

TTR is defined as time from the date of randomization to the first evidence of PR or CR and subsequently confirmed. TTR will be calculated only for subjects who achieve a confirmed best overall response of PR or CR.

Treatment Emergent adverse Event (TEAE)

TEAE are events with on set after the administration of the first dose of any study treatment and within the end of study or 30 days of the last dose of any study treatment, whichever occurs earlier.

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.

6. Analysis Sets

6.1 Full Analysis Set

The full analysis set (ITT population) will include all randomized subjects. All subjects will be analyzed according to treatment to which they are randomized. Full analysis set will be used for the primary and key secondary efficacy endpoints.

6.2 Safety Analysis Set

The safety analysis set will include subjects in the full analysis set who received at least 1 dose of investigational product. For analyses using safety analysis set, subjects will be analyzed based on actual treatment received.

6.3 Per Protocol Set

The per protocol analysis set is a subset of the full analysis set which includes subjects who do not have select important protocol deviations. The list of important protocol deviations is maintained by the sponsor on an ongoing basis and will be finalized before the PA of the study.

6.4 Health-related Quality-of-Life or Health Economics Analyses Set

Health Related Quality of Life or Health Economics Analyses Sets will be specified in the PRO Supplemental Statistical Analysis Plan.

6.5 Pharmacokinetic/Pharmacodynamic Analyses Set

The PK Analysis Set includes all subjects randomized to AMG 510 who have received at least 1 dose of AMG 510 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. The PK Analysis Set will be used to conduct the analysis of PK data, unless otherwise specified.

7. Planned Analyses

7.1 Interim Analysis and Early Stopping Guidelines

An IA of PFS for superiority of AMG 510 over docetaxel is planned when approximately 70% (160 events) of the total PFS events have been observed, or when the enrollment is finished and the Last Subject Enrolled (LSE) have had the opportunity to have 6 weeks of follow up, whichever occurs later. Early efficacy at the proposed PFS interim analysis will be claimed if the observed PFS difference meets the pre-specified statistical significance as well as being considered clinically meaningful. The clinical

meaningfulness will not be addressed in the SAP and will be deferred to the CRMD and GDL.

One administrative interim summary for OS will be performed at the PFS interim analysis if PFS interim result achieves statistical significance.

There is one OS IA planned for superiority of AMG 510 over docetaxel, either at the time when PFS achieves statistical significance or after 175 OS events (~53% maturity) have been observed, whichever is later.

Table 1: Potential Interim and Primary Analysis of PFS and OS

Analysis	PFS Information Fraction	Expected Timing from FSE	Analysis Trigger
PFS IA	70%	13 months	160 PFS events or LSE + 6 weeks, whichever occurs later
PFS PA	100%	19 months	230 PFS events or LSE + 6 weeks, whichever occurs later
OS IA at PFS PA	N.A.	19 months	175 OS events
OS PA	N.A.	22 months	198 OS events

IA = interim analysis; OS = overall survival; PA = primary analysis; PFS = progression-free survival.

The DMC will communicate their assessments on safety and efficacy, and recommendations regarding study modification or termination based on the safety and efficacy parameters to Amgen in accordance with the DMC charter.

Records of all meetings will be maintained by the DMC for the duration of the study. Records of all meetings will be transferred and stored in the TMF (in accordance with SOP-427356) at the conclusion of the study. Further details are provided in the DMC charter.

The DAP will be invoked if the DMC recommends early stopping for efficacy and/or safety. The DAP Team will make the decisions on further conduct of the study and will decide on any further action, which may include, for example, external communication to a regulatory agency, study team unblinding, and/or changes to study conduct. Further details are provided in the DAP.

7.2 Primary Analysis

The timing for the PA of PFS will be event driven and will happen when approximately 230 PFS events are reached cumulatively in 2 treatment groups. If PFS early success is achieved in the IA, the IA will serve the purpose of PA of PFS. The PFS PA may be delayed to ensure that the enrollment is finished and the delayed PA will be triggered when the last randomized subject have had the opportunity to have at least 6 weeks of follow up. The PA of OS will occur when at least 198 OS events (~60% maturity) have been observed, which is expected to be at approximately 3 months after the PFS primary analysis. There will be only one analysis for ORR (no interim analysis for ORR is planned). The primary analysis of ORR will be performed when PFS is claimed statistically significant and the last randomized subject has had the opportunity to have at least 12 weeks of follow up.

7.3 Final Analysis

The final analysis of the study will be performed when the last subject completes LTFU.

8. Data Screening and Acceptance

8.1 General Principles

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

8.2 Data Handling and Electronic Transfer of Data

The Amgen Global Study Operations-Data Management (GSO-DM) department will provide all data to be used in the planned analyses. This study will use the RAVE database.

8.3 Handling of Missing and Incomplete Data

Incomplete adverse event start dates, concomitant medications start or stop dates, and death date will be imputed and the detailed rules will be specified in Appendix A.

No imputation will be conducted for the PA of the primary and key secondary endpoints. The frequency of missing disease assessments and deviation of the actual disease assessment times from the scheduled assessment times will be summarized by treatment groups. Sensitivity analyses will be performed to assess the impact on the analysis of PFS due to any missing data/assessment, and any lost to follow-up or discontinuation of assessment of PFS not due to an event. Similar analysis will be performed for PRO endpoints and will be specified in the PRO supplemental statistical analysis plan.

Details of missing data analysis and imputation rules will be described in the SAP Appendix A.

8.4 Detection of Bias

Lack of protocol compliance and the potential for biased statistical analyses will be examined by assessing the incidence of important protocol deviations (IPD). The clinical study team will identify and document the criteria for IPD.
If applicable, the methods to detect bias are describe in the analyses of particular endpoints.

8.5 Outliers

PK 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.

Additional statistical software may be used to perform exploratory/ad-hoc analyses.

9. Statistical Methods of Analysis

9.1 General Considerations

In principle, summary statistics including mean, standard deviation, median, first and third quartiles, will be provided for continuous variables. Frequency and percentage will be summarized by treatment group for binary and categorical variables. The distribution of PFS and OS will be estimated using the Kaplan-Meier method. The HR and its 95% CI will be estimated using a Cox proportional hazards model stratified by the randomization stratification factors. The inferential comparison will be made using a

stratified log rank test. The odds ratio and difference of proportions of objective response will be calculated and the associated 95% CI will be estimated using the Clopper-Pearson method. The inferential comparison for ORR will be made using the Cochran-Mantel-Haenszel chi-square test controlling for the randomization stratification factors. For subjects who continue treatment post-progression or subjects who crossover from docetaxel to AMG 510, the first date of progression will be used for PFS analysis and subject's response post first progression or post crossover will not be used the primary analyses to evaluate objective response endpoints including PFS, ORR, DOR, TTR, and DCR.

9.2 Subject Accountability

The number and percent of subjects who were screened, randomized, received study treatment, entered long-term follow-up before disease progression and long-term follow-up for survival will be summarized by treatment group. The number and percent of subjects who discontinued study treatment, long-term follow-up before disease progression, and study will be tabulated, along with the reason for discontinuation. The number and percent of subjects randomized will be tabulated by the stratification factors. The number and percent of subjects randomized will be tabulated by study site and country. Key study dates for the first subject randomized, last subject randomized, and data cut-off date for analysis will be presented.

The crossover subjects may be tabulated separately starting from the date of first dose of AMG 510.

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.

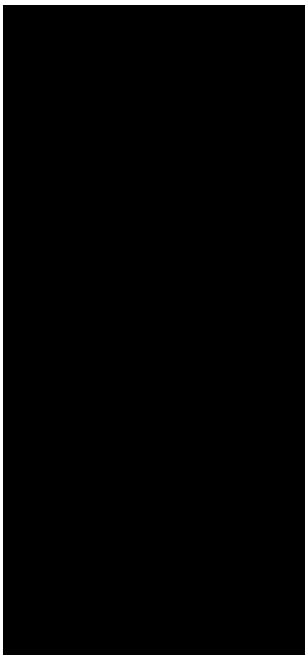
9.4 Demographic and Baseline Characteristics

Demographic and baseline disease characteristics will be summarized by treatment group and overall using descriptive statistics for the Full Analysis Set.

These include, but not limited to the following:

- Baseline demographics and characteristics:
  - Age at randomization
    - As continuous variable

    - As categorical variable: (< 65 vs >= 65)
  - Sex (Male, Female)
  - Race
    - White, Black, Asian and other
    - Asian, non-Asian
  - Ethnicity (Hispanic or Latino vs Not Hispanic or Latino)
  - Weight (kg)
  - Region (North America vs Europe vs Asia Pacific vs Other)
- Baseline disease characteristics:
  - Number of Prior Line of Therapy
    - As continuous variable
    - As categorical variable: 1vs 2 vs > 2.
  - History of CNS Involvement (yes vs no)
  - ECOG Performance Status (0 vs 1)
  - Liver Metastasis at Baseline (yes vs no)
  - Smoking History (never, current vs former)
  - Histology (Squamous vs Non-squamous)
  - Disease Stage (Locally advanced and unresectable vs Metastatic)
  - Best Response on immediate Prior Therapy: (primary refractory (progression on first scan) vs suboptimal response (stable disease) vs recurrent (initial response with subsequent growth))
  - Time from Initial Diagnosis to Randomization
  - Brain metastasis at baseline (yes vs no)
  - Presence of specific co-mutation at baseline: (yes vs no)



[redacted]

9.5 Efficacy Analyses

Estimand	Statistical Analysis Methods	Sensitivity Analysis
HR of PFS between AMG 510 and docetaxel, for subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS p.G12C mutation, before or on start of new anti-cancer therapy	Based on BICR centrally assessed outcomes on the full analysis set: - KM summaries - 1-sided p-value from stratified log-rank test. - Hazard ratio and 95% CI from stratified Cox regression.	- Unstratified analyses: 1-sided p-value from unstratified log-rank test, hazard ratio and 95% CI from unstratified Cox regression. - Per-Protocol subset: same as primary summary and analysis method for Per-Protocol subset. This analysis might be performed only if the per-protocol population is less than 90% of the ITT population. - Initiation of new anti-cancer therapy treated as PFS Event: The data censoring rules are the same as those for the primary analysis of PFS except that the use of new anti-cancer therapy will be treated as an event rather than a mechanism for censoring. The same analysis method as for primary analysis will be used. (see appendix D) - LTFU/Consent withdrew: The data censoring rules are the same as those for the primary analysis of PFS except that subjects who were lost to follow-up or withdrew consent without PD/death are treated as having an event at the next scheduled assessment time in both treatment groups - Scheduled assessment dates: same as primary summary and analysis methods, except the analysis is based on the scheduled assessment dates instead of actual assessment dates.

Estimand	Statistical Analysis Methods	Sensitivity Analysis
HR of OS between AMG 510 and docetaxel, for subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS p.G12C mutation, regardless of subsequent anti-cancer therapy and/or crossover.	Analysis based on the full analysis set: - KM summaries - 1-sided p-value from stratified log-rank test. - Hazard ratio and 95% CI from stratified Cox regression.	- Crossover adjustment: Based on methods such as Rank Preserving Structural Failure Time (RPSFT), Inverse Probability of Censoring Weighting (IPCW), Two-stage approach and other as appropriate - Unstratified analyses: 1-sided p-value from unstratified log-rank test, hazard ratio and 95% CI from unstratified Cox regression. - Per-Protocol subset: same as primary summary and analysis method for Per-Protocol subset. This analysis might be performed only if the per-protocol population is less than 90% of the ITT population. - Initiation of new anti-cancer therapy treated as a mechanism for censoring: The data censoring rules are the same as those for the primary analysis of OS except that the use of new anti-cancer therapy will be treated as an event rather than a mechanism for censoring. The same analysis method as for primary analysis will be used. - Additional OS data collected after the data cutoff date will be continuously reviewed for safety monitoring purpose.
Odds ratio of objective response between AMG 510 and docetaxel, for subjects with previously treated locally advanced and unresectable or metastatic NSCLC with *KRAS p.G12C* mutation. Subjects who discontinued treatment prior to achieving an objective response are	Based on BICR centrally assessed outcomes: - Point estimate of ORR and 95% CI by treatment group using the Clopper Pearson method. - 1-sided p-value from the CMH chi-squire test	- Investigator assessments: Same as primary summary and analysis method based on investigator assessments. - Per-Protocol subset: Same as primary summary and analysis method for Per-Protocol subset. This analysis might be

Estimand	Statistical Analysis Methods	Sensitivity Analysis
considered as non-responders.	controlling for stratification factors. - An estimate of the common odds ratio (95% CI) will be provided as a measure of the relative treatment effect.	performed only if the per-protocol population is less than 90% of the ITT population
Change from baseline to Week 12 in PRO endpoints between AMG 510 and docetaxel, for subjects with previously treated locally advanced and unresectable or metastatic NSCLC with KRAS p.G12C mutation. PRO measurements before or on start of new anti-cancer therapy including crossover will be used to estimate treatment effect.	Change from baseline over time to week 12 in symptoms of dyspnea as measured by a 4 item dyspnea domain from QLQ-C30 and QLQ-LC13, change from baseline over time to week 12 in physical functioning, global health status as measured by QLQ-C30 will be compared between treatment groups using mixed model with repeated measures (MMRM) (Mallinckrodt et al, 2008). The dependent variable of this model will be the change from baseline score over time to week 12. The model will include intercept, time, baseline score, treatment, treatment by time interaction, and randomization stratification factors as fixed effects and subject random intercept and subject random slope of time on score as random effect. Change from baseline over time to week 12 in symptoms of cough and chest pain as measured by a single question from QLQ-LC13 will be compared between treatment groups using generalized estimating equations (GEE) method for cumulative logits model. The model will include intercept, time, baseline score, treatment, treatment by time interaction and randomization stratification factors. Subgroup analysis for key secondary endpoints will be performed in subgroups defined in Section 4.2.	Multiple imputation approach with non-ignorable missing pattern will be explored as the sensitivity analysis for all the key secondary PRO endpoints.

9.5.1 Analyses of Primary Efficacy Endpoint/Estimand

The analyses of PFS will be conducted on the full analysis set, unless otherwise specified. The PA of PFS will be based on BICR centrally assessed outcomes per RECIST v1.1.

The distribution of PFS time including median and quartiles will be summarized descriptively using the Kaplan-Meier method. The corresponding 95% confidence intervals for the median and quartiles will be constructed using the method of Klein and Moeschberger (1997) with log-log transformation. PFS rates at selected landmark time points will be provided and the corresponding 95% confidence intervals will be calculated using the method of Kalbfleisch and Prentice (1980). The duration of the follow-up for PFS will be estimated by reverse Kaplan-Meier method (Schemper and Smith 1996).

The inferential comparison between treatment groups will use the log-rank test stratified by the randomization stratification factors per IXRS at level of 0.025 (1-sided). The HR and its 95% CI will be estimated using a Cox proportional hazards model stratified by the same randomization stratification factors.

The primary endpoint of PFS will be analyzed within each of the subgroups listed in section 4.2. Specifically, to determine whether the treatment effect is consistent across subgroups**,** the estimate of the hazard ratios (with 95% CI) for PFS between the treatment groups will be provided. Additionally, a treatment-by-subgroup interaction test may be provided using a Cox proportional hazards model stratified by the stratification factors.

Piecewise Cox models may be explored given evidence of non-proportional hazards (Collett, 2003). This model will allow estimation of an overall weighted hazard ratio (weights equal to fraction of total events in each interval (Lu & Pajak, 2000)) as well as within interval treatment hazard ratio. Additional analysis may be performed to explore potential sources for non-proportionality by considering baseline prognostic factors and other potential confounding factors.

9.5.2 Analyses of Secondary Efficacy Endpoint(s)

9.5.2.1 Key Secondary Endpoints

The efficacy analyses of key secondary endpoints will be conducted on the full analysis set, unless otherwise specified.

The inferential comparison between treatment groups for ORR will be made using the Cochran Mantel Haenszel chi-square test controlling for the randomization stratification factors. The ORR will be calculated by treatment group and the associated 95% CI will be estimated using the Clopper-Pearson method. An estimate of the common odds ratio (95% CI) will be provided as a measure of the relative treatment effect. The odds ratio (and 95% CI) will be estimated using the Mantel-Haenszel method. The docetaxel group will serve as the reference treatment group in the calculation of the odds ratio. The primary analysis of ORR will be based on BICR centrally assessed outcomes per RECIST v1.1. The analyses based on investigator-assessed will serve as supportive analyses. For subgroups listed in Section 4.2, the inferential comparison between treatment groups for ORR and estimates of ORR with 95% CI will be performed using the same aforementioned methods.

Overall survival will be analyzed using the same method as described for the PFS endpoints. Subgroup analysis for OS will be performed using the same method described for PFS as appropriate. If there is evidence to support non-proportional hazards, a piecewise proportional hazard model may be explored as described for PFS. OS at selected landmark time points will be provided and the corresponding 95% confidence intervals will be calculated.

To account for the potential confounding effect of subjects randomized to docetaxel who subsequently receive AMG 510, additional analyses of OS that adjust for the effect of crossover may be conducted. Methods such as Rank Preserving Structural Failure Time (RPSFT) (Robins et al 1991), Inverse Probability of Censoring Weighting (IPCW) (Robins 1993), Two-stage approach (Latimer 2014) and other may be employed. The decision to adjust the OS estimate and the final selection of the methods to be employed will be based on a review of the data, and the plausibility of the underlying assumptions for each method.

Analysis of OS based on the rank-preserving structural failure time (RPSFT) model (Robins and Tsiatis, 1991) will be conducted to correct the treatment effect estimate for bias introduced by crossover from docetaxel to AMG 510. The RPSFT model provides a randomization-based estimate of treatment effect assuming a multiplicative effect of treatment on time to event. The approach also allows reconstruction of the hypothetical docetaxel time to event curve assuming all subjects initially randomized to docetaxel would not crossover to AMG 510. The RPSFT adjusted HR (95% CI), as well as the treatment effect measured by the acceleration factor (95% CI) will be presented.

Additionally, Kaplan-Meier curves may be used to display the OS between the AMG 510 and the RPSFT model-adjusted docetaxel groups. One of the key assumptions of the RPSFT "common treatment effect", i.e, the treatment effect of AMG 510 is the same regardless of when the subject started taking AMG 510, will be verified.

The Inverse Probability of Censoring Weighting (IPCW) (Robins 1993) method censors subjects who cross over at the time of crossover (i.e., weight=0 during the intervals after crossover and, therefore, dropped from the model), and then up-weights similar subjects who do not crossover, therefore adjusts for bias of informative censoring due to crossover associated with potential time-dependent confounders. Weights for subjects that do not crossover will be estimated by fitting log logistic regression models with crossover as a dependent variable and with baseline and time-dependent covariates, which have a confounding effect on both the likelihood of switching and on the OS, as independent variables. Then the variables. weights will be used in the weighted Cox proportional hazard regression model to calculate the HR and corresponding CI of the relative OS effect of AMG 510 vs docetaxel. This method relies on the assumption that there is no unmeasured confounders.

The two-stage approach (Latimer 2014) adjusts for crossover effect by estimating the relative treatment effect of crossing over on OS by comparing subjects who crossed over from the docetaxel group to the AMG 510 group and subjects originally randomized to the AMG 510 group, by means of an accelerated failure time (AFT) model. A hypothetical time to event outcome is calculated by removing the estimated treatment effect for those subjects who crossed over. Similarly, to the IPCW method, it assumes no unmeasured confounders.

Additional analyses using alternative methods correcting for the confounding effect of crossover may also be performed.
Details regarding analyses of PRO related endpoints will be described in the PRO-related Supplemental Statistical Analysis Plan.

9.5.2.2 Other Secondary Endpoints

Duration of response (DOR) will be calculated only for subjects who achieve a confirmed best overall response of PR or CR. For those who are alive and have not experienced disease progression at the time of data cutoff for analysis, duration of response will be right-censored based on the censoring conventions defined previously for PFS (refer to Appendix D). Progression will be based on BICR centrally assessed outcomes per RECIST v1.1. The distribution of DOR, including the median and quartiles and their

corresponding 95% CIs, will be characterized using the Kaplan-Meier method based on the subjects who achieve a best response of PR or better. No inferential comparison between treatment groups will be made for duration of response.

Time to Response (TTR) will be calculated only for subjects who achieve a confirmed best overall response of PR or CR. TTR will be summarized by the non-missing sample size (n), mean, standard deviation, median, minimum, and maximum for responders by study and regimen (side-by-side) based on FAS.

Disease Control Rate (DCR) will be analyzed using the same method as described for the ORR endpoints.

The analyses of DOR, TTR and DCR are mainly performed based on BICR centrally assessed outcomes. The sensitivity analyses are performed based on the Investigator assessments.

9.5.3 Analyses of Exploratory Efficacy Endpoint(s)

[redacted]

9.6 Safety Analyses

9.6.1 Analyses of Primary Safety Endpoint(s)

There is no primary safety endpoint in Study 20190009. The safety and tolerability of AMG 510 compared to docetaxel will be assessed as a secondary endpoint. Unless otherwise specified, analyses of safety endpoints will be performed in safety analysis set.

If a subject is randomized in docetaxel group, but received at least one dose of AMG 510, even erroneously, this subject will be considered under AMG 510 group, prior to the time of crossover or if this subject did not crossover from docetaxel group to AMG 510 group. Once the subject has crossed over, their safety data should be censored at the time of crossover, i.e. for subjects who are randomized to docetaxel and crossover to AMG 510, AEs occurring after the first dose of AMG 510 will be excluded from the safety summaries for docetaxel and AMG 510 group. Exploratory analyses of select safety

endpoints starting from the date of first dose of AMG 510 will be considered for the crossover subjects separately.

9.6.2 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 24.0 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 adverse events, treatment related adverse events, grade 3 or higher TEAEs, serious adverse events, adverse events leading to withdrawal of investigational product, fatal adverse events, and adverse events of interest (EOI). If a subject experiences repeated episodes of the same AE, then the subject will be counted once within each system organ class and similarly counted once within each preferred term and the event with the highest severity grade and/or strongest causal relationship to each treatment will be used for purposes of incidence tabulations.

Subject incidence of all treatment-emergent adverse events, grade 3 or higher TEAEs, serious adverse events, adverse events leading to withdrawal of investigational product, TEAE leading to dose modification and fatal adverse events will be tabulated by system organ class in alphabetical order and preferred term in descending frequency order.
Subject incidence of all treatment-related treatment-emergent adverse events (TRAE) - grade 3 or higher, serious, adverse events leading to withdrawal of investigational product, AEs leading to dose modification and fatal adverse events will be tabulated by system organ class in alphabetical order and preferred term in descending frequency order. If relationship to treatment is missing for a treatment-emergent event, then event will be assumed as treatment-related.
Subject incidence of events of interest (standardized MedDRA queries and/or Amgen customized queries) will also be summarized according to their categories and preferred term in descending order of frequency. Time to onset and duration of select EOIs may also be summarized.

In addition, summaries of treatment-emergent AEs, grade 3 or higher TEAEs, serious adverse events, adverse events leading to withdrawal of investigational product, and fatal adverse events by preferred term in any treatment group will be provided in descending order of frequency.

Summaries of treatment-emergent and serious adverse events will be tabulated by system organ class, high level term, and preferred term.

Summaries of treatment-emergent and serious adverse events will be tabulated by system organ class, preferred term, and grade. The fatal adverse events will also be provided by system organ class in alphabetical order and preferred term in descending order of frequency.

A summary of the number of deaths and the cause of death, classified by deaths within 30 days of the last dose of study drug and deaths more than 30 days after the last dose, will be provided.

All AEs, including TEAEs, will be included in individual subject listings.

All on study deaths will be listed.

AE summaries for crossover subjects above mentioned all AE summaries may be considered separately starting from the date of first dose of AMG 510.
The severity of each adverse event will be graded using The Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 criteria.

9.6.3 Laboratory Test Results

Selected laboratory data will be summarized using descriptive statistics at each scheduled time point for each treatment group in the study. For continuous parameters, a summary of the changes from baseline to each post dose laboratory assessment will also be produced. For the summary of changes from baseline values, subjects without baseline and/or post-baseline value will be excluded.

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, urinalysis and coagulation. The parameters described in Table 11-1 of the protocol will be collected, converted to Amgen standard units and summarized. 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. A listing of alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, and total bilirubin values at each time point will be produced for the subjects suspected of Hy's law case.

9.6.4 Vital Signs

Vital signs including systolic/diastolic blood pressure, heart rate, respiratory rate, oxygen saturation and body temperature will be summarized by changes from baseline values for each treatment group using descriptive statistics.

9.6.5 Physical Measurements

Summaries of changes from baseline over time may be provided by treatment. For the summary of changes from baseline by visit, subjects without a baseline and/or post baseline value will be excluded.

9.6.6 Antibody Formation

Not applicable.

9.6.7 Exposure to Investigational Product

Drug exposure (for AMG 510, docetaxel respectively) including duration and intensity will be summarized descriptively for each treatment group.

Descriptive statistics will be produced to describe the exposure to investigational product by treatment group. The number of cycles of protocol-specified therapy administered will be summarized with an additional breakdown of the number of cycles started. In addition, the duration of therapy, the cumulative dose, and the average dose per administration and relative dose intensity will be summarized for each drug. The number and percent of subjects with dose modifications (eg, dose change/withheld, dose interruptions) and reason for modification will be summarized for both treatment groups.

Exposure for subjects who crossover may be considered separately starting from the date of first dose of AMG 510.

List of manufacturing lot number will be provided.

9.6.8 Exposure to Concomitant Medication

Number and proportion of subjects receiving therapies of interest will be summarized by preferred term for each treatment group as coded by the World Health Organization Drug (WHO DRUG) dictionary. In addition, the number and proportion of subjects receiving anti-cancer therapies while on study will be summarized by anti-cancer therapy category for each treatment group in the Full Analysis Set. The subject incidence and time to first use of subsequent anticancer therapies will be summarized.

9.7 Other Analyses

Other analyses in the study include analyses for PK, PRO, and biomarker endpoints, which may be specified in the separate analysis plans.

9.7.1 Analyses of Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Endpoints

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.

Individual concentration-time data will be tabulated and presented graphically. Summary of PK concentration over time and PK parameters will be provided. Mean concentration-time profiles for each dose will be provided.

PK parameters will include, but are not limited to, maximum observed concentration (Cmax), time to maximum concentration (tmax) and area under the plasma concentration-time curve (AUC). Other PK parameters such as AUC from time 0 to the time extrapolated to infinity (AUCinf), apparent clearance (CL/F), and terminal half-life (t1/2) may be analyzed.

For AMG 510 **and its relevant metabolites**, pharmacokinetic parameters will be determined from the time concentration profile using standard non-compartmental approaches based on the PK Analysis Set. PK parameters will be summarized using descriptive statistics including, but not limited to means, standard deviations, medians, minimums, and maximums.

Based on the review of the data, analyses to describe the relationship between AMG 510 exposure and either pharmacodynamic effect and/or clinical outcome may also be performed.

If performed, details regarding objectives, data handling, and methodology pertaining to any modeling activities will be provided in a separate population modeling analysis plan and/or a separate exposure-response analysis plan.

9.7.2 Analyses of Clinical Outcome Assessments

Details regarding analyses of PRO related endpoints will be specified in PRO Supplemental Statistical Analysis Plan.

9.7.3 Analyses of Health Economic Endpoints

Not Applicable.

[redacted]

9.7.5 Analyses of COVID-19 Impact

The following summaries to assess the impact of COVID-19 may be provided:

- Listing of all subjects impacted by COVID-19 related study disruptions
- Subject incidence of IPD and non-important protocol deviation due to COVID-19 control measures by protocol deviation category, which may include:
  - alternative IP administration process
  - alternative lab/imaging data process
  - alternative site visits
  - alternative procedures or methods not included in original study design and not identified
  - missed / partial missed visits
  - missed IP/comparator/other protocol specified treatment
  - early end of study
  - early end of treatment with AMG 510 or docetaxel
  - early end of safety follow-up
- Subject incidence of COVID-19 related dose change / withheld

10. Changes From Protocol-specified Analyses

There are no changes to the protocol-specified analyses.

11. Literature Citations / References

Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response, Jon Clinical Oncology J Clin Oncol 1983;1:710-719

Charpidou A, Vathiotis I, Gkiozos I, et al. Exceptional antitumor responses beyond immune checkpoint inhibition in non-small cell lung cancer patients: insights into optimal therapy sequencing. J Thorac Dis 2019; 11(3):E25-E31.

Clopper CJ and Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika. 1934; 26(4):404-413.

Collett D. Modelling Survival Data in Medical Research. 2nd edition. London, UK: Chapman & Hall/CRC; 2003.

Hryniuk W, Goodyear, M. The calculation of received dose intensity. Journal of Clinical Oncology 8:1935–1937, 1990

Kalbfleisch, J. D. and Prentice, R. L. The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons; 1980

Klein, J. P. and Moeschberger, M. L. (1997), Survival Analysis: Techniques for Censored and Truncated Data, New York: Springer-Verlag.Longo D, Duffey P, DeVita V, Wesley M, Hubbard S, Young R. The calculation of actual or received dose intensity: A comparison of published methods. Journal of Clinical Oncology 9:2042–2051, 1991

Korhonen, P., E. Zuber, M. Branson, N. Hollaender, N. Yateman, T. Katiskalahti, D.Lebwohl and T. Haas. (2012). Correcting Overall Survival for the Impact of Cro ssover Via a Rank-Preserving Structural Failure Time (RPSFT) Model in the RECORD-1 Trial of Everolimus in Metastatic Renal-Cell Carcinoma. Journal of Biopharmaceutical Statistics 22:1258-1271.

Latimer, N., et al. (2014) Adjusting Survival Time Estimates to Account for Treatment Switching in Randomized Controlled Trials—an Economic Evaluation Context: Methods, Limitations, and Recommendations. Medical Decision Making 34(3):387-402.

Lu, J., Pajak, T. F. Statistical power for a long-term survival trial with a time-dependent treatment effect. Control Clin Trials. 2000 Dec; 21(6): 561–573.

Mallinckrodt, C.H., Lane, P.W., Schnell, D., Peng Y, and Mancuso, J.P . Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Information Journal, 2008; 42: 303-319.

Maurer, W., Bretz, F. Multiple Testing in Group Sequential Trials Using Graphical Approaches. Statistics in Biopharmaceutical Research. 2013; 5(4): 311-320.

Mosteller RD. Simplified calculation of body surface area. N Engl J Med 1987;317(17):1098 (letter).

Robins, J. and A.Tsiatis. (1991). Correcting for Non-Compliance in Randomized Trials Using Rank Preserving Structural Failure Time Models. Communication in Statistics – Theory and Methods 20(8):2609-2631.

Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In Proceedings of the Biopharmaceutical Section, American Statistical Association, 24-33.

Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Controlled Clinical Trials 17:343–346, 1996

Simon R, Makuch RW. A non-parametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias. Statistics in Medicine 3: 35-44, 1984

US Food & Drug Administration: Guidance for Industry. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. www.fda.gov/cder/guidance/7478fnl.htm

Zhou L, Baba Y, Kitano Y, et al. KRAS, BRAF, and PIK3CA mutations, and patient prognosis in 126 pancreatic cancers: pyrosequencing technology and literature review. Med Oncol. 2016;33(4):32.

12. Prioritization of Analyses

Not Applicable.

13. Data Not Covered by This Plan

Details regarding analyses of PRO endpoints will be specified in the PRO related Supplemental Statistical Analysis Plan.

Pharmacokinetic, pharmacodynamic, exposure-response and biomarker analyses will be performed by Clinical Pharmacology Modeling and Simulation (CPMS) or biomarker group.

14. Appendices

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

Below imputation rules will be used to impute start date and stop date of AE and Concomitant Medication.

Date of prior**/post-treatment** anti-cancer therapy, PD-L1 and metastasis will be imputed using the same rule when only the date is missing (no imputation when month or year is missing).

Imputation Rules for Partial or Missing Stop Dates for AE and Concomitant Medications

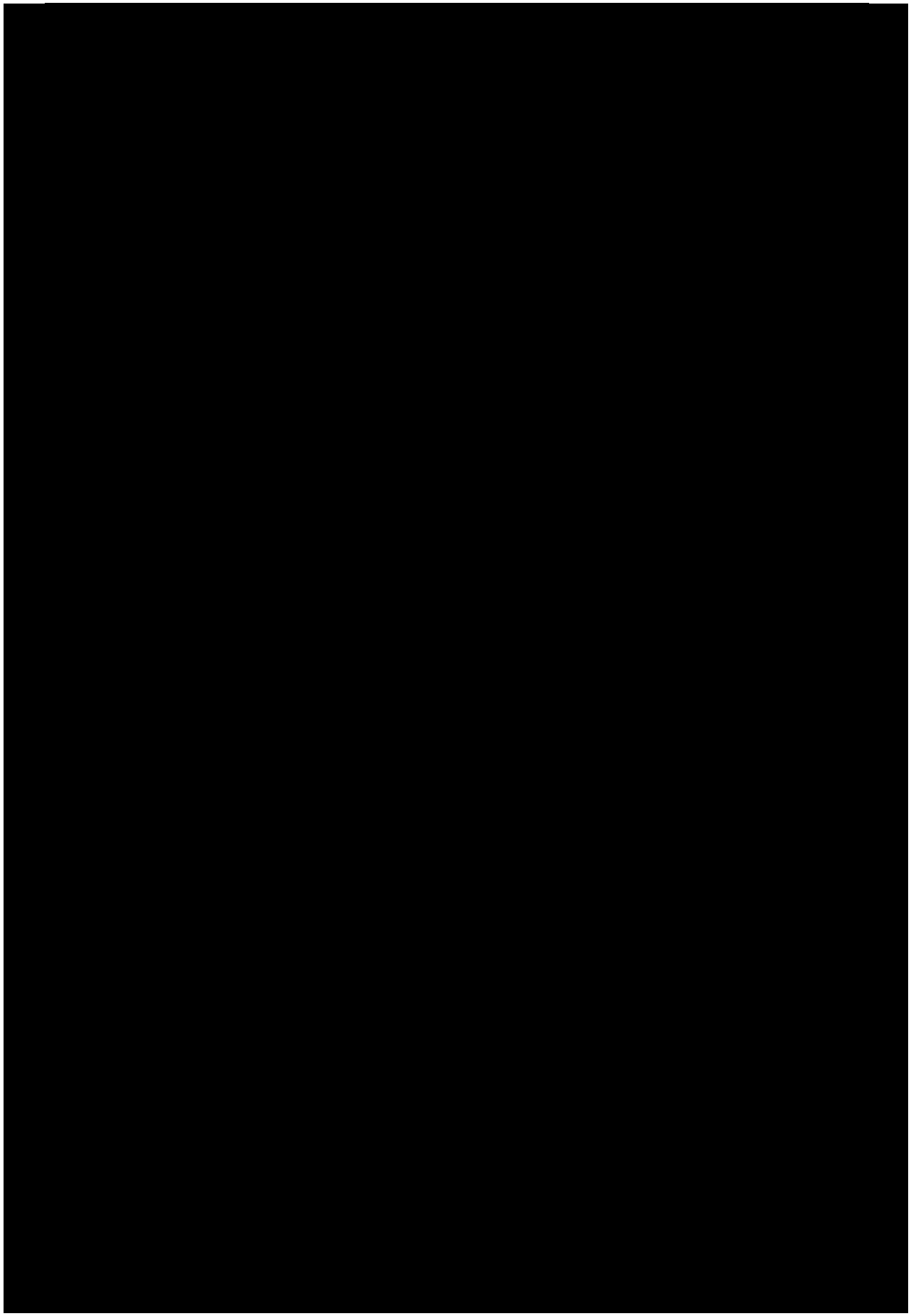
If the month and year are present, impute the last day of the month. If only the year is present, impute December 31 of that year. If the stop date is entirely missing, assume the event or medication is ongoing. If a partial or complete stop date is present and the 'ongoing' or 'continuing' box is checked,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

If imputed stop date is after end of study (EOS)/Death date then EOS/Death date will be used as stop date.

Imputation Rules for Partial or Missing Start Date for AE and Concomitant Medications

Missing	Imputation	Exception
Day	01	Default to Study Day 1 if an adverse event starts the same year and month as Study Day 1 and the flag indicates that the adverse event started on or after the first dose on the Adverse Events eCRF
Day/Month	01 JAN	Default to Study Day 1 if an adverse event started the same year as Study Day 1 and the flag indicates that the adverse event started on or after the first dose on the Adverse Events eCRF
Day/Month/Year	1. If complete stop date present a. Stop date < first dosing date: Impute January 1 of the stop year b. Stop date ≥ first dosing date: Impute the date of first dose	

	2. If partial stop date present ie, YYYYMM a. Partial Stop date < first dosing date: Impute January 1 of the stop year b. Partial Stop date ≥ first dosing date: Impute the date of first dose 3. If partial stop date present ie, YYYY a. Partial Stop date < first dosing date: Impute January 1 of the stop year b. Partial Stop date ≥ first dosing date: Impute the date of first dose 4. If stop date is completely missing: Impute the date of first dose	



Appendix D. PFS Calculation and Censoring Rule

Situation up to data cut off or EOS	Primary Analysis		Sensitivity Analysis – Initiation of new anti-cancer therapy treated as PFS Event	
	Date of Event or Censor	Out come	Date of Event or Censor	Out come
Radiological disease progression by BICR (or by investigator) prior to death	Date of first observation of radiological disease progression by BICR (or by investigator)	Event		
No radiological disease progression by BICR (or by investigator), but death record	Date of death	Event		
No evaluable post-baseline tumor assessments by BICR (or by investigator), no death recorded	Date of randomization date	Censor		
No radiological disease progression by BICR (or by investigator), no death, but start of new anti-cancer therapy recorded	Date of last evaluable assessment by BICR (or by investigator) before start of new anti-cancer therapy	Censor	Date of start of new anti-cancer therapy	Event
No radiological disease progression by BICR (or by investigator), no death recorded, no	Date of last evaluable assessment by BICR (or by investigator)	Censor	If on study IP, date of last evaluable assessment by BICR (or by investigator)	Censor

start of new anti-cancer therapy			if not on study IP, end of treatment date	Event
Death or radiological disease progression by BICR (or by investigator) immediately after more than one missed tumor assessment	Date of last evaluable assessment by BICR (or by investigator) with documented non-progression prior to missing assessment(s)*	Censor		

*This supersedes the previous rules that result in PFS event at the date of first observation of radiological disease progression by BICR (or by investigator) or death.

Appendix E. Lab Parameters for analysis

List of Lab Parameters

Local Laboratory/Local/Central[f] Laboratory Urinalysis			**Central Laboratory**
Chemistry	**Hematology**	**Coagulation**	**PK and Biomarker (If allowable under local regulations and if allowable by local**
Sodium	Hemoglobin	PT **and** INR	PK sampling
Potassium	Hematocrit	aPTT	Plasma ctDNA
Chloride	Mean corpuscular volume		Plasma cell pellet
Bicarbonate[e]	Platelets		Serum
Total CO_2[e]	RBC		PB Paxgene RNA
Triglycerides	WBC		Tumor biopsy[d]
Cholesterol	ANC		
Total protein	White blood cell Differential		
Albumin	• Total neutrophils		
Calcium	• Eosinophils		
Magnesium	• Basophils		
Phosphorous	• Lymphocytes		
Glucose	• Monocytes		
Blood urea nitrogen			
Urea[a]	**Serology[c]**	**Urinalysis[f]**	
Creatinine	HepBsAg	Specific gravity	
Creatinine clearance	HepCAb	pH	
Total creatine kinase	(if above cannot be obtained hepatitis viral load can be utilized)	Blood	
Total bilirubin		protein	
Direct bilirubin		glucose	
Alkaline phosphatase		bilirubin	
Alanine aminotransferase	Thyroid Function Tests	ketones	
Aspartate aminotransferase	• TSH	leukocyte esterase	
Other Labs	• Total T3 (or Free T3 per local standard)		
Serum or Urine Pregnancy[b]	• Free T4		

ANC = absolute neutrophil count; aPTT = activated partial thromboplastin time; BUN = blood urea nitrogen; ctDNA = circulating tumor DNA; FT4 = free thyroxine; HepBsAg = hepatitis B surface antigen; HepCAb = hepatitis C antibody; INR = international normalized ratio; PB = peripheral blood; PCR = polymerase chain reaction; PK = phgroupacokinetic; PT = prothrombin time; RBC =

red blood cell; T3 = triiodothyronine; TCO2 = total carbon dioxide; TSH = thyroid stimulating hormone; WBC = white blood cell

[a] Urea collection is acceptable in absence of BUN.

[b] these data are collected separately from the chemistry form.

[c] Hepatitis B surface antigen, hepatitis C antibody, PCR for Hepatitis C RNA (if Hepatitis C antibody is positive).

[d] Archived tumor tissue is acceptable for screening for KRAS G12C mutation status.

[e] Bicarbonate/TCO_2 can also be obtained via capillary testing

[f] Investigative sites without local capability to perform urine studies may submit samples to central laboratory.