Statistical Analysis Plan

Protocol Title:	A Phase 3 Multicenter, Randomized, Open-label, Active-controlled Study of Sotorasib and Panitumumab Versus Investigator's Choice (Trifluridine and Tipiracil, or Regorafenib) for the Treatment of Previously Treated Metastatic Colorectal Cancer Subjects with KRAS p.G12C Mutation
Short Protocol Title:	Sotorasib and Panitumumab Versus Investigator's Choice for Subjects with *KRAS p.G12C* Mutation
Protocol Number:	20190172
NCT Number:	NCT05198934
Authors:	[redacted], Study Statistician, Parexel [redacted], Biostatistics Sr Manager, GBS, Amgen Inc.
Sponsor:	Amgen, Inc. One Amgen Center Drive Thousand Oaks, CA 91320, USA
SAP Date:	**Document Version** **Date** **Amendement 2 (V3.0)** **07 December 2023**

Version Number	Date (DDMMMYYYY)	Summary of Changes, including rationale for changes
Original (v1.0)	14OCT2021	The first version of a document.
Amendment [1 (v2.0)]	24MAY2023	- To address protocol amendment 3, key changes are as follows: • Update timing of OS final analysis to 50% maturity instead of 3 months after PFS primary analysis. • Correct hypotheses statement to reflect 2-sided testing. - Update subgroups - Add definition for BPI pain severity index, pain interference index, and global BFI score. Updating analysis of BPI and BFI patient-report outcome to use MMRM instead of GEE - Add details for imputation rule when there is (partial) missing date (Appendix A) - Add PFS censoring rule for sensitivity analysis (Appendix D)
Amendment [2 (v3.0)]	**07DEC2023**	**- Address protocol amendment 4, key change for DMC termination after the PFS primary analysis (section 3.1)** **- Update definition of TEAE to account for crossover treatment (section 5)** **- Add sensitivity analysis for overall survival for crossover (section 9.5.2.1)** **- Add AE summaries excluding events that are directly related to disease progression (section 9.6.2)** **- Add AE summaries for crossed over subjects (section 9.6.2)**

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation or Term	Definition/Explanation
AE	Adverse Event
AFT	**Accelerated Failure Time**
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
AUC	Area Under the Plasma Concentration-Time Curve
BFI - SF	Brief Fatigue Inventory - Short Form
BICR	Blinded Independent Central Review
BID	Twice Daily
BIL	Bilirubin
BOR	Best Overall Response
BPI - SF	Brief Pain Inventory – Short Form
BSA	Body Surface Area
CIs	Confident Intervals
CPMS	Clinical Pharmacology Modeling and Simulation
COVID-19	Coronavirus Disease 2019
CR	Complete Response
CRC	Colorectal Cancer
CRF	Case Report Form
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DCR	Disease Control Rate
DMC	Data Monitoring Committee
DNA	Deoxyribonucleic Acid
DOR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
ED-5D-5L	Euroqol-5D Level 5
EORTC	European Organisation for Research and Treatment of Cancer Core
EOT	End of Treatment
FACT-G	Functional Assessment of Cancer Therapy – General
GP5	The single item “I am bothered by side effects of treatment,” rated on a 5-point Likert scale and part of the FACT-G

Abbreviation or Term	Definition/Explanation
HR	Hazard Ratio
IP	Investigational Product
IPCW	**Inverse Probability of Censoring Weighting**
IPD	Important Protocol Deviations
IRT	Interactive Response Technology
IV	Intravenous
KRAS	Kirsten Rat Sarcoma
LTFU	Long-Term Follow-Up
MAR	Missing at Random
MCAR	Missing Completely at Random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Effect Model for Repeated Measures
MRI	Magnetic Resonance Imaging
NCI	National Cancer Institute
NCT	National Clinical Trial
NE	Not Evaluable
ORR	Objective Response Rate
OS	Overall Survival
PA	Primary Analysis
PD	Progressive Disease
PD-1	Programmed Cell Death Protein-1
PFS	Progression-Free Survival
PGIC	Patient Global Impression Of Change
PK	Pharmacokinetic(S)
PR	Partial Response
PRO	Patient Reported Outcome
Q2W	Every 2 Weeks
QD	Once Daily
QLQ-C30	Quality of Life Questionnaire – Core Questionnaire
QOL	Quality of Life
RDI	Relative Dose Intensity
RECIST	Response Evaluation Criteria in Solid Tumors
RPSFT	**Rank Preserving Structural Failure Time**
RNA	Ribonucleic Acid
SAP	Statistical Analysis Plan

Abbreviation or Term	Definition/Explanation
SD	Stable Disease
SFU	Safety Follow-Up
TBL	Total Bilirubin
TEAE	Treatment-Emergent Adverse Event
TRAE	Treatment-Related Adverse Event
TTR	Time to Response
VAS	Visual Analog Scale
WHO	World Health Organization

1. Introduction

The purpose of this Statistical Analysis Plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol for study 20190172, Sotorasib (AMG 510) and Panitumumab dated **21 September 2023**. The scope of this plan includes the primary analysis and the final analysis that are planned and will be executed by the Amgen Global Biostatistical Science department unless otherwise specified.

2. Objectives, Endpoints and Hypotheses

2.1 Objectives and Endpoints/Estimands

Objectives	Endpoints
Primary	
• To compare progression-free survival (PFS) in previously treated subjects with *KRAS p.G12C* mutated colorectal cancer (CRC) receiving sotorasib 240 mg once daily (QD) and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib), and sotorasib 960 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib)	• PFS – defined as time from randomization until disease progression or death from any cause, whichever occurs first, for all subjects. Progression unless noted otherwise, use Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1) per Blinded Independent Central Review (BICR)
Key Secondary	
• To compare overall survival (OS) in previously treated subjects with *KRAS p.G12C* mutated CRC receiving sotorasib 240 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib), and sotorasib 960 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib)	• Overall survival - defined as time from randomization until death from any cause
• To evaluate efficacy of sotorasib 240 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib), and sotorasib 960 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib), as assessed by: • Objective response rate (ORR)	• Objective response = complete response (CR) + partial response (PR), assessed per RECIST 1.1. Response will be assessed by BICR. CR and PR require confirmatory repeat assessment at least 4 weeks after the first detection of response.

Secondary	
• To evaluate efficacy of sotorasib 240 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib), and sotorasib 960 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib), as assessed by: • Duration of response (DOR) • Time to response (TTR) • Disease control rate (DCR) • Investigator assessed ORR and PFS	• Duration of overall 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 BICR per RECIST 1.1 • Time to response – defined as time from randomization to first evidence of PR or CR based on BICR • Disease control defined as CR + PR + stable disease of at least 7 weeks measured based on BICR • PFS and ORR based on investigator assessment per RECIST 1.1
• To evaluate the safety and tolerability of sotorasib 240 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib), and sotorasib 960 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib)	• Incidence and severity of treatment-emergent adverse events, changes in vital signs, and changes in clinical laboratory tests
• To evaluate the impact of sotorasib 240 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib) and sotorasib 960 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib) on patient-reported fatigue, pain, physical function, and global health status (patient reported outcomes [PRO])	• Change from baseline over time to week 8 ○ Fatigue severity as measured by item 3 of the Brief Fatigue Inventory (BFI) ○ Pain severity as measured by item 3 of the Brief Pain Inventory (BPI) ○ Physical functioning as measured by the physical function domain of the European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire-Core Questionnaire (EORTC QLQ-C30) ○ Global health status as measured by Questions 29 and 30 of the EORTC QLQ-C30
• To evaluate and compare the effect of treatment with sotorasib 240 mg and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib), and sotorasib 960 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib) on other treatment and disease related symptoms and health related quality	• Change from baseline over time to 8 weeks for all subscales of the Brief Fatigue and Pain Inventory and subscales and domains of EORTC QLQ-C30 • Summary scores at each assessment and changes from baseline of visual analog scale (VAS) scores as measured by EuroQol-5D level 5 (EQ-5D-5L)

of life relative to investigator's choice (trifluridine and tipiracil or regorafenib)	• Summary scores on single question on symptom bother GP5 from Functional Assessment of Cancer Therapy – General (FACT-G) • Summary scores of Patient Global Impression of change at each protocol specified timepoint
• To characterize the pharmacokinetics (PK) of sotorasib and panitumumab	• Pharmacokinetic (PK) parameters of sotorasib and panitumumab including, but not limited to, maximum plasma concentration (C_{max}), area under the plasma concentration-time curve (AUC)

Exploratory	
• To evaluate the association of biomarker levels and clinical activity	[redacted] • Clinical response including but not limited to OS, PFS, DOR, objective response
• To evaluate the time to deterioration in fatigue and pain between treatment with sotorasib 240 mg QD and panitumumab relative to investigator's choice (trifluridine and tipiracil or regorafenib), and sotorasib 960 mg QD and panitumumab relative to investigator's choice (trifluridine and tipiracil or regorafenib)	• Time to deterioration in severity of pain and fatigue for the Brief Fatigue and Pain Inventory Subscales and all subscales in the EORTC QLQ-C30

2.2 Hypotheses and/or Estimations

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

Primary endpoint:

- H_{10}: PFS distribution of the 960 mg sotorasib and panitumumab arm is the same as the investigator's choice (trifluridine and tipiracil or regorafenib) arm.
- H_{11}: PFS distribution of the 960 mg sotorasib and panitumumab arm is different from the investigator's choice (trifluridine and tipiracil or regorafenib) arm.
- H_{20}: PFS distribution of the 240 mg sotorasib and panitumumab arm is the same as the investigator's choice (trifluridine and tipiracil or regorafenib) arm.
- H_{21}: PFS distribution of the 240 mg sotorasib and panitumumab arm is different from the investigator's choice (trifluridine and tipiracil or regorafenib) arm.

Key secondary endpoints:

- H_{30}: OS distribution of the 960 mg sotorasib and panitumumab arm is the same as the investigator's choice (trifluridine and tipiracil or regorafenib) arm.
- H_{31}: OS distribution of the 960 mg sotorasib and panitumumab arm is different from the investigator's choice (trifluridine and tipiracil or regorafenib) arm.
- H_{40}: OS distribution of the 240 mg sotorasib and panitumumab arm is the same as the investigator's choice (trifluridine and tipiracil or regorafenib) arm.
- H_{41}: OS distribution of the 240 mg sotorasib and panitumumab arm is different from the investigator's choice (trifluridine and tipiracil or regorafenib) arm.
- H_{50}: difference of proportions of objective response between the 960 mg sotorasib and panitumumab arm and the investigator's choice (trifluridine and tipiracil or regorafenib) arm = 0.
- H_{51}: difference of proportions of objective response between the 960 mg sotorasib and panitumumab arm and the investigator's choice (trifluridine and tipiracil or regorafenib) arm is not 0.
- H_{60}: difference of proportions of objective response between the 240 mg sotorasib and panitumumab arm and the investigator's choice (trifluridine and tipiracil or regorafenib) arm = 0.
- H_{61}: difference of proportions of objective response between the 240 mg sotorasib and panitumumab arm and the investigator's choice (trifluridine and tipiracil or regorafenib) arm is not 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 2.2-1. Maurer-Bretz Multiple Testing Procedure for PFS and OS

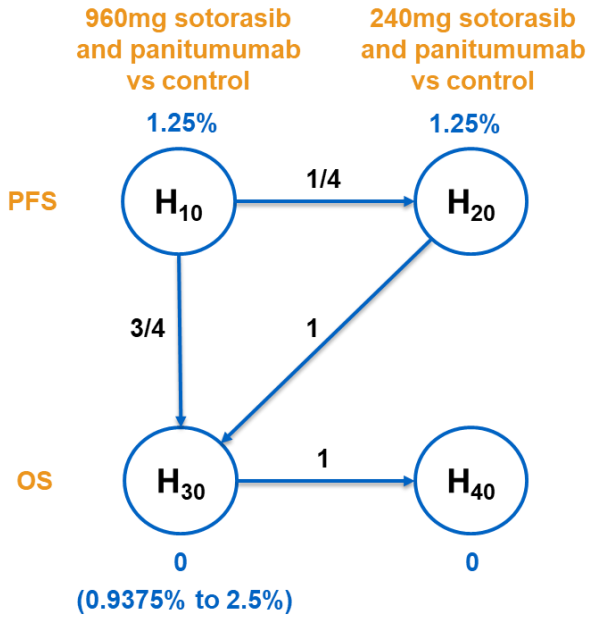


Figure 2.2-1 illustrates the Maurer-Bretz multiple testing procedure among PFS and OS at 960 mg and 240 mg dose levels vs control.

Multiplicity adjustment strategy will be performed when 960 mg dose level vs control and 240 mg dose level vs control are tested simultaneously, as well as when PFS and OS are tested by the hierarchical order in sequence within the same dose level vs control. The fractions on the directed arrows indicate the proportion of α propagated from one hypothesis to the other when its hypothesis is rejected.

One-sided alpha of 2.5% is split in half to 1.25% for both of the 960 mg and 240 mg dose levels. Starting with testing H_{10} of PFS at 960 mg dose level at the 1-sided α/2 (1.25%) level, if the hypothesis H_{10} is rejected, H_{20} of PFS at 240 mg dose level will be tested using one-sided α/2 + α/8 = 5α/8 (1.5625%) level; otherwise at the 1-sided α/2 (1.25%) level. If both hypotheses H_{10} of PFS at 960 mg dose level and H_{20} of PFS at 240 mg dose level are rejected, H_{30} of OS at 960 mg dose level will be tested using one-sided 3α/8 + 5α/8 = α (2.5%) level; otherwise, if only H_{10} of PFS at 960 mg dose level is rejected, H_{30} of OS at 960 mg dose level will be tested using one-sided 3α/8 (0.9375%) level. H_{40} of OS at 240 mg dose level will be tested at the same α level as used for H_{30} if and only if H_{30} is rejected.

If H_{30} of OS at 960 mg dose level is rejected, and H_{20} and H_{40} are also rejected, H_{50} of ORR at 960 mg dose level will be tested using one-sided α (2.5%) level, followed by

H_{60} of ORR at 240 mg dose level using one-sided α (2.5%) level. If H30 is rejected, but H_{20} is not rejected (and thus H_{40} and H_{60} will not be tested), H_{50} will be tested at one-sided 0.9375% level. If H_{30} and H_{20} are rejected, but H_{40} is not rejected, all alpha is used and neither ORR hypotheses will be tested.

If the primary endpoint is met for both 960 mg and 240 mg doses, the final dose selection for the combination will be established based on the totality of data with respect to the efficacy, safety, and clinical pharmacology endpoints.

3. Study Overview

3.1 Study Design

This is a phase 3, multicenter, randomized, open label, active-controlled study to evaluate efficacy and safety of 2 different doses of sotorasib and panitumumab versus investigator's choice (trifluridine and tipiracil, or regorafenib) in previously treated metastatic CRC subjects with *KRAS p.G12C* mutation. The study will be conducted at approximately 100 sites. The study will consist of a screening period, a treatment period, a safety follow-up (SFU) and long-term follow-up (LTFU) period. Approximately 153 subjects with previously treated metastatic CRC with the *KRAS p.G12C* mutation will be enrolled and randomized 1:1:1 to receive either sotorasib 240 mg daily (QD) and panitumumab or sotorasib 960 mg QD and panitumumab or investigator's choice (trifluridine and tipiracil, or regorafenib). Investigator's choice will need to be declared prior to randomization.

Subjects will be stratified by prior anti-angiogenic therapy (yes vs no), time from initial diagnosis of metastatic disease to randomization (≥ 18 months vs < 18 months), and ECOG status (0 or 1 vs 2).

Cycle 1 day 1 will be defined as the first day subject receives study medication. Tumor assessment will be conducted (by magnetic resonance imaging [MRI] and/or computed tomography [CT]) at baseline, at 8-week (± 7 days) intervals until blinded independent central review (BICR) assessed progression, start of another anticancer therapy, withdrawal of consent, lost to follow-up, or death, whichever occurs earliest. Tumor assessment and response will be determined by BICR using Response Evaluation Criteria in Solid Tumors (RECIST) v1.1. Subjects on investigator's choice may be allowed to cross over after the primary analysis of PFS provided the PFS shows a clinically and statistically significant improvement in PFS of the sotorasib and panitumumab arm(s) over the investigator's choice arm and the totality of the safety and efficacy data strongly favors the sotorasib and panitumumab arm(s). For subjects who

previously stopped investigator's choice at progressive disease, as determined by BICR, they may be offered cross over to sotorasib and panitumumab if Amgen determines that crossover is appropriate after the primary analysis. Subjects who discontinued therapy prior to BICR determined progressive disease will not be allowed to cross over unless they subsequently have BICR determined progressive disease before starting another systemic therapy. Subjects who withdrew consent will not be allowed to crossover. Subjects may discontinue treatment because of disease progression, intolerance of treatment leading to treatment discontinuation, initiation of another anticancer therapy, or withdrawal of consent. Continued study treatment after radiological progression may be allowed if in the opinion of the investigator, the subject is deriving clinical benefit, is clinically stable, has no unacceptable toxicity from study drug, agrees to biopsy, and approval of the Amgen Medical Monitor is obtained. For subjects who continue treatment post-progression, tumor assessments will continue as per schedule until subject stops all study drug.

For subjects who continue treatment post-progression or crossover, the date of first BICR assessed progression will be used for the primary PFS analysis and subject's tumor assessments post first progression will not be used for the primary analysis to evaluate objective response endpoints.

In all subjects, information regarding the type and duration of subsequent therapies following disease progression, 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 (e.g., due to unacceptable toxicity) will continue to be followed radiographically until disease progression, withdrawal of consent, or start of another anticancer therapy, and then further LTFU via phone or visit to clinic for assessment of survival and documentation of anticancer treatment for subjects who have not withdrawn consent. Subjects will be followed for up to 2 years from last subject enrolled, or until withdrawal of consent, loss to follow-up, or subject death, whichever occurs first.

An independent (external to Amgen) data monitoring committee (DMC) is planned for this study to review safety data as per DMC charter. Interim safety analysis will be conducted by DMC after approximately 75 subjects have been enrolled and have had the opportunity to complete at least 8 weeks of study treatment, and then at approximately 6-month intervals until last subject is off treatment and then yearly until end of the study. **The DMC reviews and meetings may terminate before the last**

subject has completed treatment and the end of the study has been achieved if the primary analysis of PFS has occurred and DMC termination was made at the recommendation of the DMC.

3.2 Sample Size

The planned enrollment is approximately 153 subjects with metastatic CRC with *KRAS p.*G12C mutation. Subjects will be randomized 1:1:1 to receive either sotorasib 240 mg QD and panitumumab or sotorasib 960 mg QD and panitumumab or investigator's choice (trifluridine and tipiracil or regorafenib). Subjects will be stratified by prior anti-angiogenic therapy (yes vs no), time from initial diagnosis of metastatic disease to randomization (≥ 18 months vs < 18 months), and ECOG status (0 or 1 vs 2).

There will be one planned PFS primary analysis for superiority of PFS between the sotorasib and panitumumab arm and the investigator's choice arm. The primary analysis of PFS will occur when approximately 60 PFS events have been observed from the sotorasib 960 mg and panitumumab arm and the investigator's choice arm. It is anticipated that approximately 90 PFS events (~60% maturity) will be observed from all three arms at the PFS primary analysis. With 90 PFS events, the study will have 90% power to demonstrate superiority with respect to PFS in combination of sotorasib and panitumumab group than the control group at the significance level of half of one-sided alpha of 2.5% at 1.25%, if the true treatment effect HR is assumed 0.4 for either of the sotorasib and panitumumab arms versus the investigator's choice arm. Assuming an enrollment rate of 20 subjects per month after a 3-month ramp-up period, with a total sample size of 153, it is estimated that approximately 11 months will be required to reach 90 PFS events (~60% maturity). This estimation is based on a median PFS of 2 months for the investigator's choice (trifluridine and tipiracil or regorafenib) arm and 5 months for either of the sotorasib and panitumumab arms, and a 10% dropout rate.

There will be 2 planned OS analyses for superiority of OS between the sotorasib and panitumumab arm and the investigator's choice arm. In the case that PFS is claimed statistically significant at primary analysis for the combination treatment group of sotorasib and panitumumab versus the control group at either 960 or 240 mg dose level, OS will be evaluated at this dose level. With 34 OS events anticipated at the PFS primary analysis, the study has ~90% probability to observe an HR < 1 when the true OS HR is 0.65. A nominal alpha of 0.01% (negligible impact on overall type 1 error rate) will be spent on this OS interim summary. Subsequently, the OS final analysis will occur when approximately 50% of subjects had OS events (approximately 77 OS events total

from all three arms), which is expected to be at approximately 6 months after the PFS primary analysis. At 77 OS events total (approximately 51 events in 2 arms), the minimum detectable difference for OS success is an HR of 0.574 (with one-sided 2.5% alpha). The probability to observe an HR < 1 and HR < 0.9 are approximately 92% and 87%, respectively, when the true OS HR is 0.65. The estimation is based on a median OS of 7.1 months for the investigator's choice (trifluridine and tipiracil or regorafenib) arm and 10.9 months for either of the sotorasib and panitumumab arms, and a 10% dropout rate.

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, prior anti-angiogenic therapy (yes vs no), time from initial diagnosis of metastatic disease to randomization (≥ 18 months vs < 18 months), and ECOG status (0 or 1 vs 2), the following additional covariates may be included:

- Region (North America vs Europe vs Asia vs rest of world)
- Age (< 65 vs ≥ 65)
- Gender (male vs female)
- Race (Caucasian, black, Asian, other)
- Presence of specific co-mutation (yes vs no)
- Differentiation (well differentiated vs moderately differentiated vs poorly differentiated)
- Sidedness (right sided vs left sided)
- Primary tumor location (colon vs rectum vs unspecified)
- Lines of prior therapy for metastatic disease (1-2 vs 3 vs 4 vs ≥ 5)
- Prior bevacizumab (yes vs no)
- Prior aflibercept (yes vs no)
- Prior ramucirumab (yes vs no)
- Prior anti PD1/PDL1 (yes vs no)
- Prior regorafenib (yes vs no)
- Prior trifluridine and tipiracil (yes vs no)
- Liver metastasis (yes vs no)

4.2 Subgroups

In addition to the stratification factors for randomization, prior anti-angiogenic therapy (yes vs no), time from initial diagnosis of metastatic disease to randomization (≥ 18 months vs < 18 months), and ECOG status (0 or 1 vs 2), primary and selected secondary endpoints will be examined in the following subgroups:

- Region (North America vs Europe vs Asia vs rest of world)
- Age (< 65 vs ≥ 65)
- Gender (male vs female)
- Race (Caucasian, Black, Asian, other)
- Presence of specific co-mutation (yes vs no)
- Differentiation (well differentiated vs moderately differentiated vs poorly differentiated)
- Sidedness (right sided vs left sided)
- Primary tumor location (colon vs rectum vs unspecified)
- Lines of prior therapy for metastatic disease (1-2 vs 3 vs 4 vs ≥ 5)
- Prior bevacizumab (yes vs no)
- Prior aflibercept (yes vs no)
- Prior ramucirumab (yes vs no)
- Prior anti PD1/PDL1 (yes vs no)
- Prior regorafenib (yes vs no)
- Prior trifluridine and tipiracil (yes vs no)
- Prior regorafenib and/or trifluridine/tipiracil (yes vs no)
- Liver metastasis (yes vs no)
- Prior oxaliplatin (yes vs no)
- Prior irinotecan (yes vs no)
- Prior fluoropyrimidine (yes vs no)
- Prior oxaliplatin and irinotecan and fluoropyrimidine (yes vs no)

If there is an insufficient number of subjects in the subgroup (i.e., less than 10% of the whole population), relevant subgroups may be combined.

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. For untreated subjects, baseline is the last non-missing assessment taken prior to randomization date.

Best Overall Response (BOR)

Best overall response (BOR) will be the best response recorded based on all post-baseline disease assessments that occur on or prior to disease progression and the initiation of subsequent anticancer treatment. At least 7 weeks from the first tumor assessment must elapse without radiological disease progression to meet the minimum criteria for stable disease (SD) duration in order to assign a best overall response of SD. In general, subjects not classifiable under the RECIST 1.1 response categories due to inadequate data or early death will be classified as non-evaluable (NE) for BOR but will be counted in the denominator of all response rate calculations.

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.

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 based on BICR **(** / Investigator) with >= 7 weeks measured.

Duration of Overall Response (DOR)

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

DOR in months = (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.

Investigational Product (IP)

The term ‘investigational product’ is used in reference to Sotorasib, Panitumumab, Trifluridine and Tipiracil, and Regorafenib.

Last Investigational Product Dose Date

The last IP date for each subject is defined as the latest date non-missing administration of Sotorasib, Panitumumab, Trifluridine and Tipiracil or Regorafenib.

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 Enrollment and 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), Hepatic Event Laboratory, Immunology, Toxicology CRFs, and in central lab data
- Start Date, Stop Date on Investigational Product Administration CRF
- Date Started and Date Ended or Resulted in Death on Events CRF
- Start date, Stop date on 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 PRO
- Liver Biopsy Results, Liver Imaging
- MUGA Scan
- COVID-19 Start Date or Breastfeeding Start Date or Date of Last Menstrual Period
- End of IP/study/safety follow-up/radiographic follow-up date
- Consent signed/withdrawn date

Objective Response by BICR (/ Investigator)

Objective response is defined as a complete response (CR) or partial response (PR) assessed per RECIST 1.1. Confirmation is determined by a repeat, consecutive assessment no less than 4 weeks from the date of first documented assessment.

Response will be assessed by BICR (/ Investigator).

Objective Response Rate (ORR)

Objective response rate 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 in months = (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 (i.e., the data cutoff date), the subject will be censored at the date of last contact through the analysis trigger date. If survival status is available from a survival sweep after the data cutoff date, the confirmed survival status post data cutoff date can be used to censor subjects at the data cutoff date.

Progression-Free Survival (PFS) by BICR (/ Investigator)

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

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

The censoring rules are available in Appendix D.

Progression unless noted otherwise, use Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1) per BICR / Investigator.

Relative Dose Intensity (RDI) of Panitumumab

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

- Cumulative actual dose (mg/kg) is defined as the total dose given during the study treatment exposure. The weight-adjusted cumulative dose (mg/kg) is defined as sum of (dose given at each visit / latest weight in kg) across visits. For subjects who did not take any drug the weight-adjusted cumulative actual dose by definition is 0 mg/kg.
- Duration of exposure (days) is defined as minimum (last non-zero dose date + 13, death date, data cutoff date) – first non-zero dose date + 1.

- Actual dose intensity (mg/kg/2 weeks) for subjects with non-zero duration of exposure is defined as: cumulative weight-adjusted actual dose (mg/kg) / round up (duration of exposure/14). For subjects who did not take any drug, the actual dose intensity is 0 mg/kg/2weeks.
- Cumulative weight-adjusted planned dose (mg/kg) is the per-protocol weight-adjusted planned dose accumulated over the duration on study treatment.
- Planned dose intensity (mg/kg/2 weeks) for subjects with non-zero duration of exposure is defined as: cumulative weight-adjusted planned dose (mg/kg) / round up (days of exposure/14).

Relative Dose Intensity (RDI) of Regorafenib

RDI of Regorafenib 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.
- Duration of exposure (days) is defined as minimum (last non-zero dose date + k, death date, data cutoff date) – first non-zero dose date + 1, where k is the number of off-days left in the cycle. Regorafenib dosing is 160mg QD on day 1-21 follows by 7 days off.
- Actual dose intensity (mg/day) for subjects with non-zero duration of exposure is defined as: cumulative actual dose (mg) / duration of exposure (days).. For subjects who did not take any drug, the actual dose intensity is 0 mg/day.
- Cumulative planned dose (mg) is the per-protocol planned dose accumulated over the duration on study treatment.
- Planned dose intensity (mg/day) for subjects with non-zero duration of exposure is defined as: cumulative planned dose (mg) / duration of exposure (days).

Relative Dose Intensity (RDI) of Sotorasib

RDI of Sotorasib 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.

- Duration of exposure (days) is defined as minimum (last dose non-zero date, death date, data cutoff date) – first non-zero dose date + 1.
- Actual dose intensity (mg/day) for subjects with non-zero duration of exposure is defined as: cumulative actual dose (mg) / duration of exposure (days). For subjects who did not take any drug, the actual dose intensity is 0 mg/day.
- Cumulative planned dose (mg) is the per-protocol planned dose accumulated over the duration on study treatment.
- Planned dose intensity (mg/day) for subjects with non-zero duration of exposure is defined as: cumulative planned dose (mg) / duration of exposure (days).

  For this study the planned dose intensity is 240 mg/day or 960 mg/day in the corresponding treatment group.

Relative Dose Intensity (RDI) of Trifluridine and Tipiracil

RDI of Trifluridine and Tipiracil 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. The cumulative BSA-adjusted actual dose (mg/m^2) is defined as sum of (dose given at each visit / latest BSA in m^2) across visits. Site-collected BSA information will be used. If site collected BSA is missing, derived BSA will be used. For subjects who did not take any drug the cumulative actual dose by definition is 0 mg/m^2.
- Duration of exposure (days) is defined as minimum (last non-zero dose date + k, death date, data cutoff date) – first non-zero dose date + 1, where k is the number of off-days before the next dose. Trifulridine/Tipiracil dosing is dosing day 1-5, then 2 days off, then dosing day 8-12, followed by 16 days off.
- Actual dose intensity (mg/ m^2/day) for subjects with non-zero duration of exposure is defined as: cumulative BSA-adjusted actual dose (mg/m^2) / duration of exposure (days). 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 duration on study treatment.

- Planned dose intensity (mg/m^2/day) for subjects with non-zero duration of exposure is defined as: cumulative BSA-adjusted planned dose (mg/m^2) / duration of exposure (days).

Time to Deterioration

Clinically meaningful deterioration from baseline will be based on meeting or exceeding a threshold for within-individual change (deterioration).

Table 5-1 below provides information about the change from baseline thresholds for clinically meaningful deterioration:

Table 5-1. Clinically Meaningful Deterioration Threshold

Scale	Deterioration Threshold
QLQ-C30 Global Health Status	Decrease of 10 points or more
QLQ-C30 Function Domains	Decrease of 10 points or more
QLQ-C30 Symptom Scales	Increase of 10 points or more
BFI-SF Item 3	Increase of 2 points or more
BPI-SF Item 3	Increase of 2 points or more

Time to deterioration will be defined as the time from randomization to the date of first clinically meaningful deterioration, or death by any cause if no clinically meaningful deterioration is observed. Death will count as an event only if it occurs before or on the last scheduled PRO assessment. Subjects who have not shown a clinically meaningful deterioration and who are alive before or on the last scheduled PRO assessment will be censored at the time of last available PRO assessment. If subject dies after two or more consecutive missed PRO assessment visits, the subject will be censored at the time of last available PRO assessment. Subjects with a baseline score that can never meet deterioration threshold (eg. < 10 for EORTC QLQ-C30 global health) will be censored at randomization date.

The censoring rules for the time to deterioration analysis are detailed in Appendix G.

Time to Response (TTR)

TTR is defined as time from the date of randomization to the first evidence of PR or CR based on BICR (/ Investigator). TTR will be calculated only for subjects who achieve a confirmed best overall response of PR or CR by BICR (/ Investigator).

Treatment-Emergent Adverse Event (TEAE)

TEAE are events with onset 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, **or prior to first dose of crossed over treatment**, whichever occurs earlier.

For summary after crossover, TEAE during crossover includes events with onset after administration of the first dose of crossover treatment and within the end of the study or 30 days of the last dose, whichever occur earlier.

Treatment-Related Adverse Event (TRAE)

A TRAE is any treatment-emergent AE with the relationship flag on the Events eCRF indicating there is a reasonable possibility that the event may have been caused by investigational medicinal product. In the unlikely event that the relationship is missing, the treatment-emergent event will be considered treatment-related (for all treatment in the combination therapy, if applicable) and documented in a footnote of the treatment-related summary.

6. Analysis Sets

The following populations are defined.

6.1 Full Analysis Set

The full analysis set (Intent-To-Treat 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(s)

The per protocol analysis set is a subset of the full analysis set which includes subjects who do not have selected 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(s)

All the by-visit tabular and graphical summary for ordinal data will be presented for the visits with at least 10 subjects in total from all treatment arms, unless otherwise specified.

6.4.1 EORTC QLQ-C30 Analysis Set

The EORTC QLQ-C30 analysis set consists of all randomized subjects with a non-missing baseline and at least one non-missing post-baseline results of any EORTC QLQ-C30 sub-scales. Subjects will be included in the analysis of each EORTC QLQ-C30 sub-scale for which they have non-missing data.

6.4.2 EQ-5D-5L Analysis Set

The EQ-5D-5L analysis set consists of all randomized subjects with a non-missing baseline and at least one non-missing post-baseline results of EQ-5D-5L VAS.

6.4.3 GP5 item of FACT-G Analysis Set

The GP5 FACT-G analysis set consists of all randomized subjects with a non-missing baseline and at least one non-missing post-baseline results of the GP5 item of FACT-G.

6.4.4 PGIC Analysis Set

The PGIC analysis set consists of all randomized subjects with at least one non-missing post-baseline results of the PGIC question.

6.4.5 BFI-SF Analysis Set

The BFI-SF analysis set consists of all randomized subjects with a non-missing baseline and at least one non-missing post-baseline results of the “fatigue at its worst” question.

6.4.6 BPI-SF Analysis Set

The BPI-SF analysis set consists of all randomized subjects with a non-missing baseline and at least one non-missing post-baseline results of the “pain at its worst” question.

6.5 Pharmacokinetic/Pharmacodynamic Analyses Set(s)

The PK analysis set includes all subjects randomized to sotorasib and panitumumab who have received at least 1 dose of sotorasib or panitumumab 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.

6.6 Interim Analyses Set(s)

Not applicable.

6.7 Study-specific Analysis Sets

Not applicable.

7. Planned Analyses

7.1 Interim Analysis and Early Stopping Guidelines

Not applicable.

7.2 Primary Analysis

The primary analysis of PFS will be event driven and occur when approximately 60 PFS events have been observed from the sotorasib 960 mg and panitumumab arm and the investigator's choice arm (i.e. approximately 90 PFS events across 3 arms). The PFS primary analysis may be delayed to ensure that the enrollment is finished and the delayed primary analysis will be triggered when the last randomized subject has had the opportunity to have at least 8 weeks of follow-up.

The analysis of OS at the 960 mg dose level for the combination treatment group of sotorasib and panitumumab versus the control group will occur when PFS is claimed statistically significant at primary analysis at the 960 mg dose level or at both the 960 mg and 240 mg dose levels. The analysis of OS at the 240 mg dose level will occur when the analysis of OS at the 960 mg dose level is claimed statistically significant.

The analysis of ORR for the combination treatment group of sotorasib and panitumumab versus the control group at either 960 mg or 240 mg dose level will occur when both PFS and OS are claimed statistically significant at primary analysis at this dose level.

7.3 Final Analysis

The OS final analysis will occur when approximately 50% of subjects having observed OS event (death), which is expected to be at approximately 6 months after the PFS primary analysis. At 50% OS events total (approximately 51 events in 2 arms), the minimum detectable difference for OS success is an HR of 0.574 (with one-sided 2.5% alpha).

Additional analyses may be conducted as required per health authority request(s).

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. Additional imputation rules, which will take the date of crossover into consideration if crossover ever occurs, may be applied as needed to access the impact of crossover on the safety summaries in the circumstance where the proportion of imputation is considered relatively large.

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 may be performed for PRO endpoints.

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

The efficacy analyses of PFS and key secondary endpoints will be conducted on the full analysis set. 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 arm for binary and categorical variables. For subjects who continue treatment post-progression or subjects who crossover, the date of first progression as assessed by BICR will be used for the primary analyses to evaluate objective response endpoints including PFS, ORR, DOR, TTR, and DCR and subject's response post first progression or post crossover will not be used.

In addition, the PRO assessments will only be considered before or on start of new anti-cancer therapy including crossover.

9.2 Subject Accountability

The number and percent of subjects who were screened, randomized, received study treatment, will be summarized by treatment group.

The number and percent of subjects who discontinued study treatment, and discontinued 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 sotorasib/panitumumab as needed if it is determined that crossover is appropriate after the primary analysis.

9.3 Important Protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first subject's initial visit and updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, subcategory codes, and descriptions will be used during the course of the study. Eligibility deviations are defined in the protocol.

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)
    - As categorical variable: (18-64 years vs 65-74 years vs 75-84 years vs >= 85 years)
  - Gender (Male vs Female)
  - Race (American Indian or Alaska Native vs Asian vs Black or African American vs Native Hawaiian or Other Pacific Islander vs White vs Multiple vs Other.)
  - Ethnicity (Hispanic or Latino vs Not Hispanic or Latino)
  - Weight (kg)
  - Height (cm)
  - Body Surface Area (m^2)
  - Region (North America vs Europe vs Asia vs Rest of world)
- Baseline disease characteristics:
  - Prior anti-angiogenic therapy (yes vs no)
  - Prior radiotherapy for current malignancy (yes vs no)
  - Prior surgery for current malignancy (yes vs no)
  - Time from initial diagnosis of metastatic disease to randomization (≥ 18 months vs < 18 months)
  - ECOG Performance Status (0 vs 1 vs 2)
  - Differentiation (well differentiated vs moderately differentiated vs poorly differentiated vs undifferentiated vs not applicable vs other)
  - Sidedness (right sided vs left sided)
  - Primary tumor location (colon (cecum, ascending colon, transverse colon, descending colon, sigmoid) vs rectum vs other)

- Liver metastasis (yes vs no)
- Lines of prior therapy for metastatic disease (1 vs 2 vs 3 vs 4 vs ≥ 5)
- Prior bevacizumab (yes vs no)
- Prior aflibercept (yes vs no)
- Prior ramucirumab (yes vs no)
- Prior anti PD1/PDL1 (yes vs no)
- Prior regorafenib (yes vs no)
- Prior trifluridine and tipiracil (yes vs no)
- Prior regorafenib and/or trifluridine/tipiracil (yes vs no)
- Prior oxaliplatin (yes vs no)
- Prior irinotecan (yes vs no)
- Prior fluoropyrimidine (yes vs no)
- Prior oxaliplatin and irinotecan and fluoropyrimidine (yes vs no)
- Prior EGFR antibody (yes vs no)
- Prior BRAF inhibitor (yes vs no)
- Presence of specific co-mutation (yes vs no)
- MSI-H result (MSI-High or MSI-Low or MSI-Stable or Not tested)
- TNM staging at initial diagnosis (as per eCRF data)
- Comutation per local data (as per eCRF data)
- TNM staging at screening (as per eCRF data)
- Brain metastasis (yes vs no)
- Histopathology type (as per eCRF data)
- Centrally confirmed KRAS G12C
- Best response to last prior line of therapy

9.5 Efficacy Analyses

Endpoint/Estimand	Primary Analysis	Sensitivity Analysis
Primary HR of PFS between the combination treatment group of sotorasib and panitumumab versus the control group at either 960 or 240 mg dose level, for subjects with metastatic CRC with *KRAS p.G12C* mutation, before or on start of new anticancer therapy	The PA of PFS will be based on BICR centrally assessed outcomes and conducted on the full analysis set. - 1-sided p-value by stratified log-rank test - HR and 95% CI by stratified Cox proportional hazards model - Kaplan Meier curve by treatment arm	- Unstratified analyses: ▪ 1-sided p-value by unstratified log-rank test ▪ HR and 95% CI by unstratified Cox regression - Investigator assessments: same primary summary and analysis method but based on investigator assessments. - Per-Protocol subset: same as primary summary and analysis method but for per-protocol subset. This analysis might be performed only if the per-protocol population is less than 90% of the full analysis set. - Initiation of new anticancer 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 anticancer 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). - Scheduled assessment dates: same primary summary and analysis methods, except the analysis is based on the scheduled assessment dates instead of actual assessment dates.
Key Secondary HR of OS between the combination treatment	The PA of OS will be conducted on the full analysis set.	- Unstratified analyses:

Endpoint/Estimand	Primary Analysis	Sensitivity Analysis
group of sotorasib and panitumumab versus the control group at either 960 or 240 mg dose level, for subjects with metastatic CRC with *KRAS p.G12C* mutation.	- 1-sided p-value by stratified log-rank test - HR and 95% CI by stratified Cox proportional hazards model - Kaplan Meier curve by treatment arm	▪ 1-sided p-value by unstratified log-rank test ▪ HR and 95% CI by unstratified Cox regression - Per-Protocol subset: same primary summary and analysis method but for per-protocol subset. This analysis might be performed only if the per-protocol population is less than 90% of the full analysis set. - Initiation of new anticancer therapy treated as OS Event: the data censoring rules are the same as those for the primary analysis of OS except that the use of new anticancer 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). - **Crossover adjustment: Based on method such as Rank Preserving Structural Failure Time (RPSFT), Inverse Probability of Censoring Weighting (IPCW), Two-stage approach and other as appropriate.**
Key Secondary Estimates of the difference of proportions of objective response between the combination treatment group of sotorasib and panitumumab versus the control group at either 960 or 240 mg dose level, for subjects with metastatic CRC with	The PA of ORR will be based on BICR centrally assessed outcomes and conducted on the full analysis set. - 1-sided p-value by stratified Cochran-Mantel-Haenszel method - Estimates of the difference of proportions of objective response and 95% CI by	Investigator assessments: same primary summary and analysis method but based on investigator assessments. - Per-Protocol subset: same primary summary and analysis method but for per-protocol subset. This analysis might be performed only if the per-

Endpoint/Estimand	Primary Analysis	Sensitivity Analysis
KRAS p.G12C mutation. Subjects who discontinued treatment prior to achieving an objective response are considered as non-responders.	stratified Mantal-Haenszel method. - The ORR of each treatment arm will be summarized along with 95% exact CI by Clopper-Pearson method.	protocol population is less than 90% of the full analysis set.

BICR = blinded independent central review; CI = confidence interval; CRC = colorectal cancer; HR = hazard ratio; ORR = objective response rate; OS = overall survival; PA = primary analysis; PFS = progression-free survival

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

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 Interactive Response Technology (IRT). 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. Both stratified and unstratified HR will be provided for subgroup analyses. 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 method controlling for the randomization stratification factors. The estimates of the difference of proportions of objective response and 95% CI will be calculated using Mantel-Haenszel method. The ORR of each treatment group will be provided along with 95% exact CI from the Clopper-Pearson method. 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 the difference of proportions of objective response 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 investigator's choice (trifluridine and tipiracil or regorafenib) who subsequently receive sotorasib and panitumumab, 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 investigator's choice (trifluridine and tipiracil or regorafenib) to sotorasib and panitumumab. 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 investigator's choice (trifluridine and tipiracil or regorafenib) time to event curve assuming all subjects initially randomized to investigator's choice (trifluridine and tipiracil or regorafenib) would not crossover to sotorasib and panitumumab. 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 sotorasib and panitumumab and the RPSFT model-adjusted investigator's choice (trifluridine and tipiracil or regorafenib) groups. One of the key assumptions of the RPSFT "common treatment effect", i.e, the treatment effect of sotorasib and panitumumab is the same regardless of when the subject started taking sotorasib and panitumumab, 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 sotorasib and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib). This method relies on the assumption that there are 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 investigator's choice (trifluridine and tipiracil or

regorafenib) group to the sotorasib and panitumumab group and subjects originally randomized to the sotorasib and panitumumab 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 if it is determined that crossover is appropriate after the primary analysis.

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, DOR will be right-censored based on the censoring conventions defined previously for PFS (refer to Appendix D). 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.

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.

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

The analyses of DOR, TTR and DCR are mainly performed based on BICR centrally assessed outcomes per RECIST v1.1. The sensitivity analyses are performed based on the Investigator assessments.

The analyses of PFS and ORR based on Investigator assessment will be analyzed using the same method as described for the primary analyses of PFS and ORR based on BICR centrally assessed outcomes.

9.5.3 Analyses of Exploratory Efficacy Endpoint(s)

Not applicable.

9.6 Safety Analyses

9.6.1 Analyses of Primary Safety Endpoint(s)

There is no primary safety endpoint in this study. The safety and tolerability of sotorasib 240 or 960 mg QD and panitumumab vs investigator's choice (trifluridine and tipiracil or regorafenib) will be assessed as a secondary endpoint. Unless otherwise specified, analyses of safety endpoints will be done in the safety analysis set.

9.6.2 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 24.0 or later will be used to code all events categorized as adverse events to a system organ class and a preferred term. The severity of each adverse event will be graded using the Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 criteria.

Subject incidence of all treatment-emergent adverse events (TEAE), grade 3 and above TEAE, serious TEAE, TEAE leading to discontinuation of IP, TEAE leading to dose reduction, dose interruption of IP and fatal TEAE will be tabulated overall and by a combination of system organ class (SOC) in alphabetical order, preferred term (PT) in order of descending frequency, and grade.

Subject incidence (excluding events that are directly related to progression of underlying cancer as identified by the Global Safety Officer from a list of MedDRA preferred terms) of all TEAE, grade 3 and above TEAE, serious TEAE, TEAE leading to discontinuation of IP, TEAE leading to dose reduction, dose interruption of IP and fatal TEAE will be tabulated overall and by a combination of system organ class (SOC) in alphabetical order, preferred term (PT) in order of descending frequency, and grade.

Subject incidence of treatment-related adverse events (TRAE), grade 3 and above TRAE, serious TRAE, TRAE leading to discontinuation of IP, TRAE leading to dose reduction, dose interruption of IP and fatal TRAE will be tabulated overall and by a combination of SOC in alphabetical order, PT in order of descending frequency, and grade.

Subject incidence of adverse events of interest (EOI) (standardized MedDRA queries and/or Amgen Medical Queries), grade 3 and above EOI, serious EOI, EOI leading to discontinuation of IP, EOI leading to dose reduction, dose interruption of IP and fatal EOI will be tabulated by EOI category and PT in order of descending frequency.

In addition, time to first onset of EOI and time to first grade 3 or higher event will be summarized for subjects with an event (days) using descriptive statistics for continuous variable. Duration of EOI and grade 3 or higher events (time of first EOI start till AE end date with AE outcome as resolved) will be summarized using descriptive statistics. Number of subjects and unresolved events will be tabulated.

Duration of EOI per subject will be the sum of all identified EOI TEAEs within the search category for the particular EOI. If TEAEs' duration overlapped, the overlapped period will be only counted once.

In addition to the incidence, the exposure-adjusted subject rates of TEAE, serious TEAE, and grade 3 and above TEAE will be tabulated by preferred term in descending order of frequency to account for different durations.

If it is determined that crossover is appropriate after the primary analysis, for subjects who cross over, the safety data will be censored at the time of crossover and select safety endpoints occurring after the first dose of sotorasib/panitumumab will be summarized separately.

TEAE summaries for crossover subjects, starting from the date of first crossover dose, may be considered. If number of crossed over subjects are small, then selected listings of AE will be provided.

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 (Analyte Listing) 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 summary.

Subject incidence of suspected Hy's law cases (Hy's law predicts potential for drug-related hepatotoxicity) will be summarized. A listing of alkaline phosphatase [ALP], aspartate aminotransferase [AST], alanine aminotransferase [ALT], total bilirubin [TBL] values at each time point will be produced for the subjects who meet Hy's law laboratory criteria.

9.6.4 Vital Signs

Vital signs including systolic/diastolic blood pressure, heart rate, respiratory rate and body temperature will be summarized by changes from baseline values for each treatment group using descriptive statistics. For the summary of changes from baseline by visit, subjects without a baseline and/or post baseline value will be excluded.

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 Electrocardiogram

Summaries and statistical analyses of ECG measurements are not planned, and these data would not be expected to be useful for meta-analysis with data from other trials.

9.6.7 Antibody Formation

Not applicable.

9.6.8 Exposure to Investigational Product

Drug exposure (for IPs) 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 (e.g., dose change/withheld, dose interruptions) and reason for modification will be summarized for all treatment groups.

List of manufacturing lot number will be provided.

9.6.9 Exposure to Concomitant Medication

The 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 anticancer therapies while on study will be summarized 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 endpoints and biomarker endpoints.

9.7.1 Analyses of Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Endpoints

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

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 (C_{max}), time to maximum concentration (t_{max}) and area under the plasma concentration-time curve (AUC). Other PK parameters such as AUC from time 0 to the time extrapolated to infinity (AUC_{inf}), apparent clearance (CL/F), and terminal half-life (t1/2) may be analyzed. Pharmacokinetic parameters will be estimated using standard non-compartmental approaches based on the PK Analysis Set and summarized by dose level using descriptive statistics including, but not limited to means, standard deviations, medians, minimums, and maximums. Above analyses will be conducted by Amgen Clinical Pharmacology Modeling and Simulation (CPMS).

Based on the review of the data, analyses to describe the relationship between sotorasib, panitumumab 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.

9.7.2 Analyses of Clinical Outcome Assessments

9.7.2.1 Compliance and Completion Rates

The completion rate for eCOAs will be presented by visit and treatment group, where each instrument is considered completed if at least one item is completed.

- Completion Rate in all analyzed subjects (at a visit) = (Number of subjects completing the respective eCOA at a visit/ Number of randomized subjects) * 100%.
- Compliance Rate = (Number of subjects completing the respective eCOA at a visit/ Number of subjects expected to have eCOA assessment at a visit) * 100%, where expected subjects i.e. still on study and still alive, excluding the subjects who are missing by design, such as death, disease progression, study discontinuation, etc.

9.7.2.2 Change from baseline over time to week 8 in EORTC QLQ-C30

Change from baseline over time to week 8 in physical functioning and global health status, and all other subscales of EORTC QLQ-C30 will be compared between treatment arms using a restricted maximum likelihood-based mixed effect model for repeated measures (MMRM) under the assumption of missing at random (MAR) (Mallinckrodt et al, 2008). The dependent variable of this model will be the change from baseline over time in disease measured every first cycle day. The model will include intercept, time, baseline score, treatment and randomization stratification factors as fixed effects, as well as an interaction term between treatment and time. Time is a continuous variable defined as days since randomization.

An estimate of the treatment difference by LS means between sotorasib and panitumumab arm and the Investigator's choice arm overall and at selected scheduled time points, the corresponding descriptive p-value and 2-sided 95% confidence intervals will be provided and used as the primary inference for the change from baseline endpoint.

MMRM model specifications are detailed in Appendix H.

9.7.2.3 Change from baseline over time to week 8 in BFI-SF and BPI-SF

Change from baseline over time to week 8 of fatigue and pain at its worst as measured by a single item #3 from BFI-SF and BPI-SF will be compared between treatment arms using MMRM method.

Change from baseline over time to week 8 for the following endpoints will also be summarized:

- Global BFI score: If a subject responds to at least 5 items on the BFI among the 9 items with numeric rating scale from 0 to 10, a mean of the non-missing items will be computed and used as a global BFI score.
- BPI pain severity index: Calculated as the mean of 4 items focused on pain intensity (i.e., pain now, average pain, worst pain, and least pain). All 4 severity items must be completed for aggregating a pain severity index.
- BPI pain interference index: Calculated as the mean of 7 items focused on pain intensity (general activity, mood, walking ability, normal work, relations with other persons, sleep, and enjoyment of life). The pain interference index is scored if more than 50%, or 4 of 7, of the items have been completed.

9.7.2.4 Descriptive Analysis

The continuous QLQ-C30 scales, EQ-5D-5L VAS score, BPI-SF (pain at its worst, severity index, interference index), and BFI-SF (fatigue at its worst, global BFI score) will be summarized descriptively by visit and by treatment group using the number of subjects with non-missing data (n), mean, standard deviation, median, minimum and maximum. Change from baseline will be summarized using the same statistics.

The ordinal GP5 item of FACT-G, EQ-5D-5L dimensions, and PGIC will be summarized descriptively by visit and by treatment group using the number and percentage of subjects with non-missing data (n) within each ordinal level. The number of missing observations will also be presented.

9.7.2.5 Time to Deterioration Analysis

Time to deterioration will be analyzed for BFI-SF, BPI-SF, and all subscales in the EORTC QLQ-C30 using a log rank test stratified by randomization factors for generation of descriptive p-value. Ties will be handled using the Breslow approach. The hazard ratio (HR) will be calculated from a Cox proportional hazard model stratified by randomization stratification factors. Kaplan-Meier curve will also be provided by treatment arm.

9.7.3 Analyses of Health Economic Endpoints

Not applicable.

9.7.4 Analyses of Biomarker Endpoints

Relationships between tumor change and levels of or change in levels of biomarkers of interest (from tissue and blood) listed as exploratory endpoints will be explored as

appropriate. Summary statistics will be provided, and graphical presentations may be used. As appropriate, details of analyses will be provided in a supplemental analysis plan for exploratory biomarker analysis.

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 with IP
  - visits occurred out of window
  - leading to early withdrawal from study
  - leading to early end of long-term follow-up
- Subject incidence who missed at least one dose of IP related to COVID-19

10. Changes From Protocol-specified Analyses

There are no changes to the protocol-specified analyses.

11. Literature Citations / References

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

Du Bois D, Du Bois EF (Jun 1916). A formula to estimate the approximate surface area if height and weight be known. Archives of Internal Medicine 17 (6): 863-71. PMID 2520314. Retrieved 2012-09-09

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

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 CH, Lane PW, Schnell D, Peng Y and Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Information Journal, 42: 303-319, 2008.

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

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

12. Prioritization of Analyses

Not applicable.

13. Data Not Covered by This 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 anticancer therapy, procedure will be imputed using the same rule when only the date is missing (no imputation when month or year is missing). Date of first post-treatment anticancer therapy will be imputed as first day of the month if only day is missing. Date of initial diagnosis of metastasis disease (in time from initial metastasis disease to randomization subgroup) will be imputed as first day of the month if only day is missing. Procedure start date will be imputed to 1st day of the month if day is missing, or January 1st if month and day are missing. Procedure end date will be imputed to last day of the month if day is missing, or December 31st if month and day are missing. If death date is not missing and imputed date is greater than death date, set imputed date equal to death date. No imputation if the whole date is missing.

Imputation Rules for Partial or Missing Start Dates

The reference date for the following rules is the date of first dose of IP.

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

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

Imputation Rules for Partial or Missing Stop Dates

Initial imputation

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

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

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

Imputation Rules for Partial or Missing Death Dates

Death data sources: End of study (EOS) deaths, survival status death, fatal AE end date. Death date sources should be reconciled before snapshot. However, if data issues fail to resolve prior to snapshot (e.g. an EOS death date is not consistent with fatal AE end date), consider the following priority to determine the (partial) death date.

- Full dates > only year and month available > only year available > total missing.
- If the source dates have the same amount of information, i.e. both or all source dates have only year and month available, but are not consistent, use the earliest (partial) date

If death occurred and the death date is partial, use the following rules:

If death year and month are available but day is missing:

- If yyyymm for the date last known to be alive equals yyyymm for death date, set death date to the day after the date last known to be alive.
- If yyyymm for the date last known to be alive is less than the yyyymm for death date, set death date to the first day of the death month.
- If yyyymm for the date last known to be alive is greater than yyyymm for death date, assume date last known to be alive is in error, set death date to the first day of the death month.

If month and day are missing and year of death is known:

- If yyyy for the date last known to be alive equals yyyy for death date, set death date to the day after last known to be alive date.
- If yyyy for the date last known to be alive is less than yyyy for death date, set death date to the first day of the death year.
- If yyyy for the date last known to be alive is greater than yyyy for death date, assume date last known to be alive is in error, set death date to the first day of the death year.

If a death date is totally missing:

Do not impute and censor the subject survival time.

Imputation Rules for Last Known Alive Dates

- If death occurred and death date is complete, then last known alive date = death date – 1.

- If death does not occur: Refer to SAP definition of Last alive date for source dates to determine this. If any of the source date is partial: Imputed last known alive date = max (complete source date, first day of the month if source date is missing, first month if only month is missing, first day of the year if source month is also missing), OS is censored at imputed last known alive.

Appendix B. Code Fragments

Provisional Code Fragments for calculating a confidence interval using the Clopper Pearson Method. The following example SAS code will be utilized as reference for the response rate analysis providing the proportion of subjects responding to treatment with corresponding 95% confidence intervals.

```
proc freq data = <indat> order = data;
  by trt;
  tables Result/binomial(exact) alpha=0.05;
  weight count/ zeros;
run;
```

Median survival & KM estimates SAS Code:

```
ods output quartiles = median (where = (percent = 50));
proc lifetest data = <dataset> timelist = (6,12) outsurv=km reduceout;
strata trtgrp;
time <time>*<status>(0) ;
        weight <weight>;
run;
```

Hazard ratio SAS Code:

```
proc phreg data = <dataset> covsandwich;
        weight <weight>;
        class trtgrp(ref='Control');
        model <time>*<status>(0) = trtgrp /rl;
ods output ParameterEstimates=hr;
run;
```

MMRM Sample SAS Code:

proc mixed data = <dataset> ;

by <grouping var 1> <grouping var 2>…<grouping var n>;

class <classification var 1> <classification var 2> …<classification var n>;

model <dependent var> = <independent var 1> <independent var 2>…<independent var n> <interaction term 1> <interaction term 2>…<interaction term n>*/ ddfm=kr*;

random intercept < visit var> */ subject* = <subject var> *type* = *un*; #if “UN” fails, use CS, AR(1), VC.

lsmeans <treatment var> <treatment var>*<visit var> */ cl alpha=0.05 pdiff*;

Appendix C. BOR Calculation Algorithm

BOR will need to be derived based on timepoint response collected on RECIST 1.1 CRF. Table 14-1 provides the BOR determination per RECIST 1.1 for trials where response confirmation is required. A BOR determined by Table 14-2 is considered a *Confirmed_BOR.*

At interim analysis, an unconfirmed response rate may be reported in addition to the confirmed response rate. The unconfirmed rate includes subjects who achieved a PR or CR and are awaiting a confirmative scan (i.e., unconfirmed responders). If these subjects have disease progression,death, next therapy, withdraw consent, or end of study before the confirmative scan, then they are no longer considered as unconfirmed responder awaiting confirmative scan. *Interim_BOR* is defined to include these subjects in addition to *Confirmed_BOR*. Table 14-2 outlines the steps to derive *Confirmed_BOR* and *Interim_BOR* given the timepoint assessments.

Table 14-1. BOR per RECIST 1.1

Criterion	Timepoint T1 Response	T1 ≥ 49* days after Baseline?	Timepoint T2 Response	T2 ≥ 49* days after Baseline?	T2 ≥ 28 days after T1?	BOR
C1	CR	Yes	CR	-	Yes	CR
C2			CR	-	No	SD
C3			PR, SD	-	-	Query data**
C4			PD	-	-	SD
C5			NE, No further evaluations			SD
C6		No	CR	-	Yes	CR
C7			CR	Yes	No	SD
C8			PR, SD	-	-	Query data**
C9			PD	-	-	PD
C10			NE, No further evaluations			NE
C11	PR	Yes	CR, PR	-	Yes	PR
C12			CR, PR	-	No	SD
C13			SD	-	-	SD
C14			PD	-	-	SD
C15			NE, No further evaluations			SD
C16		No	CR, PR	-	Yes	PR
C17			CR, PR	Yes	No	SD
C18			SD	Yes	-	SD
C19			PD	-	-	PD
C20			NE, No further evaluations			NE

C21	SD	Yes	CR, PR, SD, PD, NE, no more evaluation			SD
C22		No	CR, PR, SD	Yes	-	SD
C23			CR, PR, SD	No	-	NE
C24			PD	-	-	PD
C25			NE, No further evaluations			NE
C26	PD		-			PD
C27	NE	-	NE, No further evaluations			NE
C28		-	CR, PR, SD	Yes	-	SD
C29		-	CR, PR, SD	No	-	NE
C30		-	PD	-	-	PD

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

*CR can only be claimed if a scan at a subsequent timepoint 4 weeks or more later confirms the response of CR. PR can only be claimed if a scan at a subsequent timepoint 4 weeks or more later confirms the response of PR or better (CR).; SD will need at least 7 weeks confirmation. The threshold of 49 days accounts for the earliest available assessment by RECIST v1.1, as for the tumor assessment per SOA of every 8 weeks from cycle 1 day 1 (± 7 days).

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

Table 14-2. BOR Derivation Step for GSP

Step 1. Derive Confirmed_BOR at each visit

Derive *Confirmed_BOR* at each visit:

i) At Visit 1: Using the Visit 1 scan result and refer to Table 14-1 to derive *Confirmed_BOR*.

ii) At Visit 2 onward:

(a) If current visit is CR or PR: Find last scan that is not NE or SD and at least 28 days before current visit. Derive *BOR_temp* as below:

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

(b) If none of above fits, find last scan that is not NE, reference Table 14-1 to derive *BOR_temp*.

iii) Current visit Confirmed_BOR = best of (BOR_temp, last visit confirmed_BOR). Use rule CR > PR > SD > PD > NE	
Step 2. Derive Interim_BOR at last visit prior to DCO	
For subjects who discontinue tumor assessment (ie, had PD, death, next therapy, withdraw consent, or end of study): Assign *Interim_BOR* = *Confirmed_BOR*	For subjects with potential for more assessment (ie, no PD, death, next therapy, withdraw consent, or end of study), unconfirmed PR/CR can be considered responders. • If any scan is CR and no PD/death, then *Interim_BOR* = CR. • If any scan is PR and no CR or PD/death, then *Interim_BOR* = PR. • Else, *Interim_BOR* = *Confirmed_BOR* at the latest scan.

Appendix D. PFS Censoring Guidelines

PFS Primary Censoring Rule

Situation up to data cut off or EOS	Date of Event or Censor	Outcome
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) on or before start of new anti-cancer therapy	Censor
No radiological disease progression by BICR (or by investigator), no death recorded, no start of new anti-cancer therapy	Date of last evaluable assessment by BICR (or by investigator)	Censor
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.

PFS Sensitivity: Initiation of new anticancer therapy treated as PFS Event and ignore missed visits

Situation up to data cut off or EOS	Date of Event or Censor	Outcome
Radiological disease progression by BICR (or by investigator) prior to death	Date of first observation of radiological PD 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, no start of new anti-cancer therapy	Date of randomization	Censor
No radiological disease progression by BICR (or by investigator), no death, but start of new anti-cancer therapy recorded	Date of start of new anti-cancer therapy	Event
No radiological disease progression by BICR (or by investigator), no death recorded, no start of new anti-cancer therapy	Date of last evaluable assessment by BICR (or by investigator)	Censor
Death or radiological disease progression by BICR (or by investigator) immediately after more than one missed tumor assessment	Earliest of disease progression or death	Event

PFS Sensitivity: Scheduled assessment dates

The analysis will use the next scheduled assessment dates instead of actual assessment dates if the assessments are outside visit window. If the assessments are within visit window same rules as those of primary PFS will be used.

First, follow primary censoring rule to determine outcome and date of event or censor. If the date of PD event or censor is outside a visit window, use the following rules:

Situation up to data cut off or EOS	**Date of Event or Censor**	**Outcome**
PD in a visit window	Observed date of PD	Event
PD outside a visit window	If next scheduled visit is on or before death and DCO, set event at date of next scheduled visit	Event
	If next scheduled visit is after death date and death date is on or before DCO, set event at death date	Event
	If next scheduled visit is after death date but death date is after DCO, censor at last scheduled visit on or before DCO	Censor
	If next scheduled visit if after DCO, censor at last scheduled visit on or before DCO	Censor

No PD, but death record	Death	Event
Censoring in a visit window	Observed censoring date	Censor
Censoring outside of a visit window	Next scheduled visit (unless after DCO or start of new anticancer therapy, then it will be censored to last scheduled visit on or before DCO or start of new anticancer therapy)	Censor

If the assessment is within a visit window, same rules as those of primary PFS will be used.
Visit window is +/- 1 week of tumor assessment scan every 8 weeks.

Appendix E. Questionnaires

14.E.1 EORTC QLQ- C30

ENGLISH

EORTC QLQ-C30 (version 3)

We are interested in some things about you and your health. Please answer all of the questions yourself by circling the number that best applies to you. There are no "right" or "wrong" answers. The information that you provide will remain strictly confidential.

Please fill in your initials: _____
Your birthdate (Day, Month, Year): _____
Today's date (Day, Month, Year): 31 _____

	Not at All	A Little	Quite a Bit	Very Much
1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?	1	2	3	4
2. Do you have any trouble taking a long walk?	1	2	3	4
3. Do you have any trouble taking a short walk outside of the house?	1	2	3	4
4. Do you need to stay in bed or a chair during the day?	1	2	3	4
5. Do you need help with eating, dressing, washing yourself or using the toilet?	1	2	3	4

During the past week:	Not at All	A Little	Quite a Bit	Very Much
6. Were you limited in doing either your work or other daily activities?	1	2	3	4
7. Were you limited in pursuing your hobbies or other leisure time activities?	1	2	3	4
8. Were you short of breath?	1	2	3	4
9. Have you had pain?	1	2	3	4
10. Did you need to rest?	1	2	3	4
11. Have you had trouble sleeping?	1	2	3	4
12. Have you felt weak?	1	2	3	4
13. Have you lacked appetite?	1	2	3	4
14. Have you felt nauseated?	1	2	3	4
15. Have you vomited?	1	2	3	4
16. Have you been constipated?	1	2	3	4

Please go on to the next page

SPECIMEN

During the past week:	**Not at All**	**A Little**	**Quite a Bit**	**Very Much**
17. Have you had diarrhea?	1	2	3	4
18. Were you tired?	1	2	3	4
19. Did pain interfere with your daily activities?	1	2	3	4
20. Have you had difficulty in concentrating on things, like reading a newspaper or watching television?	1	2	3	4
21. Did you feel tense?	1	2	3	4
22. Did you worry?	1	2	3	4
23. Did you feel irritable?	1	2	3	4
24. Did you feel depressed?	1	2	3	4
25. Have you had difficulty remembering things?	1	2	3	4
26. Has your physical condition or medical treatment interfered with your family life?	1	2	3	4
27. Has your physical condition or medical treatment interfered with your social activities?	1	2	3	4
28. Has your physical condition or medical treatment caused you financial difficulties?	1	2	3	4

For the following questions please circle the number between 1 and 7 that best applies to you

29. How would you rate your overall health during the past week?

1	2	3	4	5	6	7
Very poor						Excellent

30. How would you rate your overall quality of life during the past week?

1	2	3	4	5	6	7
Very poor						Excellent

SPECIMEN

14.E.2 EQ-5D-5L

Under each heading, please tick the ONE box that best describes your health TODAY.

MOBILITY

- I have no problems in walking about ❑
- I have slight problems in walking about ❑
- I have moderate problems in walking about ❑
- I have severe problems in walking about ❑
- I am unable to walk about ❑

SELF-CARE

- I have no problems washing or dressing myself ❑
- I have slight problems washing or dressing myself ❑
- I have moderate problems washing or dressing myself ❑
- I have severe problems washing or dressing myself ❑
- I am unable to wash or dress myself ❑

USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*

- I have no problems doing my usual activities ❑
- I have slight problems doing my usual activities ❑
- I have moderate problems doing my usual activities ❑
- I have severe problems doing my usual activities ❑
- I am unable to do my usual activities ❑

PAIN / DISCOMFORT

- I have no pain or discomfort ❑
- I have slight pain or discomfort ❑
- I have moderate pain or discomfort ❑
- I have severe pain or discomfort ❑
- I have extreme pain or discomfort ❑

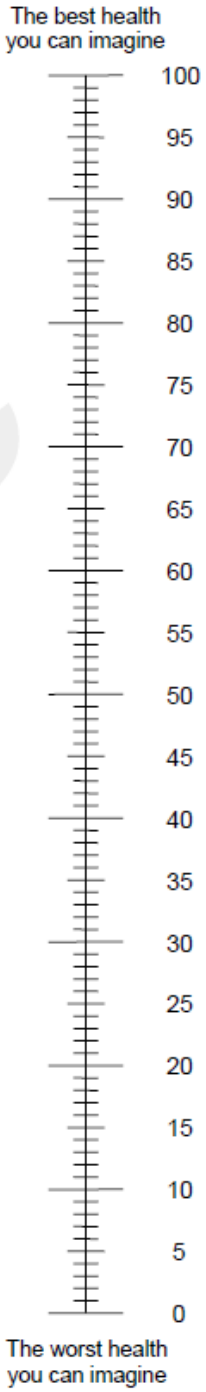
ANXIETY / DEPRESSION

- I am not anxious or depressed ❑
- I am slightly anxious or depressed ❑
- I am moderately anxious or depressed ❑
- I am severely anxious or depressed ❑
- I am extremely anxious or depressed ❑

Sample

- We would like to know how good or bad your health is TODAY.
- This scale is numbered from 0 to 100.
- 100 means the best health you can imagine.
  0 means the worst health you can imagine.
- Please mark an X on the scale to indicate how your health is TODAY.
- Now, write the number you marked on the scale in the box below.

YOUR HEALTH TODAY = ☐



Sample

14.E.3 GP5 of the FACT-G

GP5 (Version 4)

Below is a list of statements that other people with your illness have said are important. **Please select one number per line to indicate your response as it applies to the past 7 days.**

		Not at all	A little bit	Some-what	Quite a bit	Very much
GP5	I am bothered by side effects of treatment	0	1	2	3	4

English (Universal)
Copyright 1987, 1997

14.E.4 PGIC

PATIENT GLOBAL IMPRESSION OF CHANGE (PGIC)

Since the start of the study, my overall status is:

✓*one box only:*

[1] ☐ Very Much Improved
[2] ☐ Much Improved
[3] ☐ Minimally Improved
[4] ☐ No Change
[5] ☐ Minimally Worse
[6] ☐ Much Worse
[7] ☐ Very Much Worse

(US/English)

review copy
t use without perr

14.E.5 BFI-SF

Brief Fatigue Inventory

STUDY ID# _____ HOSPITAL # _____

Date: ____/____/____ Time: _____

Name: _____ _____ _____
Last First Middle Initial

Throughout our lives, most of us have times when we feel very tired or fatigued. Have you felt unusually tired or fatigued in the last week? Yes ☐ No ☐

1. Please rate your fatigue (weariness, tiredness) by circling the one number that best describes your fatigue right NOW.

0 1 2 3 4 5 6 7 8 9 10
No Fatigue — As bad as you can imagine

2. Please rate your fatigue (weariness, tiredness) by circling the one number that best describes your USUAL level of fatigue during past 24 hours.

0 1 2 3 4 5 6 7 8 9 10
No Fatigue — As bad as you can imagine

3. Please rate your fatigue (weariness, tiredness) by circling the one number that best describes your WORST level of fatigue during past 24 hours.

0 1 2 3 4 5 6 7 8 9 10
No Fatigue — As bad as you can imagine

4. Circle the one number that describes how, during the past 24 hours, fatigue has interfered with your:

A. General activity
0 1 2 3 4 5 6 7 8 9 10
Does not interfere — Completely Interferes

B. Mood
0 1 2 3 4 5 6 7 8 9 10
Does not interfere — Completely Interferes

C. Walking ability
0 1 2 3 4 5 6 7 8 9 10
Does not interfere — Completely Interferes

D. Normal work (includes both work outside the home and daily chores)
0 1 2 3 4 5 6 7 8 9 10
Does not interfere — Completely Interferes

E. Relations with other people
0 1 2 3 4 5 6 7 8 9 10
Does not interfere — Completely Interferes

F. Enjoyment of life
0 1 2 3 4 5 6 7 8 9 10
Does not interfere — Completely Interferes

Copyright 1997
The University of Texas M. D. Anderson Cancer Center
All rights reserved.

14.E.6 BPI-SF

STUDY ID #:_____ DO NOT WRITE ABOVE THIS LINE HOSPITAL #:_____

Brief Pain Inventory (Short Form)

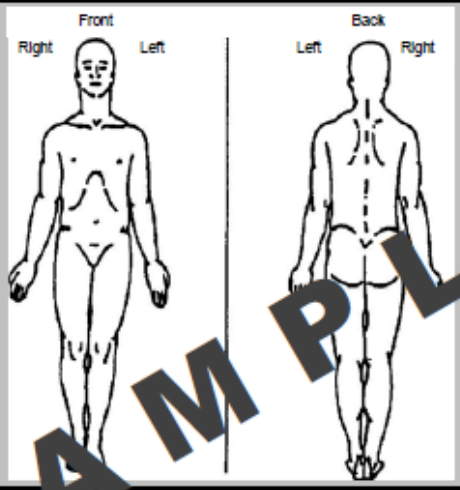
Date: ____/____/____ Time: _____

Name: _____ _____ _____
Last First Middle Initial

1. Throughout our lives, most of us have had pain from time to time (such as minor headaches, sprains, and toothaches). Have you had pain other than these every-day kinds of pain today?

1. Yes 2. No

2. On the diagram, shade in the areas where you feel pain. Put an X on the area that hurts the most.



SAMPLE

3. Please rate your pain by circling the one number that best describes your pain at its worst in the last 24 hours.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

4. Please rate your pain by circling the one number that best describes your pain at its least in the last 24 hours.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

5. Please rate your pain by circling the one number that best describes your pain on the average.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

6. Please rate your pain by circling the one number that tells how much pain you have right now.

0	1	2	3	4	5	6	7	8	9	10
No Pain										Pain as bad as you can imagine

Page 1 of 2

STUDY ID #:_____ DO NOT WRITE ABOVE THIS LINE HOSPITAL #: _____

Date: ___/___/____ **Time:** _____
Name: _____ _____ _____
Last First Middle Initial

7. What treatments or medications are you receiving for your pain?

8. In the last 24 hours, how much relief have pain treatments or medications provided? Please circle the one percentage that most shows how much relief you have received.

0% No Relief 10% 20% 30% 40% 50% 60% 70% 80% 90% 100% Complete Relief

9. Circle the one number that describes how, during the past 24 hours, pain has interfered with your:

A. General Activity
0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

B. Mood
0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

C. Walking Ability
0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

D. Normal Work (includes both work outside the home and housework)
0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

E. Relations with other people
0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

F. Sleep
0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

G. Enjoyment of life
0 Does not Interfere 1 2 3 4 5 6 7 8 9 10 Completely Interferes

SAMPLE

Copyright 1991 Charles S. Cleeland, PhD
Pain Research Group
All rights reserved

Page 2 of 2

Appendix F. EORTC QLQ-C30 Scoring

The EORTC QLQ-C30 is a self-reporting 30-item generic instrument which assesses 15 scales including 5 functional domains (physical, role, emotional, cognitive, social), 9 symptom domains (fatigue, nausea and vomiting, pain, dyspnea, insomnia, appetite loss, constipation, diarrhea, financial difficulties), and a global health status/QOL scale. The recall period is the past week. Subjects will complete the EORTC QLQ-C30 at clinic visits only.

For each of the 15 scales/items, a standardized raw score ranging between 0 and 100 will be calculated following the EORTC QLQ-C30 Scoring Manual, ver. 3 (Fayers et al., 2001). For the Global Health Status/QOL scale and functional scales of QLQ-C30, a higher score represents a better health state. However, for the symptom scales or items of QLQ-C30, a lower score represents a better health state while a higher score indicates a higher symptom burden.

For all scales, calculate the raw score (RS) of a scale using the mean of the item scores in the scale as follows

$$RS = (S1 + S2 + \ldots + Sn)/n$$

where Si, i=1,…,n, are the item scores and n is the number of items with valid scores assuming the number of items with valid scores meets the minimum requirement as specified in Table 14-3 or this scale score will be assumed missing.

Use a linear transformation to standardize the raw score in order that scores will range from 0-100:

$$Global\ Health\ Status/QOL = \frac{RS - 1}{range} \times 100,$$

$$Functional\ Scales = \left(1 - \frac{RS - 1}{range}\right) \times 100,$$

$$\text{Symptom Scales} = \frac{RS - 1}{range} \times 100$$

where range for each scale is defined in Table 14-3.

Table 14-3: QLQ-C30 Scales and Scoring Details

QLQ-C30	Number of Items	Item Range [a]	Item Numbers	Minimum Not Missing

Global Health status/QOL	2	6	29,30	1
Functional Scales				
Physical Functioning	5	3	1-5	3
Role Functioning	2	3	6,7	1
Emotional Functioning	4	3	21-24	2
Cognitive Functioning	2	3	20,25	1
Social Functioning	2	3	26,27	1
Symptom Scales / Items				
Fatigue	3	3	10,12,18	2
Nausea/vomiting	2	3	14,15	1
Pain	2	3	9,19	1
Dyspnea	1	3	8	N/A
Insomnia	1	3	11	N/A
Appetite Loss	1	3	13	N/A
Constipation	1	3	16	N/A
Diarrhea	1	3	17	N/A
Financial Difficulties	1	3	28	N/A

[a] Range is the difference between the maximum possible value of the Raw Score and the minimum possible value.

Appendix G. Censoring Rules for Time to Deterioration

Scenario	Date of Event/Censor	Status
If baseline score can never meet deterioration threshold	Date of randomization	Censor
If clinically meaningful deterioration from baseline	Earliest date to observe clinically meaningful deterioration from baseline	Event
If no clinically meaningful deterioration from baseline, but death before or on the last scheduled PRO assessment	Date of death	Event
If no clinically meaningful deterioration from baseline, and no death before or on the last scheduled PRO assessment	Date of last available PRO assessment	Censor
If no clinically meaningful deterioration from baseline, but death after two or more consecutive missed PRO assessment[a]	Date of last available PRO assessment	Censor

Footnote: [a] this rule supersedes the previous rules.

Appendix H. Model Specifications for MMRM

MMRM for change from baseline over time to week 8

For the i[th] subject, let X_i be the vector of their randomization stratification factors including prior anti-angiogenic therapy (yes vs no), time from initial diagnosis of metastatic disease to randomization (≥ 18 months vs < 18 months), and ECOG status (0 or 1 vs 2). Stratification factors with less than 5% of the whole population in one category will not be included as a covariate.

Let I_{trt} be the treatment indicator and set to 0 or 1 if subjects are randomized to the sotorasib and panitumumab arm and the Investigator's choice arm, respectively. The subject intercept and slope of time for the change from baseline will be treated as random effects and assumed to follow a multivariate normal distribution. Specifically, the change from baseline score Y_{ij} for the ith subject jth assessment measured at time t_{ij}, which is the scheduled visit time, is modeled as

$$Y_{ij} = (b_0 + b_{i0}) + (b_1 + b_{i1})t_{ij} + b_2 Q_i + X_i^T \beta + \gamma I_i^{trt} + b_3 t_{ij} \times I_i^{trt} + \varepsilon_{ij}$$

where b_0, b_1,b_2, β and γ are the fixed effects of intercept, time, baseline score stratification factors and treatment, Q_i denotes the baseline score, and the term $t_{ij} \times I_i^{trt}$ denotes the treatment-by-time interaction, b_3 is the associated fixed effect; the subject random intercept b_{i0} and the subject random slope of time b_{i1} are assumed to follow a multivariate normal distribution with mean zero and an unstructured covariance matrix (see Appendix B for SAS implementation). If the model using unstructured covariance matrix fails to converge, a structured covariance matrix will be used.