Protocol C4221015

AN OPEN-LABEL, MULTICENTER, RANDOMIZED PHASE 3 STUDY OF FIRST-LINE ENCORAFENIB PLUS CETUXIMAB WITH OR WITHOUT CHEMOTHERAPY VERSUS STANDARD OF CARE THERAPY WITH A SAFETY LEAD-IN OF ENCORAFENIB AND CETUXIMAB PLUS CHEMOTHERAPY IN PARTICIPANTS WITH METASTATIC *BRAF* V600E-MUTANT COLORECTAL CANCER

Statistical Analysis Plan
(SAP)

Version: **7.0**

Date: **07 May 2024**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. VERSION HISTORY

Table 1 Summary of Change

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
1.0/ 09 September 2020	1.0/ 15 July 2020	N/A	N/A
2.0 / 11 May 2021	3.0 / 24 February 2021	To match latest protocol amendment. Also to add analyses for Korea and Japan.	Updated various cross references that did not work in the text. Clarified that the safety set in the SLI was in fact the same as the full analysis set (FAS) in the SLI. The FAS was noted for the SLI. Also replaced treatment arm with treatment cohort for the SLI. Added in various points of clarity (confirmed and unconfirmed response in sections 3.1.2.2, 3.2.2.2, 3.2.2.3, removed a discrete summary under the continuous section 5.2.15, clarified PFS2 in section 6.2.6, noted the PRO set in section 6.3.1, noted the PK analysis set in section 6.3.2.2, clarified China in section 6.3.3.2). Clarified that secondary endpoints PFS and OS were for Arm A versus Arm B, under section 2.1. Modified PGIS to 4 categories and PGIC to 5 categories under section 3.2.3.1. Clarified that BICR was used for liver covariate, and replaced Caucasian with White in covariate section 3.3.2. Noted that the SLI is analyzed according to actual treatment arm received in section 4. Clarified the DLT evaluable set in section 4. Included new text for possible two-arm trial only (assuming Arm B was not tolerable per the safety lead-in) in section 2.2, section 5.1.1, and section 7. Inserted p-values and removed the hazard ratios for the primary analysis decision criteria for both the primary and key secondary endpoints, under section 5.1.1.2. Removed the hazard ratio for decision criteria in table 15. Removed the RECIST criteria for the secondary endpoint PFS2. Removed the plot of RMST versus time in section 5.2.16. Clarified missing dates under section 5.3.1. Corrected Table 6 to use the last scan rather than the last known alive date. Updated Tables 11, 12, and 13 in PK section 6.3.2 to be more consistent with the protocol.

Table 1 Summary of Change

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
			Clarified PK analyses in sections 6.3.2.1.1, 6.3.2.1.2, 6.3.2.1.3, and 6.3.2.2.1. Removed individual PK plots. Removed some biomarker listings under section 6.3.3.2. Clarified the race subset, the side of tumor subset, and liver metastasis subset, in section 6.4. Added in bevacizumab exposure calculation in section 6.5.3. Also changed categorical summary from months to weeks in section 6.5.3. Clarified how to handle dosing for days 1 and 2 for patients taking FOLFIRI (either Cohort 1 in the safety lead-in or Arm B in phase 3), under section 6.5.3.1. Removed the SLI cohorts from the AE sorting procedure in section 6.6.1. Added in new treatment related summaries in section 6.6.1. Also clarified that treatment related is at least one study drug related. Changed the inequality signs to strict inequality for the liver function tests, under section 6.6.4, to match the protocol and FDA guidance. Removed a listing under section 6.6.5. Removed QT parameter and a listing from section 6.6.6. Added a new section 6.7 to cover particular summaries needed for Korea and Japan. Inserted p-values and removed the hazard ratios for the futility analysis decision criteria for the primary endpoint, under section 7.2.1. In the list of abbreviations (appendix 1), added new entries for MWPC, ND, NR, and OTR. Removed entries not used in the text: FOLFOX, MMRM, REML.
3.0 / 26 April 2022	4.0 / 28 February 2022	To match protocol amendment 4: based on SLI results primary and key endpoints were modified. In addition, for Phase 3 portion, efficacy endpoints based on investigator assessment were	Following results from SLI portion, removed reference to FOLFIRI treatment for Arm B of the Phase 3 portion through the whole document. For Phase 3 portion moved the comparison of PFS by BICR between Arm A versus the Control Arm from co-primary endpoint to key secondary endpoint and updated primary and secondary estimands in sections 2.1 and 3.2. Updated the sample size and study design in sections 2.2 and 5.1.1.

Table 1 Summary of Change

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
		changed to derived investigator tumor assessment.	For Phase 3 portion specified in sections 3.2.2 and 6.2 that endpoints related to tumor assessment by investigator will be derived programmatically. Removed PRO analysis set from section 4. Substituted graphical gatekeeping procedure with hierarchical testing procedure to preserve overall alpha of the study in section 5.1.2.2. Added definitions of start of new anticancer drug therapy and of start of new anticancer therapy in sections 5.2.8 and 5.2.9. Substituted Clopper Pearson CIs by Wilson score CIs in sections 5.2.16 and 6.2.3. Specified in section 6 that for SLI portion the efficacy analyses will be performed by number of lines of prior therapies. In section 6 analyses for primary and secondary endpoints have been aligned with modified endpoints of Phase 3 portion. Added discontinuation of next-line treatment after first objective PD by investigator assessment as event in the PFS2 definition in sections 2.1, 3.2.2.6, 6.2.6. Added in section 6.6.1 tables for AEs/SAEs related to chemotherapy and AE associated to dose modification for chemotherapy. Changed the criteria to show the deaths from 30 days to 28 days after the last dose of study treatment, in section 6.6.2. Added criteria for QRS and PR interval to be included in the table for ECG, in section 6.6.6. Updated the number of events and criteria for interim analyses for PFS futility and for OS efficacy in section 7.2. Added specification for programmatic derivation of tumor response in section 10. Throught the document, included editorial changes to enahance clarity.
4.0 / 17 February 2023	5.0 / 20 December 2022	To match protocol amendment 5 dated 20 Dec 2022: based on FDA inputs/discussion: removed Arm A from key secondary endpoints, added ORR as primary endpoint for Phase	Updated primary and secondary estimands in sections 2.1, 3.2 and 3.3: for Phase 3 portion moved the comparison of PFS by BICR between Arm A versus the Control Arm from the key secondary endpoint to secondary endpoints and added ORR by BICR for Arm B versus the Control Arm as a primary endpoint. Included endpoints for Cohort 3. .

Table 1 Summary of Change

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
		3, added Cohort 3 with 2 randomized arms: Arm D and E	Added Cohort 3 to analysis set definitions and the subgroup "Full Analysis Set, ORR subset" for ORR analysi of Phase 3 in section 4. Updated the sample size and study design in sections 2.2 and 5.1.1. Modified hierarchical testing procedure to preserve overall alpha of Phase 3 and added for Cohort 3 in sections 5.1.2.2 and 5.1.2.3. In section 6 analyses for primary and secondary endpoints have been aligned with modified endpoints of Phase 3 and Cohort 3. Added some listing in PK section 6.3.2. Modified some AE tables in section 6.6. Removed interim analysis for PFS for Phase 3 and updated the number of events and criteria for interim analyses for OS efficacy in section 7.
5.0 / 15 December 2023	5.0 / 20 December 2022	Reviewed after BDR to align with requested minor changes and manage particular cases	Added a few descriptive outputs: included overall exposure for SLI, Phase 3 and Cohort 3, a table for regimens in arm C in Phase 3, a swimmer plot for subsequent therapies for Phase 3, a disposition for end of screening, a plot for OS/PFS in SLI, time to first dose modification, threshold could be applied to some in-text AE tables. In Table 16 specified that in case ORR and PFS or only PFS analyses results are statistically significant the nominal alpha level for OS analysis will be 0.023. Changed CEA, CRP and MSI categories/criteria for reporting, sections 6.4 and 6.5.1.2. Modified efficacy baseline definition and adequate baseline tumor assessment definition to consider assessments performed up to first dose date, sections 3.4, 5.2.13, 10.1.1. Specified the NE BOR includes cases with no evidence of disease at baseline (ND), section 6.1.2.2.1. Modified the endpoint definition for Phase 3 and Cohort 3, and some analyses for ctDNA biomarkers, sections 2.1, 3.2.3.3, 3.3.3.3, 6.3.3.2. Minor editorial changes.
6.0 / 08 March 2024	5.0 / 20 December 2022	Given the slower than expected number of PFS events, the required number per original study design cannot	Updated the required number of PFS events for the final analysis and as a result, decreased the power for the sample size calculation for PFS of Phase 3 portion, sections 5.1.1.2, 5.1.2.2, 6.1.2.1.1.

Table 1 Summary of Change

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
		be achieved. Therefore, the required number of PFS events for final analysis has been reduced to what could be observed in a timely manner. As a result, the power of the study has been reduced from 95% to 85% still acceptable for a Phase 3 study. Modified the definition of adequate baseline tumor assessment to reduce the number of not evaluable participants due to missing baseline tumor assessments. ctDNA biomarker changes were implemented to incorporate recent findings from ctMoniTR project. Added exploratory surgical conversion rate, to assess if study intervention enables patients to have surgical resections with curative intent. Added supplementary analysis of PFS to consider events even if they occur after long gap in assessments.	Modified the definition of adequate baseline tumor assessment to accept exams older than 35 days prior randomization for Phase 3 and Cohort 3, sections 5.2.13, 10.1.1. Modified the endpoint definition for ctDNA biomarkers for Phase 3 and Cohort 3, and some related analyses, sections 2.1, 3.2.3.3, 3.3.3.3, 6.3.3.2. Added exploratory endpoint of surgical conversion rate, sections 2.1, 6.3.4. Specified that in Cohort 3, if ORR results are not statistically significant, PFS will be analyzed descriptively, section 5.1.2.3 Added a supplementary analysis for PFS including observations/events that occur after >12 (or 16) weeks from last adequate post-baseline tumor assessment/start date, section 6.1.2.1.2. Minor editorial changes and re-wording for clarity.
7.0 / 07 May 2024	6.0 / 13 March 2024	Due to human error, the randomization for Cohort 3 was not set up correctly with a ratio 2:1 as per protocol amendment 5, but	Modified sample size for Cohort 3 from 135 to 136 to reflect the actual 1:1 randomization;, re-evaluated the power for ORR analysis and the number of PFS events by BICR required for PFS analysis, see sections 2.2, 5.1.1, 5.1.1.3, 5.1.2.3, 6.2.3.1.1

Table 1 Summary of Change

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
		with a ratio of 1:1. This was discovered after 64% of participants have been enrolled. As a consequence, randomization will continue with a 1:1 ratio and approximately 136 participants will be randomized in Cohort 3. Dose modifications for 5-FU and leucovorin will be presented separately for better clarity of each individual drug.	Dose modifications for 5-FU and leucovorin will be presented separately, see section 6.5.3.1.

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study C4221015. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint and key secondary endpoint definition or corresponding analyses will also be reflected in a protocol amendment.

As this is an open-label study, the treatment is unblinded on participant level due to different treatment regimens in each treatment arm. However, the aggregate/cumulative data summary by actual treatment arm should remain blinded to the Sponsor and the external investigators until the database lock for the primary analysis. Before the database release date for the ORR analysis of Phase 3 and Cohort 3 of the study, PK concentration and safety data will be early unblinded and provided to a restricted group of pharmacometricians/programmers in order to support preliminary modeling and simulation work ahead of formal analysis of ORR for the planned sNDA. Final exposure-safety analyses will be conducted based on the formal database snapshot for the ORR analysis.

2.1. Study Objectives, Endpoints, and Estimands

Safety Lead-in	
Objectives	**Endpoints**
Primary	
• To determine the safety and tolerability of EC + mFOLFOX6 and EC + FOLFIRI	• Incidence of DLTs
Secondary	
• To assess the overall safety and tolerability of EC + mFOLFOX6 and EC + FOLFIRI	• Incidence and severity of AEs graded according to the NCI CTCAE v4.03 and changes in clinical laboratory parameters, vital signs and ECGs • Incidence of dose interruptions, dose modifications and discontinuations due to AEs
• To estimate the efficacy of EC + mFOLFOX6 and EC + FOLFIRI	• ORR by investigator, defined as the proportion of participants who have achieved a confirmed BOR (CR or PR) per RECIST v1.1 • DOR by investigator, defined as the time from the date of first radiographic evidence of response (CR or PR) to the earliest documented disease progression per RECIST v1.1, or death due to any cause • PFS by investigator, defined as the time from the first dose to the earliest documented disease progression per RECIST v1.1, or death due to any cause • TTR by investigator, defined as the time from first dose to first radiographic evidence of response (CR or PR) per RECIST v1.1

• To estimate the efficacy of EC + mFOLFOX6 and EC + FOLFIRI	• OS defined as the time from the first dose to death due to any cause
• To characterize the PK of encorafenib, irinotecan, oxaliplatin and relevant metabolites	• PK parameters of encorafenib, irinotecan, oxaliplatin and relevant metabolites
• To assess drug-drug interaction of encorafenib with irinotecan or oxaliplatin	• Changes in exposures of irinotecan and its metabolite (SN-38) on Cycle 1 Day 15 compared to Cycle 1 Day 1 in Cohort 1 (EC + FOLFIRI) • Changes in exposures of oxaliplatin on Cycle 1 Day 15 compared to Cycle 1 Day 1 in Cohort 2 (EC + mFOLFOX6)
Tertiary/Exploratory	
• To understand the relationship between the therapeutic intervention(s) being studied and the biology of the participant's disease	• Measurements of biomarkers, consisting of DNA, RNA, protein or defined cell types, resulting from analyses of peripheral blood and/or tumor tissue biospecimen obtained at baseline, on treatment and/or at end-of-study
Phase 3	
Objectives	**Endpoints**
Primary	
• To compare the efficacy of EC + mFOLFOX6 (Arm B) vs SOC (Control Arm [Arm C]) as measured by PFS and by ORR	• PFS by BICR, defined as the time from the date of randomization to the earliest documented disease progression per RECIST v1.1, or death due to any cause • ORR by BICR
Key Secondary	
• To further compare the efficacy of Arm B vs the Control Arm as measured by OS	• OS, defined as the time from the date of randomization to death due to any cause
Secondary (Descriptive Statistics Only)	
• To further evaluate the efficacy of Arm B vs the Control Arm as measured by ORR, DOR, PFS, PFS2 and TTR • To evaluate the efficacy of EC (Arm A) vs the Control Arm as measured by ORR, DOR, PFS, PFS2, TTR and OS • To evaluate the efficacy of Arm A vs Arm B as measured by OS, PFS, PFS2, ORR, DOR and TTR	• ORR by investigator • ORR by BICR (Arm A vs Control Arm, Arm A versus Arm B) • DOR by BICR and by investigator • PFS by BICR (Arm A vs Control Arm, Arm A versus Arm B) • OS (Arm A vs Control Arm, Arm A versus Arm B) • PFS by Investigator • TTR (by BICR and by investigator), defined as the time from the date of randomization to first radiographic evidence of response (CR or PR) per RECIST v1.1 • PFS2, defined as the time from the date of randomization to the date of discontinuation of next-line treatment after first objective PD by

	investigator assessment, the second objective disease progression, or death from any cause, whichever occurs first
• To determine the safety and tolerability of EC • To determine the safety and tolerability of EC + mFOLFOX6	• Incidence and severity of AEs graded according to the NCI CTCAE v4.03 and changes in clinical laboratory parameters, vital signs, and ECGs
• To evaluate quality of life and health states, captured by PRO measures	• PRO scores as measured by the EORTC QLQ-C30, EQ-5D-5L, and anchoring instruments PGIS and PGIC
• To evaluate trough concentrations of encorafenib and its metabolite LHY746 in Arm A and Arm B	• Trough plasma concentrations of encorafenib and the metabolite LHY746 in Arm A and Arm B
• To characterize the PK of encorafenib and its metabolite LHY746 in participants randomized in mainland China (Arm A and Arm B)	• PK parameters of encorafenib and its metabolite LHY746
• To confirm the MSI-status in tumor tissue	• Summarize MSI-status as determined by retrospective central testing of baseline tumor tissue
• To determine the correlation between ctDNA, BRAF V600 alterations, and clinical outcome	• ctDNA levels and *BRAF* V600 VAF from ctDNA analysis of plasma samples collected at baseline and on treatment
Tertiary/Exploratory	
• To understand the relationship between the therapeutic intervention(s) being studied and the biology of the participant's disease	• Measurements of biomarkers, consisting of DNA, RNA, protein or defined cell types, resulting from analyses of peripheral blood and/or tumor tissue biospecimen obtained at baseline, on treatment and/or at end-of-study
• To understand the surgical conversion rate	• Surgical conversion rate, defined as the rate of participants who become eligible for surgery and undergo surgery with curative intent as a result of study intervention.
Cohort 3	
Objectives	**Endpoints**
Primary	
• To compare the efficacy of EC + FOLFIRI (Arm D) vs FOLFIRI with or without bevacizumab (Control Arm [Arm E]) as measured by ORR	• ORR by BICR
Key Secondary	
• To further compare the efficacy of Arm D vs Arm E as measured by PFS	• PFS by BICR, defined as the time from the date of randomization to the earliest documented disease progression per RECIST v1.1, or death due to any cause

Secondary (Descriptive Statistics Only)	
• To further evaluate the efficacy of Arm D vs Arm E as measured by ORR, DOR, PFS, TTR and OS	• ORR by Investigator • DOR by BICR and by Investigator, defined as the time from the date of first radiographic evidence of response (CR or PR) to the earliest documented disease progression per RECIST v1.1, or death due to any cause • PFS by Investigator, defined as the time from the date of randomization to the earliest documented disease progression per RECIST v1.1, or death due to any cause • OS, defined as the time from the date of randomization to death due to any cause • TTR (by BICR and by Investigator), defined as the time from the date of randomization to first radiographic evidence of response (CR or PR) per RECIST v1.1
• To determine the safety and tolerability of EC + FOLFIRI	• Incidence and severity of AEs graded according to the NCI CTCAE v4.03 and changes in clinical laboratory parameters, vital signs, and ECGs
• To evaluate quality of life and health states, captured by PRO measures	• PRO scores as measured by the EORTC QLQ-C30, EQ-5D-5L, and anchoring instruments PGIS and PGIC
• To evaluate trough concentrations of encorafenib and its metabolite LHY746 in Arm D	• Trough plasma concentrations of encorafenib and the metabolite LHY746 in Arm D
• To confirm the MSI status in tumor tissue	• Summarize MSI-status as determined by retrospective central testing of baseline tumor tissue
• To determine the correlation between ctDNA levels, BRAF V600 alterations, and clinical outcome	• ctDNA levels and BRAF V600 VAF from ctDNA analysis of plasma samples collected at baseline and on treatment •
Tertiary/Exploratory	
• To understand the relationship between the therapeutic intervention(s) being studied and the biology of the participant's disease	• Measurements of biomarkers, consisting of DNA, RNA, protein or defined cell types, resulting from analyses of peripheral blood and/or tumor tissue biospecimen obtained at baseline, on treatment and/or at end-of-study
• To understand the surgical conversion rate	• Surgical conversion rate, defined as the rate of participants who become eligible for surgery and undergo surgery with curative intent as a result of study intervention.

2.1.1. Primary Estimands

Estimands are defined below for the primary endpoint of incidence of dose-limiting toxicities (DLTs) in the Safety Lead-in (SLI) as well as the primary and key secondary endpoints in the Phase 3 portion and Cohort 3 of the study.

Safety Lead-in

The primary estimand for the SLI portion of the study is the DLT rate estimated based on data from DLT-evaluable participants during the DLT-evaluation period, which is the first 28 days after the first dose of study intervention in the SLI. It includes the following attributes:

- Population: all participants with local or central laboratory confirmed *BRAF* V600E- metastatic colorectal cancer (mCRC), as defined by the screening inclusion/exclusion criteria in Section 5.1.2 and Section 5.2.2 in the protocol to reflect the targeted population of the SLI portion of the study.

- Variable: DLT rate during the DLT-evaluation period, which is the first 28 days after the first dose of study intervention in the SLI. DLTs are defined in Section 6.1.1.1 in the study protocol.

- Intercurrent events: hypothetical strategy will be applied for the intercurrent events. The intercurrent event is treatment discontinuation for reasons other than treatment-related toxicity that leads to <75% of the planned dose of each study intervention during the DLT evaluation period. Participants without DLTs and with the intercurrent event will not be included in the DLT rate calculation. See Section 4 for more details on the DLT evaluable definition.

- Population-level summary measure: DLT rate defined as the number of DLT-evaluable participants with DLTs in the DLT-evaluation period divided by the number of DLT-evaluable participants.

Phase 3, Two Primary Estimands

The primary estimands in the Phase 3 portion of the study are the treatment effect in progression-free survival (PFS) by blinded independent central review (BICR) and in ORR by BICR per Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1) for Arm B vs Arm C. They are the hypothetical estimands and include the following 4 attributes.

PFS for Arm B vs Arm C:

- Population: all participants who are randomized in the Phase 3 portion of the study in Arm B and Arm C. This analysis population will be based on FAS defined in Section 4.

- Variable: PFS by BICR, defined as the time from the date of randomization to the earliest documented disease progression per RECIST v1.1, or death due to any cause.

- Intercurrent events: Details of the intercurrent events and censoring rules for the primary analysis are summarized in Table 4.

- Population-level summary measure: HR for PFS and corresponding 2-sided 95.4% CI based on Cox proportional hazard model stratified by randomization strata. PFS will be compared between the two treatment arms using a 1-sided stratified log-rank test.

OR for Arm B vs Arm C:

- Population: the first 110 participants who are randomized in the Phase 3 portion of the study into each of Arm B and Arm C (Full Analysis Set, ORR Subset defined in Section 4).

- Variable: Objective response defined as complete response (CR), or partial response (PR) according to RECIST v1.1 based on BICR assessment, from the date of randomization until the date of the first documentation of progression of disease (PD). Both CR and PR must be confirmed by repeat assessments performed no less than 4 weeks after the criteria for response are first met. Only tumor assessments performed on or before the start date of subsequent anticancer therapies will be considered in the assessment.

- Intercurrent events: All data collected after an intercurrent event of subsequent anticancer therapy or first PD will be excluded.

- Population-level summary measure: stratified (by the randomization strata) odds ratio in terms of OR defined as the odds of OR with Arm B divided by the odds of OR with Arm C, with its 2-sided 99.8% CI. ORR treatment effect will be compared using Cochran-Mantel-Haenszel test stratified by the randomization strata.

Cohort 3

The primary estimand in the Cohort 3 portion of the study is the treatment effect in OR by BICR per RECIST v1.1 for Arm D vs Arm E. It includes the following 4 attributes:

- Population: all participants who are randomized in Cohort 3. This analysis population will be based on FAS defined in Section 4.

- Variable: Objective response defined as CR, or PR according to RECIST v1.1 based on BICR assessment, from the date of randomization until the date of the first documentation of PD, death or start of new anticancer therapy. Both CR and PR must be confirmed by repeat assessments performed no less than 4 weeks after the criteria for response are first met. Only tumor assessments performed on or before the start date of subsequent anticancer therapies will be considered in the assessment.

- Intercurrent events: All data collected after an intercurrent event of subsequent anticancer therapy or first PD will be excluded.

- Population-level summary measure: stratified (by the randomization strata) odds ratio in terms of OR defined as the odds of OR with Arm D divided by the odds of OR

with Arm E, with its 2-sided 95% CI. ORR treatment effect will be compared using Cochran-Mantel-Haenszel test stratified by the randomization strata.

2.1.2. Secondary Estimand

Phase 3

The key secondary estimand in the Phase 3 portion of the study is the treatment effect in overall survival (OS) for Arm B vs Arm C. It is the hypothetical estimand and includes the following 4 attributes:

- Population: all participants who are randomized in the Phase 3 portion of the study in Arm B and Arm C. This analysis population will be based on FAS defined in Section 4.
- Variable: OS, defined as the time from date of randomization to death due to any cause.
- Intercurrent events: Participants not known to have died are censored at the date of last contact when participant was known to be alive.
- Population-level summary measure: HR for OS and corresponding 2-sided CI based on Cox proportional hazard model stratified by randomization strata. Level of CI will depend on alpha spent for the analysis. OS will be compared between the two treatment arms using a 1-sided stratified log-rank test.

Cohort 3

The key secondary estimand in the Cohort 3 portion of the study is the treatment effect in PFS by BICR. It is the hypothetical estimand and includes the following 4 attributes:

- Population: all participants who are randomized in the Cohort 3. This analysis population will be based on FAS defined in Section 4.
- Variable: PFS by BICR, defined as the time from the date of randomization to the earliest documented disease progression per RECIST v1.1, or death due to any cause.
- Intercurrent events: details of the intercurrent events and censoring rules are summarized in Table 4.
- Population-level summary measure: HR for PFS and corresponding 2-sided 95% CI based on Cox proportional hazard model stratified by the randomization strata. PFS will be compared between the two treatment arms using a 1-sided stratified log-rank test.

2.1.3. Additional Estimands

The following supportive estimands are defined for the Phase 3 and for the Cohort 3 portion of the study:

Supportive Estimand 1 (OR):

- Population: all participants with a central laboratory confirmed result of *BRAF* V600E-mutant mCRC, as defined by the screening inclusion/exclusion criteria in Section 5.1.2 and Section 5.2.2 in the study protocol to reflect the targeted population of the Phase 3 portion and Cohort 3 of the study.
- Variable: same as primary estimand of OR.
- Intercurrent events: same as primary estimand of OR.
- Population-level summary measure: same as primary estimand of OR.

Supportive Estimand 1 (PFS):

- Population: all participants with a central laboratory confirmed result of *BRAF* V600E-mutant mCRC, as defined by the screening inclusion/exclusion criteria in Section 5.1.2 and Section 5.2.2 in the study protocol to reflect the targeted population of the Phase 3 portion and Cohort 3 of the study.
- Variable: same as primary estimand of PFS.
- Intercurrent events: same as primary estimand of PFS.
- Population-level summary measure: same as primary estimand of PFS.

Supportive Estimand 2 (PFS):

- Population: same as primary estimand of PFS.
- Variable: same as primary estimand of PFS.
- Intercurrent events: treatment policy strategy will be applied for the intercurrent event of starting new anticancer therapy. All other intercurrent events will be addressed in the same approach as the primary estimand of PFS. Table 7 summarizes the details of the intercurrent events and censoring rules for this supportive analysis.
- Population-level summary measure: same as primary estimand of PFS.

Supportive Estimand 3 (OS):

- Population: all participants with a central laboratory confirmed result of *BRAF* V600E-mutant mCRC, as defined by the screening inclusion/exclusion criteria in Section 5.1.2 and Section 5.2.2 in the study protocol to reflect the targeted population of the Phase 3 portion and Cohort 3 of the study.
- Variable: same as key secondary estimand of OS.
- Intercurrent events: same as key secondary estimand of OS.

- Population-level summary measure: same as key secondary estimand of OS.

2.2. Study Design

C4221015 study is an open-label, multicenter, randomized Phase 3 study of encorafenib plus cetuximab (EC) with or without chemotherapy vs standard of care chemotherapy in participants with previously untreated *BRAF* V600E-mutant mCRC. It includes a SLI portion, a randomized Phase 3 portion, and a randomized Cohort 3.

Prior to the Phase 3 portion, a SLI will be conducted at a limited number of sites to evaluate the safety/tolerability and PK of EC in combination with either mFOLFOX6 or FOLFIRI. Approximately 60 participants will be enrolled in the SLI portion (up to 30 per regimen). The results of the SLI will inform which chemotherapy regimen is used in Arm B of the Phase 3 portion of the study.

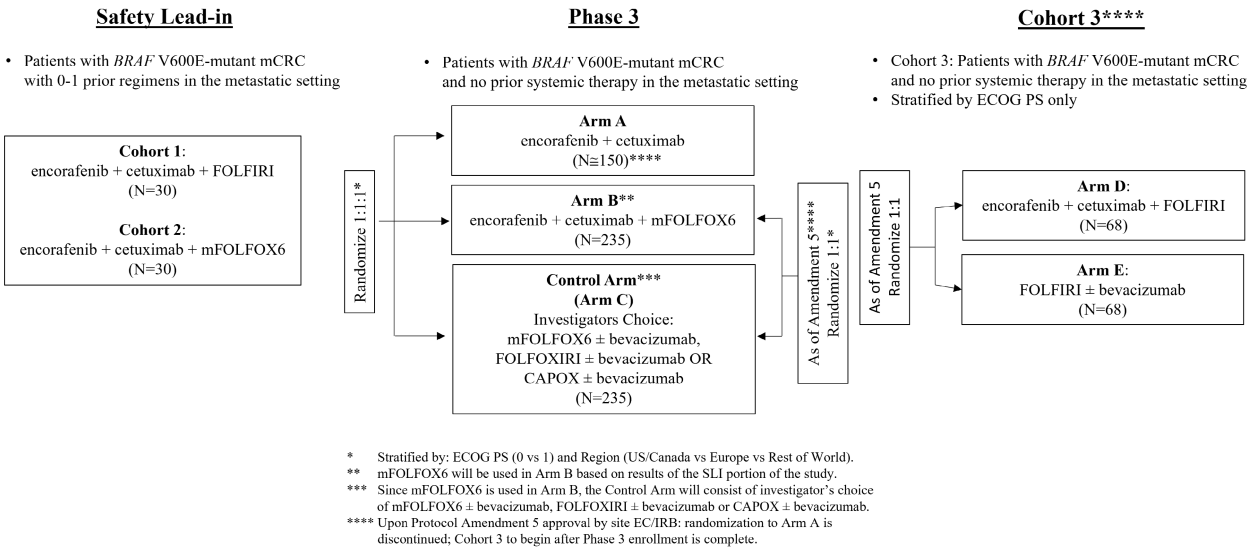
After reviewing the totality of the SLI data, the E-DMC and SC determined that there were no new safety signals and that the study can proceed as planned. As a result of the SLI data review, Arm B will consist of EC in combination with mFOLFOX6. In addition, Phase 3 treatment Arm C will consist of investigator's choice of mFOLFOX6 ± bevacizumab, FOLFOXIRI ± bevacizumab or CAPOX ± bevacizumab.

In the Phase 3 portion, approximately 620 participants will be randomized overall, initially at a ratio of 1:1:1 to receive EC (Arm A), EC + chemotherapy (Arm B) or standard of care (SOC) chemotherapy (Arm C), and then 1:1 to receive EC + chemotherapy (Arm B) or SOC chemotherapy (Arm C) after the approval of Protocol Amendment 5, with a total of approximately 150 participants for Arm A and approximately 235 participants each for Arm B and Arm C. The primary objective in the Phase 3 portion of the study is to compare the efficacy, as measured by the primary endpoints of PFS by BICR and ORR by BICR, of Arm B versus Arm C. Phase 3 portion will be considered positive (ie, demonstrated evidence of effectiveness) if the hypothesis test of either primary endpoint is statistically significant.

With protocol amendment 5, a Cohort 3 with 2 randomized arms will be included. Approximately 136 participants will be enrolled in Cohort 3 at a ratio of 1:1 to receive EC + FOLFIRI (Arm D, n=68) or FOLFIRI with or without bevacizumab (Arm E, n=68). Cohort 3 will start after the enrollment of the Phase 3 portion is complete. The primary objective in the Cohort 3 portion of the study is to compare the efficacy, as measured by the primary endpoint of ORR by BICR, of Arm D versus Arm E. Cohort 3 portion will be considered positive (ie, demonstrated evidence of effectiveness) if the hypothesis test of the primary endpoint is statistically significant.

A schema of the study design is presented in Figure 1.

Figure 1 Study Design



* Stratified by: ECOG PS (0 vs 1) and Region (US/Canada vs Europe vs Rest of World).
** mFOLFOX6 will be used in Arm B based on results of the SLI portion of the study.
*** Since mFOLFOX6 is used in Arm B, the Control Arm will consist of investigator's choice of mFOLFOX6 ± bevacizumab, FOLFOXIRI ± bevacizumab or CAPOX ± bevacizumab.
**** Upon Protocol Amendment 5 approval by site EC/IRB: randomization to Arm A is discontinued; Cohort 3 to begin after Phase 3 enrollment is complete.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Safety Lead-in

3.1.1. Primary Endpoint

The primary endpoint for the SLI is incidence of DLTs estimated based on data from DLT-evaluable participants during the DLT-evaluation period, which is the first 28 days after the first dose of study intervention in the SLI. The criteria of DLTs are specified in Section 6.1.1.1 in the protocol.

3.1.2. Secondary Endpoints

3.1.2.1. Safety Endpoints

- Incidence and severity of adverse events (AEs) graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events version 4.03 (NCI CTCAE v4.03) and changes in clinical laboratory parameters, vital signs and ECGs.
- Incidence of dose interruptions, dose modifications and discontinuations due to AEs.

3.1.2.2. Efficacy Endpoints

Objective Response Rate

Objective response rate (ORR) is defined as the proportion of participants who have achieved a best overall confirmed response (BOR) of complete response (CR) or partial response (PR) per RECIST v1.1. The BOR is the best response obtained among all tumor assessment visits after the date of first dose until documented disease progression, death, or start of subsequent

anticancer therapy. Clinical deterioration or clinical progression noted on the completion eCRF will not be considered as documented disease progression.

ORR based on investigator assessment will be calculated for the SLI participants.

Duration of Response

Duration of response (DOR) is defined as the time from the date of first radiographic evidence of response (CR or PR) to the earliest documented disease progression per RECIST v1.1, or death due to any cause.

DOR based on investigator assessment will be calculated for the SLI participants with a confirmed response (CR or PR).

Progression-free Survival

PFS is defined as the time from date of the first dose to the earliest documented disease progression per RECIST v1.1, or death due to any cause.

The censoring and event date options to be considered for the PFS analysis are described in details in Section 6.1.2.1.1.

PFS based on investigator assessment will be calculated for the SLI participants.

Time to Response

Time to response (TTR) is defined as the time from the date of first dose to first radiographic evidence of response (CR or PR) per RECIST v1.1. TTR will be calculated for participants with a confirmed response (CR or PR).

TTR based on investigator assessment will be calculated for the SLI participants.

Overall Survival

OS is defined as the time from the date of first dose to death due to any cause. If a participant is not known to have died at the time of the cutoff for analysis, then OS will be censored at the date of last contact.

3.1.2.3. PK Endpoints

- PK parameters of encorafenib, irinotecan, oxaliplatin and relevant metabolites.
- Changes in exposures of irinotecan and its metabolite (SN-38) on Cycle 1 Day 15 compared to Cycle 1 Day 1 in Cohort 1 (EC + FOLFIRI).
- Changes in exposures of oxaliplatin on Cycle 1 Day 15 compared to Cycle 1 Day 1 in Cohort 2 (EC + mFOLFOX6).

3.1.3. Exploratory Endpoints

Measurements of biomarkers, consisting of DNA, RNA, protein or defined cell types, resulting from analyses of peripheral blood and/or tumor tissue biospecimen obtained at baseline, on treatment and/or at end-of-study.

3.2. Phase 3

3.2.1. Primary Endpoints

- PFS by BICR is defined as the time from date of randomization to the earliest documented disease progression per RECIST v1.1 based on BICR assessment, or death due to any cause.
  The censoring and event date options to be considered for the PFS primary analysis are described in details in Section 6.1.2.1.1.

- ORR by BICR is defined as the proportion of participants who have achieved a BOR of confirmed (by repeat assessments performed no less than 4 weeks after the criteria for response are first met) CR or PR per RECIST v1.1. The BOR, as assessed by BICR, is the best response obtained among all tumor assessment visits after the date of randomization until documented disease progression, death, or start of subsequent anticancer therapy. Participants who do not have a post-baseline tumor assessment due to early progression of disease, who receive anti-cancer therapies other than the study treatments prior to reaching a CR or PR, or who die, have PD, or stop tumor assessments for any reason prior to reaching a CR or PR will be counted as non-responders in the assessment of OR. Each participant will have an objective response status (0: no OR; 1: OR). Clinical deterioration or clinical progression noted on the completion eCRF will not be considered as documented disease progression.

3.2.2. Efficacy Secondary Endpoints

3.2.2.1. Overall Survival

The key secondary endpoint of OS is defined as the time from the date of randomization to death due to any cause. If a participant is not known to have died at the time of the cutoff for analysis, then OS will be censored at the date of last contact.

3.2.2.2. Objective Response Rate by Derived Investigator

ORR by derived investigator assessment (response will be derived programmatically from the target lesion measurements, non-target lesion status, and new lesions recorded on the CRF, see APPENDIX 2 for details) is defined as the proportion of participants who have achieved a BOR of confirmed (by repeat assessments performed no less than 4 weeks after the criteria for response are first met) CR or PR per RECIST v1.1. The BOR is the best response obtained among all tumor assessment visits after the date of randomization until documented disease progression, death, or start of subsequent anticancer therapy. Clinical deterioration or clinical progression noted on the completion eCRF will not be considered as documented disease progression.

3.2.2.3. Duration of Response

DOR in the Phase 3 portion of the study has the same definition of DOR as described in Section 3.1.2.2. Confirmed responses will be used.

Both DOR by BICR and DOR by derived investigator assessment will be calculated.

3.2.2.4. Progression-free Survival by Derived Investigator

PFS by derived investigator assessment is defined as the time from date of randomization to the earliest documented disease progression per RECIST v1.1 based on derived investigator assessment (PD will be derived programmatically from the target lesion measurements, non-target lesion status, and new lesions recorded on the case report form (CRF)), or death due to any cause.

3.2.2.5. Time to Response

TTR is defined as the time from the date of randomization to first radiographic evidence of response (CR or PR) per RECIST v1.1. TTR will be calculated for participants with a confirmed response (CR or PR).

Both TTR by BICR and TTR by derived investigator assessment will be calculated.

3.2.2.6. PFS after Next Line of Treatment (PFS2)

PFS after next line of treatment (PFS2) is defined as the time from the date of randomization to the date of discontinuation of next-line treatment after first objective PD by investigator assessment, to second objective disease progression, or death from any cause, whichever occurs first. Second objective disease progression is progressive disease (PD2) after the start of subsequent anticancer therapy based on investigator assessment.

3.2.3. Other Secondary Endpoints

3.2.3.1. Patient-Reported Outcomes Endpoints

Patient-reported outcomes (PROs) as measured by the following instruments: European Organization for Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30), the EuroQol EQ-5D-5L (EQ-5D-5L) and visual analogue scale (VAS), patient global impression of disease severity (PGIS), and patient global impression of change (PGIC).

The EORTC QLQ-C30 is a 30-question survey, which can be grouped into 5 functional domain subscales, including a physical functioning subscale, a role functioning subscale, an emotional functioning subscale, a cognitive functioning subscale and a social functioning subscale. Higher scores on the functional domains are indicative of higher levels of functioning. Oncology related symptoms assessed by the EORTC QLQ-C30 include fatigue (3 items), pain (2 items), nausea and vomiting (2 items), and dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial impact (1 item each). Higher scores are reflective of a greater presence of symptoms.

The EQ-5D-5L is a 6-item patient-completed questionnaire designed to assess health status in terms of a single index value or utility score. There are 2 components, a Health State Profile which has individuals rate their level of problems (none, slight, moderate, severe, extreme/unable) in 5 areas (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), and a VAS in which participants rate their overall health status from 0 (worst imaginable) to 100 (best imaginable). Published weights are available that allow for the creation of a single summary score. Overall scores range from 0 to 1, with lower scores representing a higher level of dysfunction.

The PGIS is a single-item PRO designed to facilitate an anchor-based methodology for establishing meaningful within person change (MWPC) of other PRO-based endpoints such as the EORTC QLQ-C30. There are 4 response options, ranging from "none" to "severe".

The PGIC is a single-item PRO designed to assess the participant's overall impression of the degree of change they have experienced since the start of study treatment. There are 5 response options, ranging from "much better" to "much worse".

3.2.3.2. Pharmacokinetic Endpoints

- Trough plasma concentrations of encorafenib and the metabolite LHY746 in Arm A and Arm B.
- PK parameters of encorafenib and its metabolite LHY746 in participants randomized in mainland China (Arm A and Arm B, as appropriate).

3.2.3.3. Biomarker Endpoints

- Microsatellite instability (MSI)-status as determined by retrospective central testing of baseline tumor tissue.
- ctDNA levels
- BRAF V600 VAF in ctDNA

3.2.4. Exploratory Endpoints

Measurements of biomarkers, consisting of DNA, RNA, protein or defined cell types, resulting from analyses of peripheral blood and/or tumor tissue biospecimen obtained at baseline, on treatment and/or at end-of-study.

3.3. Cohort 3

3.3.1. Primary Endpoint

ORR by BICR is defined as the proportion of participants who have achieved a BOR of confirmed (by repeat assessments performed no less than 4 weeks after the criteria for response are first met) CR or PR per RECIST v1.1. The BOR, as assessed by BICR, is the best response obtained among all tumor assessment visits after the date of randomization until documented disease progression, death, or start of subsequent anticancer therapy.

Participants who do not have a post-baseline tumor assessment due to early progression of disease, who receive anti-cancer therapies other than the study treatments prior to reaching a CR or PR, or who die, have PD, or stop tumor assessments for any reason prior to reaching a CR or PR will be counted as non-responders in the assessment of OR. Each participant will have an objective response status (0: no OR; 1: OR). Clinical deterioration or clinical progression noted on the completion eCRF will not be considered as documented disease progression.

3.3.2. Efficacy Secondary Endpoints

3.3.2.1. Progression-free Survival

The key secondary endpoint of PFS by BICR is defined as the time from date of randomization to the earliest documented disease progression per RECIST v1.1 based on BICR assessment, or death due to any cause.

PFS by derived investigator assessment is defined as the time from date of randomization to the earliest documented disease progression per RECIST v1.1 based on derived investigator assessment (PD will be derived programmatically from the target lesion measurements, non-target lesion status, and new lesions recorded on the case report form (CRF)), or death due to any cause.

The censoring and event date options to be considered for the PFS primary analysis are described in detail in Section 6.1.2.1.1.

3.3.2.2. Objective Response Rate by Derived Investigator

ORR by derived investigator assessment is defined as described in Section 3.2.2.2.

3.3.2.3. Duration of Response

DOR in the Cohort 3 portion of the study has the same definition of DOR as described in Section 3.1.2.2. Confirmed responses will be used.

Both DOR by BICR and DOR by derived investigator assessment will be calculated.

3.3.2.4. Time to Response

TTR is defined as described in Section 3.2.2.5.

Both TTR by BICR and TTR by derived investigator will be calculated.

3.3.2.5. Overall Survival

OS is defined as described in Section 3.2.2.1.

3.3.3. Other Secondary Endpoints

3.3.3.1. Patient-Reported Outcomes Endpoints

Patient-reported outcomes (PROs) as measured by the following instruments: European Organization for Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30), the

EuroQol EQ-5D-5L (EQ-5D-5L) and visual analogue scale (VAS), patient global impression of disease severity (PGIS), and patient global impression of change (PGIC), see Section 3.2.3.1 for additional details.

3.3.3.2. Pharmacokinetic Endpoints

Trough plasma concentrations of encorafenib and the metabolite LHY746 in Arm D.

3.3.3.3. Biomarker Endpoints

- Microsatellite instability (MSI)-status as determined by retrospective central testing of baseline tumor tissue.
- ctDNA levels
- BRAF V600 VAF in ctDNA

3.3.4. Exploratory Endpoints

Measurements of biomarkers, consisting of DNA, RNA, protein or defined cell types, resulting from analyses of peripheral blood and/or tumor tissue biospecimen obtained at baseline, on treatment and/or at end-of-study.

3.4. Baseline Variables

The date of first dose of study treatment is the earliest date of non-zero dosing of any study drug. The date of last dose of study treatment is the latest date of non-zero dosing of any study drug.

For all endpoint assessments in the SLI and all safety assessments in the Phase 3 portion and in Cohort 3 of the study, baseline is defined as the last completed assessment prior to date of first dose of study treatment in the SLI, the Phase 3, and Cohort 3, respectively. If an assessment that is planned to be performed prior to the first dose of study treatment in the protocol is performed on the same day as the first dose of study treatment and the time is unknown, it will be assumed that it was performed prior to study treatment administration and will be considered as baseline assessment.

For efficacy assessments in the Phase 3 portion and in Cohort 3 of the study, baseline is defined as the last assessment prior to randomization. If the assessment that is planned to be performed before randomization per the protocol is performed on the same day as the date of randomization and assessment time point is missing, it will be assumed that it was performed prior to randomization and will be considered as baseline assessment. If no tumor assessment was performed prior to randomization, any tumor assessment conducted between randomization and prior or on the first study intervention date will be considered as baseline tumor assessment.

For PRO the last measurement prior to or on the date of first dose of study treatment will be used as the baseline measurement. If there are no measurements meeting these criteria, then baseline is considered missing.

Unscheduled assessments will be used in the determination of baseline. Data reported at the End of Treatment visit are not eligible for baseline selection.

3.4.1. Stratification

Randomization for Phase 3 is stratified by the following factors as recorded in the Interactive Response Technology (IRT) system:

- ECOG performance status: 0 vs 1.
- Region: US/Canada vs Europe vs Rest of World.

Randomization for Cohort 3 is stratified by ECOG performance status (0 vs 1) as recorded in the IRT system.

All stratified analysis will utilize the strata defined in the IRT system.

3.4.2. Covariates

In addition to the randomization strata, the following baseline characteristics variables will be included as covariates in the Cox regression model to explore the potential prognostic impact on PFS per BICR and OS:

- Age (<65 years vs ≥65 years).
- Gender (Male vs Female).
- Race (White vs non-White).
- Race (Asian vs non-Asian).
- Number of organs involved at baseline (≤2 vs ≥3) by BICR.
- *BRAF* V600E mutation per central assessment (Detected vs Not Detected/Indeterminate).
- Prior anticancer adjuvant/neoadjuvant therapy (yes vs no).
- Removal status of primary tumor (no resection/partial resection vs complete resection).
- Side of tumor (left/right vs left vs right).
- Presence of liver metastases at baseline, based on target and non-target lesion assessment (yes vs no) (based on BICR).

3.5. Safety Endpoints

Safety endpoints will be summarized based on the on-treatment period (defined in Section 5.2.7) unless otherwise specified. Safety data collected outside the on-treatment period will be listed but not summarized.

3.5.1. Adverse Events

An AE is considered a treatment-emergent adverse event (TEAE) if the event starts on or after the first dosing day and time/start time, if collected, but before the last dose +28 days, or start of subsequent anticancer therapy minus 1 day, whichever occurs first. Any events that started prior to the first dose date are not considered TEAEs. If an AE starts on the same day as the first dose date, it will be considered treatment emergent unless the CRF data indicates otherwise via explicitly recording time for AE onset and treatment dosing.

Adverse events will be coded using current Medical Dictionary for Regulatory Activities (MedDRA). Severity of AEs will be graded using the NCI-CTCAE v 4.03 toxicity grading scale.

3.5.2. Laboratory Data

Hematology, chemistry and coagulation tests results will be programmatically graded according to the NCI CTCAE version 4.03 for relevant parameters. Additional details are provided in Section 6.6.4.

3.5.3. Vital Signs Data

Vital signs data includes weight, pulse, systolic blood pressure (BP), and diastolic BP will be summarized. Additional details are provided in Section 6.6.5.

3.5.4. Electrocardiograms Data

Clinically notable ECG values during the on-treatment period will be summarized. Additional details are provided in Section 6.6.6.

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

Data for all participants will be assessed to determine if participants meet the criteria for inclusion in each analysis population prior to unblinding and releasing the database and classifications will be documented per standard operating procedures.

Participant Analysis Set	Description
Enrolled	All participants who sign the Screening informed consent/assent document.
DLT-Evaluable Analysis Set	All participants who receive at least 1 dose of study drug in the SLI and either experience DLT during the DLT evaluation period or complete the DLT evaluation period without DLT.

Participant Analysis Set	Description
	Participants who complete the DLT evaluation period without DLTs before receiving at least 75% of the planned dose of each of the three study drugs (encorafenib, cetuximab, and mFOLFOX6 or FOLFIRI) during the DLT evaluation period for reasons other than treatment-related toxicity are not evaluable for DLT. Examples of that include non-compliance, protocol deviations, and other issues (e.g. car accident, COVID issues). Patients who complete the DLT evaluation period without DLTs and received less than 75% of the planned dose on one or two study drugs (not all three) due to an unrelated toxicity are DLT evaluable. Patients who complete the DLT evaluation period and without DLTs who receive less than 75% of the planned dose for one or two study drugs, due to treatment-related toxicity, are DLT evaluable. Also such patients who receive less than 75% of the planned dose for all three study drugs, due to treatment-related toxicity, are also DLT evaluable (since this event would constitute an actual DLT). The DLT evaluation period is the first 28 days after the first dose of study intervention. Participants in the SLI will be analyzed according to the actual study treatment received.
Full Analysis Set (FAS)	For participants in the SLI, the FAS includes all participants who receive at least 1 dose of study drug. Note for the SLI, this set plays the role of the safety analysis set. Participants in the SLI will be analyzed according to the actual study treatment received. For participants in the Phase 3 portion and Cohort 3, the FAS includes all participants who are randomized in the Phase 3 portion and Cohort 3 of the study, respectively. Participants will be analyzed according to the study treatment assigned at randomization.
Full Analysis Set, ORR Subset	For primary analysis of ORR/DOR/TTR in the Phase 3 portion of the study, "Full Analysis Set, ORR Subset" comprises the first 110 participants randomized in each Arm B and Arm C.
Central *BRAF* V600E Positive Analysis Set	All participants who are randomized in the Phase 3 portion and in Cohort 3 of the study with a confirmed central laboratory result of *BRAF* V600E mutation. Participants will be analyzed according to the study treatment assigned at randomization.
Safety Analysis Set	All participants who receive at least 1 dose of study drug. For the Phase 3 portion and Cohort 3 of the study, participants will be analyzed according to the study treatment assigned at

Participant Analysis Set	Description
	randomization unless the incorrect treatment(s) was/were received throughout the dosing period in which case participants will be analyzed according to the actual study treatment received in the first cycle.
PK Analysis Set	The PK concentration set is defined as all enrolled participants who are treated and have at least 1 analyte concentration. The PK parameter analysis set is defined as all enrolled participants treated who have sufficient information to estimate at least 1 of the PK parameters of interest and have no major protocol deviations affecting PK assessment. Participants will be analyzed according to the actual study treatment received.
Biomarker Analysis Set	All participants who are in the safety set and who had at least 1 of the pharmacodynamics or biomarkers evaluated at pre and/or post dose. Analysis sets will be defined separately for tumor tissue-based and blood-based biomarkers and will also be defined separately for non-paired and paired biomarkers. The tumor tissue analysis set is defined as participants from the safety analysis set who had a tumor biomarker assessment at baseline. The cfDNA analysis set is defined as participants from the safety analysis set who have a blood-based biomarker assessment at baseline. The paired cfDNA analysis set is defined as all patients in the safety analysis set who have blood-based biomarker data at baseline and a corresponding post-treatment timepoint (eg, C1D15, C2D1, C2D15, EOT). The cfDNA diagnostics concordance analysis set is defined as participants who have both a central tumor tissue result and a central plasma result.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

5.1.1. Hypotheses and Sample Size Calculation

The planned sample size of the study is approximately 816 participants including approximately 60 participants in the SLI, approximately 470 participants in the Phase 3

portion of the study (Arm B and Arm C), and approximately 136 participants in Cohort 3. In addition, up to the time of Protocol Amendment 5 approval, approximately 150 participants are estimated to be randomized in Arm A of the Phase 3 portion of the study.

5.1.1.1. Safety Lead-in

There is no statistical hypothesis for the SLI portion of the study.

The primary endpoint of the SLI is incidence of DLTs. Up to 60 participants will be enrolled on a rolling basis in an alternating manner to receive either EC + mFOLFOX6 or EC + FOLFIRI (up to 30 participants per cohort). The safety data will be evaluated by the Steering Committee and External Data Monitor Committee (E-DMC) after the first 9 evaluable participants in each cohort have been followed for a minimum of 28 days with EC + mFOLFOX6 or EC + FOLFIRI. The committees will also review the cumulative safety and PK data at the end of the SLI when all evaluable participants in each of the two cohorts have been followed for a minimum of 28 days.

5.1.1.2. Phase 3

A total of approximately 620 participants will be randomized, initially in a 1:1:1 ratio, stratified by ECOG performance status (0 vs 1) and region (US/Canada vs Europe vs Rest of World), to one of the three treatment arms, and then 1:1 to receive EC + chemotherapy (Arm B) or SOC chemotherapy (Arm C) after the approval of Protocol Amendment 5, with a total of approximately 150 participants for Arm A and 235 participants each for Arm B and Arm C. It is assumed that participant accrual will reach a maximum of 36 participants per month for an accrual duration of approximately 26 months.

Primary Endpoint

The primary objective in the Phase 3 portion of the study is to compare the efficacy, as measured by the primary endpoints of PFS by BICR and ORR by BICR, of Arm B versus Arm C. The following statistical hypotheses will be tested to address the primary objective:

$$H_{01}: HR_{PFS\ (B\ versus\ Control)} \geq 1 \text{ vs } H_{11}: HR_{PFS\ (B\ versus\ Control)} < 1$$

$$H_{02}: OR_{OR\ (B\ versus\ Control)} \leq 1 \text{ versus } H_{12}: OR_{OR\ (B\ versus\ Control)} > 1$$

where $HR_{PFS\ (B\ versus\ Control)}$ is the HR for PFS of Arm B versus the Control Arm and $OR_{OR\ (B\ versus\ Control)}$ is the odds ratio for objective response of Arm B versus the Control Arm.

Approximately 230 PFS events by BICR will be required to have at least 85% power to detect a hazard ratio of 0.67 using a one-sided stratified log-rank test at a significance level of 0.023.

The sample size of 235 participants per arm was determined based on the assumption of a hazard ratio of 0.67 under the exponential model assumptions with median PFS of 7 months on the control arm and 10.4 months on Arm B. Further assumptions include a drop-out hazard rate of 0.07 for each treatment arm and a non-uniform participants accrual over 26 months (for 2 arms: 6 participants per month for the first 5 months, 11 participants a month

for months 6 through 7, 16 participants a month for months 8 through 11, 23 participants a month for months 12 through 16, and 24 participants per month thereafter).

The sample size of 220 participants (110 per arm) will provide 90% power for test of odds ratio of 2 proportions between the Arm B and Arm C using a 1-sided chi-square test at a significance level of 0.001 assuming an ORR by BICR of 35% and 65% for Arm C and Arm B, respectively.

Key Secondary Endpoint

The key secondary objective is to compare the efficacy, as measured by the key secondary endpoint of OS, of Arm B versus Arm C. The following statistical hypothesis will be tested to address the key secondary objectives:

$$H_{03}: HR_{OS\ (B\ versus\ Control)} \geq 1 \text{ vs } H_{13}: HR_{OS\ (B\ versus\ Control)} < 1$$

where $HR_{OS\ (B\ versus\ Control)}$ is the HR for OS of Arm B versus the Control Arm.

For the comparison of OS for Arm B versus the Control Arm, 297 OS events will achieve 85% power to detect an OS HR of 0.70 using a 1-sided stratified log-rank test at a significance level of 0.023, and a 2-look group-sequential design with a Lan-DeMets (O'Brien-Fleming) spending function to determine the efficacy boundary[1]. An exponential distribution for OS was assumed, that corresponds to an improvement in the median OS of 6.4 months (21.4 months in Arm B versus 15 months in the Control Arm). The sample size further assumes a dropout hazard rate of 0.011 within each treatment arm, a non-uniform participants accrual over 26 months (same accrual over time as assumed for the primary endpoint), interim analyses after 80% of OS events, and follow-up after the last participant is randomized of about 24 months.

5.1.1.3. Cohort 3

A total of approximately 136 participants will be randomized in a 1:1 ratio, stratified by ECOG performance status (0 vs 1), to receive EC + FOLFIRI (Arm D, n=68) or FOLFIRI with or without bevacizumab (Arm E, n=68). It is assumed that participant accrual will reach about 20 participants per month for an accrual duration of approximately 7 months.

Primary Endpoint

The primary objective in Cohort 3 is to compare the efficacy, as measured by ORR by BICR, of Arm D versus Arm E. The following statistical hypothesis will be tested:

$$H_{01}: OR_{OR\ (D\ versus\ E)} \leq 1 \text{ versus } H_{11}: OR_{OR\ (D\ versus\ E)} > 1$$

where $OR_{OR\ (D\ versus\ E)}$ is the odds ratio for objective response of Arm D versus Arm E.

The sample size of 136 participants (68 in Arm D and 68 in Arm E) will provide 93.1% power for test of odds ratio of 2 proportions between the Arm D and Arm E using a 1-sided chi-square test at a significance level of 0.025 assuming an ORR by BICR of 35% and 65% for Arm E and Arm D, respectively.

Key Secondary Endpoint

The key secondary objective is to compare the efficacy, as measured by the key secondary endpoint of PFS by BICR, of Arm D versus Arm E. The following statistical hypothesis will be tested to address the key secondary objective:

H_{02}: $HR_{PFS\ (D\ versus\ E)} \geq 1$ versus H_{12}: $HR_{PFS\ (D\ versus\ E)} < 1$

where $HR_{PFS\ (D\ versus\ E)}$ is the hazard ratio for PFS of Arm D versus Arm E

For the comparison of Arm D versus Arm E PFS, 136 participants and 73 PFS by BICR events in Arm D + Arm E will achieve 80% power to detect a PFS HR of 0.519 using a 1-sided stratified log-rank test at a significance level of 0.025. An exponential distribution for PFS was assumed, that corresponds to an improvement in the median PFS of 6.5 months (13.5 months in Arm D versus 7 months in Arm E). The sample size further assumes a dropout hazard rate of 0.015 within each treatment arm, uniform participants accrual of 20 participants a month.

5.1.2. Decision Rules

5.1.2.1. Safety Lead-in

The target DLT rate in the first 28 days in the SLI is <33% (ie, <3 out of 9 evaluable participants with DLTs).

Table 2 provides a comparison of the characteristics of this DLT evaluation rule and the traditional 3 + 3 rules based on 9 participants.

Table 2 Characteristics of DLT Probability During the DLT Evaluation Period

True DLT Rate During DLT Evaluation Period	Probability of Dose Declared Toxic Using 3 + 3 Rules	Probability of Observed DLT Rate ≥33% in 9 Participants During DLT Evaluation Period
10%	0.094	0.053
20%	0.291	0.262
30%	0.506	0.537
40%	0.691	0.768
50%	0.828	0.910

The DLT evaluation period is the first 28 days after the first dose of study intervention in the SLI, and the DLT-evaluable is defined in Section 4.

If the doses are determined to be tolerable in the first 9 evaluable participants in either or both regimens based on the observed DLT rate (<33%) and evaluation of the overall toxicity profile, the SLI will be expanded for a total of up to 30 evaluable participants in each cohort. If the DLT rate is ≥33% in either cohort, that combination regimen will not be evaluated further in the SLI and will not be used in the Phase 3. Table 3 summarizes the probability of

observing at least one instance of a toxicity that has a true incidence rate of 10% and 15% for the given number of participants.

Table 3 Probability of Observing ≥ 1 Instance of Toxicity at 10% and 15% Incidence Rate

True Toxicity Incidence Rate	Number of Participants	Probability of Observing ≥1 Instance of Toxicity
10%	15	0.79
	20	0.88
	30	0.96
15%	15	0.91
	20	0.96
	30	0.99

If any participant is deemed non-evaluable for DLT, additional participants may be enrolled to ensure there are a sufficient number of evaluable participants in the SLI.

5.1.2.2. Phase 3

For the Phase 3 portion of the study a hierarchical testing procedure will be used to control the family-wise type I error rate[2].

Eight months after randomization of the first 110 participants each in Arm B and in Arm C, or after the completion of enrollment of the Phase 3 portion of the study, whichever occurs later, the ORR by BICR will be compared for Arm B versus Arm C, on these first 220 participants, using a 1-sided alpha of 0.001.

If the results of ORR by BICR analysis are statistically significant, an interim analysis for OS on all participants will be conducted at this time using a portion of nominal 1-sided alpha of 0.001.

Once at least 230 PFS events by BICR will be observed for Arm B + Arm C and at least 12 months after the completion of enrollment of the Phase 3 portion of the study, the PFS analysis for the comparison of Arm B versus Arm C will be tested using a 1-sided alpha of 0.023.

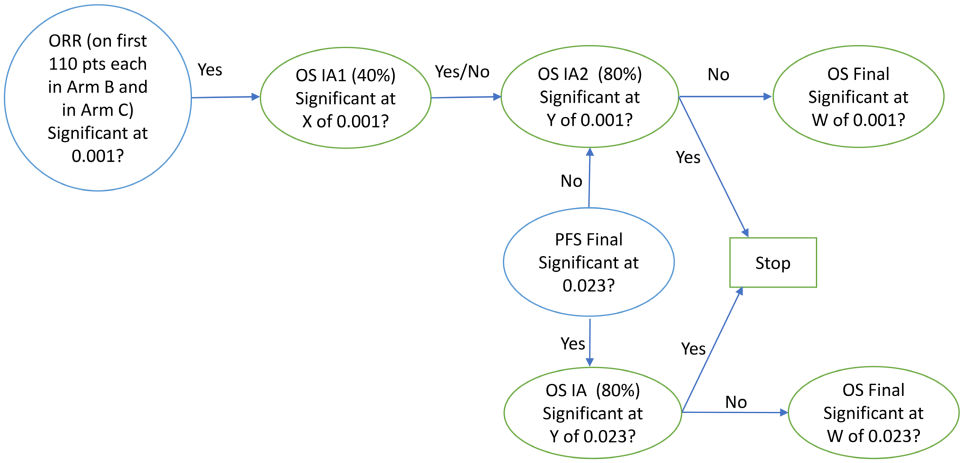
If the results of PFS by BICR analysis are statistically significant, an interim analysis for efficacy OS will be conducted at this time using a portion of nominal 1-sided alpha of 0.023. The boundaries for the efficacy analysis will be derived using a Lan-DeMets alpha-spending function[1] that approximates O'Brien-Fleming stopping boundaries for overall type I error rate at 0.023 level (1-sided).

If the results of PFS by BICR analysis are not statistically significant, but the ORR by BICR on 220 participants resulted statistically significant, an interim analysis for efficacy OS will be conducted at the time of PFS analysis using a portion of nominal 1-sided alpha of 0.001.

If the value of the test statistic for interim OS of Arm B versus Arm C at the time of PFS analysis exceeds the efficacy boundary no further test will be performed.

If the results of the interim analysis of OS are not statistically significant, a final OS analysis will be performed once 297 events are observed for Arm B + Arm C. Details of the OS interim efficacy analysis are described in Section 7.2. Figure 2 outlines the steps of multiple testing strategy.

Figure 2 Multiple Testing Strategy



Proportions of events for OS Interim Analyses are shown only for illustration purposes, since the observed number of events at IAs may be different from planned;
X, Y and W will be determined based on the proportion of information fraction at the time of OS Interim Analyses
STOP means no additional test for OS

The overall significance level for Phase 3 portion of the study is 0.025 (1-sided) and will be used as follows:

- 1-sided alpha of 0.001 for the comparison of ORR by BICR of Arm B versus Arm C;
- 1-sided alpha of 0.023 for the comparison of PFS by BICR of Arm B versus Arm C;
- alpha for the comparison of OS of Arm B versus Arm C will depend on the outcome of ORR and PFS;
- 1-sided alpha of 0.001 for the comparison of PFS by BICR of Arm A versus Arm C, even if there is no plan for formal testing.

No formal statistical testing will be performed on the secondary and exploratory endpoints (with the exception of the key secondary endpoint described above), sensitivity analyses and subgroup analyses. These results will be reported with descriptive statistics only.

Confidence intervals will be 2-sided with a confidence level of 95%, if not otherwise specified.

5.1.2.3. Cohort 3

In Cohort 3, the analysis of ORR by BICR comparing Arm D versus Arm E will be performed at least 8 months after the randomization of the last participant using 1-sided alpha of 0.025. A hierarchical testing procedure will be used to control the family-wise type I error rate at 0.025 (1-sided) level for ORR and PFS. PFS by BICR will be tested when at least 73 events will be observed, only if the analysis for ORR is statistically significant. If ORR results are not statistically significant, PFS will be analyzed descriptively.

No formal statistical testing will be performed on the secondary and exploratory endpoints (with the exception for key endpoint described above), sensitivity analyses and subgroup analyses. These results will be reported with descriptive statistics only.

Confidence intervals will be 2-sided with a confidence level of 95%, if not otherwise specified.

5.2. General Methods

Unless otherwise specified, all analyses will be performed separately for the SLI, for the Phase 3 portion, and for the Cohort 3 of the study.

In general, descriptive summaries will be presented for the efficacy and safety variables collected. Continuous variables will be summarized using mean, standard deviation, minimum, median, and maximum. Categorical variables will be summarized using frequency counts and percentages.

All statistical analyses will be conducted using SAS®, version 9.4 or higher (SAS Institute, Inc.). Noncompartmental PK analyses will be performed using an internally validated software system, openNCA. Sample size calculation for the primary and key secondary efficacy endpoints were performed with East version 6.5 (Cytel)[3].

5.2.1. Data Handling After the Cut-off Date

Data after the cut-off date may not undergo the cleaning process and will not be displayed in any listings or used for summary statistics, statistical analyses or imputations.

5.2.2. Pooling of Data by Center

In order to provide overall estimates of treatment effects, data will be pooled across trial centers. The “center” factor will not be considered in statistical models or for subgroup analyses due to the high number of participating centers in contrast to the anticipated small number of participants randomized at each center.

5.2.3. Analyses to Assess the Impact of COVID-19 Pandemic

The study enrollment started during the COVID-19 pandemic period. The following data summaries and analyses may be performed to assess the impact of COVID-19 on the trial population and study data. Additional analyses may be added in a SAP amendment if they

are considered necessary to evaluate the outcome of the trial. Details of these summaries and analyses are included in the respective sections.

- A listing of all participants affected by COVID-19 related study disruption.
- A listing of protocol deviations related to COVID-19.
- COVID-19 related AEs and deaths.
- Summary of missing tumor assessments due to COVID-19.

5.2.4. COVID-19 Anchor Date

An anchor date will be used as a start date for COVID-19 pandemic related periods based on Pfizer guidance and standard operating procedure (SOP):

- For global pandemic reference date: Use the date the World Health Organization designated COVID-19 as a global pandemic - March 11, 2020.
- For China reference date: Use the date COVID-19 was identified as the causative agent of outbreak in Wuhan by the China Center for Disease Control and Prevention - January 9, 2020.

A different anchor date may be used for purposes of regulatory submission should the regulatory authority requests.

5.2.5. Definition of Start Date

The start date is defined as date of the first dose of any study drug for the SLI and date of randomization for the Phase 3 portion and for the Cohort 3 of the study.

5.2.6. Definition of Study Day

The study day for assessments occurring on or after the first dose of study treatment (eg, adverse event onset, laboratory date) will be calculated as:

Study day = Date of the assessment/event - start date of study treatment +1.

The study day for assessments occurring prior to the first dose of study treatment (eg, baseline characteristics, medical history) will be negative and calculated as:

Study day = Date of the assessment/event - start date of study treatment.

For efficacy endpoints and for tumor assessments in the Phase 3 portion and in the Cohort 3 the study day will be calculated respect to the randomization date and not to the date of first dose of study treatment.

The study day will be displayed in all relevant data listings.

5.2.7. On-treatment Period

The on-treatment period is defined as the time from the first dose of study treatment through minimum of (last dose of study treatment +28 days, start day of subsequent anticancer systemic therapy - 1 day).

5.2.8. Definition of Start of New Anti-cancer Drug Therapy

Start date of new anti-cancer systemic therapy is used to determine the end of the on-treatment period (see Section 3.5).

The start date of new anti-cancer drug therapy is the earliest start date of anti-cancer drug therapy recorded in the 'Follow-up Cancer Therapy' eCRF pages that is after the first dose of study treatment. When start date of anti-cancer drug therapy is missing or partially missing, the imputation rules described in Section 5.3.1 should be applied using only data from the 'Follow-up Cancer Therapy' eCRF pages.

5.2.9. Definition of Start of New Anti-cancer therapy

Start date of new anti-cancer therapy (drug, radiation, surgery, including metastasectomy) is used for censoring in efficacy analyses (see Section 6.1.2.1.1).

The start date of new anti-cancer therapy is the earliest date after randomization amongst the following:

- Start date of anti-cancer drug therapy recorded in the 'Follow-up Cancer Therapy' eCRF pages
- Start date of radiation therapy recorded in 'RADIATION TREATMENT', 'RADIATION TREATMENT FOR PHASE 3 ' eCRF pages with 'Subcategory Intent' = 'Curative in intent'
- Surgery date recorded in 'NON-DRUG TREATMENTS - CANCER SURGERY', 'NON-DRUG TREATMENTS - CANCER SURGERY FOR PHASE 3', eCRF pages when 'Outcome of Procedure' = 'Resected' or 'Partially Resected'.

When start date of anti-cancer therapy is missing or partially missing, the imputation rules described in Section 5.3.1 should be applied using 'Follow-up Cancer Therapy', 'RADIATION TREATMENT', 'RADIATION TREATMENT FOR PHASE 3, 'NON-DRUG TREATMENTS - CANCER SURGERY', 'NON-DRUG TREATMENTS - CANCER SURGERY FOR PHASE 3 ' eCRF pages.

5.2.10. Date of Last Contact

The date of last contact will be derived for participants not known to have died at the data cutoff date using the latest complete date (ie, imputed dates will not be used in the derivation) among the following:

- All participant assessment dates (eg, blood draws (laboratory, Pharmacokinetics (PK), vital signs, performance status, ECG, tumor assessments);
- Start and stop dates of concomitant therapies including non-drug treatments or procedures;
- Completion dates for PRO Questionnaires;

- Start and end dates of anticancer therapies administered after study treatment discontinuation including systemic therapy, radiation, and surgeries;
- AE start and end dates;
- Last date of contact collected on the 'Survival Follow-up' eCRF (do not use date of survival follow-up assessment unless status is 'alive' or 'subject remains in follow-up');
- Study treatment start and end dates;
- Randomization and enrollment date; and
- Date of discontinuation on disposition eCRF pages (do not use if reason for discontinuation is lost to follow-up or death).

Only dates associated with actual examinations of the participant will be used in the derivation. Dates associated with a technical operation unrelated to participant status such as the date a blood sample was processed or dates data were entered into the eCRF will not be used. For participants with assessment dates after the data cutoff date the last contact date will be derived as the cutoff date for the analysis.

5.2.11. Unscheduled Assessments

Unless otherwise specified, unscheduled assessments will not be displayed in summary tables by nominal visit/timepoint. Unscheduled assessments will be used when deriving baseline and worst case on-treatment for safety and PRO analyses (except where noted for baseline ECGs). Additionally, unscheduled tumor assessments will be used for efficacy analyses (eg, defining date of progression/censoring, best overall response, date of last contact).

5.2.12. Tumor Assessment Date

If there are multiple scan dates associated with a tumor evaluation, ie, radiological assessments occur over a series of days rather than the same day, the earliest scan date associated with the evaluation will be used as the date of the assessment. The same convention should be used for investigator, derived investigator and BICR assessments, noting however that in the event that the BICR provides the date of PD, that date should be used instead of deriving the date of PD.

5.2.13. Adequate Baseline Tumor Assessment

An adequate baseline tumor assessment is defined as follows (see additional details in Section 10.1.1):

- Baseline assessments of all lesions (target, and non-target) must be prior to first dose date (SLI) or date of randomization (Phase 3 and Cohort 3) per Schedule of Activities in Section 1.3 of the protocol. For Phase 3 and Cohort 3 lesions assessed after randomization, but before or on the first dose date will also be considered adequate.

- All lesions (target and non-target) must have non-missing assessments.

  Note: for target lesions, an actual measurement should be recorded and it should meet the criteria for being measurable and for non-target lesions the actual status at baseline should indicate that the lesion was present.

5.2.14. Adequate Post-baseline Tumor Assessment

An adequate post-baseline tumor assessment is defined as an assessment where a response of CR, PR, SD, non-CR/non-PD, or PD can be determined. Time points where the response is NE, ND (no disease at baseline) or no assessment was performed will not be used for determining the censoring date.

5.2.15. Standard Derivations and Reporting Conventions

The following conversion factors will be used to convert days into weeks, months or years: 1 week = 7 days, 1 month = 30.4375 days, 1 year = 365.25 days.

Percentages will be reported to one decimal place. The rounding will be performed to closest integer/first decimal using the common mid-point between the two consecutive values. Eg, 5.1 to 5.4 will be rounded to an integer of 5, and 5.5 to 5.9 will be rounded to an integer of 6.

5.2.16. Analyses for Binary and Categorical Endpoints

Binary and categorical endpoints will be summarized by frequency counts and percentages along with corresponding 2-sided Wilson score[4] 95% CI. Unless otherwise specified, the calculation of proportions will include the missing category. Therefore counts of missing observations will be included in the denominator and presented as a separate category.

In case the analysis refers only to certain visits, percentages will be based on the number of participants with an assessment at that visit, unless otherwise specified.

5.2.17. Analyses for Continuous Endpoints

Continuous endpoints will be summarized using descriptive statistics ie, number of non-missing values and number of missing values [ie, n (missing)], mean, median, standard deviation, minimum, maximum and first and third quartile (Q1 and Q3). PK summaries will also include coefficient of variation percent (%CV).

5.2.18. Analyses for Time-to-Event Endpoints

The stratified log-rank test will be used for comparing treatments. Hazard ratios and the associated 2-sided 95% confidence intervals (CIs) are estimated by stratified Cox proportional hazards model, including treatment as a covariate. To account for the required alpha spending of the group sequential design, 2-sided repeated CIs[5] will also be provided.

Time to event endpoints will be summarized using the Kaplan-Meier method and estimated survival curves will be displayed graphically when appropriate. Graphs will describe the number of participants at risk over time. The median, quartiles, and probabilities of an event at particular points in time will be estimated. CIs for medians and quartiles are based on the

Brookmeyer-Crowley[6] method. CIs for the estimated probability of event at a particular timepoint will be generated using the log(-log) method with back transformation to a confidence interval on the untransformed scale. Summaries of the number and percentages of participants with an event and reason for censoring will also be provided on summary tables and/or figures.

Assessment of proportional hazards

Schoenfeld residuals for the stratified Cox proportional regression model may be plotted to investigate graphically violations from the proportional hazards (PH) assumption; a non-zero slope is evidence of departure from PH. The PH assumption may be formally tested using Schoenfeld's residual test (Schoenfeld, 1980[7], Therneau & Grambsch, 2000[8]). Large departures from PH will be evidenced by a p-value <0.05. Note however that the test will not be sensitive to detect non-linear deviations from PH.

In addition, the PH assumption will be checked visually within each stratum by plotting

$$\log(-\log(S(t))) \text{ vs } \log(t),$$

where S(t) is the estimated survival function at time t.

Restricted Mean Survival Time

The restricted mean survival time (RMST) methodology (Zhang, 2013)[9] is applicable independently of the PH assumption and can be used, at a minimum, as a sensitivity analysis to explore the robustness of the primary analysis result on time-to event endpoint.

In particular, as it pertains to the **cut-off point (τ)** to evaluate the RMST it is noted that the cut-off point should not exceed the minimum of the largest observed time for both treatment arms so that the RMST of all treatment arms being evaluated can be adequately estimated and comparison between treatments is feasible; τ should be clinically meaningful and closer to the end of the study follow-up so that the majority of survival outcomes will be covered by the time interval. The RMST up to time τ can be interpreted as the expected survival time restricted to the common follow-up time τ among all participants. Analyses will be repeated using the following criteria to define τ:

- τ_1 = minimum of (largest observed survival time for experimental arm, largest observed survival time for control arm);
- τ_2 = minimum of (largest survival event time for experimental arm, largest survival event time for control arm);
- $\tau 3$ = midpoint between $\tau 1$ and $\tau 2$.

RMST can be estimated consistently by the area under the Kaplan-Meier curve over $[0, \tau]$.

The treatment effect between the experimental arm and the control arm will be assessed based on the difference in RMST. The associated 95% CI for the difference in means and 1-sided p-value will be generated.

5.3. Methods to Manage Missing Data

Unless otherwise specified, all data will be evaluated as observed, and no imputation method for missing values will be used.

Any imputations will occur at the analysis dataset level. Additionally, in all participant data listings imputed values will be presented and flagged as imputed.

Missing statistics, eg, when they cannot be calculated, should be presented as 'ND' for not done, 'NR' for not reached or 'NA' for not applicable. For example, if N=1, the measure of variability cannot be computed and should be presented as 'ND' or 'NA'.

5.3.1. Missing Dates

For purposes of data listings, dates will reflect only the information provided by the investigator on the eCRF.

If start dates for adverse events or concomitant medications are completely missing a worst case approach will be taken whereby the events will be considered treatment emergent and the medications will be considered concomitant. If only partial information are available (eg, only a month and year or only a year) and the partial information provide sufficient information to indicate the dates are prior to the start of study treatment (eg, month/year less than month/year of first dose) then these will be considered to have started prior to treatment; otherwise a similar worst case approach will apply and these will be considered to have started after treatment.

Missing or Partial Death Dates

Missing or partial death dates will be imputed based on the last contact date:

- If the entire date is missing it will be imputed as the day after the date of last contact (see derivation of date of last contact in Section 5.2.10).
- If the day or month is missing, death will be imputed to the maximum of the full (non-imputed) day after the date of last contact and the following:
  - Missing day: 1st day of the month and year of death, or
  - Missing day and month: January 1st of the year of death.

Date of First and Last Dose of Study Treatment

No imputation will be done for first dose date.

Date of last dose of study drug for encorafenib and capecitabine, if unknown or partially unknown, will be imputed as follows:

- If the last date of study drug is completely missing and there is no End of Treatment eCRF page and no death date, the participant should be considered to be ongoing and

use the cutoff date for the analysis as the last dosing date. **Note:** the study team should confirm that the participant is actively receiving dose at the time of the data cutoff.

- If the last date of study drug is completely or partially missing and there is EITHER an End of Treatment eCRF page OR a death date available (within the data cutoff date), then impute as follows:
    - If only Year is available and Year < Year of min (EOT date, death date), impute the date as 31DECYYYY.
    - If both Year and Month are available and Year = Year of min (EOT date, death date) and Month < Month of min (EOT date, death date), impute as last day of the month.
    - For all other cases, impute as min (EOT date, death date).

For 5-FU infusion, if unknown, the date of last dose is assumed equal to the last dose start date plus 2 days. For all other study drugs administered by infusion or bolus the date of last dose is assumed equal to the last dose start date.

Date of Start of New Anticancer Therapy

Incomplete dates for start date of new anticancer therapy (systemic therapy, radiation and surgery, etc.) will be imputed as follows and will be used for determining censoring dates for efficacy analyses. PD date below refers to PD date by investigator assessment (as reported on Investigator Overall Tumor Assessment CRF page). If the imputation results in an end date prior to the imputed start date then the imputed start date should be set to the end date.

- The end date of new anticancer therapy will be included in the imputations for start date of new anticancer therapy. If the end date of new anticancer therapy is:
    - completely missing then it will be ignored in the imputations below;
    - partially missing with only year (YYYY) available then the imputations below will consider 31DECYYYY as the end date of the new anticancer therapy;
    - partially missing with only month and year available then the imputations below will consider the last day of the month for MMMYYYY as the end date of the new anticancer therapy.
- For participants who have not discontinued study treatment at the analysis cut-off date, last dose of study treatment is set to the analysis cut-off date in the imputations below.

- If the start date of new anticancer therapy is completely or partially missing then the imputed start date of new anticancer therapy is derived as follows:
  - Start date of new anticancer therapy is completely missing:

    Imputed start date = min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy]
  - Only year (YYYY) for start of anticancer therapy is available:

    IF YYYY < Year of min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy] THEN imputed start date = 31DECYYYY;

    ELSE IF YYYY = Year of min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy]

    THEN imputed start date = min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy]

    ELSE IF YYYY > Year of min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy]

    THEN imputed start date = 01JANYYYY
  - Both Year (YYYY) and Month (MMM) for start of anticancer therapy are available

    IF

    YYYY = Year of min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy], AND

    MMM < Month of min [max(PD date +1 day, last dose of study treatment +1 day), end date of new anticancer therapy]

    THEN

    imputed start date = DAY (Last day of MMM) MMM YYYY ;

    ELSE IF

    YYYY = Year of min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy], AND

    MMM = Month of min [max(PD date +1 day, last dose of study treatment +1 day), end date of new anticancer therapy]

    THEN

    imputed start date = min [max(PD date +1 day, last dose of study treatment +1 day), end date of new anticancer therapy]);

    ELSE IF

    YYYY = Year of min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy], AND

MMM > Month of min [max(PD date +1 day, last dose of study treatment +1 day), end date of new anticancer therapy]

THEN

imputed start date = 01 MMM YYYY;

ELSE IF

YYYY < Year of min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy]

THEN

imputed start date = DAY (Last day of MMM) MMM YYYY;

ELSE IF

YYYY > Year of min [max(PD date +1, last dose of study treatment +1), end date of new anticancer therapy]

THEN

imputed start date = 01 MMM YYYY.

AE Dates

AE Onset Date:

If completely missing, the onset date will be set to first dose date if the first dose date is less than AE stop date. Otherwise if the first dose date is after AE stop date, then set the onset date to the earliest of non-missing AE stop date or informed consent date.

AE Stop Date:

If completely missing, the stop date will be imputed as the latest of the end of study date, death date, last dose date of the study treatment, or onset date.

Partial AE Date:

Partial AE date will be imputed bases on the imputation rule for "Other Missing or Partial Dates". If the AE onset date is imputed from a partial AE date and the first dose date falls in the same month as the AE onset date, the following will be done:

- The AE onset date is reset to the first dose date.

If AE stop date is imputed, and less than the first dose date, set the AE stop date to the first dose date.

Partial Date of Progression after Subsequent Therapy

Partial date of progression after subsequent therapy will be imputed as follows:

- If the entire date is missing or if only year is provided, no imputation will be performed.
- If month and year are provided and only the day is missing, the date will be imputed as the 15th of the month.

Other Missing or Partial Dates

Imputation methods for other partial dates are as follows:

- If the day of the month is missing for a start date used in a calculation, the 1st of the month will be used to replace the missing date.
- If both the day and month are missing, the first day of the year is used.
- For stop dates, the last day of the month, or last day of the year is used if the day or day and month are missing, respectively.
- If the date is completely missing, no imputation will be performed.

These rules are used unless the calculations result in negative time durations (eg, date of resolution cannot be prior to date of onset). In these cases, the resolution and onset dates will be the same and the duration will be set to 1 day.

5.3.2. Missing Patient-Reported Outcomes Data

For the EORTC QLQ-C30 and EQ-5D-5L, missing items will be handled per the respective scoring manuals of each questionnaire. For the PGIS and PGIC, there will be no adjustments for missing data.

5.3.3. Missing Pharmacokinetic Data

Deviations, Missing Concentrations and Anomalous Values

In summary tables and plots of median profiles, concentrations will be set to missing if one of the following cases is true:

- A concentration has been reported as ND (ie, not done) or NS (ie, no sample);
- A deviation in sampling time is of sufficient concern or a concentration has been flagged as anomalous by the clinical pharmacologist.

Summary statistics will not be presented at a particular timepoint if more than 50% of the data are missing. For analysis of pharmacokinetic concentrations, no values will be imputed for missing data.

5.3.4. Missing Biomarker Data

No missing data will be imputed.

6. ANALYSES AND SUMMARIES

All efficacy analyses will be performed by cohort and number of lines of prior therapies on the FAS for the SLI, and by treatment arm on the FAS for the Phase 3 and for Cohort 3.

6.1. Primary Endpoints

6.1.1. Safety Lead-in

The number and proportion of participants experiencing DLTs during the DLT-evaluation period will be summarized by System Organ Class (SOC) and preferred term (PT) based on maximum toxicity grade by treatment cohort. Analyses of DLT will be performed on DLT-Evaluable Set.

All DLTs and their attributes will be presented in data listings sorted by participant identifier, AE and date of onset of the AE.

6.1.2. Phase 3

The primary objective is to compare the efficacy, as measured by the primary endpoints of PFS by BICR and ORR by BICR, of Arm B versus Arm C.

6.1.2.1. PFS by BICR

6.1.2.1.1. Main Analysis

The primary efficacy analysis will compare PFS time by BICR between Arm B and Arm C using a 1-sided stratified log-rank test at the significance level of 0.023. The analysis will be performed based on the FAS when at least 230 PFS events by BICR will be observed in Arms B + C and at least 12 months after the completion of enrollment of the Phase 3 portion of the study. This will allow for adequate follow-up based on PFS assumptions in Arm B. Strata will be based on those specified in the IRT system as described in Section 3.4.1.

PFS will be calculated in months as follows:

PFS (months) = (date of event or censoring – date of randomization +1)/30.4375

Participants without an event or with an event more than 12 weeks (for the first 18 months after randomization) or 16 weeks (after the first 18 months of randomization) after the last adequate tumor assessment that documented no progression will be censored on the date of the last adequate tumor assessment that documented no progression. In addition, if a new anticancer therapy is started prior to an event, the participant will be censored on the date of the last adequate tumor assessment that documented no progression prior to the start of the new anticancer therapy.

Adequate baseline assessment and adequate post-baseline assessment are defined in Section 5.2.13 and Section 5.2.14, respectively.

Participants with no baseline tumor assessment (including participants with an inadequate baseline assessment) or with no adequate post-baseline tumor assessments within 12 weeks after the date of randomization will be censored on the day of randomization, unless the participant dies within 12 weeks of randomization, in which case, death will be an event on date of death.

The censoring and event date options to be considered for the PFS primary analysis are presented in Table 4.

Table 4 PFS Outcome and Event Dates – Primary Analysis

Situation	Date of Progression/Censoring	Outcome
No adequate baseline assessment, including no disease at baseline	Date of randomization[a]	Censored[a]
PD or death ≤12 (or 16)[b] weeks after last adequate tumor assessment or ≤12 weeks after date of randomization	Date of PD or death	Event
PD or death >12 (or 16)[b] weeks after the last adequate tumor assessment[c]	Date of last adequate tumor assessment[c] documenting no PD prior to new anticancer therapy, or missed assessments	Censored
No PD		
New anticancer therapy given[d]		

a If the participant dies ≤12 weeks after date of randomization, the death is an event with date on death date.

b Durations are equal to 2 times the length of the tumor assessment interval, which is 12 weeks for the first 18 months after randomization, and 16 weeks thereafter.

c If there are no adequate post-baseline assessments prior to the PD or death, then the time without adequate assessment should be measured from the date of randomization; if the criteria is met, the censoring will be on the date of randomization.

d New anticancer therapy includes systemic therapy, radiation and surgery

The treatment effect will be estimated using a Cox's proportional hazards model stratified by randomization strata to calculate the HR for PFS of Arm B versus Arm C, along with the corresponding 2-sided 95% CI and 2-sided 95.4% CI.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median PFS time with 2-sided 95% CI. The PFS at 3, 6, 9, 12, 15, 18 and 21 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley[6] (conftype=linear option in SAS PROC LIFETEST) and the CIs for the survival function estimates at the timepoints defined above will be derived using the log(-log) method according to Kalbfleisch and Prentice[10] (conftype=loglog default option in SAS Proc LIFETEST). The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of participants with each event type (PD or death) and censoring reasons will be presented by treatment arm along with the overall event and censor rates.

Reasons for censoring will be summarized according to the categories in Table 5. If a participant meets multiple definitions for censoring the list will be used to define the hierarchy.

Table 5. PFS Censoring Reasons and Hierarchy - Primary Analysis

Hierarchy	Condition	Censoring Reason
1	No adequate baseline assessment, including no disease at baseline	No adequate baseline assessment
2	Start of new anticancer therapy[a] before event	Start of new anticancer therapy
3	Event >12 (or 16)[b] weeks from last adequate post-baseline tumor assessment/start date	Event after missing assessments or inadequate assessments[c]
4	No event and [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Subject refused further follow-up]	Withdrawal of consent
5	No event and lost to follow-up in any disposition page	Lost to follow-up
6	No event and [EOS present OR disposition page for any phase after screening says participant will not continue into any subsequent phase of the study] and no adequate post-baseline tumor assessment	No adequate post-baseline tumor assessment
7	No event and none of the conditions in the prior hierarchy are met	Ongoing without an event

a New anticancer therapy includes systemic therapy, radiation and surgery
b Durations are equal to 2 times the length of the tumor assessment interval, which is 12 weeks for the first 18 months after date of first dose (SLI) or date of randomization (Phase 3 and Cohort 3), and 16 weeks thereafter.
c see section 5.2.14 for definition of adequate post baseline tumor assessment

The PFS time or censoring time and the reasons for censoring will also be presented in a patient listing.

Schoenfeld residuals for the stratified Cox proportional model may be plotted to investigate graphically violations from the proportional hazards (PH) assumption (see Section 5.2.18 for details).

Time of Follow-Up for PFS by BICR

A Kaplan-Meier plot for PFS follow-up duration will also be created to assess the follow-up time in each treatment arm using reversed Kaplan-Meier method as shown in Table 6. The Kaplan-Meier estimate of the median PFS follow-up time with its corresponding 2-sided 95% CI will be calculated and presented in the plot.

Table 6 Reverse Censoring Rules for PFS Follow-up Duration

	Date of event/censoring	Censoring
Participant without a PFS event by BICR	Time from randomization to last tumor assessment scan	No
Death or PD	Time from randomization to date of first PD or death	Yes

6.1.2.1.2. Sensitivity/Supplementary Analyses

The following sensitivity and supplementary analyses will be performed to explore the robustness of the primary analysis results and the supportive estimands. These analyses are regarded as purely exploratory.

Sensitivity Analyses:

- A stratified analysis will be performed to compare the PFS time by derived investigator assessment.
- An unstratified analysis will be performed to compare the PFS time by BICR.
- PFS by BICR will be analyzed based on Central *BRAF* V600E Positive analysis set.
- PFS by BICR will also be analyzed based on RMST differences (see Section 5.2.18 for details).
- Multivariable Cox regression analysis will be performed to explore the potential influences of baseline participant or disease characteristics on PFS. The baseline characteristics variables specified in Section 3.4.2 will be included as covariates in the Cox regression model in addition to the randomization strata which will be included in the model during the selection procedure.

  A stepwise selection procedure will serve to identify these explanatory variables of potential prognostic values. Variables are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further variable can be added to the model or if the variable just entered into the model is the only variable removed in the subsequent backward elimination. The level of significance for an explanatory variable to enter the model is set to 0.15 (p-value of Score test) and the significance level for removing it is set to 0.40 (p-value of Wald

test). This analysis will be performed using the stepwise selection method in SAS (PROC PHREG). Once this procedure stops, the variable for treatment arm will be added to the last selected model in order to evaluate the effect of treatment on PFS time when adjusted for the selected explanatory variables. The HRs of all selected explanatory variables and of treatment effect will be reported including 2-sided 95% CIs. No interactions will be considered.

Supplementary Analyses:

- PFS by BICR will be analyzed using the treatment policy strategy (Section 2.1.3). The analysis will use the same methodology and summary as the main analysis but will include observations that occur after the intercurrent events of starting new anticancer therapy. Reasons for censoring will be summarized according to the categories in Table 5, excluding number 2 - start of new anticancer therapy. All other intercurrent events will be addressed in the same approach as the primary estimand of PFS. Table 7 summarizes the details of the intercurrent events and censoring rules for this supplementary analysis.

Table 7 PFS Outcome and Event Dates – Supplementary Analysis

Situation	Date of Progression/Censoring	Outcome
No adequate baseline assessment, including no disease at baseline	Date of randomization[a]	Censored[a]
PD or death ≤12 (or 16)[b] weeks after last adequate tumor assessment or ≤12 weeks after date of randomization	Date of PD or death	Event
PD or death >12 (or 16)[b] weeks after the last adequate tumor assessment[b]	Date of last adequate tumor assessment[c] documenting no PD prior to missed assessments	Censored
No PD		

a If the participant dies ≤12 weeks after date of randomization, the death is an event with date on death date.
b Durations are equal to 2 times the length of the tumor assessment interval, which is 12 weeks for the first 18 months after randomization, and 16 weeks thereafter.
c If there are no adequate post- baseline assessments prior to the PD or death, then the time without adequate assessment should be measured from the date of randomization; if the criteria were met the censoring will be on the date of randomization.

- PFS by BICR will be analyzed using the same methodology and summary as the main analysis, but will include observations/events that occur after >12 (or 16) weeks from last adequate post-baseline tumor assessment/start date. Reasons for censoring will be summarized according to the categories in Table 5, excluding number 3 – "Event >12 (or 16) weeks from last adequate post-baseline tumor assessment/start date". All other intercurrent events will be addressed in the same approach as the primary estimand of PFS.

Discordance between BICR and Derived Investigator Assessment on PFS

A summary of discordances of events (PD or death) between BIRC and derived investigator assessment will be provided as shown in Table 8:

Table 8 Discordance Summary for Derived Investigator vs BICR Assessment on PFS

		BICR	
Derived Investigator		**Event**	**No Event**
	Event	a = a1 + a2 + a3	b
	No Event	c	d

a1: number of agreements on timing and occurrence of event;
a2: number of times agreement on event but INV declares event later than BICR;
a3: number of times agreement on event but INV declares event earlier than BICR;
N= a+b+c+d.
The timing agreement of event is defined as a window of ±7 days.

The following **measures of discordances** will be calculated:

- Total Event Discrepancy Rate: (b+c)/N.
- Early Discrepancy Rate (EDR): (a3+b)/(a+b).
- Late Discrepancy Rate (LDR): (a2+c)/(a2+a3+b+c).
- Overall Discrepancy Rate: (a2+a3+b+c)/N.

The EDR represents the positive predictive value of investigator assessment and quantifies the frequency with which the Investigator declares PFS event earlier than BICR within each treatment arm as a proportion of the total number of investigator-assessed PFS events.

The LDR quantifies the frequency with which the investigator declares PFS event later than BICR as a proportion of the total number of participants with discrepancies within the treatment arm.

Total event discrepancy rate represents the frequency with which the investigator and BICR disagreed on overall censoring/event status as a proportion of the number of participants within the treatment arm (N).

Overall discrepancy rate represents the frequency with which the investigator and BICR disagreed on overall censoring/event status or the timing of an event as a proportion of the number of participants within the arm (N).

Discordance metrics will be calculated for each treatment arm and, for each metric, the difference in discordance between Arm B and Arm C is used to evaluate potential bias. If the

discordance is similar across the treatment arms then it suggests the absence of evaluation bias favoring a particular arm. A negative differential discordance for EDR and/or a positive differential discordance for LDR may be indicative of investigator evaluation bias in favor of the experimental arm, ie, Arm B (Amit et al, 2011)[11].

6.1.2.2. ORR by BICR

6.1.2.2.1. Main Analysis

ORR is defined as the proportion of participants who have achieved a BOR of confirmed CR or PR per RECIST v1.1 as assessed by BICR. The BOR is the best response obtained among all tumor assessment visits after date of randomization until documented PD by BICR, or start of subsequent anticancer therapy, or death. Confirmation of the response will be no sooner than 4 weeks after the initial documentation of CR or PR.

The following rules will be applied for the confirmation:

- CR = at least two determinations of CR at least 4 weeks apart and documented before progression, or start of new anticancer therapy.
- PR = at least two determinations of PR or better (and not qualifying for a CR) at least 4 weeks apart and before progression, or start of new anticancer therapy.
- SD (for participants with at least one measurable lesion at baseline)= at least one SD assessment (or better and not qualifying for CR or PR) ≥6 weeks after the study start date and before progression, or start of new anticancer therapy .
- Non-CR/Non-PD (for participants with only non-target disease at baseline) = at least one Non-CR/Non-PD assessment (or better and not qualifying for CR) ≥6 weeks after the study start date and before progression, or the start of new anticancer therapy.
- PD = progression ≤12 weeks after the study start date not qualifying for CR, PR or SD.
- Not Evaluable (NE) = all other cases, including ND (no evidence of disease at baseline).

Clinical deterioration or clinical progression noted on the completion eCRF will not be considered as documented disease progression.

The frequency (number and percentage) of participants with BOR of CR, PR, SD, PD, non-CR/non-PD (applicable only to participants with non-measurable disease at baseline), and NE will be tabulated.

Participants with a BOR of "Not Evaluable" based on confirmed responses will be summarized by reason for having unknown status. The following reasons will be used:

- No adequate baseline assessment.
- No evidence of disease at baseline.

- No post-baseline assessments due to early death (ie, death ≤12 weeks after the start date).
- No post-baseline assessments due to COVID-19 (ie, participants miss tumor assessment visits due to COVID-19 pandemic).
- No post-baseline assessments due to other reasons.
- All post-baseline assessments have overall response NE.
- New anticancer therapy started before first post-baseline assessment.
- SD of insufficient duration (<6 weeks after the start date without further evaluable tumor assessments).
- PD too late (>12 weeks after the start date, ie, tumor assessment of PD was >12 weeks after start of treatment and there was no tumor assessment in between).

Special and rare cases where BOR is NE due to both SD of insufficient duration ('too early') and late PD will be classified as "SD of insufficient duration".

The ORR will be calculated along with corresponding 2-sided Wilson Score[4] 95% CI.

The treatment effect of the Arm B compared to Arm C, as measured by stratified (by the randomization strata) odds ratio and its 95% CI and its 99.8% CI in terms of OR defined as the odds of OR with Arm B divided by the odds of OR with Arm C will be tested using Cochran-Mantel-Haenszel statistics stratified by the randomization strata at 1-sided 0.001 level of significance considering the first 110 randomized participants each in Arm B and in Arm C. This analysis will be performed eight months after randomization of the first 110 participants each in Arm B and in Arm C, or after the completion of enrollment of the Phase 3 portion of the study, whichever occurs later.
ORR may be evaluated descriptively only at the time of PFS analysis on all participants randomized in the Phase 3.

6.1.2.2.2. Sensitivity/Supplementary Analyses

The following sensitivity and supplementary analyses will be performed to explore the robustness of the primary analysis results. These analyses are regarded as purely exploratory.

Sensitivity Analyses:

- An updated OR by BICR analysis at the time of PFS analysis may be performed including all randomized participants.

- OR by BICR will be analyzed based on Central *BRAF* V600E Positive analysis set. A stratified analysis will be performed to compare the OR by derived investigator assessment.

Concordance between BICR and derived investigator assessment on OR will be presented.

Table 9. Definition of Agreement Used to Determine Total Agreement Rates in Response

	Derived Investigator	BICR	Variable Name
Agreement	No response	No response	a
	Response	Response	b
Disagreement	No response	Response	c
	Response	No response	d

N = (a + b + c + d)

The total agreement rate measures the proportion of participants for whom there is a concordance between the BICR and derived investigator as to whether the participant was a responder or a non-responder with both the BICR and the derived investigator, among all participants who are evaluated by both BICR and investigator.

- Total Agreement Rate = [(a+b) / N] × 100%.

Concordance metrics will be calculated for each treatment arm and the difference in concordance between Arm B and Arm C is used to evaluate potential bias. If the concordance is similar across the treatment arms then it suggests the absence of evaluation bias favoring a particular arm.

6.1.3. Cohort 3

The primary objective is to compare the efficacy, as measured by the primary endpoint of ORR by BICR, of Arm D versus Arm E.

6.1.3.1. ORR by BICR

6.1.3.1.1. Main Analysis

ORR is defined as in Section 6.1.2.2.1 and will be calculated along with corresponding 2-sided Wilson Score[4] 95% CI.

The treatment effect of the Arm D compared to Arm E, as measured by stratified (by the randomization strata) odds ratio and its 95% CI in terms of OR defined as the odds of OR with Arm D divided by the odds of OR with Arm E will be tested using Cochran-Mantel-Haenszel statistics stratified by the randomization strata at 1-sided 0.025 level of significance. This analysis will be performed at least eight months after randomization of the last participant in Cohort 3.

6.1.3.1.2. Sensitivity/Supplementary Analyses

The following sensitivity and supplementary analyses will be performed to explore the robustness of the primary analysis results. These analyses are regarded as purely exploratory.

Sensitivity Analyses:

- OR by BICR will be analyzed based on Central *BRAF* V600E Positive analysis set.
- A stratified analysis will be performed to compare the OR by derived investigator assessment.

Concordance between BICR and Derived Investigator Assessment on OR will be presented as described in Section 6.1.2.2.2.

6.2. Efficacy Secondary Endpoints

6.2.1. Safety Lead-in

6.2.1.1. Objective Response Rate by Investigator

The OR is defined as described in Section 6.1.2.2.1, but the study start date is date of the first dose and tumor assessment is based on response reported by investigator on eCRF. ORR will be calculated by dividing the number of participants with BOR of CR or PR per RECIST v1.1 by the number of participants dosed in each cohort and number of lines of prior therapy.

The ORR will be calculated along with corresponding 2-sided Wilson Score[4] 95% CI.

6.2.1.2. Duration of Response by Investigator

DOR will be calculated for participants with a confirmed response (CR or PR) as follows:

DOR (months) = (date of event or censoring – date of first CR or PR + 1)/30.4375

The same censoring rules specified for PFS in Section 6.1.2.1.1 apply to DOR. DOR will be estimated using the same Kaplan-Meier method as described for PFS. DOR rates at 3, 6, 9, 12 and 15 months will be estimated with corresponding 2-sided 95% CIs.

Frequency of participant with DOR ≥ 6 months, ≥ 9 months, ≥ 12 months will be reported.

In addition, a plot of time to and duration of response for participants with a confirmed response (swimmer plot) will be created.

DOR will be calculated and presented by cohort and number of lines of prior therapies.

6.2.1.3. Time to Response by Investigator

TTR will be calculated for participants with a confirmed response (CR or PR) in weeks as follows:

TTR (in weeks) = [date of first CR or PR – date of first dose +1]/7

TTR will be summarized using simple descriptive statistics (mean, standard deviation, median, minimum, maximum, first quartile, third quartile).

TTR will be calculated and presented by cohort and number of lines of prior therapies.

6.2.1.4. Progression-free Survival by Investigator

For the SLI, PFS will be calculated in months as follows:

PFS (months) = (date of event or censoring – date of first dose +1)/30.4375

The same censoring rules specified for PFS in the Phase 3 portion in Section 6.1.2.1.1 apply to PFS in the SLI, with the date of randomization replaced by the date of first dose in the SLI and tumor response as assessed by the investigator. Kaplan-Meier estimates will be presented in a table and a plot by cohort and number of lines for prior therapies together with a summary of associated statistics including the median PFS time with 2-sided 95% CI. The PFS at 3, 6, 9, 12, 15, 18 and 21 months will be estimated with corresponding 2-sided 95% CIs.

Frequency (number and percentage) of participants with each event type (PD or death) and censoring reasons will be presented by cohort and number of lines for prior therapies along with the overall event and censor rates. Reasons for censoring will be summarized according to the categories in Table 5.

6.2.1.5. Overall Survival

OS will be calculated in months as follows:

OS (months) = (date of death or censoring – date of first dose +1)/30.4375

Kaplan-Meier estimates will be presented in a table and a plot by cohort and by line of prior therapies together with a summary of associated statistics including the median OS time with 2-sided 95% CI. The OS at 6, 12, 18 and 24 months will be estimated with corresponding 2-sided 95% CIs.

Frequency (number and percentage) of participants with death events and censoring reasons will be presented by cohort and by line of prior therapies along with the overall event and censor rates.

6.2.2. Phase 3

6.2.2.1. Overall Survival

6.2.2.1.1. Main Analysis

The key secondary efficacy analysis will compare OS time between Arm B and Arm C using a 1-sided stratified log-rank test. The nominal significance level (0.001 or 0.023) will depend on significance of ORR/PFS analyses. The actual alpha spent at the interim and final analysis will be updated based on the actual number of events and the information fraction (see details in Section 7.2). The analysis will be performed based on the FAS. Strata will be based on those specified in the IRT system as described in Section 3.4.1.

OS will be calculated in months as follows:

OS (months) = (date of death or censoring – date of randomization +1)/30.4375

The treatment effect will be estimated using a Cox's proportional hazards model stratified by randomization strata to calculate the HR for OS of Arm B versus Arm C.

In order to account for the group sequential design on this endpoint, the repeated CI method[5] will be used to construct the 2-sided repeated CI for the HR at the interim and the final analyses of OS, in addition to the unadjusted 2-sided 95% CI.

OS time will be estimated using the same Kaplan-Meier method as described for PFS in Section 6.1.2.1.1. The OS at 6, 12, 18, 24, 30 and 36 months will be estimated with corresponding 2-sided 95% CIs.

Frequency (number and percentage) of participants with death events and censoring reasons will be presented by treatment arm along with the overall event and censor rates. The event and censoring reasons are as follows:

- Death:
  - Death due to COVID-19;
- Ongoing and no death;
- Withdrawal of consent;
- Lost to follow-up.

The OS time or censoring time and the reasons for censoring will also be presented in a patient listing.

Schoenfeld residuals for the stratified Cox proportional model may be plotted to investigate graphically violations from the proportional hazards (PH) assumption (see Section 5.2.18 for details).

Time of Follow-Up for OS in Phase 3

A Kaplan-Meier plot for OS follow-up duration will also be created to assess the follow-up time in each treatment arm using the reversed Kaplan-Meier method as shown in Table 10. The Kaplan-Meier estimate of the median OS follow-up time with its corresponding 2-sided 95% CI will be calculated and presented in the plot.

Table 10 Reverse Censoring Rules for OS Follow-up Duration

	Date of event / censoring	Censoring
Participant didn't die	Time from randomization to last date known to be alive	No
Participant died	Time from randomization to date of death	Yes

6.2.2.1.2. Sensitivity/Supplementary Analysis

The following sensitivity and supplementary analyses will be performed to explore the robustness of the key secondary endpoint analysis results in the Phase 3 portion of the study. These analyses are regarded as purely exploratory.

Sensitivity Analyses:

- An unstratified analysis will be performed to compare the OS time.
- OS will be analyzed based on Central *BRAF* V600E Positive analysis set.
- OS will also be analyzed based on RMST differences (see Section 5.2.18 for details).
- The same multivariable Cox regression analysis that is performed on PFS as described in Section 6.1.2.1.2 will also be conducted on OS to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact.

Supplementary Analyses:

Since majority of the patients enrolled in this study will likely receive additional anticancer systemic therapy upon PD by BICR, it is unclear if/how this additional therapy may impact the treatment effect estimate on OS. Based on type/frequency/distribution of follow-up therapy, analyses to correct for follow-up systemic therapy on treatment effect estimate for OS (eg 2-stage or other appropriate method) may be implemented. A swimmer plot may be produced to show OS time course from randomization, for each patient, denoting duration of randomized treatment and duration of subsequent anti-cancer systemic therapies.

6.2.2.2. Duration of Response

DOR will be calculated as described in Section 6.2.1.2 at the time of ORR analysis and at the time of PFS analysis.

Both DOR by BICR and DOR by derived investigator will be calculated and presented by arm.

6.2.2.3. Time to Response

TTR will be calculated at the time of ORR analysis for participants with a confirmed response (CR or PR) in weeks as follows:

TTR (in weeks) = [date of first CR or PR – date of randomization +1]/7

TTR will be summarized using simple descriptive statistics (mean, standard deviation, median, minimum, maximum, first quartile, third quartile).

Both TTR by BICR and TTR by derived investigator will be calculated and presented by arm.

6.2.2.4. Progression-free Survival 2

PFS2 will be calculated in months for Phase 3 portion of the study as follows at the time of PFS analysis and final OS analysis:

PFS2 (months) = [minimum of (date of death, date of second objective PD, date of discontinuation of next-line treatment after first objective PD by investigator assessment) or censoring – date of randomization +1]/30.4375

As defined in Section 3.2.2.6, PD2 is PD after the start of subsequent anticancer therapy based on investigator assessment. A participant will be considered to have an event if

1. the participant had PD by investigator on or prior to start of next-line anticancer treatment, AND started next-line anticancer treatment, AND (had objective PD after start of next-line anticancer treatment OR discontinued next-line anticancer treatment).

   OR

2. the participant died.

The censoring and event date options to be considered for PFS2 analysis are presented in Table 11.

PFS2 time will be estimated using the same Kaplan-Meier method as described for PFS.

Table 11 Outcome, Event Dates, Censoring Reasons and Hierarchy for PFS2 Analyses

Scenario	Date of event/ censoring	Event/ Censoring reason/ Censoring hierarchy
(No PD[a]) and (no death)	Date of last adequate tumor assessment[b] documenting no PD	Censored/ No PD/ 1

Table 11 Outcome, Event Dates, Censoring Reasons and Hierarchy for PFS2 Analyses

Scenario	Date of event/ censoring	Event/ Censoring reason/ Censoring hierarchy
(No PD[a]) and death	Date of death	Event (Death)
(PD[a] and no NTX[c]) and death	Date of death	Event (Death)
(PD[a] date > NTX[c] start date) and (no death)	Start date of NTX[c]	Censored/ Start of new anticancer treatment before PD/ 2
(PD[a] date > NTX[c] start date) and death	Date of death	Event (Death)
(PD[a] date ≤ NTX[c] start date) and (PD2[d] date > NTX[c] start date)		
• If PD2 date ≤ NTX[c] end date	Date of PD2	Event (PD after start of next line anticancer treatment)
• If PD2 date > NTX[c] end date	End date of NTX[c]	Event (Discontinuation of next line anticancer treatment)
(PD[a] date ≤ NTX[c] start date) and (no PD2[d]) and (no death)		
• If NTX[c] end date is non-missing	End date of NTX[c]	Event (Discontinuation of next line anticancer treatment)
• Else if [withdrawal of consent date ≥ date of randomization OR End of study (EOS) = Participant refused further follow-up]	(Withdrawal of consent date) or (EOS visit date where participant refusal of further follow-up is recorded)	Censored/ Withdrawal of consent/ 3
• Else if [lost to follow-up in any disposition page]	Last contact date	Censored/ Lost to follow-up/ 4
• Else if no prior conditions are met	Last contact date	Censored/ Ongoing without PFS2 event/ 5
(PD[a] and no NTX[c]) and (no death)	Last contact date	Censored/ Ongoing without PFS2 event/ 5
(PD[a] date ≤ NTX[c] start date) and (no PD2[d]) and death		
• If NTX[c] end date is non-missing	End date of NTX[c]	Event (Discontinuation of next line anticancer treatment)
• Else if the prior condition is not met	Date of death	Event (Death)

a. PD is the first PD by investigator, without considering any censoring rules

Table 11 Outcome, Event Dates, Censoring Reasons and Hierarchy for PFS2 Analyses

Scenario	Date of event/ censoring	Event/ Censoring reason/ Censoring hierarchy

b. If there are no adequate post-baseline assessments, then the censoring date is the date of randomization. If a participant has initiated next-line anticancer treatment, the last adequate post-baseline assessment on or prior to start date of next-line anticancer treatment will be considered.

c. NTX is the next-line anticancer treatment.

d. PD2 is the first PD by investigator assessment after initiation of next-line anticancer treatment, without considering any censoring rules.

6.2.2.5. Efficacy Endpoints of Arm A Versus Arm C and Versus Arm B Comparison

For the Phase 3 portion of the study, the following efficacy endpoints will also be analyzed between Arm A and Arm C/Arm B using the same analysis methods as those used for the corresponding endpoint of Arm B versus Arm C at the time of PFS analysis. As these comparisons are not part of the multiple testing strategy as outlined in Section 5.1.2.2, no formal statistical testing will be performed on these endpoints between Arm A and Arm C/Arm B:

- PFS by BICR and by derived investigator assessment.
- OS.
- ORR by BICR and by derived investigator assessment.
- DOR by BICR and by derived investigator assessment.
- TTR by BICR and by derived investigator assessment.
- PFS2.

6.2.3. Cohort 3

6.2.3.1. Progression-free Survival

6.2.3.1.1. Main Analysis

The key secondary efficacy analysis will compare PFS time by BICR between Arm D and Arm E using a 1-sided stratified log-rank test at the significance level of 0.025. The analysis will be performed based on the FAS when approximately 73 PFS events by BICR will be observed only if the results for ORR by BICR analysis are statistically significant. Strata will be based on those specified in the IRT system as described in Section 3.4.1.

PFS will be calculated as described in Section 6.1.2.1.1.

The treatment effect will be estimated using a Cox's proportional hazards model stratified by randomization strata to calculate the HR for PFS of Arm D vs. Arm E, along with the corresponding 2-sided 95% CI.

Kaplan Meier estimates will be presented as described in Section 6.1.2.1.1, including the time to follow-up for PFS by BICR.

6.2.3.1.2. Sensitivity/Supplementary Analyses

Sensitivity and supplementary analyses will be conducted as described in Section 6.1.2.1.2.

6.2.3.2. Duration of Response

DOR will be calculated as described in Section 6.2.1.2 at the time of ORR analysis and at the time of PFS analysis.

Both DOR by BICR and DOR by derived investigator assessment will be calculated and presented by arm.

6.2.3.3. Time to Response

TTR will be calculated and summarized at the time of ORR analysis as described in Section 6.2.2.3.

Both TTR by BICR and DOR by derived investigator assessment will be calculated and presented by arm.

6.2.3.4. Overall Survival

OS for Cohort 3 will be calculated at the time of ORR and PFS analysis as described in Section 6.2.2.1.1, but it will be reported descriptively only.

6.3. Other Secondary Endpoints

6.3.1. Patient-Reported Outcomes (PRO)

The following PRO analyses will be performed on the full analysis set at the time of PFS analysis to support the CSR development for Phase 3 portion and for Cohort 3 of the study. All other PRO analyses that are not included in this SAP will be described in detail in a separate PRO analysis plan.

6.3.1.1. Scoring

The EORTC QLQ-C30 and EQ-5D-5L will be scored according to their respective validation papers and User's Guides (Fayers et al.)[12]. The current version at the time of analysis of UK weights will be used to construct the index values of the EQ-5D-5L.

PGIS and PGIC questionnaire items will be analyzed separately. Items within each questionnaire will not be combined or scored.

6.3.1.2. Instrument Completion Rates

For each treatment arm and at each time point, the number and percentage of participants who complete each of the EORTC QLQ-C30, EQ-5D-5L, PGIS, and PGIC will be summarized, as will the reasons for non-completion of these measures.

6.3.1.3. Descriptive Summary

For the Global Health Status/QoL scale, the subscales of the EORTC QLQ-C30 and EQ-5D-5L VAS, summary tables including the mean, median, SD, range and 95% CIs, and line charts depicting the means along with standard error bars over time will be provided for each scale value and change from baseline at each time point by treatment arm.

6.3.1.4. Health Status

For the EQ-5D-5L, the number and percentage of participants reported having "none", "slight", "moderate", "severe", or "extreme/unable" problems at each time point will be summarized.

6.3.2. Pharmacokinetic Analysis

6.3.2.1. Safety Lead-in

Plasma concentrations and PK parameter estimates will be determined for encorafenib, its metabolite (LHY746), irinotecan and its metabolite SN-38, and oxaliplatin (total platinum in plasma and platinum in plasma ultrafiltrate). PK summaries will be presented by cohort by visit for participants in the PK analysis set. Cycle 1 Day 15 is defined as steady state of encorafenib.

6.3.2.1.1. Plasma Concentrations of Encorafenib and LHY746, Irinotecan and SN-38, and Oxaliplatin

In the SLI EC + FOLFIRI cohort, plasma concentrations of encorafenib and its metabolite (LHY746), and irinotecan and its metabolite (SN-38) will be quantified at the time points indicated in Table 12. In the SLI EC + mFOLFOX6 cohort, plasma concentrations of encorafenib and its metabolite (LHY746), and oxaliplatin (total platinum in plasma and platinum in ultrafiltrate) will be quantified at the time points indicated in Table 13.

Table 12 PK Sampling Schedule for Cohort 1 (EC + FOLFIRI) in the Safety Lead-in

Visit	Sampling Time	Irinotecan PK Sample (3 mL Blood)	Encorafenib PK Sample (2 mL Blood)
Cycle 1 Day 1	0 hr (predose[b])	X	
	0.75 hr	X	
	1.5 hr	X[c]	
	2.5 hr	X	
	3.5 hr	X	
	5.5 hr	X	
	7.5 hr	X	
Cycle 1 Day 3	0 hr (predose[d])	X	
Cycle 1 Day 15[a]	0 hr (predose[d])	X	X
	0.75 hr	X	X
	1.5 hr	X[c]	X
	2.5 hr	X	X
	3.5 hr	X	X
	5.5 hr	X	X
	7.5 hr	X	X
Cycle 1 Day 17	0 hr (predose[d])	X	X
Day 1 of Cycle 2 – 6	0 hr (predose[d])	X	X

a. Encorafenib will be administered within 5 minutes before the start of irinotecan infusion
b. Predose is within 30 minutes prior to irinotecan infusion
c. Immediately before irinotecan infusion ends
d. Predose is within 30 minutes prior to encorafenib

Note: the time points are related to the start time of administration of respective drugs. Collection of samples up to and including 8 hours after dose administration should be obtained within ~ 10% of the nominal time relative to dosing (eg, within 6 minutes of a 1-hour sample)

Table 13 PK Sampling Schedule for Cohort 2 (EC + mFOLFOX6) in the Safety Lead-in

Visit	Sampling Time	Oxaliplatin PK Sample (5 mL Blood)	Encorafenib PK Sample (2 mL Blood)
Cycle 1 Day 1[a]	0 hr (predose[b])	X	X
	1 hr	X	X
	2 hr	X[c]	X
	3 hr	X	X
	4 hr	X	X
	6 hr	X	X
	8 hr	X	X
Cycle 1 Day 3	0 hr (predose[b])	X	X
Cycle 1 Day 15[a]	0 hr (predose[b])	X	X
	1 hr	X	X
	2 hr	X[c]	X
	3 hr	X	X
	4 hr	X	X
	6 hr	X	X
	8 hr	X	X
Cycle 1 Day 17	0 hr (predose[b])	X	X
Day 1 of Cycle 2 – 6	0 hr (predose[b])	X	X

a. Encorafenib will be administered within 5 minutes before the start of oxaliplatin infusion
b. Predose is within 30 minutes prior to encorafenib
c. Immediately before oxaliplatin infusion ends

Note: the time points are related to the start time of administration of respective drugs. Collection of samples up to and including 8 hours after dose administration should be obtained within ~ 10% of the nominal time relative to dosing (eg, within 6 minutes of a 1 hour sample)

All plasma concentration values for each participant in the safety set will be included in the bioanalytical plasma concentration listings, and participants will be identified as being in the PK set, as applicable. Individual concentration records will be flagged for the affected visit if any of the following occur:

a. Participant vomits within 4 hours following encorafenib dosing on the day of PK sampling (Cycle 1 Day 1 and Cycle 1 Day 15).

b. Encorafenib plasma levels are not considered to be at steady-state (ie, encorafenib dosing was not performed for at least 3 consecutive days prior to Cycle 1 Day 15).

c. Participant receives a higher or lower dose of encorafenib compared to planned treatment on the day of PK sampling (Cycle 1 Day 1 and Cycle 1 Day 15).

d. Participant receives a higher or lower dose of irinotecan compared to planned treatment on the day of PK sampling (Cycle 1 Day 1 and Cycle 1 Day 15).

e. Participant receives a higher or lower dose of oxaliplatin compared to planned treatment on the day of PK sampling (Cycle 1 Day 1 and Cycle 1 Day 15).

f. PK sampling time deviation >10%

The plasma concentrations of encorafenib and LHY746, and irinotecan and SN-38, and oxaliplatin will be summarized for all nominal time points, including predose (trough) concentrations for Day 1 of Cycle 2 through Cycle 6, using the following descriptive statistics: N (number of participants in the population), n (number of participants with non-missing values), arithmetic mean, standard deviation, coefficient of variation (CV), geometric mean, geometric standard deviation, geometric CV, minimum, median and maximum. Exclusion of an individual concentration-time data point from the calculation of summary statistics will follow the rules below:

- For encorafenib and LHY746 on Cycle 1 Day 1, if any of the above flags [a, c, and f] apply; for encorafenib and LHY746 on Cycle 1 Day 15, if any of the above flags [a, b, c, and f] apply.
- For irinotecan and SN-38 on Cycle 1 Day 1, if any of the above flags [d and f] apply; for irinotecan and SN-38 on Cycle 1 Day 15, if any of the above flags [a, b, d and f] apply
- For oxaliplatin on Cycle 1 Day 1, if any of above flags [a, e and f] apply; for oxaliplatin on Cycle 1 Day 15, if any of above flags [a, b, e, and f] apply.

Concentrations reported as below the limit of quantification (BLQ) will be set to zero for the calculation of arithmetic mean, standard deviation, CV, minimum, median and maximum. For the calculation of geometric mean and geometric CV, BLQ values will be set as missing. All summary statistics below the limit of quantification will be presented as "< X.XX" where X.XX represents LLOQ.

The median plasma concentration-time profiles on Cycle 1 Day 1 and Cycle 1 Day 15 will be presented graphically for each analyte by cohort, as appropriate, using both linear and semi-logarithmic scales and nominal time. In the median plasma concentration-time profiles of irinotecan and SN-38, or oxaliplatin, the time of the predose concentrations on Cycle 1 Day 3 and Cycle 1 Day 17 will be imputed as nominal time 48 hrs post-start of IV infusion on Cycle 1 Day 1, and Cycle 1 Day 15, respectively. Only patients who have at least 1 matched pair of estimated pharmacokinetic parameters available on both Cycle 1 Day 1 and Cycle 1 Day 15 will be included in the median concentration profiles except for encorafenib and LHY746 in the cohort 1 because the first dose of encorafenib will not be administered till Cycle 1 Day 3 in the cohort 1 (EC + FOLFIRI). The median trough plasma concentration-visit plots will be presented for each analyte by cohort, as appropriate, using both linear and semi-logarithmic scales.

6.3.2.1.2. Plasma Pharmacokinetic Parameters for Encorafenib and LHY746, Irinotecan and SN-38, and Oxaliplatin

Pharmacokinetic parameters defined in Table 14 for participants in the PK set will be determined, as data permit and when appropriate, with noncompartmental analysis (NCA) using an internally validated software system, open NCA. Actual blood collection times with respect to the actual dose time of corresponding study drug will be used for PK parameter

calculation. If actual times are not recorded/available, nominal times will be used. For irinotecan and SN-38, and oxaliplatin, the time of predose samples on Cycle 1 Day 3 and Cycle 1 Day 17 will be imputed as the actual time intervals from start of infusion on Cycle 1 Day 1 and Cycle 1 Day 15, respectively, and included for calculation of PK parameters. All BLQ values before the observed maximum plasma concentration (C_{max}) will be set to 0; all BLQ values after C_{max} will be considered as missing.

The AUC parameters will be calculated according to the linear-up log-down trapezoidal rule. Additional PK parameters may be calculated at the discretion of the pharmacokineticist.

All PK parameter values will be presented in data listings by analyte and visit and cohort. Pharmacokinetic parameters (and if appropriate, dose-normalized exposure parameters) (Table 14) will be summarized in tables by analyte and visit and cohort using the following descriptive statistics: N, n, arithmetic mean, standard deviation, CV, geometric mean, geometric CV, minimum, median and maximum. For T_{max} values, median, minimum and maximum will be presented.

Exclusion of pharmacokinetic parameters from the calculation of summary statistics will follow the rules below:

- For encorafenib and LHY746 on Cycle 1 Day 1 (only for the cohort 2, EC + oxaliplatin), if any of the above flags [a and c] apply; for encorafenib and LHY746 Cycle 1 Day 15, if any of the above flags [a, b, and c] apply.

- For irinotecan and SN-38 on Cycle 1 Day 1, if any of the above flags [d] apply; for irinotecan and SN-38 on Cycle 1 Day 15, if any of the above flags [a, b and d] apply.

- For oxaliplatin on Cycle 1 Day 1, if any of above flags [a and e] apply; for oxaliplatin on Cycle 1 Day 15, if any of above flags [a, b and e] apply.

When dose-normalized pharmacokinetic parameters are calculated, exclusion of dose-normalized pharmacokinetic parameters from the calculation of summary statistics will follow the rules below:

- For encorafenib and LHY746 on Cycle 1 Day 1 (only for the cohort 2, EC + oxaliplatin), if any of the above flags [a] apply; for encorafenib and LHY746 Cycle 1 Day 15, if any of the above flags [a and b] apply.

- For irinotecan and SN-38 on Cycle 1 Day 15, if any of the above flags [a and b] apply.

- For oxaliplatin on Cycle 1 Day 1, if any of above flags [a] apply; for oxaliplatin on Cycle 1 Day 15, if any of above flags [a and b] apply.

In addition, some samples may be excluded from the summary analysis based on open NCA group assessment.

Table 14 Pharmacokinetic Parameters

Analytes	PK Parameters
Oral Encorafenib and the metabolite LHY746	C_{max}, C_{min}, C_{trough}, AUC_{last}, AUC_{tau}, $R_{ac,Cmax}$, $R_{ac,AUC}$, MR_{Cmax}, MR_{AUC}, T_{max}, AUC_x, CL/F, V_z/F If appropriate, dose normalized exposure parameters to the planned dose: C_{max}, C_{min}, C_{trough}, AUC_{last}, AUC_{tau}, and AUC_x
IV infusion Irinotecan and the metabolite SN-38 Oxalipatin (platinum in plasma and platinum in plasma ultrafiltrate)	C_{max}, C_{min}, C_{trough}, AUC_{last}, AUC_x, $R_{ac,Cmax}$, $R_{ac,AUC}$, MR_{Cmax}, MR_{AUC}, T_{max}, AUC_{inf}, k_{el}, $t_{1/2}$, CL, V_z If appropriate, dose normalized exposure parameters to the planned dose: C_{max}, C_{min}, C_{trough}, AUC_{last}, AUC_x, and AUC_{inf}

Note: Pharmacokinetic parameters will be determined, as data permit and when appropriate, with noncompartmental analysis (NCA) using an internally validated software system, openNCA.

6.3.2.1.3. Drug-drug Interaction

Effect of steady-state encorafenib on irinotecan and SN-38 or oxaliplatin PK will be evaluated based on overall assessment of the geometric mean ratios (Cycle 1 Day 15/Cycle 1 Day 1) and associated 90% CI for PK parameter estimates of irinotecan and SN-38 or oxaliplatin, respectively, (ie, C_{max} and AUC, CL). Only participants who have a matched pair of estimated pharmacokinetic parameters available Cycle 1 Day 1 and Cycle 1 Day 15 will be included in this analysis. Exclusion is described as in Section 6.3.2.1.2.

The effect of a single dose of encorafenib on oxaliplatin PK, as appropriate, may be evaluated by comparing exposures of oxaliplatin (ie, C_{max}, AUC) on Cycle 1 Day 1 in EC + mFOLFOX6 cohort to historical data. The effect of irinotecan or oxaliplatin on encorafenib and LHY746 may be evaluated by comparison of encorafenib and LHY746 PK results to historical results. Comparisons to historical PK results will be qualitative, exploratory, and only done as data permit.

The final PK analysis may vary at the discretion of the pharmacokineticist.

6.3.2.2. Phase 3

The central laboratory, analytical laboratories, and Pfizer clinical assay group colleagues will be unblinded. If the need arises for early analysis of the PK data (before database lock and release of the randomization codes for the study), a PK unblinding plan will be developed. A PK analyst, who is not associated with the study team, will conduct the analysis to avoid unblinding of the study team. The PK analysis set will be used.

Trough plasma concentrations of encorafenib and LHY746 on Day 1 of Cycle 1 through Cycle 6 will be determined in Arm A and Arm B. Individual concentration records will be flagged as in Section 6.3.2.1.1. All plasma concentration values for each participant in the safety set will be included in the bioanalytical plasma concentration listings, and participants will be identified as being in the PK set, as applicable. The values will be summarized by analyte and visit and arms descriptively (n, mean, standard deviation, CV, median, minimum, maximum, geometric mean, and its associated CV). The median trough plasma concentration-visit plots will be presented for each analyte by arm, as appropriate, using both linear and semi-logarithmic scales.

6.3.2.2.1. Pharmacokinetics of Encorafenib and the Metabolite LHY746 in Mainland Chinese Population

Plasma concentration and pharmacokinetic parameters of encorafenib and LHY746 for Day 1 of Cycle 1 and steady state (Day 15 of Cycle 1, or Day 1 or Day 15 of Cycle 2 if a dose interruption occurs prior to Day 15 of Cycle 1) will be summarized for participants randomized to Arm A or Arm B in China who underwent intensive PK.

Individual concentration records will be flagged as in Section 6.3.2.1.1. All plasma concentration values for each participant in the safety set will be included in the bioanalytical plasma concentration listings, and participants will be identified as being in the PK set, as applicable. Plasma concentration will be summarized descriptively (n, mean, standard deviation, CV, median, minimum, maximum, geometric mean, and its associated CV) by analyte and nominal time for Day 1 of Cycle 1 and steady state. The median plasma concentration-time plots will be presented for each analyte, as appropriate, using both linear and semi-logarithmic scales, for Day 1 of Cycle 1 and steady-state.

Plasma pharmacokinetic parameters listed below, for Cycle 1 Day 1 and steady state, will be estimated using noncompartmental analysis if data permits or if considered appropriate:

- C_{trough}: trough plasma concentration.
- C_{max}: maximum plasma concentrations.
- T_{max}: time to maximum plasma concentration.
- AUC_t: area under the plasma concentration versus time profile from time 0 to time t.
- AUC_{tau}: area under the plasma concentration versus time profile within a dose interval.
- AUC_{inf}: area under the plasma concentration versus time curve to infinity.
- $t_{1/2}$: terminal elimination half-life.
- CL/F: oral plasma clearance.

- V_z/F: apparent volume of distribution.
- R_{ac}: accumulation ratio.
- MR: metabolite ratio.

All PK parameter values will be presented in data listings by analyte for Cycle 1 Day 1 and steady state. The PK parameters will be summarized by analyte, descriptively (n, mean, standard deviation, CV, median, minimum, maximum, geometric mean, and its associated CV).

Participants who vomit within 4 hours following encorafenib dosing on the day of PK sampling (Cycle 1 Day 1 and steady state) or who receive a higher or lower dose of encorafenib compared to planned treatment on the day of PK sampling (Cycle 1 Day 1 and steady state) will be excluded from the summary of plasma concentration and pharmacokinetic parameters and median plasma concentration-time plots. When appropriate, dose normalized concentrations and PK parameters may be used for the analysis. The final PK analysis may vary at the discretion of the pharmacokineticist.

6.3.2.3. Cohort 3

Trough plasma concentrations of encorafenib and LHY746 on Day 1 of Cycle 1 through Cycle 6 will be determined in Arm D. Individual concentration records will be flagged as in Section 6.3.2.1.1. All plasma concentration values for each participant in the safety set will be included in the bioanalytical plasma concentration listings, and participants will be identified as being in the PK set, as applicable. The values will be summarized by analyte and visit descriptively (n, mean, standard deviation, CV, median, minimum, maximum, geometric mean, and its associated CV). The median trough plasma concentration-visit plots will be presented for each analyte by arm, as appropriate, using both linear and semi-logarithmic scales.

6.3.3. Biomarker Analyses

The secondary biomarker endpoint data will be analyzed for the Phase 3 and Cohort 3 portion of the study based on the biomarker analysis sets as defined in Section 4. Data will be presented by treatment arm. Summaries of baseline levels, changes from baseline (where appropriate), expression or genetic alterations/status will be reported. For continuous variables, summary statistics may include the mean, ratio to baseline, standard deviation, median, first and third quartiles, %CV, and minimum/maximum levels; for categorical variables, summary may include number and percentage, odds ratio, and frequency statistics, as appropriate.

No duplicate results for biomarker or pharmacodynamic markers are expected. However, if more than one record is received for a particular time point, the duplicate records might be averaged or most recent result will be used for continuous data. For non-continuous data, the results will be reviewed by the study team and a representative sample will be selected. When participants provide both archival and de novo tumor tissue samples at screening, data from the de novo sample will be used in the summary analysis.

Data from biomarker assays may be analyzed using graphical methods and descriptive statistics such as Kaplan-Meier estimates of efficacy parameters (eg, PFS, OS) with biomarkers as a covariate, or Box plots. The statistical approach may examine correlations of biomarker results with pharmacokinetic parameters and measures of anti-tumor efficacy.

Regarding exploratory biomarkers, the data analysis will be conducted with the goal of identifying predictive and/or pharmacodynamic (ie, baseline and change from baseline values) biomarkers associated with clinical outcome, encompassing both safety and efficacy. These analyses may not be reported in the Clinical Study Report. Candidate biomarkers will be validated in subsequent trials.

6.3.3.1. Tumor Tissue Analyses

6.3.3.1.1. MSI Status Analyses

Retrospective analyses for tumor-based MSI status performed at a central laboratory will be based on the tumor tissue analyses set.

MSI status in tumor tissue is measured at baseline (pre-screening/screening). Participants are expected to be MSI-H, MSS/MSI-L (presented combined in the tables) or MSI-unknown.

A descriptive summary table at baseline for MSI status will be provided. When appropriate, a descriptive summary table comparing results from local MSI/dMMR testing and central MSI testing within the tumor tissue analyses set will be provided. The number and percentage of participants in each category will be tabulated. All data will be listed.

6.3.3.1.2. *BRAF* V600E Analyses

Analyses for tumor-based *BRAF* V600E will be based on the full analysis set.

Diagnostic concordance assessment of *BRAF* V600E status between local tumor tissue/plasma results (as reported in the eCRF) and the reference, central PCR-based QIAGEN *BRAF* V600E FDA-approved test will be presented. The reference, the central test is the tumor tissue PCR testing performed at screening as part of the enrollment criteria, and all participants are therefore considered *BRAF* V600E positive (ie, detected). Overall concordance will be summarized by treatment arm and in total.

6.3.3.2. ctDNA Analyses

Retrospective analyses of ctDNA alterations will be based on either the cfDNA analysis set or the paired cfDNA analysis set. Note that mainland China uses a different vendor than the rest of the world for this biomarker; hence data from mainland China will be reported separately.

ctDNA alterations will be measured at defined timepoints including baseline, on-treatment and End of Treatment. Biomarkers will be summarized at defined timepoints when appropriate. Participants may have detectable ctDNA before, during and/or after treatment. Participants with detectable ctDNA are expected to have either none or one or more measurable tumor biomarkers including *BRAF* V600 alterations.

The following definitions will be used for subgroups based on whether ctDNA was detected or not at C1D1 and C1D15:

- • "ND/ND" = ND ctDNA at C1D1 and C2D15
- • "D/ND" = D at C1D1 and ND at C2D15
- • "ND/D" = ND at C1D1 and D at C2D15
- • "D/D" = D at C1D1 and C2D15

PFS and OS analyses will be stratified by ctDNA status at baseline (detected or not-detected) and/or changes in ctDNA status (ND/ND, D/ND, ND/D, D/D [C1D1/C2D15]) and will use the same Kaplan-Meier method as described for PFS. Kaplan-Meier curves with appropriate summary statistics including median time and associated 2-sided 95% CIs will be calculated.

A summary table of measurable *BRAF* V600 VAF (mean, maximum, minimum and standard deviation) at available timepoints will be provided. A summary table will be provided for subgroups defined by BRAF V600 VAF at baseline as measurable or not-measurable, or as ND/ND, D/ND, ND/D, D/D based on ctDNA detection at C1D1 vs C2D15.

Summaries of objective response (yes=objective response, and no=no objective response) will be provided for subgroups defined by the *BRAF* V600 alteration status (measurable or not-measurable) at baseline or as ND/ND, D/ND, ND/D, D/D based on ctDNA detection at C1D1 vs C2D15. Graphical analyses (eg stick plots, box plots, etc) will be provided when appropriate.

PFS and OS analyses will be stratified by BRAF V600 status at baseline (measurable or not-measurable) and/or changes in ctDNA detection status ND/ND, D/ND, ND/D, D/D [C1D1/C2D15])) and will use the same Kaplan-Meier method as described for PFS. Kaplan-Meier curves with appropriate summary statistics including median time and associated 2-sided 95% CIs will be calculated.

Diagnostic concordance assessment of *BRAF* V600E status between the plasma-based *BRAF* V600E test and the reference, central PCR-based QIAGEN *BRAF* V600E FDA-approved test will be presented for cfDNA diagnostics concordance analysis set. The reference, the central test is the tumor tissue PCR testing performed at screening as part of the enrollment criteria, and all participants are therefore considered *BRAF* V600E positive (ie, detected). Overall concordance will be summarized by treatment arm and in total.

6.3.4. Surgical conversion rate (exploratory endpoint)

For the Phase 3 and Cohort 3 portion of the study the surgical conversion rate, defined as the rate of participants who become eligible for surgery and undergo surgery with curative intent as a result of study intervention, will be evaluated in the safety analysis set. Summary data will be presented by treatment arm.

6.4. Subset Analyses

All the subset analyses will be exploratory and thus descriptive in nature; no adjustment for multiplicity will be performed. Analyses will only be performed if there is sufficient sample size. The determination of whether or not there is sufficient sample size will be defined after enrollment is complete and prior to database lock. As a general rule, time to event analyses will not be performed on subgroups unless there are at least 10 events or 10% of total events, the largest, within the defined subset. Analyses of response rates and safety will only be performed if there are at least 10 participants or 10% of total participants within the defined subset.

The following subset analyses will be performed for ORR by BICR, PFS by BICR and OS for Phase 3 FAS participants (Arm B vs. Arm C) and for ORR by BICR and PFS by BICR for Cohort 3 FAS participants:

- Age (<65 years vs ≥65 years).
- Gender (Male vs Female).
- ECOG performance status based on the IRT system (0 vs 1).
- Number of organs involved at baseline (≤2 vs ≥3) according to BICR.
- Side of tumor (left vs right). Classifications are below:
  - Descending (Left)
  - Rectosigmoid (Left)
  - Sigmoid (Left)
  - Rectum (Left)
  - Splenic Flexure (Left)
  - Cecum (Right)
  - Ascending (Right)
  - Transverse (Right)
  - Hepatic Flexure (Right)
- Presence of liver metastases at baseline, based on Target and Non-target lesion assessment (yes vs no), according to BICR.

In addition, the following subset analyses may be performed for ORR by BICR, PFS by BICR and OS for Phase 3 and Cohort 3 FAS participants to support country specific regulatory submission, if required. PK will also be analyzed (see Section 6.7). Results of these subset analyses may not be included in the main CSR.

- Japanese subgroup (participants randomized in Japan).
- Korean subgroup (participants randomized in Korea).
- Chinese subgroup (participants randomized in China).

Subset analyses for PFS, OS will use the censoring rules described in Section 6.1.2.1.1. The unstratified HR and its corresponding 95% CI will be calculated per subset level and will be presented in a forest plot.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

The following analyses will be performed by cohort on the FAS for the SLI, and by treatment arm for the Phase 3 and for the Cohort 3 on the FAS.

6.5.1.1. Demographic Summary

The following demographic parameters will be listed and summarized by number and percentage:

- Gender (male, female).
- Age (<18; 18-<65; ≥65; ≥75).
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Other, Not Reported).
- Racial designation for Asian (Japanese, Korean, Chinese, other).
- Ethnicity (Hispanic/Latino, not-Hispanic/Latino).

Age (continuous) will be summarized with descriptive statistics (mean, median, standard deviation, minimum, and maximum).

6.5.1.2. Baseline and Disease Characteristics

The following baseline characteristics and disease characteristics will be listed and summarized:

- Body site as collected on the "Primary Diagnosis" eCRF page.
- Stage at initial diagnosis.
- Primary tumor removed prior to study entry (No, Complete resection, Partial resection).
- Side of tumor (left vs right).
- Number of organs involved based on baseline target and non-target lesion assessment by investigator (SLI) or BICR (Phase 3 and Cohort 3).
- Presence of liver metastases at baseline, based on target and non-target lesion assessment by investigator (SLI) or BICR (Phase 3 and Cohort 3).

- Time since initial diagnosis, defined as [date of first dose (SLI) or date of randomization (Phase 3 and Cohort 3) – date of initial diagnosis]/30.4375.
- Time since recurrence/metastatic disease (months), defined as [date of first dose (SLI) or date of randomization (Phase 3 and Cohort 3) – date of recurrence/metastatic disease]/30.4375.
- ECOG Performance status.
- Line of therapy (for SLI only): first line if no prior treatment; second line if participant received a prior treatment defined as advanced/metastatic or locoregional disease or maintenance or if neoadjuvant/adjuvant with a disease recurrence occurred during or within 6 months of the last dose of therapy.
- Local *BRAF* V600E status as collected on the "Local BRAF Results" eCRF page combining results from tumor tissue and plasma sample.
- Central *BRAF* V600E status based on tumor tissue.
- *KRAS* and *NRAS* results as collected on the "Local KRAS Results" eCRF page combining results from tumor tissue and plasma sample.
- MSI/MRR results as collected on the "Local MSI/MRR Results" eCRF page combining results from tumor tissue and plasma sample.
- Baseline CEA values (≤ 5 ug/L vs. > 5 ug/L) and CRP values (≤ 10 mg/L vs. > 10 mg/L) and descriptive summary.

6.5.1.3. Stratification Factors

The number of participants per strata will be summarized based on strata reported in the IRT system and strata reported in CRF as described in Section 3.4.1.

6.5.1.4. Medical History

Medical history will be coded using the most current version of MedDRA and summarized by MedDRA's SOC and PT. Each participant will be counted only once within each PT or SOC. Summaries will be ordered by primary SOC and PT by decreasing frequency based on mFOLFOX6 Cohort for SLI, the Arm B for Phase 3, and Arm D for Cohort 3.

6.5.1.5. Prior Anticancer Treatments

Prior anticancer treatments include systemic therapy, radiation, and surgery.

The number and percentage of participants in each of the following anticancer therapy categories will be tabulated:

- Participants with at least one prior anticancer systemic therapy (SLI).
- Participants with at least one prior anticancer neoadjuvant/adjuvant therapy (Phase 3 and Cohort 3).
- Participants with at least one prior anticancer radiotherapy.

- Participants with at least one prior anticancer surgery, excluding procedure=biopsy.

All prior anticancer therapy will be presented in a listing.

6.5.2. Study Conduct and Participant Disposition

The following analyses will be performed by cohort on the FAS for the SLI, and by treatment arm for the Phase 3 and for the Cohort 3 on the FAS.

6.5.2.1. Patient Disposition

The disposition summary will be created for the SLI, the Phase 3 portion and the Cohort 3 of the study. A listing will also be presented.

The following disposition categories will be summarized:

- Total number of participants screened.
- Primary reasons for screening discontinuation.

Number (%) of participants who screened failed overall and grouped by the main reason

- Number (%) of participants who were randomized (Phase 3 and Cohort 3 only).
- Number (%) of participants who were randomized but not treated (Phase 3 and Cohort 3 only).
- Number (%) of participants who received at least one dose of study drug.
- Number (%) of participants who are still on treatment.
- Number (%) of participants who discontinued the treatment.
    - Primary reasons for treatment discontinuation.
- Number (%) of participants who discontinued the treatment but are still in long-term follow-up.
- Number (%) of patients who discontinued the study.
    - Primary reasons for study discontinuation.

The number and percent of participants enrolled by region, country and site, and the number and percent of participants in each analysis set as described in Section 4 will also be summarized. A listing of analysis sets will be provided. A listing will show the randomized/assigned treatment along with actual treatment received.

COVID-19 Related Disposition

A listing of all participants affected by COVID-19 related study disruption will be created. The listing will present subject number identifier by investigational site, and a description of COVID-19 related events including:

- Protocol deviations.
- Adverse events treatment.
- Treatment discontinuation.
- Study discontinuation.
- Death.

Ukraine/Russia War Related Disposition

A listing of all participants affected by Ukraine/Russia war related study disruption may be created.

6.5.2.2. Protocol Deviations

Potentially important protocol deviations (PIPDs) will be compiled prior to database closure and will be summarized by category and also presented in a listing. Categories will be assigned by the study Clinician.

In addition, PIPDs related to COVID-19 will be presented in a separate table and listing.

6.5.3. Study Treatment Exposure

The following analyses will be performed by cohort for the SLI and by treatment arm for the Phase 3 and the Cohort 3 on safety analysis set.

Treatment Regimens

- Safety Lead-in.
    - Participants enrolled in the SLI will receive either a regimen of EC + mFOLFOX6 or a regimen of EC + FOLFIRI in 28 day cycles. Encorafenib is dosed at 300 mg daily; cetuximab, mFOLFOX6 and FOLFIRI are dosed once every 2 weeks with an IV infusion.
- Phase 3.
    - Arm A: Participants randomized to Arm A will receive encorafenib + cetuximab in 28 day cycles. Encorafenib is dosed at 300 mg daily and cetuximab is dosed once every 2 weeks with an IV infusion.

- Arm B: Participants randomized to Arm B will receive a regimen of EC + mFOLFOX6 in 28-day cycles. The dosing schedule is the same as the SLI.

- Arm C: Participants randomized to the Arm C will receive one of the standard-of-care chemotherapy per physician choice as described in Table 13 in the protocol. mFOLFOX6, and FOLFOXIRI with or without bevacizumab are dosed once every 2 weeks with an IV infusion; and CAPOX with or without bevacizumab is dosed once every 3 weeks with an IV infusion.
  Within Arm C, the number of participants receiving different regimens (mFOLFOX6, FOLFOXIRI, or CAPOX with or without bevacizumab) will be summarized.

- Cohort 3.

  - Arm D: Participants randomized to Arm D will receive a regimen of EC + FOLFIRI in 28-day cycles. The dosing schedule is the same as the SLI.

  - Arm E: Participants randomized to Arm E will receive FOLFIRI with or without bevacizumab and are dosed once every 2 weeks with an IV infusion, see Table 15 in the protocol.

Duration of Exposure

The date of receiving the last non-zero dose is considered as last dose date.

- Encorafenib: daily dosing.

  duration of exposure (weeks) = (last dose date – first dose date +1)/7

- Cetuximab, mFOLFOX6, FOLFIRI, bevacizumab: once every two weeks on 28-day Cycle for all drugs except 5-FU:

  duration of exposure (weeks) = (last dose date – first dose date +14)/7

  - 5-FU if the last dose is a bolus:

    duration of exposure (weeks) = (last dose date – first dose date +14)/7

  - 5-FU if the last dose is an infusion:

    duration of exposure (weeks) = (last dose date – first dose date +12)/7

- CAPOX (oxaliplatin + capecitabine):

  - Oxaliplatin: once every three weeks on 21-day cycle

    duration of exposure (weeks) = (last dose date – first dose date + 21)/7

- Capecitabine: twice daily for 2 weeks followed by a 1-week rest period once every 3 weeks

  duration of exposure (weeks) = (last dose date – first dose date + 7)/7

- Bevacizumab: once every three weeks on 21-day cycle

  duration of exposure (weeks) = (last dose date – first dose date +21)/7

- Overall regimen:

  duration of exposure (weeks) = (last dose date of the last drug (including + 1 or 7 or 14 or 12 or 21 depending on last drug type) – first dose date of the first drug)/7

Cumulative Dose

Cumulative dose is defined as:

- **Intended cumulative dose** (mg or mg/kg or mg/m^2) = sum of all protocol specified doses across each intended day of dosing, as per protocol.
- **Actual cumulative dose** (mg or mg/kg or mg/m^2) = sum of all actual doses that the participant received. The actual doses reported in CRF are converted to mg/kg or mg/m^2 based on height and weight at most recent collection on or before dosing. BSA is derived using the DuBois & DuBois formula

  $$BSA = 0.007184 * Height^{0.725} * Weight^{0.425}.$$

For participants who did not take any drug the actual cumulative dose is by definition equal to zero.

Dose Intensity

- **Intended dose intensity** (mg/day or mg/kg/day or mg/m^2/day) = intended cumulative dose/duration of exposure.
- **Actual dose intensity** (mg/day or mg/kg/day or mg/m^2/day) = actual cumulative dose/duration of exposure.
- **Relative dose intensity** (%) = 100* (actual dose intensity/intended dose intensity).

A summary of exposure, including duration of exposure, cumulative dose, actual dose intensity, and relative dose intensity (including categories <50%, 50%-<75%, 75%-<90%, 90%-<110%, and ≥110%, if applicable), will be presented for each study drug. Duration of exposure will also be categorized by time intervals (eg, < 4 weeks, 4-12 weeks, 12-48 weeks, etc. as appropriate for the protocol) for which frequency counts and percentages of participants will be provided.

Actual doses administered will also be listed for each participant by treatment arm.

6.5.3.1. Dose Modification

Dose modification will be summarized based on the dosing data collected on the study treatment CRF page. Dose reduction and interruption will be summarized for each study drug. Time to first dose modification (reduction and/or interruption) will be calculated and summarized with descriptive statistics for participant with at least one dose modification.

Dose Reduction

A dose reduction is defined as a decrease in dose of at least 10%, from the protocol-planned dose and a decrease from the previous non-zero dose, even if this decrease has been directly preceded by an interruption. For encorafenib and capecitabine, to qualify as a dose reduction, it should last for 2 or more days. For example, for encorafenib, in the sequence of total daily dose 300 mg – 0 mg – 200 mg (for at least 2 days), the 200 mg dose will be counted as a reduction. Also, for patients taking FOLFIRI in Cohort 1 in the SLI, the encorafenib dose is not given for Day 1 and Day 2, with the first encorafenib dose planned on Day 3. Days 1 and 2 will not count toward any reductions. A day 3 encorafenib dose of 200 mg for example would constitute a dose reduction since 300 mg was planned at Day 3.

The number and percentage of participants with at least one dose reduction as well as a breakdown of dose reductions (1, 2, ≥3) will be summarized by cohort or treatment arm.

Reasons for dose reductions will also be summarized. Participants can contribute to more than one reason if multiple dose reductions occurred for different reasons, but will only be counted once per reason. Percentages will be calculated based on the total number of participants in the safety analysis set.

Dose Interruption

For daily dosing of encorafenib and capecitabine, an interruption is defined a 0 mg dose administered for one or more days. For intermittent dosing of IV drugs, such as cetuximab, Oxaliplatin, Leucovorin, 5-FU, Irinotecan and bevacizumab, an interruption is defined as more than 20 days (delay of at least 7 days) between successive start dates with non-zero actual doses for regimens on 28-day cycles (treatment on D1 and D15 of every cycle) or as more than 27 days (delay of at least 7 days) between successive start dates with non-zero actual doses for regimens on 21-day cycles (treatment on D1 of every cycle). For capecitabine the off-week, that is part of the regimen, will not be counted as an interruption. For patients taking FOLFIRI in Cohort 1 of the safety lead-in, the planned encorafenib doses of 0 mg on Days 1 and 2 will not constitute a dose interruption.

The number and percentage of participants with dose interruptions will be summarized. Percentages will be calculated based on the total number of participants in the safety analysis set.

6.5.4. Concomitant Medications

Concomitant medications will be summarized by cohort on the FAS set for the SLI, and by treatment arm on the FAS for the Phase 3 and for the Cohort 3.

Concomitant medications are medications, other than study medications, which started prior to first dose of study treatment and continued during the on-treatment period (see Section 5.2.7) as well as those started during the on-treatment period. Concomitant medications will be coded in the WHO Drug coding dictionary and will be tabulated by ATC classification level 2 and PT by decreasing frequency based on mFOLFOX6 Cohort for SLI, the Arm B for Phase 3, and Arm D for Cohort 3. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used. A participant will be counted only once within a given drug name, even if he/she received the same medication at different times.

6.5.5. Subsequent Anticancer Therapies/Procedures

Subsequent anticancer therapies/procedures will be summarized by cohort on the FAS set for the SLI, and by treatment arm on the FAS for the Phase 3 and for the Cohort 3.

Subsequent anticancer therapies are defined as therapies entered on the “Follow-up Cancer Therapy” CRF page. Subsequent anticancer radiation and surgery will be derived based on the data collected on the "Radiation Treatment” and “Non-Drug Treatments - Cancer Surgery” CRF page, respectively. The radiation treatments and the surgeries will be considered as the subsequent anticancer procedures if the date of the radiation/surgery is after the last dose date of all study drugs.

The number and percentage of participants within each category (medication therapy, radiation therapy, and surgeries) will be provided.

Medications will be coded using the WHO Drug coding dictionary and will be tabulated by drug category and PT by decreasing frequency based on mFOLFOX6 Cohort for SLI, the Arm B for Phase 3, and Arm D for Cohort 3. The final list of drug category will be provided upon medical review of all subsequent anti-cancer systemic treatment before database lock.

All subsequent anticancer therapy will be presented in a listing.

6.6. Safety Summaries and Analyses

All safety analyses will be performed on the safety analysis set by cohort for the SLI and by treatment arm for the Phase 3 and for the Cohort 3.

Safety data will be presented in tabular and/or graphical format and summarized descriptively, where appropriate.

6.6.1. Adverse Events

All analyses will be based on TEAEs as defined in Section 3.5.1 unless otherwise specified. AEs not considered treatment emergent will be flagged in data listings.

Seriousness, toxicity grade, action taken (interruption, reduction, and withdraw) of AEs are as reported by the investigator on the "Adverse Events" CRF.

A high level summary of AEs will include the number and percent of participants with:

- Any TEAEs.
- Serious TEAEs.
- TEAEs with CTCAE Grade 3 or 4.
- TEAEs with CTCAE Grade 5.
- Related TEAEs to:
  - Any study drug.
  - Encorafenib.
  - Cetuximab.
  - Other study intervention (i.e capecitabine, 5-FU, irinotecan, leucovorin, oxaliplatin, and bevacizumab).
- Related TEAEs to any study drug with CTCAE Grade 3 or 4.
- Related TEAEs to any study drug with CTCAE Grade 5
- Related serious TEAEs to:
  - Any study drug.
  - Encorafenib.
  - Cetuximab.
  - Other study intervention.
- TEAEs leading to dose interruption of:
  - Any study drug.
  - Encorafenib.
  - Cetuximab.
  - Other study intervention.

- TEAEs leading to dose reduction of:
  - Any study drug.
  - Encorafenib.
  - Cetuximab.
  - Other study intervention.
- TEAEs leading to permanent discontinuation from:
  - Any study drug.
  - Encorafenib.
  - Cetuximab.
  - Other study intervention.
- TEAEs leading to dose reduction of any study drug due to COVID-19.
- TEAEs leading to permanent discontinuation from any study drug due to COVID-19.

Each unique AE at the PT level in each cohort/treatment arm of the study for a participant is included in the count.

Adverse Events by SOC and PT

The following summaries will be created by SOC and PT by decreasing frequency based on mFOLFOX6 Cohort for SLI, the Arm B for Phase 3, and Arm D for Cohort 3:

- TEAEs by maximum toxicity grade (all causality).
- TEAEs by maximum toxicity grade (treatment related).
- COVID-19 related TEAEs (all causality).
- Serious TEAEs by maximum toxicity grade (all causality).
- Serious TEAEs by maximum toxicity grade (treatment related).
- TEAEs leading to dose reduction of any study drug due to COVID-19.
- TEAEs leading to permanent discontinuation from any study drug due to COVID-19.

An event will be considered treatment related if the investigator considered the event related to at least one study drug or this information is unknown.

Adverse Events by PT Only

The following summaries will be created by AE PT (ie, summaries will not include SOC) by decreasing frequency based on mFOLFOX6 Cohort for SLI, the Arm B Phase 3, and Arm D for Cohort 3:

- TEAEs (all causality) experienced by ≥10% of participants in at least one cohort (SLI) or treatment arm (Phase 3 and Cohort 3).
- Treatment related TEAEs experienced by ≥10% of participants in at least one cohort (SLI) or treatment arm (Phase 3 and Cohort 3).
- TEAEs with CTCAE Grade ≥3 (all causality) by PT and maximum toxicity grade.
- Serious TEAEs (all causality).
- COVID-19 related serious TEAEs (all causality).

A cutoff may be applied to some of the above AE tables for intext display.

Adverse Events Leading to Dose Modification by PT

The following summaries will be created by PT of AEs by decreasing frequency of PTs based on mFOLFOX6 Cohort for SLI, the Arm B for Phase 3, and Arm D for Cohort 3:

- TEAEs leading to dose interruptions by maximum toxicity grade (all causality).
- TEAEs leading to dose reductions by maximum toxicity grade (all causality).
- TEAEs leading to permanent discontinuation by maximum toxicity grade (all causality).
- TEAEs leading to dose interruptions by maximum toxicity grade (treatment related).
- TEAEs leading to dose reductions by maximum toxicity grade (treatment related).
- TEAEs leading to permanent discontinuation by maximum toxicity grade (treatment related).

The AE by PT summary will be created separately for:

- Any study drug.
- Encorafenib.
- Cetuximab.

- Other study intervention.

A cutoff may be applied to the tables for AE leading to dose modification. The treatment related tables, for AE leading to dose modification, will include AEs treatment related to a specific drug for which the dose modification was implemented (i.e. treatment related to encorafenib for AEs leading to dose reduction of encorafenib). Each participant will be counted only once within each PT.

All AEs and their attributes will be presented in data listings sorted by participant identifier, AE and date of onset of the AE.

6.6.2. Deaths

All deaths, deaths within 28 days after last dose of study drug, deaths within 30 days after first dose of study drug as well as the primary reason for death, will be tabulated based on information from the "Death Details" and "Disposition Long-term Follow-Up" eCRFs.

- Number of deaths.
- Number of deaths within 28 days after *last* dose of study treatment.
- Number of deaths within 30 days after *first* dose of study treatment.
- Primary cause of death.
    - Disease under study.
    - Study treatment toxicity.
    - Unknown.
    - Other.

In addition, if there are ≥5 deaths due to COVID-19, a separate death summary will be created.

Date and cause of death will be provided in patient data listing together with selected dosing information (study treatment received, date of first/last administration, dose).

6.6.3. Adverse Events of Special Interest

Specific groupings of AEs of special interest (AESI) will be considered and the number of patients with at least one event in each grouping will be reported. Such groups consist of AEs for which there is a specific clinical interest in connection with encorafenib treatment (ie, where encorafenib may influence a common mechanism of action responsible for triggering them) or AEs which are similar in nature (although not identical), and there is potential overlap with the safety profile of cetuximab and/or chemotherapy.

All AESI groupings are defined through the use of PT, High Level Terms (HLT), SOC, Standardized MedDRA Queries (SMQ) or through a combination of these components. The MedDRA terms that define each AESI grouping will be outlined in the Case Retrieval Strategy (CRS), which will be used to map reported AEs to the notable AEs groupings and will be maintained by the Sponsor. The list of AESIs may be updated during the course of the trial based on accumulating safety data.

A high level summary of AESIs will include the number and percent of participants with:

- Any AESIs.
- Serious AESIs.
- AESIs with CTCAE Grade 3 or 4.
- AESIs with CTCAE Grade 5.
- Related AESIs.
- Related AESIs with CTCAE Grade 3 or 4.
- Related AESIs with CTCAE Grade 5.
- Related serious AESIs.

In addition, AESIs will be summarized by grouping and PT as follows:

- AESIs by maximum toxicity grade (all grade, Grade 3, Grade 4, Grade 5, all causality).
- AESIs by maximum toxicity grade (all grade, Grade 3, Grade 4, Grade 5, treatment related).
- Serious AESIs all grade (all causality).
- Serious AESIs all grade (treatment related).

A cutoff may be applied to some of the above AESI tables, for intext display.

6.6.4. Laboratory Data

Laboratory results will be converted to International System of Units (SI) units which will be used for applying toxicity grades and for all summaries.

As described in Section 3.4, baseline is defined as the last completed assessment prior to date of first dose for the safety assessments. If there are multiple assessments that meet the baseline definition on the same day without the ability to determine which was truly last, then the worst grade will be assigned as the baseline grade.

Laboratory results will be programmatically classified according to NCI-CTCAE version 4.03 grade. Non-numerical qualifiers will not be taken into consideration in the derivation of

grade (eg, hypokalemia Grade 1 and Grade 2 are only distinguished by a non-numerical qualifier and therefore Grade 2 will not be derived). In summary statistics the number and percentage of participants corresponding to grades that only include non-quantitative criteria will be displayed as a blank or NA (not assessed) rather than 0. If there is any overlap between grade criteria (eg, CTCAE grading criteria for Creatinine Increased – a value can fall into one range based on comparison to ULN and another range based on comparison to baseline), the highest (worst) grade would be assigned to that record. Grade 5 is defined in the CTCAE criteria guidance as an event with an outcome of death. Since laboratory data does not collect an outcome, Grade 5 is not used when programmatically grading laboratory data.

Grade 0 or Outside Toxicity Reference (OTR) is not defined specifically in the CTCAE guidance. However, programmatically this is used as a category to represent those participants who did not meet any of the Grades 1 to 4 criteria. If the laboratory value is evaluable for CTCAE criteria grading (numeric value is present, valid units and ranges are present as required to allow conversion to standard units and grading), and does not qualify for any of the Grade 1-4 criteria for a given lab test, then the value is assigned as Grade 0 or OTR.

Abnormalities will be described using the worst grade post-baseline. When determining the maximum post-baseline grade for a given participant and CTCAE test, the maximum across all analytes and assessments contributing to that CTCAE test will be used. Several laboratory tests have bi-directional grading criteria defined so that both low (hypo) and high (hyper) values can be graded separately. Each criterion will be summarized separately. In the cases where a value is graded as a Grade 1, 2, 3, or 4 for one of the directions, that value will also be assigned as a Grade 0 for the opposite direction for that test. For example, a value meeting the criteria for Grade 3 Hypercalcemia will be classified as a Grade 0 Hypocalcemia. For CTCAE terms that can be derived using one of several laboratory tests, the maximum post-baseline grade for a given participant and CTCAE term will be the maximum across all possible laboratory tests.

For **WBC differential counts** (total neutrophil [including bands], lymphocyte, monocyte, eosinophil, and basophil counts), the absolute value will be used when reported by the lab. When only percentages are available (this is mainly applicable for neutrophils and lymphocytes, because the CTCAE grading is based on the absolute counts), the absolute value is derived as follows:

$$\text{Derived differential absolute count} = (\text{WBC count}) \times (\text{Differential \%value}/100)$$

If the investigator reports both the absolute and % value for Neutrophils or Lymphocytes from the same laboratory sample date and participant, ONLY the absolute value will be graded. The % value will not be graded in this scenario.

If the % value is converted to the differential absolute count for grading and the LLN for the differential absolute count is not available (only LLN for % is available) then Grade 1 will be assigned if the following conditions are met:

- Lymphocyte count decreased:
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥800/mm^3.
- Neutrophil count decreased
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥1500/mm^3.

For **calcium**, CTCAE grading is based on Corrected Calcium and Ionized Calcium. Corrected Calcium is calculated from Albumin and Calcium as follows:

- Corrected Calcium (mg/dL) = measured total calcium (mg/dL) – 0.8 [serum albumin (g/dL)-4].

Laboratory toxicities will be tabulated using descriptive statistics (number of participants and percentages). If both central and local labs are collected for a subject, summaries of worst on-treatment abnormalities will be based on both local and central lab data.

Additional laboratory results that are not part of NCI-CTCAE will be presented according to the categories: below normal limit, within normal limits and above normal limit (according to the laboratory normal ranges). The following summary tables will be created:

- Shift summary of laboratory parameters during the on-treatment period by maximum CTCAE grade.
- Shift summary of laboratory parameters from ≤ Grade 2 at baseline to ≥ Grade 3 post-baseline.
- Shift summary of laboratory test results with no CTCAE criteria by worst (below, within or above normal range) on-treatment assessment.

All laboratory test results will be presented in a data listing sorted by patient identifier, laboratory test, and date/time of collection. The CTCAE grades and the classifications relative to the laboratory reference ranges will be presented. Values outside laboratory

normal ranges will be flagged where appropriate and the central laboratory data and local laboratory data will be flagged accordingly.

Drug Induced Liver Toxicity

Alanine aminotransferase (ALT), aspartate aminotransferase (AST), Alkaline Phosphatase (ALP), and total bilirubin (TBILI) are used to assess possible drug induced liver toxicity. The ratios of test result over the ULN will be calculated and classified for these three parameters during the on-treatment period.

Summary of liver function tests will include the following categories. The number and percentage of participants with each of the following during the on-treatment period will be summarized:

- ALT ≥3×ULN, ALT ≥5xULN, ALT ≥10×ULN, ALT ≥20×ULN.
- AST ≥3×ULN, AST ≥5xULN, AST ≥10×ULN, AST ≥20×ULN.
- (ALT or AST) ≥3×ULN, (ALT or AST) ≥5×ULN, (ALT or AST) ≥10×ULN, (ALT or AST) ≥20×ULN.
- TBILI ≥2×ULN.
- Concurrent ALT ≥3×ULN and TBILI ≥2×ULN.
- Concurrent AST ≥3×ULN and TBILI ≥2×ULN.
- Concurrent (ALT or AST) ≥3×ULN and TBILI ≥2×ULN.
- Concurrent (ALT or AST) ≥3×ULN and TBILI ≥2×ULN and ALP ≥2×ULN.
- Concurrent (ALT or AST) ≥3×ULN and TBILI ≥2×ULN and ALP <2×ULN or missing.

Concurrent measurements are those occurring on the same date.

Categories will be cumulative, ie, a participant with an elevation of AST ≥10×ULN will also appear in the categories ≥5×ULN and ≥3×ULN. Liver function elevation and possible Hy's Law cases will be summarized using frequency counts and percentages.

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will also be created, with different symbols for different treatment arms, by graphically displaying:

- Peak serum ALT (/ULN) vs peak total bilirubin (/ULN) including reference lines at ALT=3×ULN and total bilirubin=2×ULN.

- Peak serum AST(/ULN) vs peak total bilirubin (/ULN) including reference lines at AST=3×ULN and total bilirubin=2×ULN.

In addition, a listing of all TBILI, ALT, AST and ALP values for participants with a post-baseline TBILI ≥2×ULN, ALT ≥3×ULN or AST ≥3×ULN will be provided.

6.6.5. Vital Signs

Vital signs data includes weight, pulse, systolic blood pressure (BP), and diastolic BP. Measurements were only to be provided once per timepoint. If multiple assessments are provided per timepoint, the maximum value will be used for reporting.

The following criteria define clinically notable vital sign abnormalities:

- Clinically notable elevated values.
    - Systolic BP: ≥160 mmHg and an increase ≥20 mmHg from baseline.
    - Diastolic BP: ≥100 mmHg and an increase ≥15 mmHg from baseline
    - Heart rate (collected as pulse rate in the vital signs eCRF): ≥120 bpm with increase from baseline of ≥15 bpm
    - Weight: increase from baseline of ≥10%
    - Body temperature [C]: ≥37.5 C
- Clinically notable low values
    - Systolic BP: ≤90 mmHg with decrease from baseline of ≥20 mmHg;
    - Diastolic BP: ≤50 mmHg with decrease from baseline of ≥15 mmHg;
    - Heart rate (collected as pulse rate in the vital signs eCRF): ≤50 bpm with decrease from baseline of ≥15 bpm;
    - Weight: ≥20% decrease from baseline;
    - Body temperature [C]: ≤36 C.

All assessments, including unscheduled assessments will be considered. A participant can be included in multiple categories if different criteria are met at different timepoints.

Number and percentage of participants with at least one post-baseline vital sign abnormality will be summarized.

6.6.6. Electrocardiograms

As per protocol, baseline ECG will be obtained as the mean of the triplicate measurements collected at Cycle 1 Day 1 predose.

Heart rate, QT, PR and QRS data will be collected on the ECG eCRF page. QTcF (Fridericia correction) will be derived as follows:

- Fridericia's formula: QTcF= QT/ $\sqrt[3]{RR}$, where RR = 60/heart rate

Number and percentages of participants with clinically notable ECG values during the on-treatment period will be summarized. All participants reporting a clinical notable ECG value will be presented in data listings sorted by participant identifier and ECG date. The clinically notable ECG values are defined in Table 15.

Table 15 Clinical Notable ECG Criteria

Parameter	Criterion
QTcF	increase from baseline >30 ms
	increase from baseline >60 ms
	new >450 ms
	new >480 ms
	new >500 ms
PR Interval	new > 280 ms
QRS	new > 120 ms
Heart rate	increase from baseline >25% and to a value >100 bpm
	decrease from baseline >25% and to a value <50 bpm

ms: milliseconds

6.7. Summaries for Korea, Japan and China

The following analyses by country (Korea, Japan, China) may be performed to support country specific regulatory submission, if required. Results of these analyses may not be included in the main CSR. The appropriate SLI, Phase 3 and Cohort 3 portions will be noted below.

6.7.1. Efficacy Analyses

ORR by BICR, PFS by BICR and OS will be analyzed by country for Phase 3 and for Cohort 3, and potentially for SLI (if required for submission in Japan). The full analysis set or full analysis set, ORR subset will be used. See Section 6.4 for details.

Additionally, ORR and PFS by derived investigator, DOR, TTR (by BICR and by derived investigator assessment) and PFS2 will also be analyzed by country for Phase 3 and Cohort 3 (except PFS2), and potentially for SLI (if required for submission in Japan), utilizing the full analysis set or full analysis set, ORR subset.

6.7.2. Safety Analyses

The following analyses will be performed. See Sections 6.6.1, 6.6.2, 6.6.4, and 6.6.5 for details. The SLI, the Phase 3 and Cohort 3 utilizing the safety analysis set will be reported.

- A high-level summary of AEs (see Section 6.6.1)
- AEs by SOC and PT (all causalities)
- AEs leading to dose modification by PT
    - Dose interruption summary (all causalities), done for any drug, for encorafenib, for cetuximab, and for other study intervention
    - Dose reduction summary (all causalities), done for any drug, for encorafenib, for cetuximab, and for other study intervention
    - Dose permanently discontinued summary (all causalities), done for any drug, for encorafenib, for cetuximab, and for other study intervention
- Deaths
- Laboratory shift tables
- Vital signs

6.7.3. Other Analyses

The following other analyses will be performed for SLI, Phase 3 and Cohort 3:

- Demographic summary (full analysis set)
- Baseline and disease characteristics (full analysis set)
- Patient disposition (full analysis set)
- Protocol deviations (full analysis set)
- Study treatment exposure (safety analysis set)
- Subsequent anticancer therapies/procedures will be summarized (full analysis set)

- Pharmacokinetic analyses (PK analysis set). See Sections 6.3.2.2 and 6.3.2.3 for details of Phase 3 and Cohort 3, respectively.

The following other analyses will be performed for the SLI when data are appropriate:

- Pharmacokinetic analyses (PK analysis set). See Section 6.3.2.1 for details.

7. INTERIM ANALYSES

7.1. Introduction

Depending on the results of ORR and PFS analysis for the Phase 3, there will be two, one or no interim for efficacy analyses of OS.

7.2. Interim Analyses and Summaries for Efficacy OS

At each analysis timepoint, the critical boundaries for the group sequential test will be derived from the predefined spending functions as described below for OS of the Phase 3 portion of the study. The calculations of boundaries will be performed using EAST software (Cytel).[3]

No futility analysis will be performed for OS. Let u(t1) and u(tF) denote the upper critical boundaries based on the test statistics Z_1 and Z_F for efficacy at the interim and the final analysis, respectively.

In what follows P_0 denotes the probability under the null hypothesis, and $\alpha(t_1)$ denotes the α spent based on the predefined α-spending function at information fraction t1 (t1 is calculated as the ratio of the number of OS events observed at the time of the data cutoff for the interim analysis and the total number of OS events targeted for the final analysis).

The critical value $u(t_1)$ for the interim analysis of OS is determined such as

$$P_0(Z_1 \geq u(t_1)) = \alpha(t_1).$$

The boundary for the final efficacy analysis is derived such that

$$\alpha(t_2) + P_0(Z_1 < u(t_1), Z_F \geq u_F) = 0.001 \text{ or } 0.023.$$

As described below, if the number of OS events in the final analysis deviates from the target number of OS events, the final analysis criteria will be determined as above taking into account the actual alpha spent at the interim analysis and the actual correlation between the 2 test statistics Z_1 and Z_F, so that the overall 1-sided significance level across all analyses and comparisons is preserved at 0.001 or 0.023.

If the hypothesis test of the ORR by BICR Arm B versus Arm C (on 110 participants for each arm) is statistically significant, an interim analysis for superiority will be conducted at that time on the corresponding OS comparison (on all participants of the 2 arms) using a portion of overall nominal alpha=0.001. It is anticipated that this will occur at approximately

40% information in the Arm B versus Arm C OS comparison (ie, approximately 119 OS events in these two arms).

If the hypothesis test of the PFS Arm B versus Arm C is statistically significant, an interim analysis for superiority will be conducted at that time on the corresponding OS comparison using a portion of overall nominal alpha=0.023. If the results of PFS Arm B versus Arm C is not statistically significant, but results for ORR are statistically significant, an interim analysis for superiority will be conducted at that time of PFS analysis on the corresponding OS comparison using a portion of overall nominal alpha=0.001. It is anticipated that this will occur at approximately 80% information in the Arm B versus Arm C OS comparison (ie, approximately 238 OS events in these two arms).

A Lan-DeMets alpha-spending function[1] that approximates O'Brien-Fleming stopping boundaries will be used to control the overall type I error rate at 0.001/0.023 level (1-sided).

Table 16 summarizes the operating characteristics and efficacy boundaries for the OS analysis. The exact efficacy boundaries and the actual alpha spent at the interim analyses will be updated at the time of the analyses based on the actual number of events and the information fraction.

The interim analysis of OS will be performed using the methodology described in Section 6.2.2.1.1.

Table 16 OS – Summary of Planned Efficacy Boundaries

Formal Testing	Analysis	Information Fraction[a]	Cumulative OS Events	Efficacy Boundaries on P-scale (1-sided)
In case only ORR analysis results are statistically significant				
Arm B vs Arm C	Interim Efficacy 1	40%	119	<0.0001
	Interim Efficacy 2	80%	238	0.0002
	Final	100%	297	0.0009
In case ORR and PFS[b] or only PFS analyses results are statistically significant				
Arm B vs Arm C	Interim Efficacy	80%	238	0.0111
	Final	100%	297	0.0198

The efficacy boundaries are determined using Lan-DeMets alpha spending function that approximates O'Brien-Fleming stopping boundaries.

a. The exact efficacy boundaries and the actual alpha spent on each analysis will be updated based on the actual number of events observed.
b. In case ORR is significant, interim efficacy 1 is conducted at p-value <0.0001

Interim analysis results may be used for decisions regarding stopping for early success in OS. Participants may be discontinued from the study as a result of the interim analysis.

8. REFERENCES

1. Lan K, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70(3):659-63.

2. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine 2009; 28: 586-604.

3. East® version 6.5 manual (2018). Cytel Inc. Cambridge, MA USA.

4. Wilson EB. Probable Inference, the Law of Succession and Statistical Inference. J Amstat Assoc. 1927;22(158):209-12.

5. Jennison C, Turnbull BW. Group sequential methods with applications to clinical trials: CRC Press; 1999.

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

7. Schoenfeld D. Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika 1980; 67:145-53.

8. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health 2000.

9. Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. DOI 10.1515/bams-2013-0101 Bio-Algorithms and Med-Systems 2013; 9(4): 183–9.

10. Kalbfleisch JD, Prentice, RL. Statistical Analysis of Failure Time Data, 2nd Edition.Hoboken, Wiley Interscience.

11. Amit O, et al. Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis and recommendation from a PhRMA working group. European Journal of Cancer 47:1772-1778, 2011.

12. Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. EORTC QLQ-C30 Scoring Manual, 3rd edn. Brussels, EORTC, 2001.

9. APPENDIX 1 LIST OF ABBREVIATIONS

Abbreviation	Term
AE	adverse event
AESI	adverse event of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	Anatomic Therapeutic Chemical
AUC	area under the curve
BICR	blinded independent central review
BP	blood pressure
BLQ	below the limit of quantification
BOR	Best overall response
BRAF	B-RAF proto-oncogene, serine/threonine-protein kinase
CRS	case retrieval strategy
cfDNA	circulating free DNA
ctDNA	circulating tumor DNA
CI	confidence interval
C_{max}	maximum observed concentration
CR	complete response
CSR	clinical study report
CV	coefficient of variation
CTCAE	Common Terminology Criteria for Adverse Events
DLT	dose-limiting toxicity
DOR	duration of response
EC	encorafenib plus cetuximab
ECOG	Eastern Cooperative Oncology Group
ECG	electrocardiogram
eCRF	electronic case report form
E-DMC	external data monitoring committee
EDR	early discrepancy rate
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire for Cancer Patients – 30 Item Core Questionnaire
EOT	end of treatment
EQ-5D-5L	EuroQol-5D-5L
EQ VAS	EuroQol-visual analogue scale
FAS	full analysis set
FOLFIRI	fluorouracil/leucovorin/irinotecan

Abbreviation	Term
FOLFOXIRI	fluorouracil/leucovorin/oxaliplatin/irinotecan
HLT	High level terms
HR	hazard ratio
IRT	interactive response technology
IV	intravenous
LDR	late discrepancy rate
LLOQ	Lower limit of quantification
mCRC	metastatic colorectal cancer
MedDRA	Medical Dictionary for Regulatory Activities
mFOLFOX6	modified fluorouracil/leucovorin/oxaliplatin
MWPC	meaningful within person change
NA	not applicable
NCI	National Cancer Institute
ND	not done or no disease at baseline
NE	non-evaluable
NR	not reached
ORR	objective response rate
OS	overall survival
OTR	outside toxicity reference
PD	progressive disease
PD2	second objective disease progression
PFS	progression-free survival
PFS2	PFS after next line of treatment
PIPDs	potentially important protocol deviations
PK	pharmacokinetic(s)
PR	partial response
PRO	patient-reported outcome
PT	preferred term
Q1	first quartile
Q3	third quartile
QT	QT interval
QTcF	corrected QT (Fridericia method)
RECIST	Response Evaluation Criteria in Solid Tumors
RMST	restricted mean survival time
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SLI	Safety Lead-in

Abbreviation	Term
SMQ	standardized MedDRA queries
SOC	standard of care or system organ class
SOP	standard operating procedure
TEAE	treatment-emergent adverse event
TTR	Time to response
ULN	upper limit of normal
vs	versus
WBC	white blood cell
WHO	World Health Organization

10. APPENDIX 2 SPECIFICATIONS FOR PROGRAMMATIC DERIVATION OF TUMOR RESPONSE USING RECIST 1.1

For Phase 3 portion and Cohort 3, the primary efficacy analysis will be based on blinded independent central review (BICR) interpretation of tumor assessment scans from study sites, performed by Calyx. The details of this process are described in the study specific Calyx charter. Analyses based on BICR will use data as transferred from BICR without any further derivation (i.e. using the response at each timepoint provided by Calyx) except to account for censoring.

For SLI portion, the efficacy analysis will be based on investigator tumor assessment reported in the eCRF. Analyses based on investigator will use data as reported without any further derivation (i.e. using the response at each timepoint provided by the investigator) except to account for censoring.

For Phase 3 portion and Cohort 3, a secondary efficacy analysis will be based on a programmatic approach to derive tumor response/progression, using RECIST 1.1 and the investigator tumor assessment data recorded on the target, non-target, and new lesion electronic case report forms (eCRF), as described in this appendix.

The tumor response criteria are adapted from RECIST 1.1 (Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline [Version 1.1]. Eur J Cancer. 2009;45(2):228-47).

10.1. Study Specific Information

10.1.1. Adequate Baseline

The following must be met to qualify for "Adequate Baseline" assessment:

- Baseline assessments of all lesions (target, and non-target) must be prior to the first dose date (SLI) or date of randomization (Phase 3 and Cohort 3) per Schedule of Activities in Section 1.3 of the protocol. For Phase 3 and Cohort 3 lesions assessed after randomization, but before or on the first dose date will also be considered adequate.

- All lesions recorded at baseline must have an associated status recorded on the eCRF. For target lesions, an actual measurement should be recorded and it should meet the criterion for being measurable; for non-target lesions the actual status at baseline should be checked on the eCRF. For example, if 3 target lesion sites of disease are recorded all 3 must have associated measurements recorded; otherwise, the sum of longest dimensions (SOD) cannot be calculated.

- Baseline lesions must be assessed with an acceptable method of tumor assessment as specified in the protocol and could include: Conventional CT Scan, Spiral CT Scan, X-ray, MRI, PET/CT, physical examination and Bone Scan.

10.1.2. Interchangeable Methods of Tumor Assessment

Following tumor assessment methods are considered "interchangeable": Conventional CT, Spiral CT, Enhanced CT, MRI, PET/CT.

10.1.3. Period for Derivation of Best Overall Response (BOR)

Is defined as the time from the start date until progressive disease (PD) or the day before initiation of new anticancer therapy in the absence of PD or death, whichever occurs first.

Assessments done after PD or starting with the day of new anticancer therapy will not be considered for evaluation of BOR.

10.1.4. New Anticancer Therapy

Includes any systemic anticancer therapy (other than study medication), radiation therapy to target or non-target lesions with curative intent, or surgery (including metastasectomy) for target or non-target lesions with outcome of "resected" or "partially resected" provided on or after the start date, see also Section 5.2.9.

10.2. Lesion Evaluation at Assessment Times

Often, at pre-planned and unplanned tumor assessment times (visits or weeks), there are multiple dates associated with evaluation of the same target, non-target and new lesions. In addition, tumor evaluations are not always recorded on the eCRF under a specific visit (week) that would allow identifying the lesions/investigator overall tumor assessment (IOTA) that were evaluated at that particular visit (week). In order to address this and group lesion (target, non-target and new) evaluations and IOTA assessment to the corresponding actual post-baseline tumor assessment visit (week), a clustering algorithm is applied. Each cluster represents an actual unique assessment date. For each patient, the number of clusters is equal to the maximum number of unique assessment dates available among all lesions (target, non-target, and new) and IOTA assessments. The SAS procedure, Proc Fastclus, is applied to a variable that represents the days from the start date to the unique date of the evaluation/assessment for each lesion/IOTA (date of scan – start date +1). Then the evaluations/assessments of lesions/IOTA that occurred "close to each other in time" will be assigned to the same cluster.

10.2.1. Response Evaluation of Target Lesions

Complete Response (CR) is defined by the disappearance of all non-lymph node target lesions (where all target lesions are recorded with a length of 0 mm on the "Target Lesions" eCRF). Any pathological lymph nodes (recorded as target lesion) must have reduction in short axis to <10 mm.

- Note: the SOD may not be zero if lymph nodes are included as target lesions.

Partial Response (PR) is defined by a 30% or more decrease in SOD of target lesions, taking as reference the baseline SOD.

PD is defined by a 20% or more increase in the SOD of target lesions relative to nadir (smallest SOD considering baseline and all assessments prior to the time point under evaluation), with a minimum absolute increase of 5 mm relative to nadir.

- Note: if only a subset of target lesions are assessed and the sum of the non-missing lesion diameters results in an increase above of at least 5 mm and at least 20% above the nadir, then the SOD will still be calculated and the progression criteria will have been otherwise met. If the sum of non-missing lesions does not indicate PD, then the SOD will be left as missing.

Stable Disease (SD) is assigned when neither sufficient shrinkage to qualify for CR or PR, nor sufficient increase to qualify for PD is observed, taking as reference the nadir.

No Target Lesion at Baseline (NB) is assigned if "No Target Lesion" is checked, on the "Target Lesions" eCRF at baseline.

Not All Evaluated (NAE) is assigned if, in the absence of PD based on evaluated target lesions:

- Any individual target lesion is evaluated as "Indeterminate" on the "Target Lesions" eCRF.
- Inconsistent methods (unless considered "interchangeable") are used for any target lesions after start date.
- One or more target lesions are not assessed.
- One or more target lesions were excised or irradiated and have not reappeared or increased.

Determination of target lesion response in case of reappearance of one or more target lesion(s) that have previously disappeared:

- If the previous target lesion response was CR, and a non-lymph node target lesion reappears, then the response is always PD.
- If the previous target lesion response was CR, and a lymph node target lesion reappears, then the response (whether CR or PD) is assessed based on PD criteria noted above. The response will be PD only if SOD criterion for PD is met and if the lymph node returns to pathologic size (≥10 mm) and meets the absolute requirement of 5 mm increase over nadir for the reappearing lesion. Otherwise, the response is CR.
- If the previous target lesion response was PR, then the response should be evaluated based on the SOD.

Notes:

- The SOD is only considered if the methods of assessment are consistent with baseline. The "interchangeable" methods noted in Section 10.1.2 above are all considered consistent methods.

- In the SOD, the longest diameter will be used for non-nodal lesions and the short axis dimension will be used for each lymph node included in the sum.

10.2.2. Response Evaluation of Non-Target Lesions

CR is defined by the complete disappearance of all non-target lesions (where all non-target lesions are marked "Absent" on the "Non-Target Lesion" eCRF). All lymph nodes must be non-pathological in size (<10 mm in short axis).

Non-CR/Non-PD is defined by persistence of one or more non-target lesions (ie, if any non-target lesions are marked "Present/Not Increased" on the "Non-Target Lesion" eCRF).

No NB is assigned if "No Non-Target Lesions" is marked on the "Non-Target Lesions" eCRF at baseline.

NAE is assigned if, in the absence of PD based on evaluated non-target lesions:

- Any individual non-target lesion is evaluated as "Indeterminate" (marked as "Indeterminate" on the "Non-Target Lesions" eCRF).

- Inconsistent methods (unless considered "interchangeable" as noted in Section 10.1.2) are used for any non-target lesions after start date.

- One or more non-target lesions are not assessed.

PD is assigned if any non-target lesion is marked "Increased" on the "Non-Target Lesion" eCRF. However, in an effort to programmatically define "unequivocal progression" of non-target lesions, the derived non-target lesion response will also take into account the "Non-Target Lesions" assessment from the IOTA eCRF, as noted in Table 17.

- Note: the lesions assessed are only considered for CR, Non-CR/Non-PD and PD if the methods of assessments are consistent with baseline. The "interchangeable" methods noted in Section 10.1.2 above are all considered consistent methods.

Table 17. Derivation of Non-Target Lesion Assessment

Non-Target lesion assessment status	"Non-Target lesions" assessment status on IOTA eCRF	Derived non-target lesion response
PD ("Increased" is marked for ≥1 lesion)	PD	PD
	Non-CR/Non-PD	Non-CR/Non-PD
	CR	Non-CR/Non-PD
	Not Assessed (NA)	PD
	No Baseline (NB)	PD
	Indeterminate (IND)	NAE
	Missing	PD

Instruction: Relevant discrepancies between the non-target lesion assessment status and information on the IOTA page must be queried.
Abbreviations: IOTA = investigator overall tumor assessment; eCRF = electronic case report forms; CR = Complete Response; PD = Progressive Disease; NA = Not Assessed; NB = No Baseline; IND = Indeterminate; NAE= Not All Evaluated.

10.2.3. New Lesion Evaluation Criteria

A new lesion is defined by the appearance of 1 or more new lesions (where any lesion is marked "New" on the "New Lesions" eCRF).

Any lesion that is recorded for the first time after the start date without being marked as "New" on the "New Lesions" eCRF must be queried. In case the inconsistency is not resolved at time of database snapshot/lock, the lesion will be considered as new and the patient in progression at that time point.

The requirement for consistent methods of assessment with baseline does not apply for new lesions.

If a new lesion is equivocal at one assessment (classified as "Equivocal") and verified at the next assessment, (classified as "Unequivocal"), the earlier date is used as the progression date.

10.3. Objective Response Status at Each Assessment

Objective response status is determined from the derived target and derived non-target lesion data using the conventions in Table 18 under the assumption that there are no new lesions identified at the visit.

Objective status after a change in modality that is not considered interchangeable is classified as Not Evaluable (NE).

If there are any new lesions at a time point, then the response is PD at that time point regardless of target or non-target lesion response.

Table 18. Derivation of Objective Status Based on Target and Non-Target Lesion Response Assuming No New Lesions*

Target lesion response	Non-Target lesion response	Objective response status
CR	CR	CR
CR	Non-CR/Non-PD	PR
CR	PD	PD
CR	NAE	PR
PR	CR	PR
PR	Non-CR/Non-PD	PR
PR	PD	PD
PR	NAE	PR
SD	CR	SD
SD	Non-CR/Non-PD	SD
SD	PD	PD
SD	NAE	SD
PD	Any	PD
NAE	PD	PD
NAE	CR Non-CR/Non-PD NAE	NE
NB	CR	CR
	Non-CR/Non-PD	Non-CR/Non-PD
	PD	PD
	NAE	NE
CR	NB	CR
PR		PR
SD		SD
PD		PD
NAE		NE

*If there are any new lesions at a time point, then the response is PD at that time point regardless of target or non-target lesion response.
Abbreviations: CR = Complete Response; PR = Partial Response; SD = Stable Disease; PD = Progressive Disease; NAE = Not All Evaluated; NB = No Baseline; NE = Not Evaluable.
Note: If non-target (or target) lesions are not collected at baseline, then the overall response is equivalent to the target (or non-target) lesions response, respectively.

10.4. Date of Response at Each Assessment/Overall:

The date of CR, PR, PD, non-CR/non-PD, SD, NE is derived as the date of the first radiographic evaluation included in the cluster.

The date of first response for purposes of calculating duration of response and time to response is defined as the first date a CR or PR was documented.

10.4.1. Best Overall Response Evaluation for Each Patient

BOR is derived from the sequence of objective responses reported during the "Period for Derivation of Best Overall Response" according to the rules specified in Section 4.2.1.5 of the Oncology Rulebook.

Table 19 presents derivation of BOR for specific cases assuming that confirmation of response is not required.

Table 19. Derivation of Best Overall Response Based on Objective Status at Assessment Time Points Assuming That Confirmation of Response is Not Required

Week 6	Week 12	Week 18	Best overall response
Early Death			NE
CR	PD		uCR
CR	NE		uCR
CR	PR		uCR
PR	PD		uPR
PR	CR	PD	uCR
PR	SD		uPR
PR	NE		uPR
SD	PD		SD[a]
SD	NE		SD[b]
Non-CR/non-PD	PD		Non-CR/non-PD[c]
Non-CR/non-PD	PD		PD[d]
NE	PD		PD
NE	SD	PD	SD
NE	NE	PD	NE

Abbreviations: CR = Confirmed Response; PR = Partial Response; SD = Stable Disease; PD = Progressive Disease; NE = Not Evaluable; uCR = unconfirmed CR; uPR = unconfirmed PR.

a. If SD was documented less than 42 days after start date, the best overall response will be PD.

b. If SD was documented less than 42 days after start date, the best overall response will be NE.

c. Applicable when only non-target lesions are present at baseline. If non-CR/non-PD documented more than 41 days after start date.

d. Applicable when only non-target lesions are present at baseline. If non-CR/non-PD documented less than 42 days after start date.

Table 20 presents derivation of BOR for specific cases assuming that confirmation of response is required.

Table 20. Best Overall Response When Confirmation of Response is Required

Overall response first time point	Overall response subsequent time point	Best overall response
CR	CR	CR
CR	PR	SD, PD, or PR[a]
CR	SD	SD provided minimum criteria for SD duration met, otherwise PD
CR	PD	SD provided minimum criteria for SD duration met, otherwise PD
CR	NE	SD provided minimum criteria for SD duration met, otherwise NE
PR	CR	PR
PR	PR	PR
PR	SD	SD
PR	PD	SD provided minimum criteria for SD duration met, otherwise PD
PR	NE	SD provided minimum criteria for SD duration met, otherwise NE
NE	NE	NE

Note that best overall response of CR or PR requires confirmation ≥4 weeks after response is first observed. Minimum duration criteria for SD must be met for a best overall response of SD.
Abbreviations: CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease; NE = not evaluable.
a. 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 21 presents derivation of BOR for specific examples of overall response at sequential time points.

Table 21. Best Overall Response When Confirmation of Response is Required: Further Examples

Overall response sequential time points	Best overall response
SD - CR - CR - PR	CR if the two CRs were ≥4 weeks apart; else, SD if at any time on or before the 3rd time point evaluation the minimum criteria for SD duration was met; else PD
SD - PR - CR - PR	PR if the 2nd and 3rd evaluations (PR, CR) were ≥4 weeks apart; else SD if at any time on or before the 3rd time point evaluation the minimum criteria for SD duration was met; else PD
SD - CR - SD - CR	SD if at any time on or before the 2nd time point evaluation the minimum criteria for SD duration was met; else PD
PR - CR - NA - CR - PR	CR if the 2nd and 4th evaluations (CR, CR) were ≥4 weeks apart; else PR if the 1st and 2nd evaluations (PR, CR) or the 1st and 4th evaluations (PR, CR) were ≥4 weeks apart; else SD if at any time on or before the 4th time point evaluation the minimum criteria for SD duration was met; else PD
PR - CR - CR - PR	CR if the 2nd and 3rd evaluations (CR, CR) were ≥4 weeks apart; else PR if the 1st and 2nd evaluations (PR, CR) or the 1st and 3rd evaluations (PR, CR) were ≥4 weeks apart; else SD if at any time on or before the 3rd time point evaluation the minimum criteria for SD duration was met; else PD
PR - CR - CR - SD	CR if the 2nd and 3rd evaluations (CR, CR) were ≥4 weeks apart; else PR if the 1st and 2nd evaluations (PR, CR) or the 1st and 3rd evaluations (PR, CR) were ≥4 weeks apart; else SD if at any time on or before the 3rd time point evaluation the minimum criteria for SD duration was met; else PD
CR - NE - CR - SD	CR if the 1st and 3rd evaluations (CR, CR) were ≥4 weeks apart; else SD if at any time on or before the 3rd time point evaluation the minimum criteria for SD duration was met; else PD
PR - PR - CR - PR	PR if 2 of the first 3 evaluations (1st to 2nd, or 1st to 3rd, or, 2nd to 3rd) were ≥4 weeks apart; else SD if at any time on or before the 3rd time point evaluation the minimum criteria for SD duration was met; else PD
PR - PR - CR - SD	PR if 2 of the first 3 evaluations were ≥4 weeks apart (1st to 2nd, or 1st to 3rd, or, 2nd to 3rd); else SD if at any time on or before the 3rd time point evaluation the minimum criteria for SD duration was met; else PD
CR - SD - SD - CR	SD if at the 1st time point evaluation the minimum criteria for SD duration was met; else PD
PR - CR - PR - CR	PR if the 1st and 2nd time point evaluations (PR, CR) were ≥4 weeks apart; else SD if at any time on or before the 2nd time point evaluation the minimum criteria for SD duration was met; else PD

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.
Abbreviations: CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease; NE = not evaluable; NA = not assessed.