Signature Page for ISO_CC_007_SAP
Study ISO-CC-007 v8.0

eSignature for Approval	Karin Gedda Sponsor 21-Apr-2022 15:52:06 GMT+0000

Signature Page for ISO_CC_007_SAP
Study ISO-CC-007 v8.0

Confidential

Statistical Analysis Plan

Sponsor:	ISOFOL
Protocol Title:	A randomized, multicenter, parallel-group, Phase III study to compare the efficacy of arfolitixorin versus leucovorin in combination with 5-fluorouracil, oxaliplatin, and bevacizumab in patients with advanced colorectal cancer
Study Code:	ISO-CC-007
Protocol Version	V4.0 20Aug2021
	Of note, this version of the SAP was written after the interim analysis.

For IDDI:

Author: Nicolas Dubois, Biostatistician

Reviewers:

- Perrine Bamps, Senior Biostatistician

- Koenraad D'Hollander, Senior Biostatistician

For Sponsor:

Reviewer: Karin Gedda, VP Clinical Operations

Table of Contents

1. List of Abbreviations and Definition of Terms

Abbreviation	Term
5-FU	5-fluorouracil
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALT	Alanine Transaminase
AST	Aspartate Transaminase
ATC	Anatomical Therapeutic Chemical
AUC	Area Under the Curve (=AUClast)
BICR	Blinded Independent Central Review
BOR	Best Overall Response
BUN	Blood Urea Nitrogen
C_{max}	Maximum (or peak) plasma concentration
CMH	Cochran-Mantel-Haenszel test
CP	Conditional Power
CR	Complete Response
CRC	Colorectal Cancer
CRF	Case Report Form
CSP	Clinical Study Protocol
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
DOR	Duration of Response
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Forms
EOT	End of Treatment Visit
gGT	Gamma-Glutamyl Transferase
HR	Hazard Ratio
ICH	International Council for Harmonisation
IMP	Investigational Medicinal Product
ITT	Intention-To-Treat analysis set

Abbreviation	Term
ITTE	Intention-to-Treat Excluding Additional Japanese Patients analysis set
ITTJ	Intention-To-Treat Japanese analysis set
IWRS	Interactive Web Response System
LDH	Lactodehydrogenase
LLOQ	Lower Limit of Quantification
LV	Leucovorin
MedDRA	Medical Dictionary for Regulatory Activities
MRI	Magnetic Resonance Imaging
Methyl THF	Methyl Tetrahydrofolate, metabolite
MTHF	Methylene Tetrahydrofolate, parent drug
NCI	National Cancer Institute
NE	Not Evaluable
ORR	Overall Response Rate
OS	Overall Survival
PD	Progressive Disease
PFS	Progression Free Survival
PK	Pharmacokinetic
PP	Per-Protocol analysis set
PPJ	Per-Protocol Japanese analysis set
PR	Partial Response
PRO	Patient Reported Outcome
RECIST	Response Evaluation Criteria In Solid Tumours
RFS	Recurrence Free Survival
RR	Response Rate
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Stable Disease
SOC	System Organ Class
SOP	Standard Operating Procedure
THF	Tetrahydrofolate, metabolite
Tlast	Time at which last plasma concentration is observed

Abbreviation	Term
Tmax	Time at which maximum plasma concentration is observed
VAS	Visual Analogue Scale
WBC	White Blood Cell
WHO	World Health Organization

2. Introduction

This Statistical Analysis Plan was written for the clinical trial IS0-CC-007 conducted in the EU, Canada, USA, Australia and Japan. The ICH guideline E3 "Structure and Content of Clinical Study Reports" was used as a guide to the writing of the plan. This Statistical Analysis Plan is based on the current version of the clinical study protocol. This current version of the SAP was written after the interim analysis.

3. Study Design and Objectives

3.1 Study Objectives

3.1.1 Primary Objective

The primary objective is to compare the effect of treatment with arfolitixorin versus Leucovorin (LV), in combination with 5-FU, oxaliplatin, and bevacizumab in patients with advanced Colorectal cancer (CRC), on the **Overall Response Rate (ORR)**. This is defined as the best confirmed overall response recorded from the start of the study treatment until the end of treatment or the last available assessment at database cutoff. All responses will be confirmed 8 weeks after onset of response. All assessments including confirmation of response will be performed by BICR.

3.1.2 Secondary Objectives

The key secondary objectives are

- To estimate the **Progression-free survival (PFS)**, defined as the time from randomization to first occurrence of tumor progression assessed by BICR (RECIST 1.1) based on CT-scans/MRIs conducted every 8 weeks after start of treatment, or death from any cause. Patients undergoing metastasis resection will not be censored for PFS. Recurrence of disease after surgery will be counted as a progression event.
- To estimate the **Duration of Response (DOR).** The duration of overall response (DOR) is measured from the first time point at which criteria are met for complete response (CR) or partial response (PR) through the last time point when overall response has been objectively documented.

The other secondary objectives are

- To estimate the **Overall survival (OS)**, defined as time from randomization to death from any cause.
- To assess the **Quality of Life**, assessed using the EQ-5D patient reported outcome questionnaire (PRO).
- To assess the **Safety and tolerability** by counting number and severity of adverse events (AEs), including clinically significant abnormal laboratory findings, regardless of causal relationship to arfolitixorin or LV. Specific AEs will be followed using PRO (NCI PRO-CTCAE).
- To assess **Patients undergoing curative metastasis resection** after treatment with study drug, defined as the number of patients qualifying for curative metastasis resection after treatment with study drug.

3.1.3 Exploratory Objectives

The exploratory objectives*[1] are

- To assess the **Daily living abilities**, as assessed by the Eastern Cooperative Oncology Group (ECOG) Performance Status Scale.
- To determine the **pharmacokinetic (PK)** characteristics of arfolitixorin and LV in patients with advanced CRC. The plasma concentration of methylene tetrahydrofolate ([6R]-MTHF), methyl-tetrahydrofolate (methyl-THF), and tetrahydrofolate (THF) will be determined in a limited number of patients in both treatment arms.
- **Recurrence Free Survival (RFS)** for patients undergoing metastatic resection. Defined as the time between the first surgery with complete removal of the metastasis and recurrence of the disease or death from any cause.
- To investigate folate metabolism- and transportation-related gene expression levels in patients with advanced CRC. **Gene expression levels** of folate metabolism- and transportation-related genes, analyzed by sub-group low/high gene expression levels.

3.2 Study Design

This is a randomized, multicenter, parallel-group, Phase III study to compare the efficacy of arfolitixorin versus LV in patients with advanced CRC treated by 5-FU, oxaliplatin, and bevacizumab. To be eligible for the study, patients must have advanced CRC, without indication for resection at study inclusion, with at least one measurable lesion of metastatic disease according to RECIST 1.1 criteria within 28 days of randomization.

3.2.1 Study intervention description

490 eligible and consenting patients were randomized 1:1 to one of two treatment groups (A and B) and administered one of the following treatment regimens starting within 3 days of randomization. Following treatment cycles should be started 14 (+ 7) days after the previous administration.

- Group A "**Experimental**": ARFOX (Arfolitixorin + 5-FU + Oxaliplatin) + Bevacizumab
- Group B "**Comparator**": mFOLFOX-6 (Leucovorin + 5-FU + Oxaliplatin) + Bevacizumab

Initially, approximately 440 were planned to be randomized. However, Japanese authorities requested that approximately 12.7% of the total study population are enrolled from Japan. For this reason, additional patients from Japan were randomized. In the end a total of 490 patients were randomized (cf. section 3.3.6). This decision to randomize extra Japanese patients was taken before the moment of the interim analysis, and without looking at the data. To prevent a potential bias, the statistician of the final analysis used a dummy randomization file at the time of the interim analysis. Only an independent statistician and the DSMB members had access to the unblinded interim analysis report.

[1] * This SAP does not apply to gene expression level exploratory objectives.

3.2.2 Measures to minimize bias: randomization and blinding

a) Randomization

Patients are randomized in a 1:1 ratio to either the experimental arm (arfolitixorin) or the comparator arm (Leucovorin), following the completion of all screening assessments and after confirmation of their eligibility, using a stratified permuted block randomization. Randomization is stratified for the following baseline factors:

- Geographical region (Europe / North America / Australia / Japan),
- Primary tumor location (left colon/right colon/rectal cancer),
- Previous neo-adjuvant/adjuvant CRC treatment (yes/no)

Randomization is performed via an interactive web response system (IWRS). The system enables time and date stamped entries that in turn allow real time compliance monitoring.

b) Blinding

Patient and Investigator blinding of the study treatment is not possible as there is a difference in the color of the injection fluids and the method of administration. To minimize any bias arising either from the inability to blind patients and Investigators or from potential inconsistencies in local tumor response assessments across study sites, analyses of the tumor response endpoints is based on blinded, retrospective, centrally-adjudicated assessment of CT-scans/MRIs according to a pre-specified Imaging Review Charter (Banook 2021).

3.2.3 Visits

The study consists of a screening visit, a randomization visit during which patients receive their first course of IMP, regular treatment visits every 2 weeks thereafter, and CT/MRI assessments every 8 weeks from baseline. Treatment visits are to continue until PD, as identified by the Investigator based on locally-performed assessment of CT-scans/MRIs. In the event of PD, Investigators are to discontinue IMP and manage the patient according to local routine. Patients are followed thereafter until death.

a) Treatment period

Treatment visits occur every 14 days (+7 days) until PD.

b) Assessment visits (every eight weeks until PD)

Assessment visits are to occur every 8 weeks (+/-7 days) after the baseline visit (1st IMP administration day). Assessment visits should continue until PD or premature discontinuation of IMP.

c) End of treatment visit (EOT)

Patients prematurely discontinuing IMP and those reaching PD should complete an end of treatment visit within 30 days after last dose of IMP. The EOT visit should be conducted before initiation of new anti-cancer therapy.

d) Post treatment visits: Follow-up visits until centrally confirmed progressive disease

Patients who discontinue IMP for reasons other than centrally confirmed PD move into the Follow-up phase and are assessed every 8 weeks.

Every effort should be made to collect information regarding disease status until disease progression, death, or the end of the study.

e) Post treatment visits: Survival follow-up

Patients discontinuing study treatment and/or with centrally confirmed PD move into the survival follow-up phase. Patient survival status is assessed every 12 weeks until death or the end of the study, whichever occurs first. Patient survival data can be collected by phone, visit, medical records, contact with relatives etc. Date of death, or the Investigators best estimate of the date of death, should be entered in the eCRF.

Information regarding post-study anti-cancer therapies is collected if new treatment is initiated.

The schedule of activities is shown in the table below.

Table 1: Schedule of activities

		PRE-TREATMENT	TREATMENT PERIOD					END OF TREATMENT	FOLLOW-UP	
		Screening	Administer IMP		Assessment visit	Administer IMP	Assessment visit	End of treatment visit	Follow-up visits until centrally confirmed PD	Survival follow-up
			Cycle 1	Cycles 2-4		Cycles 5 - PD or discontinuation			*Only patients prematurely discontinuing IMP prior to centrally confirmed PD*	*All patients*
	Timing		Day 1 Baseline	14 days after previous admin visit	Week 8 from day 1	14 days after previous admin visit	Weeks 16, 24, etc. from day 1	Within 30 days after last dose of IMP and before start of new anti-cancer treatment	Every 8 weeks from end of treatment visit until centrally confirmed PD	Every 12 weeks after centrally confirmed PD until death or end of study
CSP Section	**Visit window**	Day -28 to Day -1	-3 days[a,f]	+7 days	±7 days	+7 days	±7 days		±7 days	±7 days
Screening procedures										
10.1.1	Informed consent	X								
8.3	Patient Study ID	X								
5.1/5.2	Eligibility criteria	X	X							
8.3	Demographics/medical history	X								
6.4	Randomization		X[a]							
General procedures										
6.6	Concomitant therapy	X	X	X	X	X	X	X		
8.2	Physical examination	X	X		X		X	X		
8.2	Vital signs	X	X	X	X	X	X	X		
8.2	Height	X								

Confidential

		PRE-TREATMENT	TREATMENT PERIOD					END OF TREATMENT	FOLLOW-UP	
		Screening	Administer IMP		Assessment visit	Administer IMP	Assessment visit	End of treatment visit	Follow-up visits until centrally confirmed PD	Survival follow-up
			Cycle 1	Cycles 2-4		Cycles 5 - PD or discontinuation			*Only patients prematurely discontinuing IMP prior to centrally confirmed PD*	*All patients*
	Timing		Day 1 Baseline	14 days after previous admin visit	Week 8 from day 1	14 days after previous admin visit	Weeks 16, 24, etc. from day 1	Within 30 days after last dose of IMP and before start of new anti-cancer treatment	Every 8 weeks from end of treatment visit until centrally confirmed PD	Every 12 weeks after centrally confirmed PD until death or end of study
CSP Section	**Visit window**	Day -28 to Day -1	-3 days[a,f]	+7 days	±7 days	+7 days	±7 days		±7 days	±7 days
8.2	Weight[e]	X	X	X	X	X	X	X		
8.2	Biopsy available for Pharmacogenetics	X								
6.1/6.2	Administer and record study treatment		X	X		X				
6.6	Record other anti-cancer treatment							X	X	X
8.1	ECOG	X	X		X		X	X	X	
Efficacy assessments										
8.1	Plasma PK samples		X[b]	X[b]						
8.1	CT-scan/MRI[c]	X			X		X	X	X	
8.1	PRO EQ-5D		X		X		X	X		
8.1	Survival Status									X[g]
Safety assessments										
8.4	AEs		X	X	X	X	X	X[d]	X[d]	
8.2	PRO CTCAE		X	X	X		X	X		
8.2	Hematology and clinical chemistry[e]	X[f]		X		X		X		

Confidential

		PRE-TREATMENT	TREATMENT PERIOD					END OF TREATMENT	FOLLOW-UP	
		Screening	Administer IMP		Assessment visit	Administer IMP	Assessment visit	End of treatment visit	Follow-up visits until centrally confirmed PD	Survival follow-up
			Cycle 1	Cycles 2-4		Cycles 5 - PD or discontinuation			*Only patients prematurely discontinuing IMP prior to centrally confirmed PD*	*All patients*
	Timing		Day 1 Baseline	14 days after previous admin visit	Week 8 from day 1	14 days after previous admin visit	Weeks 16, 24, etc. from day 1	Within 30 days after last dose of IMP and before start of new anti-cancer treatment	Every 8 weeks from end of treatment visit until centrally confirmed PD	Every 12 weeks after centrally confirmed PD until death or end of study
CSP Section	**Visit window**	Day -28 to Day -1	-3 days[a,f]	+7 days	±7 days	+7 days	±7 days		±7 days	±7 days
8.2	Urinalysis[e]	X[f]		X		X		X		
8.2	ECG[h]		X	X						
8.2	Pregnancy test	X[f]						X		

[a] Randomisation can occur on the same day as study treatment initiation, or up to 3 days before study treatment initiation. Study treatment must be initiated within 28 days from start of screening.

[b] Plasma PK samples will be collected during Cycle 1 and Cycle 2.

[c] CT-scans/MRIs must be performed on thorax, abdomen and pelvis. Post-randomisation CT-scans/MRIs should be performed every 8 weeks (±7 days) from Day 1 to PD confirmed by BICR. End of Treatment visit scans will be taken if the previous scan is >14 days.

[d] AEs are to be recorded until the day of the End of Treatment visit. After the End of Treatment visit, only SAEs related to study specific procedures or considered to be at least possibly related to study drug should be recorded.

[e] Laboratory and urine samples and weight are to be collected according to local practice (7 days before study treatment). The laboratory and urinalysis reports must be obtained and reviewed prior to administration of chemotherapy for each cycle.

[f] Haematology, clinical chemistry, urinalysis, and pregnancy testing must be performed within 7 days prior to study treatment initiation and must be assessed before randomization

[g] Survival follow-up visit could be conducted either by phone. ordinary visits, hospital records or other means found suitable.

[h] 12-lead ECG should be performed during administration visits 1-4. Directly prior to and 10 min (±5 min) after each arfolitixorin dose, and directly prior to and 10 min (±5 min) after initiation of Leucovorin infusion, during administration visits 1-4.

3.3 Sample Size Justification

3.3.1 Hierarchical testing approach

A fixed sequence hierarchical test procedure will be applied, with a pre-defined order in which the hypotheses are to be tested. Each test will be performed without a multiplicity adjustment for multiple endpoints, i.e. at the nominal α-level foreseen by the group sequential design. The secondary endpoint PFS can only be tested formally after the first hypothesis, i.e. the null hypothesis for ORR, is rejected at the final analysis.

3.3.2 Accrual rate

The accrual rate was assumed to ramp up as follows: 2, 3, 7, 11, 14, 22 and 25 patients respectively in the 1st to 7th months after study start, and 28.8 patients monthly thereafter.

3.3.3 Primary endpoint - Overall Response Rate (ORR)

The primary endpoint is the Overall Response Rate (ORR). Null and alternative hypotheses are expressed in terms of ΔORR, the difference in ORR between the randomized treatment arms:

- H0: ΔORR=0
- H1: ΔORR≥15%

For this adaptive design, the initial sample size calculation is based on an improvement by 15% in ORR; specifically, ORR is assumed to be 45% in the control arm vs. 60% in the experimental arm.

It was assumed that up to 10% of patients would be non-evaluable for response in both treatment arms. These patients are to be considered non-responders and do not contribute to the difference between experimental and control (hence the expected difference is 13.5% rather than 15%).

For a two-sided test with α=0.05 and a power of 80% to detect this difference, at least 440 patients were needed to be randomized. Of note, the addition of extra Japanese patients leads to a statistical power greater than 80%. An interim calculation of the conditional power for ORR as well as PFS took place when 16-week BICR evaluation had been performed for the 330th patient. At this analysis, the DSMB did not recommend a 50% sample size increase (from 440[2] to 660 patients). Statistical considerations regarding the interim analysis are explained in Section 4.2.1.

3.3.4 Key secondary endpoint - Progression-free survival (PFS)

The study was initially powered to detect a hazard ratio of 0.725, which corresponds to a clinically meaningful difference in PFS (median PFS equal to 10 months in the control arm vs. 13.8 months in the experimental arm).

For a power of 80%, 300 PFS events had to be observed using a 1-sided alpha=0.025. This number allowed performing one interim analysis. The statistical design allowed for a 50% sample size increase (to 660 patients), should interim results be promising enough to warrant such a sample size increase[2]. In case the sample size had had to be increased, the target number of events would have been increased to 450 PFS events, which provides

[2] The interim analysis was performed before the decision to add the 50 extra Japanese patients to the efficacy and safety analyses.

80% power to detect a hazard ratio of 0.77 (median PFS equal to 10 months in the control arm vs. 13 months in the experimental arm).

Due to a number of patients censored because of new anti-cancer therapies higher than expected, it became clear during the study that the 300 PFS events could not be reached.

The following table shows, for 300 and 235 PFS events, the hazard ratio for which the trial would have 80% power, the hazard ratio for which the trial would just reach statistical significance, and the corresponding median PFS in the treatment arm, assuming a 10-month median PFS in the control arm.

Table 2: Hazard ratio for significance in function of the number of PFS events

		Design		Final analysis	
Number of PFS events	**Median PFS Control arm**	**Hazard ratio**	**Median PFS Treatment arm**	**Hazard ratio for significance**	**Median PFS Treatment arm**
300[3]	*10 months*	*0.725*	*13.8 months*	*0.80*	*12.5 months*
235	*10 months*	0.692	14.5 *months*	0.77	12.9 *months*

A hazard ratio of 0.77 was considered a clinically relevant maximal threshold for statistical significance. The 235 number of PFS events is expected to have occurred on 10 April 2022. The final PFS analysis will therefore be performed when at least 235 PFS events will be reached. Adjustment of the significance level in the interim analysis and final analysis has been performed using a Rho family ($\rho = 5$) boundary.

[3] Initial version of the protocol

3.3.5 Other secondary endpoints

An analysis of OS at the same time as the final PFS analysis, with 440 patients, had a 30% power to demonstrate a hazard ratio of 0.8 (corresponding to a median OS equal to 25 months in the control arm vs. 31.25 months in the experimental arm), when approximately 180 events are observed, using a 1-sided α=0.025. If the sample size had been increased to 660 patients, analyzed at the same time as the final PFS analysis , this power would have increased to 45%, when approximately 270 events are observed. The addition of the extra Japanese patients to the 440 initial patients, leads therefore to a statistical power slightly greater than 30%. No interim efficacy analysis is foreseen for OS. Note that this trial is underpowered for OS, and that there is no intention to claim efficacy for OS.

Given the key relevance of this endpoint in an overall assessment of benefit/risk, it is important to power the OS analysis to exclude a significant detrimental effect of arfolitixorin as compared to leucovorin in terms of OS. The OS analysis is based on the following considerations:

1) Collection of vital status will continue in all randomized patients after the final PFS analysis, until the OS analysis. Of note, the aim for this continued follow-up is to monitor the patients' safety, rather than showing superior efficacy of arfolitixorin in terms of OS.

2) Arfolitixorin is not expected to cause serious toxicity. It is therefore reasonable to assume that if superior efficacy of arfolitixorin is demonstrated, OS is unlikely to be inferior in the arfolitixorin arm compared to the leucovorin arm. The magnitude of the OS difference is hard to predict, for example because of the multiple lines of active therapies currently available for the treatment of patients with advanced colorectal cancer. Assuming similar post-progression survival in both trial arms, a reasonable assumption is that the difference in median PFS between arms will "carry over" to the difference in median OS.

3) A "significant detrimental effect on OS" will be taken equal to the benefit of leucovorin estimated in the meta-analysis of all trials comparing 5-fluorouracil alone with 5-fluorouracil + leucovorin (Meta-Analysis Group In Cancer 2004). The overall OS HR of leucovorin was equal to 0.9 in these trials; hence, an OS hazard ratio equal to 1/0.9 = 1.11 in the current trial may be considered a significant detriment, to the extent that it will outweigh the benefit observed historically with leucovorin.

4) Collection of vital status will continue after the PFS analysis, until 60% of the patients have had a death event. In function of the results of the interim analysis 264 deaths (60% of 440 patients) or 396 deaths (60% of 660 patients) would have been required. However, with the addition of the extra patients 294 deaths (60% of 490 patients) will be required instead of 264 (60% of 440 patients).

3.3.6 Japanese study

Japanese enrollment continued beyond completion of the global enrollment. Approximately 12.7% of the total study population had to be enrolled from Japan to obtain required probability of at least 70% that the ratio (effect in Japanese patients)/(effect in all patients) is at least 0.5 i.e. $\frac{D_{Japan}}{D_{All}} > 0.5$, with D_{Japan} the difference in ORR between randomized treatment arms, exclusively in Japanese patients, and D_{All} the difference in ORR between randomized treatment arms (Ministry of Health 2007). The proportion of Japanese patients needed was calculated according to the methods described in (Ikeda and Bretz, 2010).

4. General Analysis Definitions

Data will be analyzed using SAS (Version 9.4 or higher). Graphs will be produced in SAS or R.

Descriptive statistics will be tabulated as follows:

- Categorical data will be summarized in contingency tables presenting frequencies and percentages. By default, those percentages will be calculated on the number of patients in the analysis population. If this is not the case, it will specifically be mentioned in the footer of the table.
- Continuous data will be summarized using number of non-missing values (n), mean, standard deviation, median, minimum and maximum values.

There will be 3 columns (i.e. overall population, arfolitixorin arm and Leucovorin arm).

Listings with individual values will be provided for all data presented in the tables.

Significance tests will be performed at the $\alpha = 0.05$ significance level for two-tailed tests and at $\alpha = 0.025$ for one-tailed tests. When a confidence interval is to be calculated, Wald confidence intervals will be used, unless otherwise specified.

Adjustment of the significance level in the interim analysis and final analysis has been done according to group sequential methods, as explained in Section 3.3.1.

4.1 Study Period and Visit Window Definitions

Visit windows will not be used, tables will assume that observations are from the recorded visit irrespective of the date specified.

4.2 Planned analyses

The trial design will include the following analyses (and adaptations if indicated):

- An interim calculation of the conditional power for ORR and PFS took place when 16-week BICR evaluation had been performed for the 330th patient. At this analysis, the DSMB could have recommend a 50% sample size increase (from 440 to 660 patients), according to the guidelines in Table 3[4].
- In the current case where the sample size is not increased, the final analysis of ORR and PFS of the main study will take place when at least 235 PFS events are observed.
- The final analysis of OS of the main study will be performed when 60% of the patients have had a death event.

4.2.1 Interim analysis[4]

a) Interim analysis of Overall Response Rate (ORR)

An interim analysis took place when 16-week BICR evaluation had been performed for the 330th patient. For the interim analysis of ORR, the significance level was determined using a Rho spending function, with $\rho = 5$, resulting in a very conservative boundary. Assuming an information fraction of 75% for ORR at this interim analysis, the 1-sided significance level for efficacy was 0.006, and it would have been reached if the difference in ORR had been larger than 13.8%. This small alpha spending was foreseen to take into account the interim look at the efficacy primary endpoint, i.e. ORR. However, there was no intention

[4] The interim analysis was performed before the decision to add the extra Japanese patients to the efficacy and safety analyses.

to claim efficacy for ORR at this stage. The interim results, i.e. the conditional power for both ORR and PFS, only served to guide the decision rules regarding sample size re-estimation. Results of this analysis remained blinded to everyone, except the DSMB. Regardless of the efficacy results of the interim analysis, the trial continued to accrue patients, until the total planned sample size is reached.

The interim ORR analysis was performed on a subset of the Intent-to-treat analysis set (cf. section 4.3) for subjects who were evaluable for the 16-week BICR assessment or who discontinued the study before any evaluation by the BICR. Patients who discontinued the study before the 16-week assessment were considered non-responder. They were counted in the denominator of the ORR but not in the numerator. All scans evaluated by the BICR until the end of treatment available at the time of the interim analysis were used.

b) Interim analysis of Progression Free Survival (PFS)

As for ORR, there was no intention to claim efficacy for PFS at this early stage either. A small alpha spending was foreseen to take into account the interim look. For this, a very conservative Rho spending function boundary, with $\rho = 5$, was used. Assuming an information fraction of 37% for PFS at this interim analysis, the 1-sided significance level for efficacy was 0.000173, and it would have been reached if the hazard ratio had been less than 0.51.

The interim analysis of the PFS was performed on the ITT analysis set (cf. section 4.3) and used all scans evaluated by the BICR and available at the time of the interim analysis.

c) Sample size adaptation

The conditional power for both ORR and PFS were calculated at the interim analysis, assuming that the estimated treatment difference at interim analysis was the true effect. Depending on the conditional powers, the trial could have either continue as planned to accrue 440 patients, or the sample size of the trial would have been increased by 50%, for a total sample size of 660 patients. The guidelines provided in Table 3 were used by the DSMB to make a recommendation to the Sponsor. Two possible decisions could be communicated to the Sponsor by the DSMB: "sample size increase" or "no sample size increase".

The conditional power for ORR would be equal to approximately 80% if the observed difference in ORR, at the time of the interim analysis (75% information fraction) was equal to 11.5%. The conditional power for PFS would be equal to 25% if the observed hazard ratio, at the time of the interim analysis (37% information fraction) was equal to 0.85. Note that a HR=0.85 corresponds with a Median survival time Control/Experiment of 10/11.75, which is considered clinically meaningful. The conditional power for PFS would be equal to 50% if the observed hazard ratio, at the time of the interim analysis (37% information fraction) was equal to 0.80. The conditional power for PFS would be equal to 75% if the observed hazard ratio, at the time of the interim analysis (37% information fraction) was equal to 0.75.

ORR		PFS		Recommendation	CP660 for	
CP440	Observed ORR difference	CP440	Observed HR		ORR	PFS
≥ 80%	≥11.5%	≥ 75%	≤0.75 (≥3.3 m)	Continue trial unchanged	NA	NA
		50-75%	>0.75; ≤0.8 (≥2.5<3.3m)	Increase sample size	≥ 90%	70-90%
		25-50%	>0.8; ≤0.85 (≥1.8;<2.5m)	Increase sample size*	≥ 90%	>40 - <70%
		<25%	>0.85 (<1.8m)	Continue trial unchanged	NA	NA
60-80%	≥ 10%;< 11.5%	≥ 75%	≤0.75 (≥3.3 m)	Increase sample size	80-90%	≥90%
		50-75%	>0.75; ≤0.8 (≥2.5<3.3m)	Increase sample size	80-90%	70-90%
		25-50%	>0.8; ≤0.85 (≥1.8;<2.5m)	Increase sample size*	80-90%	>40 - <70%
		<25%	>0.85 (<1.8m)	Continue trial unchanged	NA	NA
< 60%	<10%	≥ 75%	≤0.75 (≥3.3 m)	Continue trial unchanged	NA	NA
		50-75%	>0.75; ≤0.8 (≥2.5<3.3m)	Continue trial unchanged	NA	NA
		25-50%	>0.8; ≤0.85 (≥1.8;<2.5m)	Continue trial unchanged	NA	NA
		<25%	>0.85 (<1.8m)	Continue trial unchanged	NA	NA

Table 3: Guidelines for sample size increase at interim analysis

ORR = overall response rate; PFS = progression-free survival S=Success, F=Failure, CP=Conditional Power assuming that the estimated treatment difference at interim analysis is the true effect, CP440= conditional power at N=440, CP660= conditional power at N=660. m=months, NA=not applicable
** Including type I error rate adjustment, using the Cui, Hung and Wang method (Cui et al., 1999)*

4.2.2 Final analysis of ORR and PFS

Initially, the final analysis of ORR and PFS should have taken place when 300 events are observed, if the sample size was not increased. While if the sample size was increased, then the final analysis of ORR and PFS should have taken place when 450 PFS events are observed.

Given the decision of the DSMB at the interim analysis to continue the study unchanged, and given the updated timing of the final analysis, the final analysis of ORR and PFS will take place when 235 PFS events are observed (see also Sections 3.2.1 & 3.3.4).

a) Overall Response Rate (ORR)

The 1-sided significance level for efficacy will be 0.024. If the sample size would have been increased to 660 patients, the 1-sided significance level for efficacy would have been be 0.025. More details about the calculation of the ORR are provided in section 4.5.

All scans evaluated by the BICR at the time of the final analysis of the ORR and PFS will be used.

b) Progression Free Survival (PFS)

The 1-sided significance level for efficacy will be 0.025.

If the sample size would have been increased to 660 patients, the 1-sided significance level for efficacy would have been 0.025.

c) Overall Survival (OS)

The Overall Survival analysis will be presented at the time of the final analysis of ORR and PFS as supportive analysis.

d) Other efficacy endpoints

All other efficacy endpoints will be presented at the time of the final analysis of ORR and PFS.

e) Safety analysis

The safety analyses will be performed on the Safety analysis set (cf. section 4.3).

4.2.3 Final analysis of Overall Survival (OS)

The final analysis of OS will be performed when 60% of the patients have had a death event.

The overall survival analysis will be presented. The other efficacy endpoints and the safety assessment will be re-evaluated based on the data available at the time of the overall survival analysis. The efficacy endpoints will be assessed both on the ITT and the PP analysis sets. The safety analyses will be performed on the Safety analysis set (cf. section 4.3).

4.2.4 Analysis of Japanese patients

The analysis of the results for the Japanese patients will be performed at the time of the final ORR and PFS analysis.

For this analysis, all analyses planned in this SAP will be repeated on the corresponding Japanese analysis subsets (cf. section 4.3) except if specified otherwise.

4.3 Definition of Populations

All enrolled patients who were randomized will be included in the analyses. The following analysis sets will be considered:

- **Intention-to-Treat (ITT) analysis set** consists of all randomized participants. Analyses will be based on the randomized treatment.
- **Intention-to-Treat Japanese (ITTJ) analysis set** consists of all randomized Japanese participants. Analyses will be based on the randomized treatment.
- **Intention-to-Treat Excluding Additonal Japanese Patients (ITTE) analysis set** consists of randomized participants excluding additional Japanese patients. Analyses will be based on the randomized treatment.
- **Per-Protocol (PP) analysis set** defines a subset of the participants in the ITT analysis set without major protocol deviations. The definition of major protocol deviations will be agreed upon, and all cases of such major deviations adjudicated prior to database lock. Analyses will be based on the randomized treatment.
- **Per-Protocol Japanese (PPJ) analysis set** defines a subset of the participants in the ITTJ analysis set without major protocol deviations. The definition of major protocol deviations will be agreed upon, and all cases of such major deviations adjudicated prior to database lock. Analyses will be based on the randomized treatment.
- **Safety analysis set** defines the subset of participants for whom safety analyses will be conducted. It includes all patients who received at least one dose of study medication. Analysis will be based on the actual treatment.
- **Safety Japanese analysis set** defines the subset of Japanese participants for whom safety analyses will be conducted. It includes all Japanese patients who received at least one dose of study medication. Analysis will be based on the actual treatment.
- **PK analysis set** defines a subset of the participants in the Safety analysis set for whom PK samples were analysed. Analyses will be based on the actual treatment.
- **PK Japanese analysis set** defines a subset of the participants in the Safety Japanese analysis set for whom PK samples were analysed. Analyses will be based on the actual treatment.

4.4 Subgroup Definitions

Subgroup analyses for the efficacy endpoints: ORR and PFS (all estimands) with respect to the trial's stratification factors (Geographic region, Primary tumor location and previous adjuvant CRC treatment) and the prognostic factors described hereunder will be performed. Of note, all factors involved in the subgroup definition will be based on baseline information. The confidence intervals and the p-values will be displayed only for exploratory purposes with no intention to claim efficacy.

The consistency of the treatment effect across prognostic subgroups will be assessed visually using forest plots analyses based on observed results.

Table 4: Subgroup definitions

Factors	Subgroup levels
Primary tumor location	Left colon, Rectal cancer, Right colon
Previous adjuvant CRC treatment	Yes, No
Primary tumor resected	Yes, No
ECOG performance status	0, 1
Tumor mutation status BRAF	Wild-type, Mutant
Tumor mutation status KRAS	Wild-type, Mutant
Tumor mutation status NRAS	Wild-type, Mutant
Liver metastasis	Yes, No
Lung metastasis	Yes, No
Peritoneal metastasis	Yes, No
Other metastasis	Yes, No
Measurable lesion burden	< 25^{th} percentile [25^{th} perc. – 75^{th} perc.] > 75^{th} percentile

Factors	Subgroup levels
Geographic region	North America, Europe, Australia, Japan
Sex	Male, Female
Age	<65 years ≥65 years >75 years
Body Mass Index (BMI)	<25 25+
Albumin	Low Normal High
Alanine Aminotransferase	Low Normal High
Aspartate Aminotransferase	Low Normal High
Gamma Glutamyl Transferase	Low Normal High
Leukocytes	Low Normal High
Hemoglobin	Low Normal High
Platelets	Low Normal

Factors	Subgroup levels
	High
Neutrophils	Low Normal High
Neutrophils/Leukocytes ratio	Low Normal High

4.5 Estimand definitions

According to the FDA guidance document E9(R1) *Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials* (May 2021), and the EMA guideline document ICH E9 (R1) *addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials* (30 July 2020), the primary endpoint ORR and the secondary endpoint PFS are formulated using the estimand framework, taking into account relevant intercurrent events.

4.5.1 Estimand definitions for ORR

The primary estimand for ORR is defined by:

- Treatments: experimental vs comparator (see Section 3.2.1)
- Population: ITT analysis set (see Section 4.3)
- Variable: proportions of patients with ORR as BOR (See Section 4.6 for definitions)
- Intercurrent events and strategies to address those:
    - Early death, metastasis resection, new anticancer treatment and treatment discontinuation causing an non-evaluable or missing result for ORR: a composite strategy will be used, by imputing any missing or unevaluable response data as failure (non-responder imputation).
- Population-level summary: The common difference in ORR between the two arms and its associated 95% confidence interval will be estimated using the Mantel-Haenszel estimate of the common risk difference by using Mantel-Haenszel stratum weights (Mantel and Haenszel 1959) and the Sato variance estimator of this difference (Sato 1989). (see Section 9.1.1).

No supplemental estimands are defined for ORR.

4.5.2 Estimand definitions for PFS

The primary estimand for PFS is defined by:

- Treatments: experimental vs comparator (see Section 3.2.1)
- Population:
    - Population defined through the inclusion/exclusion criteria
    - Analysis set: ITT analysis set (see Section 4.3)
- Variable: time from randomization to event time (see Section 4.6 PFS definition)
- Intercurrent events and strategies to address those:
    - Start of a new anticancer treatment (except metastasis resection) before any documented progressive disease or death: a hypothetical strategy will be used, by censoring patients with this ICE at the last tumor assessment with documented non-progression prior to this ICE (see Section 4.6, Table 5).
    - Metastasis resection: patients with metastasis resection continue to be scanned until disease progression, defined as new lesion appearance or significant increase of residual lesions (See Section 4.6 under BOR). For this ICE a treatment policy is followed, i.e. the ICE will be ignored.
- Population-level summary: hazard ratio, estimated by a stratified Cox proportional hazards model (see Section 9.1.2).

The first supplemental estimand for PFS is defined by:

- Treatments: experimental vs comparator (see Section 3.2.1)
- Population:
    - Population defined through the inclusion/exclusion criteria
    - Analysis set: ITT analysis set (see Section 4.3)

- Variable: time from randomization to event time (see Section 4.6 PFS definition)
- Intercurrent events and strategies to address those:
  - Start of a new anticancer treatment (except metastasis resection) before any documented progressive disease or death: a treatment policy strategy will be followed, by ignoring this ICE (see Section 4.6, Table 6).
  - Metastasis resection: patients with metastasis resection continue to be scanned until disease progression, defined as new lesion appearance or significant increase of residual lesions (See Section 4.6 under BOR). For this ICE a treatment policy is followed, i.e. the ICE will be ignored.
- Population-level summary: hazard ratio, estimated by a stratified Cox proportional hazards model (see Section 9.1.2).

The second supplemental estimand for PFS is defined by:

- Treatments: experimental vs comparator (see Section 3.2.1)
- Population:
  - Population defined through the inclusion/exclusion criteria
  - Analysis set: ITT analysis set (see Section 4.3)
- Variable: time from randomization to event time (see Section 4.6 PFS definition)
- Intercurrent events and strategies to address those:
  - Start of a new anticancer treatment (except metastasis resection) before any documented progressive disease or death: a composite strategy will be used: patients with this ICE will be considered to have a progression event on the date of start of the ICE (see Section 4.6, Table 7).
  - Metastasis resection: patients with metastasis resection continue to be scanned until disease progression, defined as new lesion appearance or significant increase of residual lesions (See Section 4.6 under BOR). For this ICE a treatment policy is followed, i.e. the ICE will be ignored.
- Population-level summary: hazard ratio, estimated by a stratified Cox proportional hazards model (see Section 9.1.2).

4.6 Calculated Variables

- **Additional Japanese patients** are defined as the additional patients enrolled to address the Japanese authorities' request that approximately 12.7% of the total study population are enrolled from Japan. A flag variable in a SDTM dataset will be used to distinguish these additional Japanese participants.

- **Baseline** is defined as the last non-missing value measured or collected before the first dose of any study treatment. If time not available, assessments done on the date of study treatment administration are assumed to take place before the administration, unless specified otherwise.

- **Best confirmed Overall Response (BOR)**

The estimand related to this outcome is specified in Section 4.5.1.

The evaluations of the Overall Response and the Best confirmed Overall Response (BOR) are performed by the BICR using RECIST 1.1 (Eisenhauer et al., 2009).

When there is a discordance between assessments of the two radiologists then the adjudicated assessment will be used. More details about the evaluation of the confirmed response is provided in the BICR Imaging Review Charter (Banook 2021).

BOR will not be derived by IDDI, instead the BICR BOR results, based on RECIST 1.1, will be used.

Patients with metastatic resection are continued to be scanned until disease progression. The BICR assessments of the overall response of patients with a metastasis resection will be as follows:

- Patients having complete metastasis surgery will have the overall response as NE until new lesion appearance leading to PD (Progressive Disease);
- Patients having partial metastasis surgery will have the overall response as NE too until new lesion appearance or significant increase of the residual lesion(s) leading to PD (Progressive Disease) as per RECIST 1.1 criteria.

The process of the BICR evaluation of the CT-scans/MRIs is described in detail in the BICR Imaging Review Charter (Banook 2021).

- **Overall response rate (ORR) RECIST 1.1**
  The overall response rate (ORR) is defined as the proportion of subjects for whom the best confirmed overall response (BOR), as per BICR, is complete response (CR) or partial response (PR) as determined by the BICR using RECIST 1.1.

$$\frac{Number\ of\ patients\ with\ a\ confirmed\ BOR\ of\ CR\ or\ PR}{Number\ of\ Patients\ included\ in\ the\ ORR\ analysis}$$

- **Progression-free survival (PFS)**
  = (date of death due to any cause or date of PD, where PD is defined as the first occurrence of tumor progression based on CT-scans/MRIs, whichever comes first, as per BICR) – randomization date + 1. The PFS will be expressed in months.

  In addition, some anticipated special cases or intercurrent events will be taken into account. Three estimands concerning PFS are specified in Section 4.5.1: the primary estimand, for which Table 5 is to be used, the first supplemental estimand, for which Table 6 is to be used, and the second supplemental estimand, for which Table 7is to be used.

Table 5: Censoring rules for the primary PFS estimand

Situation / Intercurrent Event	Handling
No documented progression and no death (no observed event)	Censored at last tumor assessment with documented non-progression
Missing (or incomplete) baseline tumor assessments	Censored at randomization
Progression or death documented immediately after 2 or more consecutive missed assessment visits	Censored at last tumor assessment with documented non-progression (before missing visits)
Death occurring before any documented progressive disease (and not immediately following 2 consecutive missed assessment visits)	Considered as an event at the date of death
New anticancer treatment (except metastasis resection) started before any documented progression or death	Censored at last tumor assessment with documented non-progression (before start of new treatment)

Situation / Intercurrent Event	Handling
Metastasis resection, started before any documented progression or death	These patients will continue to be followed up until death or PD, as documented by BICR. Considered as an event at date of progression or death.

Table 6: Censoring rules for the first supplemental estimand for PFS

Situation / Intercurrent Event	Handling
No documented progression and no death (no observed event)	Censored at last tumor assessment with documented non-progression
Missing (or incomplete) baseline tumor assessments	Censored at randomization
Progression or death documented immediately after 2 or more consecutive missed assessment visits	Censored at last tumor assessment with documented non-progression (before missing visits)
Death occurring before any documented progressive disease (and not immediately following 2 consecutive missed assessment visits)	Considered as an event at the date of death
New anticancer treatment (except metastasis resection) started before any documented progression or death	These patients will continue to be followed up until death or PD, as documented by BICR.
Metastasis resection, started before any documented progression or death	These patients will continue to be followed up until death or PD, as documented by BICR. Considered as an event at date of progression or death.

Table 7: Censoring rules for the second supplementary estimand for PFS

Situation / Intercurrent Event	Handling
No documented progression and no death (no observed event)	Censored at last tumor assessment with documented non-progression
Missing (or incomplete) baseline tumor assessments	Censored at randomization
Progression or death documented immediately after 2 or more consecutive missed assessment visits	Censored at last tumor assessment with documented non-progression (before missing visits)
Death occurring before any documented progressive disease (and not immediately following 2 consecutive missed assessment visits)	Considered as an event at the date of death
New anticancer treatment (except metastatic resection) started before any documented progression or death	Considered as a progression event at the date of start of new anticancer treatment.
Metastasis resection, started before any documented progression or death	These patients will continue to be followed up until death or PD, as documented by BICR.

Situation / Intercurrent Event	Handling
	Considered as an event at date of progression or death.

- **Overall Survival** (OS)
  = date of death due to any cause - randomization date + 1.
  The OS will be expressed in months.

  In addition, some anticipated special cases or intercurrent events will be taken into account as described in the Table 8.

Table 8: Censoring rules for OS

Situation / Intercurrent Event	Handling
No documented death (no observed event)	Censored at the last contact date or the randomization date (whichever comes later)
Lost to follow-up	Censored at the last contact date or the randomization date (whichever comes later)

- **Duration of Response (DOR)**
  = (date of death due to any cause or date of PD, where PD is defined as the first occurrence of tumor progression based on CT-scans/MRIs as per BICR, whichever comes first) – date of first confirmed response of CR or PR (the start date of response, not the date when response was confirmed) + 1.
  The DOR will be expressed in months.

  Patients with no response of CR or PR or in whom CR or PR is never confirmed will not be included for this endpoint.

  The anticipated special cases or intercurrent events described in Table 5: Censoring rules for the primary PFS estimand also apply for DOR. Note: patients with metastatic resection are continued to be scanned until disease progression.

- **Recurrence Free Survival (RFS) for patients undergoing metastatic resection**
  = (date of death due to any cause or date of PD, where PD is defined as the first occurrence of tumor progression based on CT-scans/MRIs as per BICR, whichever comes first) – date of first surgery with removal of the metastasis + 1. The RFS will be expressed in months.

  The anticipated special cases or intercurrent events described in Table 5: Censoring rules for the primary PFS estimand also apply for RFS. Note: patients with metastatic resection are continued to be scanned until disease progression.

4.7 Partial Dates

Partial adverse event (AE) start date will not be imputed.

If a medication date is missing or partially missing, in a way that it cannot be determined if it was taken prior or concomitantly (using the date and the eCRF variable that indicates if the medication was taken prior to the date of first dose of study treatment), it will be considered as concomitant medication.

4.8 Methods To Be Used For Handling Missing Data

As general rules, missing data will not be imputed.

Per definition of ORR in section 4.5, any subjects with missing response information to treatment will be counted as non-responders for ORR.

For PK data, the values under the LLOQ will be imputed to ½ times the concentration value of the LLOQ.

5. Study Patients

5.1 Disposition of Patients

The study disposition of all enrolled patients will be summarized with descriptive statistics. It will present the number of patients:

- Screened,
- Screen failure,
- ITT analysis set (see definitions in Section 4.3),
- Safety analysis set (see definitions in Section 4.3),
- Per-Protocol analysis set (see definitions in Section 4.3).

Note: for the analysis of Japanese patients the number of patients in the ITTJ, the PPJ and the Safety Japanese analysis sets will be presented.

For treatment discontinuation, the following categorical outcomes will be presented:

- Patients with ongoing treatment,
- Patients who discontinued the treatment, and the reason for discontinuation.

The frequency and primary reason for terminating the study will also be summarized, as will be death and cause of death.

5.2 Protocol Deviations

Major protocol deviations will be summarized for the ITT analysis set. Deviations to the protocol will be classified as major or minor by Isofol in a list sent to IDDI before database lock.

5.3 In- and Exclusion Criteria

Listing of all in- and exclusion criteria not met will be provided for all screened patients.

6. Demographic and other Baseline Characteristics

Descriptive statistics with respect to patient characteristics at baseline will be displayed for the ITT analysis set. Intervention groups (Overall, Arfolitixorin and Leucovorin) will be compared on baseline characteristics, including demographics and laboratory measurements. No inferential statistics will be used.

The variables to be summarized are:

- Gender, Age (continuous),

- Race, ethnicity (If race and ethnicity are collected),
- Primary tumor location (left colon/right colon/rectal cancer) from IWRS,
- Previous neo-adjuvant/adjuvant CRC treatment (yes/no) from IWRS,
- Geographic region (Europe / North America / Japan / Australia) from IWRS,
- Pregnancy test.

7. Medical history

Colorectal Cancer History will summarize by treatment arms:

- The primary tumor resected (Yes, No missing),
- The primary tumor still in place (Yes, No, missing),
- The location of metastatic cancer (Liver, Lung, Other),
- The time since original diagnosis (in months),
- The time since metastatic diagnosis (in months).

Other medical history will be tabulated by system organ class and preferred term (per MedDRA), with three separate tables:

- Overall,
- 'ongoing' medical history,
- 'not ongoing' medical history.

8. Prior and Concomitant Treatment and procedures

Prior and Concomitant medications will be classified according to World Health Organization Drug Dictionary, WHODrug Global C3. The number and percentage of participants receiving a prior or concomitant medication will be displayed by Anatomical Therapeutic Chemical (ATC) level 1 and 4 for the ITT analysis set. Medications will be reported as prior when they start before the first day of study treatment and if they end before the first day of study treatment. Medications will be reported as concomitant when they start before, on or after first day of study treatment and continue afterwards.

There will be two separate tables, one for prior medications and one for concomitant medications.

Prior and Concomitant medication summaries will be sorted alphabetically by ATC level 1 and 4. A listing of all medications recorded on the prior and concomitant medications CRF page will provide details including indication, dose, route, frequency, and start and stop dates.

Concomitant procedures and surgeries will be classified using the Medical Dictionary for Regulatory Activities (MedDRA) and will be tabulated by system organ class and preferred term for the ITT analysis set.

Two tables will report the other anti-cancer therapies, one for the metastasis resection and a second one for the anti-cancer therapies excluding metastasis resection. The metastasis resection will be displayed by MedDRA System Organ Classes and preferred terms. The number and percentage of participants receiving other anti-cancer therapies will be displayed by WHODrug ATC level 1 and 4 for ITT analysis set as the other medication tables. The other anti-cancer therapies and the metastatic resection are entered by investigator in a distinct CRF form than the other concomitant treatments and procedures. Variables in the SDTM datasets will allow identifying other anti-cancer therapies and metastasis resections.

9. Efficacy Evaluation

All efficacy analyses will be performed on both the ITT and the per-protocol analysis sets. The analyses performed on the ITT analysis set are considered as primary.

9.1 Response assessment

9.1.1 Overall Response Rate (ORR)

ORR will be analyzed using a Cochran-Mantel-Haenszel test (CMH), stratified for the stratification factors used for randomization (geographic region, primary tumor location and previous neo-adjuvant/adjuvant CRC treatment).

The ORR will be estimated along with 95% confidence limits for proportion.

The table for ORR will show the following results

- The best confirmed overall response (CR/PR/SD/PD/NE),
- Overall response rate with 95% confidence limits and p-value from a stratified Cochran-Mantel-Haenszel test.
- The common difference in ORR between the two arms and its associated 95% confidence interval will be estimated using the Mantel-Haenszel estimate of the common risk difference by using Mantel-Haenszel stratum weights (Mantel and Haenszel 1959) and the Sato variance estimator of this difference (Sato 1989).

If the ORR response is coded 0 for failure and 1 for response then the option "column=2" must be used and the treatment variable must be coded "1"= experimental treatment and "2"=control treatment, so that p1-p2 represents the desired difference.

This test will be performed using the SAS PROC FREQ with statements

```
PROC FREQ data=dataset;

      TABLES strata*trt*orr  / COMMONRISKDIFF(CL=MH TEST=MH column=2);
RUN;
```

The 1-sided significance level for the final analysis of ORR will be 0.024 if no sample size increase and 0.025 in case of sample size increase.

This ORR analysis will be repeated on the Intention-to-Treat Excluding additional Japanese patients (ITTE) analysis set in order to account for potential bias due to non-concurrent enrollment of the additional Japanese patients.

A listing will be provided for

- Time point response information (sum of target lesions, target lesion response, non-target lesion response, overall response), as entered in the CRF at each tumor assessment, including baseline (only for sum of target lesions) and unscheduled assessments.

Note: For the analysis of the Japanese patients:

- In addition to the analyses planned above, the ratio (effect in Japanese patients)/(effect in all patients) will be compared to 0.5 i.e. $\frac{D_{Japan}}{D_{All}} > 0.5$ as mentioned in section 3.3.6.

- The tests mentioned above will be performed without use of stratification factors.

9.1.2 Progression free Survival (PFS), Duration of Response (DOR) and Overall Survival (OS)

PFS, DOR and OS will be analyzed using the same methodology:

- The stratified logrank test will be produced with the SAS procedure PROC LIFETEST using the stratification factors applied for randomization. Descriptive statistics will also be presented. Kaplan-Meier curves will be created with SAS PROC LIFETEST or with R.
- A stratified Cox proportional hazards model (using the same stratification factors as for randomization and the option "ties=discrete") will be performed with the SAS procedure PROC PHREG. The assumption of proportional hazards will be assessed.
- The 1-sided significance level for the final analysis of PFS, DOR and OS will be 0.025.

The following results will be presented:

- Kaplan-Meier Graphical display (including numbers at risk),
- Number of patients,
- Number of events,
- Median months to event + 95% confidence interval (Brookmeyer and Crowley confidence interval),
- Stratified logrank test statistic and its p-value and
- Hazard Ratio (with 95%CI) from the Cox regression.

Note: For the analysis of the Japanese patients, the tests mentioned above will be performed without use of stratification factors.

Sensitivity analyses (not for the analysis of Japanese patients)

In addition to the primary analysis of PFS, two sensitivity analyses will be conducted, following the censoring rules as displayed in Table 6 and Table 7 of section 4.5. The results of those sensitivity analyses will be presented similarly to the primary analysis of PFS.

Analyses on the ITTE analysis set

The Progression Free Survival (PFS), the Duration of Response (DOR) and the Overall Survival (OS) analyses will be repeated on the Intention-to-Treat Excluding additional Japanese patients (ITTE) analysis set.

9.1.3 Recurrence Free Survival (RFS)

For patients having metastasis resection done during the study, treatment with the study regimen may be continued at the investigator discretion.

The exploratory endpoint Recurrence Free Survival (RFS) will be analyzed using the same methodology as PFS, DOR and OS (cf. section 9.1.2).

The exploratory endpoint Recurrence Free Survival (RFS) will not be done for the analysis of Japanese patients.

9.2 PRO EQ-5D questionnaire

The PRO EQ-5D questionnaire results will be analyzed using descriptive statistics and graphical representation.

PRO EQ-5D has five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems and extreme problems (Van Reenen et al., 2015). The number and the proportion of patients will be presented by treatment arm in a table for each level, for each dimension and for each visit.

The EQ VAS records the respondent's self-rated health on a vertical, visual analogue scale where the endpoints are labelled 'Best imaginable health state' and 'Worst imaginable health state'. EQ VAS data will be described by the mean, the median, the standard deviation, the first quartile and the second quartile as recommended by EQ-5D-5L user guide (Van Reenen et al., 2015).

9.3 Patients undergoing metastasis resection

Defined as the number of patients qualifying for curative metastasis resection after treatment with study drug will be compared between the treatment groups. This will be summarized by treatment groups using descriptive statistics.

10. Safety Evaluation

All safety analyses will be done on the safety analysis set.

10.1 Extent of Exposure

The extent of exposure will be summarized descriptively for each arm.

10.2 Adverse Events

Adverse events (AEs) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA 22.1) and will be graded according to the National Center Institute Common Terminology Criteria for AEs (NCI-CTCAE criteria [v5.0]). Uncoded terms will be presented under "SOC uncoded", with their reported term.

Adverse events will be analyzed in terms of their type, incidence, severity and relationship to the study treatment. Related AEs are defined as events with a relationship to study treatment equal to 'Possible', 'Probable', or 'Related' or with missing relationship.

A summary table will present by treatment arm the number and percentage of patients with at least one:

- AE
- Serious AE
- Adverse event of CTC grade ≥3
- Related AE
- AE leading to permanent discontinuation of the study treatment
- Fatal AE
- AE of special interest (according to current version of List of AESI PT)

Of note, the list of preferred terms to be considered as AESI will be provided by sponsor to IDDI for programming purposes.

In addition, tabulations of the number of patients who experienced AEs as well as severity of the events will be presented by system organ class and preferred term.

Patients will only be counted once for each preferred term. In case a patient experienced the same event more than once, the worst severity will be presented.

The following tabulations will be presented:

- All AEs
- AEs leading to permanent discontinuation of the study treatment
- AEs related to the study treatment
- AE of special interest (according to current version of List of AESI PT)

Listing of all adverse events by treatment arm will be provided, flagging the ones that are AESI, including the patient identifier, age, race, sex, verbatim, preferred term, duration of the event, severity, action taken, outcome, causality, and date of onset.

In addition, a listing of AESIs will be provided similarly to the listing of all AEs.

10.3 Deaths and Serious Adverse Events

Serious adverse events (SAEs), fatal AEs and *NCI/CTC grade ≥3 AEs* will be summarized grouped by system organ class and preferred term. In addition, listings of SAEs, fatal AEs and grade ≥3 AEs will be provided, similarly to the listing of all AEs.

The number of deaths will be tabulated together with the primary cause of death. The details of the 'other cause' will be included in the listing.

The number of deaths on IMP treatment or within 30 days of study drug discontinuation will be displayed by treatment arm.

The number of deaths occurring in the follow-up after 30 days of study drug discontinuation will be displayed by treatment arm.

10.4 NCI PRO-CTCAE questionnaire

The NCI PRO-CTCAE questionnaire results will be analyzed using descriptive statistics and graphical representations.

The number and the proportion of patients will be presented by treatment arm in a table for each categorical response level of each question and for each visit.

10.5 Clinical Laboratory Determination

The laboratory safety samples (hematology, clinical chemistry, and urinalysis) will be collected according to the protocol schedule of activities and in accordance with local practice and analyzed by local laboratories.

The observed values will be recorded and assessed as "normal" or "abnormal" in the eCRF. Abnormal values will be assessed by the Investigator as "clinically significant" or "not clinically significant".

All required laboratory tests are specified in Table 9.

Table 9: Laboratory tests

HEMATOLOGY	CLINICAL CHEMESTRY	URINALYSIS
Hemoglobin	Sodium	Protein
Hematocrit	Potassium	Glucose

HEMATOLOGY	CLINICAL CHEMESTRY	URINALYSIS
WBC	Calcium	Blood
Neutrophils	Glucose	
Lymphocytes	BUN or urea	
Monocytes	Creatinine	
Eosinophils	AST	
Basophils	ALT	
Platelet count	Alkaline phosphatase	
	Total bilirubin	
	Total protein	
	Albumin	
	LDH	
	gGT	

Laboratory tests will be summarized using descriptive statistics for actual values and change from baseline for continuous laboratory test. For categorical laboratory test, frequencies and percentages of patients in each category will be used. Summaries will be presented separately for hematology, blood chemistry and urinalysis parameters.

Grading of laboratory values will be assigned programmatically as per the NCI-CTCAE criteria [v5.0]. The calculations of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account. CTCAE grade 0 will be assigned for all non-missing values not graded as 1 or higher. For laboratory tests where grades are not defined by the CTCAE, results will be categorized as normal/abnormal based on laboratory normal ranges.

For all laboratory tests where grades are defined by the CTCAE grading, shift tables from the baseline CTCAE grade to the worst CTCAE grade during study will be produced.

For the calculation of the worst CTCAE grade, both scheduled and unscheduled values available post baseline will be used.

Shift tables from baseline abnormality grade (Normal/Abnormal) to worst abnormality grade during study will be displayed for all laboratory tests where grades are not defined by the CTCAE grading.

10.6 Vital Signs, Physical Findings and Other Observations Related to Safety

Descriptive statistics of systolic and diastolic blood pressure, heart rate and body weight will be presented by visit. Both absolute values and change from baseline will be presented. In case of multiple measurements, the highest for post-baseline blood pressure and heart rate data and the lowest for post-baseline weight will be kept for the calculations.

Descriptive statistics of ECGs results (normal, abnormal not clinically significant, abnormal clinically significant) will be presented for the worst outcome (normal< abnormal not clinically significant< abnormal clinically significant). The ECG results will also be presented as a shift table comparing baseline vs worst.

Data from physical examination and neurological examination will only be listed.

Descriptive statistics of ECOG results will be presented by a shift table, comparing the worst value over the study versus baseline.

11. Pharmacokinetics evaluations

All individual plasma concentrations of MTHF, methyl-THF and THF will be listed by treatment group. Summary statistics will include geometric mean, median, arithmetic mean, standard deviation, coefficient of variation, minimum, maximum and the number of measurements under the LLOQ.

The values under the LLOQ will be imputed to ½ times the concentration value of the LLOQ.

The concentration/time curves will be presented in linear/linear or log/linear scale individual (spaghetti plots) and also summarized by treatment group including error bars.

The following PK parameters will be listed as individual data divided in treatment groups as well as summary statistics by treatment group, for MTHF (parent drug) and each metabolite (methyl-THF and THF):

- AUC (=AUClast)
- C_{max}
- t_{max}
- t_{last}

The pharmacokinetics evaluation will be performed on the PK analysis set.

12. References

Banook, Image Review Charter ISO_CC-007, (version 9.0), 24Mars2021

Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. Biometrics—1999—; 55:853–7

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:S228-47

Ikeda K, Bretz F. Sample size and proportion of Japanese patients in multi-regional trials. Pharmaceut. Statist. 2010;9:207-216

Mantel, N., and Haenszel, W. (1959). Statistical Aspects of Analysis of Data from Retrospective Studies of Disease. *Journal of the National Cancer Institute* 22:719–748, 1959

Meta-Analysis Group In Cancer (2004). Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: an updated meta-analysis. J Clin Oncol 22: 3766-3775

Ministry of Health, Labour and Welfare. Basic Principles on Global Clinical Trials, 2007.

Sato, T. (1989). On the Variance Estimator of the Mantel-Haenszel Risk Difference. *Biometrics* 45:1323–1324. Letter to the editor

Van Reenen M., Janssen B. EQ-5D-5L User Guide (version 1.2), EuroQol Research Foundation, 2015