Integrated Analysis Plan

Clinical Trial Protocol Identification No.	MS200647_0055
Title:	A Phase II/III, Multicenter, Randomized, Placebo-controlled Study of Gemcitabine Plus Cisplatin With or Without Bintrafusp alfa (M7824) as First-line Treatment of Biliary Tract Cancer
Study Phase	Phase II/III
Investigational Medicinal Product(s)	Bintrafusp alfa
Clinical Study Protocol Version	14 July 2021/Version 5.0
Integrated Analysis Plan Author	Coordinating Authors PPD
Integrated Analysis Plan Date and Version	1st February 2022, Version 4
Integrated Analysis Plan Reviewers	

Function	Name
PPD Merck Healthcare	PPD
PPD Merck Healthcare	PPD
PPD	PPD
PPD, Merck Healthcare	PPD
PPD EMD Serono	PPD
PPD	PPD
PPD, EMD Serono	PPD
PPD EMD Serono	PPD
PPD, Merck Healthcare	PPD
PPD, Merck Healthcare	PPD
PPD	PPD
PPD	PPD
PPD EMD Serono	PPD
PPD, Merck Healthcare	PPD
PPD	PPD
PPD EMD Serono	PPD
PPD EMD Serono	PPD

Approval Page

Integrated Analysis Plan: MS200647_0055

A Phase II/III, Multicenter, Randomized, Placebo-controlled Study of Gemcitabine Plus Cisplatin With or Without Bintrafusp alfa (M7824) as First-line Treatment of Biliary Tract Cancer

Approval of the IAP by all Merck Data Analysis Responsible has to be documented within ELDORADO via eSignature. With the approval, the Merck responsible for each of the analyses also takes responsibility that all reviewers' comments are addressed adequately.

By using eSignature, the signature will appear at the end of the document.

1 Table of Contents

2 List of Abbreviations and Definition of Terms

Abbreviation	Definition
1L	First-line
CCI	[redacted]
AdaM	Analysis Data Model
AE	Adverse Event
AESI	Adverse Events of Special Interest
ALT	Alanine Aminotransferase
ALP	Alkaline Phosphatase
aPTT	Activated Partial Thromboplastin Time
AST	Aspartate Aminotransferase
AUC	Area under the concentration-time curve
AUC_{0-t}	Area under the concentration-time curve from time zero to the last sampling time at which the concentration is at or above the lower limit of quantification
$AUC_{0-\infty}$	Area under the concentration-time curve from time zero extrapolated to infinity
ATC	Anatomical Therapeutic Chemical
BLQ	Below the lower limit of quantification
BMI	Body Mass Index
BSA	Body Surface Area
BTC	Biliary Tract Cancer
CCA	Cholangiocarcinoma
C_{EOI}	The concentration observed immediately at the end of infusion
CI	Confidence Interval
COVID-19	Coronavirus Disease 2019
C_{max}	Maximum observed concentration
CMH	Cochran-Mantel-Haenszel
CR	Complete Response
CrCl	Creatinine Clearance
CSR	Clinical Study Report
ctDNA	Circulating tumor DNA
CPI	Checkpoint Inhibitors
C_{trough}	The concentration observed immediately before next dosing (corresponding to pre-dose or trough concentration for multiple dosing)
CV	Coefficient of Variation (%)
DBP	Diastolic Blood Pressure
DCR	Disease Control Rate
DI	Dose Intensity
DLT	Dose-limiting Toxicities
DOR	Duration of Response
DRR	Durable Response Rate

ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	Electronic Case Report Form
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
EDR	Early Discrepancy Rate
[redacted]	
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HGB	Hemoglobin
HLGT	High Level Group Terms
HLT	High Level Term
IA	Interim Analysis
IAP	Integrated Analysis Plan
IC	Immune Cells
ICH	International Council for Harmonization
IDMC	Independent Data Monitoring Committee
IPD	Important Protocol Deviations
irAE	Immune-related Adverse Event
IRC	Independent Review Committee
[redacted]	
ITT	Intent-to-treat
IXRS	Interactive Response System
λ_z	Terminal first order (elimination) rate constant
LDR	Late Discrepancy Rate
Max	Maximum
LLN	Lower Limit of Normal
MCH	Mean Corpuscular Hemoglobin

MCHC	Mean Corpuscular Hemoglobin Concentration
MCV	Mean Corpuscular Volume
MedDRA	Medical Dictionary for Regulatory Activities
Min	Minimum
MN	Miettinen & Nurminen
MSI	Microsatellite Instability
MSS	Microsatellite Stable
CCI	
NC	Not Calculated
NCI-CTCAE	National Cancer Institute – Common Terminology Criteria for Adverse Events
nd	Not determined
ND	No Disease
CCI	
ORR	Objective Response Rate
ORR2L	Objective Response Rate on Subsequent Line of Anticancer Treatment
OS	Overall Survival
PA	Primary Analysis
PD	Progressive Disease
CCI	
PFS	Progression-free Survival
CCI	
PK	Pharmacokinetic(s)
PR	Partial Response
CCI	
PT	Preferred Term
Q1	First Quartile
Q3	Third Quartile
RBC	Red Blood Cell
RDI	Relative Dose Intensity
RECIST 1.1	Response Evaluation Criteria in Solid Tumors Version 1.1
RMST	Restricted Mean Survival Time
RNA	Ribonucleic Acid
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAFRI	Safety Run-in
SAP	Statistical Analysis Plan
SBP	Systolic Blood Pressure
SD	Stable Disease
SDTM	Study Data Tabulation Model
SI	International System of Units
SMC	Safety Monitoring Committee
SMQ	Standardized MedDRA Queries

SoC	Standard of Care
SOC	System Organ Class
StD	Standard Deviation
$t_{1/2}$	Apparent terminal (elimination) half-life
T4	Thyroxine
TBILI	Total Bilirubin
TC	Tumor Cells
TEAE	Treatment-emergent Adverse Event
TGF-β	Transforming Growth Factor-Beta
TMB	Tumor Mutational Burden
TMTB	Total Measured Tumor Burden
TNM	Tumor, Lymph Nodes, Metastasis
TSH	Thyroid-Stimulating Hormone
ULN	Upper Limit of Normal
VAS	Visual Analog Scale
WBC	White Blood Cells

3 Modification History

Unique Identifier for IAP Version	Date of IAP Version	Authors	Changes from the Previous Version
1	PPD	PPD	Not Applicable
2	PPD	PPD	- Change clinical study protocol version to version 4 - In section 6, updated overall study design description as per protocol update version 3 - In section 6, updated the analysis population for interim analysis adaptation to first 150 randomized antibiotics naïve participants - Updated overview of planned analysis in Table 2 - Added section 7.2 COVID-19 Impact CCI - In section 8.2, for subgroup definition, updated 'disease status at initial diagnosis' by 'disease status at study entry', and group 'unknown' category with 'locally advanced category' CCI CCI - In section 9.1, specified outcomes of adverse events are not affected by the data cut-off rules - In section 9.2 and 9.3, updated definition of baseline measurement to consider in the randomized part and added Table 5 - In section 9.8, added definition of on-treatment period for immune-related adverse events CCI - Updated sections 10.1 "Disposition of participants and Discontinuations", 10.2.1 "Important Protocol Deviations" "Adverse Events" including COVID-19 analyses details - Added Kaplan-Meier plots of time to study treatments discontinuation and time to study discontinuation in section 10.1 - In section 11.1, updated definition of BMI with height in cm converted in m^2 - In section 11.3.1, added category 'others' for molecular abnormality listing CCI

Unique Identifier for IAP Version	Date of IAP Version	Authors	Changes from the Previous Version
			- In section 11.4, updated previous anticancer therapy description - In section 12.4, updated list of type of anticancer drug therapy and added appendix 6 - Updated the whole section 13 especially by removing description by cycle - In section, 14.2.1, updated 'No baseline assessment' definition to consider only cases with no tumor scan recorded before randomization - In section 14.2.1, updated definition of 'Lost to follow-up' to add participants who discontinued/completed the study without an event - In sections 14.1.2, 14.2.2 and 14.3.2, added sensitive analysis for stratified test to be performed using randomization strata from eCRF - In section 14.3, changed 'more than 28 days after' by "at least for 4 weeks apart' in the definition of Best Overall Response based on confirmed responses (for CR and PR responses) - In section 14.3.1, updated 'No baseline assessment' definition to consider only cases with no tumor scan recorded before randomization and removed 'No IRC review' category CCI - In section 14.6, updated 'No baseline assessment' definition to consider only cases with no tumor scan recorded before randomization and removed 'No IRC review' category - In section 15.1.1, replaced 'Treatment-related Anemia" by "Anemia" and added specific listings - In section 15.1.1, added chemotherapy related TEAEs bleeding events analysis - In sections 15.1, 15.2.3 and 8.1, clarified definition of Dose-limiting Toxicity (DLT) to considered evaluation from investigator and SMC review - In section 15.2.5.1, updated 'time related to first onset' to 'time related to first onset of each study treatment' - In section 15.2.5.2, updated definition of Immune-related adverse events - Update title of 15.2.5.3 from "Potential TGFβ-mediated skin adverse events" to "TGF-β inhibition mediated skin adverse events" - In section 15.2.5.3 added list of pre-specified PT for potential TGF-β inhibition mediated skin Aes, and

Unique Identifier for IAP Version	Date of IAP Version	Authors	Changes from the Previous Version
			changed the frequency table by PT to frequency table by worst grade - In section 15.2.5.4, added frequency table for anemia and treatment-related anemia - In section 15.3, updated Table 11 for NCI-CTC gradable parameters, Table 12 for non NCI-CTC parameters add added definition of corrected calcium, WBC differential counts for absolute value and estimated creatinine clearance - Updated section "Three-Tier Approach to Summarizing and Analyzing Aes" [redacted]
3	PPD	PPD	- In section 7.2, added description for COVID-19 vaccination [redacted] - Update IAP with additional expansion criterion introduced in protocol version 5.0
4	PPD	PPD	- Section 6: added "Overview of Planned Analyses after Trial Discontinuation" - SRI analysis set renamed as SAFRI across the document for clarity - The description of best overall response was partially changed to ORR to be consistent with the study protocol - Abbreviation BOR has been replaced by Best Overall Response as per guidance v.3.0 on objective response definition - Section 8.2: removed subgroups that will not be included anymore in the analyses for the abbreviated CSR due to data not available

Unique Identifier for IAP Version	Date of IAP Version	Authors	Changes from the Previous Version
			- Section 12.1: added description of listing on COVID-19 vaccination - Section 15.1.2: modified table of TEAEs leading to treatment discontinuation/interruption/modification adding treatment dose reductions for chemotherapy regimens and providing frequencies of TEAEs and treatment related TEAEs leading to infusion rate reductions of any treatment. - Section 15.2.5.4: added overview table of anemia events - Section 15.3: removed creatinine clearance from gradable parameters. Added Activated Partial Thromboplastin Time/standard thromboplastin time as gradable parameter. Removed creatinine clearance derivation as there is no need to derive this parameter at analysis level given that all the creatinine clearance units can be converted in this study, therefore there is no issue of missing creatinine clearance values.

4 Purpose of the Integrated Analysis Plan

The purpose of this integrated analysis plan (IAP) is to document technical and detailed specifications for the interim and primary analysis of data collected for protocol MS200647_0055. Especially, the purpose of this version 3 is to update the analysis plan as per protocol version 5.0.

Results of the analyses described in this IAP will be included in the Clinical Study Report (CSR). Additionally, the planned analyses identified in this IAP may be included in regulatory submissions or future manuscripts. Any post-hoc or unplanned analyses performed to provide results for inclusion in the CSR but not identified in this prospective IAP will be clearly identified in the CSR.

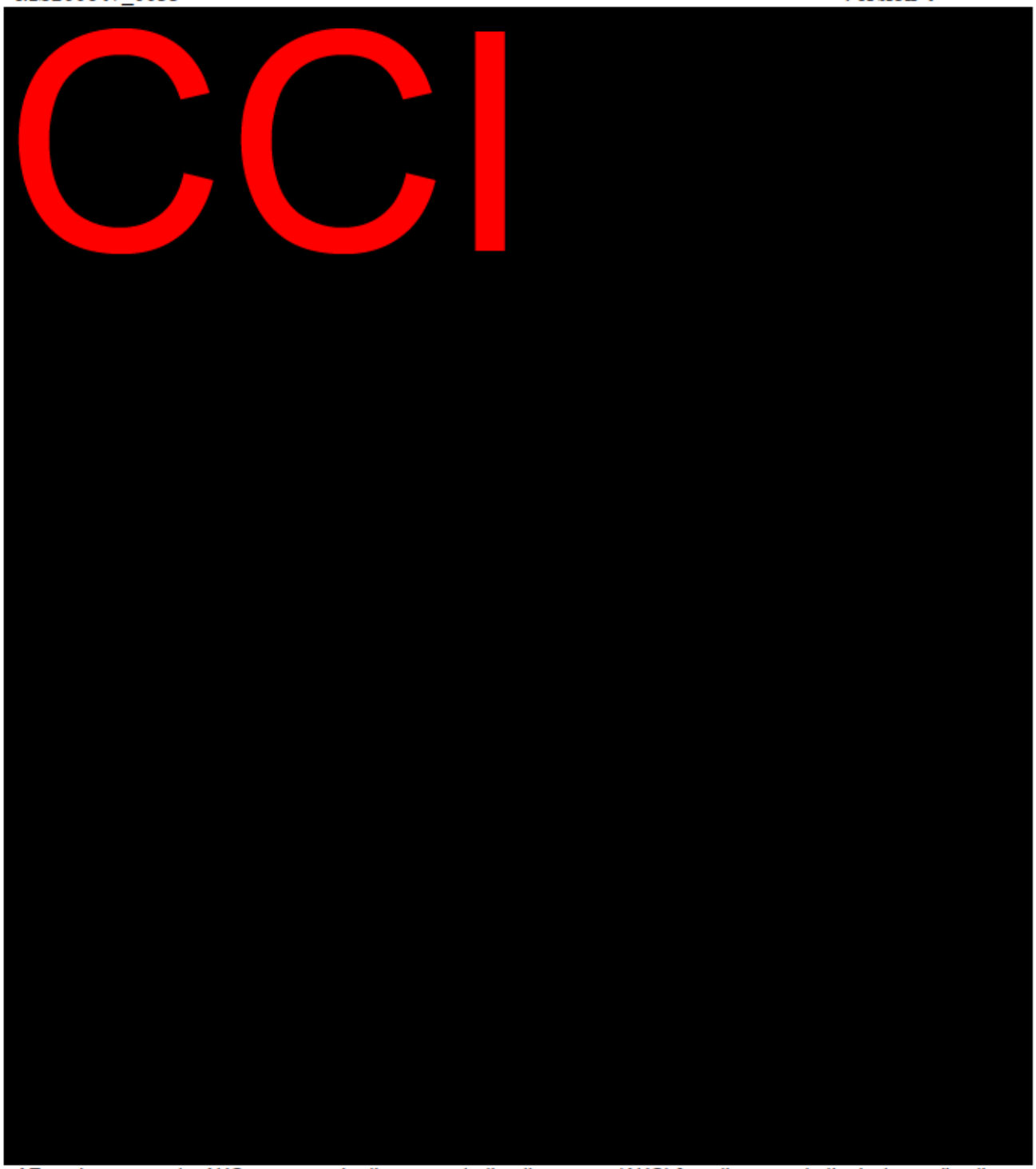
The IAP is based upon Section 9 (Statistical Considerations) of the study protocol and is prepared in compliance with International Council for Harmonization (ICH) Guideline E9. It describes analyses planned in the protocol. Details of the Independent Data Monitoring Committee (IDMC) analysis will be developed in a separate statistical analysis plan (SAP).

5 Objectives and Endpoints

Table 1 Study Objectives and Endpoints

Objectives	Endpoints (Outcome Measures)	IAP Section
Open-label, Safety Run-in		
To assess the following items with bintrafusp alfa 2400 mg Q3W in combination with gemcitabine and cisplatin in locally advanced or metastatic BTC		
Primary		

Objectives	Endpoints (Outcome Measures)	IAP Section
To assess if bintrafusp alfa 2400 mg Q3W is safe and tolerable and to confirm this dose as the recommended Phase II dose for the randomized, double-blind part of the study	Occurrence of DLTs during the DLT evaluation period	15.2.3
Secondary		
To assess the safety profile of bintrafusp alfa in combination with gemcitabine and cisplatin	• Occurrence of TEAEs and treatment-related AEs • Occurrence of abnormalities (Grade ≥ 3) in laboratory tests	15.1, 15.2, 15.3
Randomized, Double-blind Part		
To assess the following items with bintrafusp alfa in combination with gemcitabine plus cisplatin versus placebo with gemcitabine plus cisplatin in participants with advanced or metastatic BTC who have not received chemotherapy/immunotherapy in the advanced/metastatic setting		
Primary		
To assess OS	OS	14.1
Secondary		
To assess PFS	PFS according to RECIST 1.1 as assessed by the Investigator[a]	14.2
To assess ORR	Confirmed objective response according to RECIST 1.1 as assessed by the Investigator[a]	14.3
To assess DOR	DOR assessed by confirmed complete response or partial response until death or progression of disease according to RECIST 1.1 as assessed by the Investigator[a]	14.4
To assess DRR	Durable confirmed response of at least 6 months according to RECIST 1.1 as assessed by the Investigator[a]	14.5
To assess the safety profile of bintrafusp alfa or placebo in combination with gemcitabine plus cisplatin	Occurrence of TEAEs and treatment-related AEs, including adverse events of special interest	15.1
To characterize the PK profile of bintrafusp alfa	• PK profile of bintrafusp alfa in terms of C_{eoi} and C_{trough} for participants in the M7824 arm • PK profile of bintrafusp alfa in terms of AUC_{0-t}, $AUC_{0-\infty}$, C_{max}, t_{max}, and $t_{½}$ for participants in the safety run-in part of the study only	16.1
To evaluate the immunogenicity of bintrafusp alfa and to correlate it to exposure	Immunogenicity as measured by antidrug antibody (ADA) assays at baseline and on-treatment for participants in the bintrafusp alfa arm	16.4
CCI [redacted]		



AEs=adverse events, AUC_{0-t}=area under the concentration-time curve (AUC) from time zero to the last sampling time at which the concentration is at or above the lower limit of quantification, $AUC_{0-\infty}$=AUC from time zero extrapolated to infinity, based on the predicted value for the concentration at the last sampling time, BTC=biliary tract cancer, C_{eoi}=concentration observed immediately at the end of infusion, C_{max}=maximum observed concentration, C_{trough}=concentration observed immediately before next dosing, DLT=dose-limiting toxicity, DOR=duration of response, DRR=durable response rate, CCI IL=Item Library, IRC=Independent Review Committee, CCI, ORR=objective response rate, OS=overall survival, PFS=progression-free survival, PK=pharmacokinetics, CCI Q3W=once every 3 weeks, CCI RECIST

1.1=Response Evaluation Criteria in Solid Tumors Version 1.1, $t_{½}$=apparent terminal half-life, TEAEs=treatment-emergent adverse events, t_{max}=time to reach the maximum observed concentration collected during a dosing interval.

a If the study is not expanded to Phase III, tumor-based efficacy endpoints will also be assessed as per the IRC.

6 Overview of Planned Analyses

The study consists of an open-label, safety run-in part followed by a randomized, double-blind, placebo-controlled Phase II/III part:

- The multicenter, open-label safety run-in part will measure the safety and tolerability of first-line (1L) bintrafusp alfa in combination with gemcitabine and cisplatin. A total of 6 to 12 dosed participants will be enrolled in each of two separate regional cohorts (Asian sites and non-Asian sites).

The randomized, multicenter, double-blind, placebo-controlled Phase II/III part will evaluate whether 1L bintrafusp alfa in combination with the current standard of care (SoC) (gemcitabine plus cisplatin) improves overall survival (OS) compared to placebo, gemcitabine, and cisplatin. Up to 300 participants will be randomized and the analysis for the adaptation decision will be conducted in the first 150 randomized antibiotics naïve participants in Phase II. Thereafter, in the analysis for the adaptation decision, if the IDMC recommends expansion based on the pre-specified criteria, the study will be expanded into Phase III with a total sample size of 500 participants. If the study is not expanded to Phase III, it will be completed as a Phase II study (see Appendix 4).

Randomization will be stratified according to the following factors:

- Type of biliary tract cancer (BTC) (based on 3 anatomical locations, i.e., intrahepatic cholangiocarcinoma (CCA), extrahepatic CCA including ampulla of Vater's cancer, and gallbladder cancer)
- Metastatic at diagnosis vs others, where "others" includes participants with unresectable, locally advanced disease at diagnosis and participants with resectable disease at diagnosis who have undergone prior surgical resection with curative intent
- Asia sites versus non-Asia sites

Statistical analyses will be performed on the basis of CDISC Study Data Tabulation Model (SDTM) data, used as source for CDISC Analysis Data Mode (ADaM) data creation. These SDTM data contain as clean as possible eCRF data, as well as external data including laboratory data, biomarker data, electrocardiogram data, and tumor assessment by the Independent Review Committee (IRC).

Except for the open-label safety run-in analyses, all data will be included up to a clinical cut-off date which is determined by the number of events required for the analysis. The clinical cut-off date will be the date on which approximately the number of events is expected based on the event projections. The final number of events might deviate from the planned number, e.g. due to cleaning activities. The data cut-off date will not be adjusted retrospectively in this case.

Table 2 and Table 3 display an overview of the analyses to be provided for this study.

Table 2 **Overview of Analyses**

Part	Analysis	Data Cut-Off Point
Open-label safety run-in	**Safety/Efficacy analysis**	When the last participant of the open-label part of the study has reached a minimum follow-up time of 12 months, The analysis may be conducted at time of phase II PA or phase III IA
Randomized, Phase II	**Interim analysis adaptation**	When 80 PFS events per IRC have occurred and the first 150 randomized antibiotics naïve participants have at least 19 weeks of follow-up
Randomized Phase II	**Primary analysis***	When 60% of total Phase II participants have meet OS events
Randomized Phase III	**Interim analysis**	When 210 OS events have occurred
Randomized, Phase III	**Primary analysis**	When 334 OS events have occurred

* Only if study is not expanded into Phase III with N=500.

PFS=progression-free survival; OS=overall survival.

Table 3 **Overview of Endpoints for Each Analysis**

Endpoints	Open-label Safety Run-in	Phase II		Phase III	
	Safety/Efficacy Analysis	Interim Analysis Adaptation	Primary Analysis[a]	Interim Analysis	Primary Analysis
Disposition	✓	✓	✓	✓	✓
Protocol deviations	✓	✓	✓	✓	✓
Randomization details		✓	✓	✓	✓
Demographics and others baseline characteristics	✓	✓	✓	✓	✓
Medical history	✓		✓	✓	✓
Previous and concomitant medications	✓		✓	✓	✓
Previous anticancer treatment	✓	✓	✓	✓	✓
Subsequent anticancer treatment	✓	✓	✓	✓	✓
Treatment exposure and compliance	✓	✓	✓	✓	✓
OS	✓[b]	✓	✓	✓	✓
PFS RECIST 1.1 per IRC		✓	✓		
PFS RECIST 1.1 per Investigator	✓[b]		✓	✓	✓
CCI [redacted]					
ORR RECIST 1.1 per IRC		✓	✓		
ORR RECIST 1.1 per Investigator	✓[b]	✓	✓	✓	✓
ORR2L			✓	✓	✓

Endpoints	Open-label Safety Run-in	Phase II		Phase III	
	Safety/Efficacy Analysis	Interim Analysis Adaptation	Primary Analysis[a]	Interim Analysis	Primary Analysis
CCI					
DOR RECIST 1.1 per IRC		✓[b]	✓	✓	✓
DOR RECIST 1.1 per Investigator			✓	✓	✓
CCI					
DRR			✓	✓	✓
CCI					
Follow-up time PFS		✓	✓	✓	✓
Follow-up time OS			✓	✓	✓
Safety (TEAE, SAE, AESI…)	✓	✓	✓	✓	✓
3-Tier approach				✓	✓
Laboratory evaluation	✓	✓	✓	✓	✓
Vital signs	✓		✓	✓	✓
Other safety (ECG/ECOG)	✓		✓	✓	✓
PK	✓		✓	✓	✓
CCI					

AESI=adverse events of special interest; ORR=objective response rate; ORR2L=objective response rate on subsequent line of anticancer treatment; DOR=duration of response; DRR=durable response; ECG=electrocardiograms; IA=interim analysis; CCI IRC=independent review committee; CCI progression-free survival; PA=primary analysis; PFS=progression-free survival; PK=pharmacokinetics; CCI; OS=overall survival; RECIST 1.1=response evaluation criteria in solid tumors version 1.1; TEAE= TEAE=treatment-emergent adverse events; SAE=serious adverse events.

[a] Analysis will be performed only if study is not expanded into Phase III.

[b] Due to the small number of participants or events expected, analysis of these endpoints will be based on figures and listings as described in the corresponding sections below. The list of tables, listings and figures to be provided at each analysis is described in a separate document.

6.1 Safety Monitoring Committee

A Safety Monitoring Committee (SMC) will be responsible for safety evaluations of the open label safety run-in part. Details can be found in the SMC Charter.

The SMC will evaluate safety based on the dose-limiting toxicities (DLTs) observed in each of the Asian and non-Asian cohorts during the safety run-in part of the study once the 6th evaluable participant of each cohort has completed the DLT observation period. Once all participants in a cohort have completed the DLT period or discontinued from study prematurely, a data snapshot will be taken for provision of SMC outputs. There will be no data cut-off applied.

Data results for SMC will be provided using patient profiles. No Statistical Analysis Plan will be written.

6.2 Safety and Efficacy Analysis for Open-label Safety Run-in Part

Analysis of safety and efficacy endpoints will be performed for the open-label safety run-in part based on Table 3 above and triggered by the data cut-off point displayed in Table 2.

Due to the small number of participants expected in the open-label safety run-in part, efficacy data will not be summarized by time-to-event analyses. Instead, data will be presented through graphs (swimmer, waterfall and spider plots) with the relevant information.

Note that the following data will not be analyzed at the time of this safety/efficacy analysis. However, a pooled analysis together with the randomized participants will be performed.

CCI

Note that some endpoints (e.g.: protocol deviations, medical history, etc) may only be detailed via listings.

6.3 Independent Data Monitoring Committee

To ensure participants' safety, an Independent Data Monitoring Committee (IDMC) will be established for periodic review of safety data throughout the randomized double-blind part of the study. Both data from the safety run-in part and the randomized part will be summarized. Results from safety run-in part will be presented overall and will not be split by Asian/non-Asian cohort.

In addition, the IDMC will review efficacy and safety data from the first 150 participants of the randomized, double-blind part of the study during interim analysis (IA) for adaptation to Phase III sample size extension (see Section 6.4). The IDMC will recommend to the study team whether to continue the study as a Phase II study only or to expand into Phase III.

For each IDMC meeting, the independent statistician will prepare the outputs (using programs prepared by the blinded team based on dummy treatment arms) in agreement with the IDMC Charter and transmit the analyses, tabulations, and listings to the IDMC only for the meeting. The independent statistician will be available at the IDMC meeting should any questions from the IDMC members arise regarding the data and/or analyses.

The full membership, mandate, and processes of the IDMC will be detailed in the IDMC Charter.

A specific statistical analysis plan will be prepared for the IDMC.

6.4 Interim Analysis for Adaptation Decision

The analysis for adaptation decision will be conducted in the first 150 randomized antibiotics naive participants when 80 PFS events have occurred and once at least 19 weeks of follow-up for the first 150 randomized antibiotics naive participants is reached.

In case multiple participants are randomized on the same day as the 150th participant, the analysis for adaptation decision will include all antibiotics naive participants randomized up to this day.

Both data from the safety run-in part and the randomized part will be summarized. Results from safety run-in part will be presented overall, and will not be split by Asian/non-Asian cohort.

Efficacy analysis will be performed on the first 150 randomized antibiotics naïve participants, all others analysis including safety will be performed on all participants randomized till cut-off date.

This IA will trigger the decision to expand the study into Phase III if the following criteria are met:

- Odds ratio of a confirmed ORR ($\mathbf{OR}_{ORR}$) is ≥ 1.6 or PFS hazard ratio ($\mathbf{HR}_{PFS}$) is < 0.75, and,
- the probability of success exceeds a prespecified threshold (success is defined as projected OS HR below a prespecified threshold at the primary analysis). Details will be specified in an appendix to the IDMC charter and IDMC statistical analysis plan.

Others analysis will be performed as summarized in Table 3. Data relative to participants' duration of response will not be summarized in a table but displayed through a swimmer plot and a spider plot.

6.5 Interim and Primary Analysis

For Phase II PA (if applicable), and Phase III IA and PA (if applicable), analyses will be provided as summarized in Table 3, triggered by the data cut-off point displayed in Table 2.

7 Changes to the Planned Analyses in the Clinical Study Protocol

7.1 Overview of Planned Analyses after Trial Discontinuation

The external IDMC reviewed the data concerning 12 participants in safety run-in and 297 randomized participants in the Phase II part of study for adaptation decision, with a data cut-off date of 20th May 2021. Based on the totality of the efficacy and safety data available, the IDMC did not support the continuation of the clinical trial. Based on this recommendation, the Sponsor has made the decision to discontinue this clinical trial.

Therefore, the analyses described in Section 6 will not be performed as planned, but one primary analysis will be performed for an abbreviated CSR. The analyses considered relevant for this purpose are listed below:

- Participant disposition

- Important protocol deviations
- Enrollment details (Number or randomized by region and country, number or randomized by stratification factor at randomization and difference between stratification factor collected in e-CRF vs IXRS)
- Demographics, medical history and other baseline characteristics (disease history, ECOG performance status, height, weight, BSA and BMI at baseline, skin status history)
- Previous and concomitant medications, procedures, follow-up treatments
- Treatment exposure and compliance (e.g. duration of treatments, number of infusions/fractions, cumulative dose, chemotherapy dose reductions, therapy delays, infusion rate reductions, infusion temporary interruption)
- Efficacy analyses:
    - Overall survival (OS)
        - Overall Survival – Forest plot of hazard ratio by subgroups
    - PFS per RECIST 1.1 as assessed by IRC
        - PFS per RECIST 1.1 as assessed by IRC - Forest plot of hazard ratio by subgroups
    - ORR per RECIST 1.1 as assessed by IRC
        - Cross tabulation Best Overall Response per RECIST 1.1 as assessed by IRC vs Investigator
        - Tumor shrinkage in target lesions
        - ORR per RECIST 1.1 as assessed by IRC - Forest plot of odds ratio by subgroups
    - Follow-up duration (PFS and OS)
    - Duration of response
- Safety analyses (overview of treatment-emergent adverse events (TEAEs), serious TEAEs (SAE), treatment-related TEAEs, TEAEs leading to treatment discontinuation/interruption/modification, TEAEs Leading to death, adverse events of special interest (AESI), TEAEs associated to COVID-19, vital signs, ECOG, clinical laboratory evaluations)

Data from the cutoff of 20th May 2021 were the basis for the decision to discontinue the trial and will be used for the primary analysis for abbreviated CSR. Efficacy analysis will be performed on

the first 150 randomized antibiotics naïve participants, all others analyses including safety will be performed on all participants randomized till cut-off date.

In case a new safety signal is observed, an additional analysis will be done at the end of study and included in CSR addendum. In the other cases, no further analyses will be planned.

7.2 COVID-19 Impact

No changes to the planned analysis of the efficacy endpoints will be performed due to the impact of COVID-19 outbreak.

Instead, additional outputs (summary tables and listings) will be generated to assess potential impacts of COVID-19 in this study, including:

- Overview table of the impact by COVID-19
- Listing of participants affected by COVID-19
- Table of indirect COVID-19 impact
- Table of COVID-19 related protocol deviations
- Listing of COVID-19 related protocol deviations
- Table of treatment-emergent adverse events (TEAEs) associated to COVID-19
- Listing of AEs related to COVID-19
- Listings of vaccines for COVID-19

Participants will be categorized based on the COVID-19 study period defined as:

- Pre-pandemic time period: participants who started the treatment before the start of COVID-19 pandemic.
- During-pandemic time period: participants who started the treatment on the same date or after the start of COVID-19 pandemic
- Post-pandemic time period: the end of pandemic is considered not yet reached anywhere and will be defined later

The start of COVID-19 pandemic is defined by country as the earliest date of either the date of the first death from COVID-19 occurred in each country according to the published data by European Centre for Disease Prevention and Control on 26th June 2020 (https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide) or 11th March 2020 (when the WHO declared COVID 19 pandemic).

Details of the analyses are provided in Sections 10.1, 10.2, 15.1.1.

8 Analysis Populations and Subgroups

8.1 Definition of Analysis Populations

Screening Analysis Set

The screening analysis set includes all participants, who signed the informed consent, regardless of the participant's randomization and study intervention status in the study.

Safety Run-In Analysis Set

The safety run-in (SAFRI) analysis set includes all participants from the safety run-in part who were administered any dose of any study intervention.

Dose-limiting Toxicity Analysis Set

The dose-limiting toxicity (DLT) analysis set includes all participants who experienced at least one DLT (either by Investigator or by SMC) or who completed the safety run-in, i.e., the 21-day DLT evaluation period, receiving at least one infusion of bintrafusp alfa and of both gemcitabine and cisplatin and not being withdrawn during the DLT evaluation period for reasons other than toxicity.

Intent-to-Treat Analysis Set

The intent-to-treat (ITT) analysis set includes all randomized participants. Participants will be analyzed according to the treatment assigned at randomization as per the intent-to-treat principle.

The ITT analysis set will be used for all analyses of demographics and baseline characteristics, efficacy, PROs, and biomarkers.

Safety Analysis Set

The safety (SAF) analysis set includes all participants from the ITT analysis set who received at least one dose of study treatment.

Participants will be classified according to the treatment assigned at randomization unless the incorrect treatment(s) was/were received throughout the dosing period in which case participants will be classified according to the first study treatment received.

Pharmacokinetic Analysis Set

The pharmacokinetic (PK) analysis set (PKAS) includes all participants (from safety run-in part and randomized part) who received at least one dose of bintrafusp alfa, and who provided at least one sample with a measurable concentration of bintrafusp alfa after start of treatment, without important protocol deviations or events deemed to affect PK evaluation.

All PK analyses will be based on this analysis set. Refer to Section 16.1 for protocol deviations and handling relevant to PK.

CCI

CCI

Table 4 summarizes the use of the analysis sets in the different analyses.

Table 4 Overview of the Analysis Sets Used in the Analyses

Analyses	Screening	SAFRI	DLT	ITT	SAF
Disposition	✓				
Protocol deviation		✓		✓	
Demographics		✓		✓	
Baseline assessments		✓		✓	
Previous and concomitant therapies		✓		✓	
Compliance and exposure		✓			✓
Efficacy		✓		✓	
Safety and tolerability: dose-limiting toxicity		✓	✓		
Safety and tolerability: other endpoints					✓
Pharmacokinetics					

CCI

CCI

DLT=dose-limiting toxicity; CCI ITT=intention-to-treat; PK=pharmacokinetic; SAF=safety; SAFRI=Safety Run-in.

CCI

[redacted]

9 General Specifications for Data Analyses

This section describes any general specifications not included in subsequent sections.

The specifications for PK data handling/analysis are presented in Section 16.1. Unless otherwise indicated, all general (i.e., disposition of participants and discontinuations, protocol deviations, demographics and other baseline characteristics, and previous or concomitant medication), safety and efficacy analyses will be presented separately for the two treatment arms (bintrafusp alfa and placebo, short for the combinations of bintrafusp alfa + chemotherapy and placebo + chemotherapy) in the randomized, double-blind part and per cohort (Asian/non-Asian) for the safety run-in part.

Throughout the document, the term "cohort" (Asian, Non-Asian) will be used for the safety run in whereas the term "arm" (bintrafusp alfa, placebo) will be used for the randomized, double-blind part.

Unless otherwise specified, comparisons will be presented for bintrafusp alfa versus placebo.

9.1 Data Handling After Cut-off Date

By its nature, data after cut-off may be incomplete and subject to further change and will not be used for summary statistics, statistical analyses, listings or imputations.

Stop dates and outcomes of adverse events are not affected by this rule, e.g., a stop date of an AE and its related outcome which starts prior to the cut-off but stops after date of cut-off, will not be changed.

9.2 Definition of Baseline and Change from Baseline

In the safety run-in part, the last non-missing measurement prior to the first study treatment administration will be used as the "Baseline" measurement. If an assessment that is planned to be performed before treatment per protocol is performed on the same day as the start of treatment, respectively, but the assessment time is not available, it will be assumed that it was performed prior and will be considered as baseline.

In the randomized part, the last non-missing measurement prior to randomization or prior to the start of study treatment will serve as the baseline measurement, depending of the type of analysis. Details are given in the table below:

Table 5 Baseline Date to be Used in the Analyses for Randomization Part

Analysis	Date of start of study treatment	Date of randomization[a]
Disposition		✓
Protocol deviations		✓
Demographics and others baseline characteristics		✓
Medical history		✓
Previous and concomitant medications, Previous/ Subsequent anticancer therapy	✓	
Treatment exposure and compliance	✓	
Efficacy (OS, PFS, ORR…)		✓
Safety (TEAE, SAE, AESI…)	✓	
Laboratory evaluation	✓	
Vital signs	✓	
Other safety (ECG/ECOG)	✓	
PK	✓	

Analysis	Date of start of study treatment	Date of randomization[a]
CCI		

[a] If the last non-missing measurement prior to randomization is not available, the last measurement prior to the first study treatment administration will be used with the exception of pre-randomization assessments used for the derivation of efficacy endpoints (e.g. tumor assessment at baseline, which will be set to missing, if not done prior to randomization).

For the purposes of the description of baseline characteristics on the ITT Analysis Set, the baseline value of all safety parameters such as vital signs will be defined as the last non-missing measurement prior to randomization. Similarly, subgroup classification based on baseline values will be derived using the last non-missing measurement prior to randomization.

If an assessment that is planned to be performed before randomization, or treatment per protocol is performed on the same day as the randomization or start of treatment, respectively, but the assessment time is not available, it will be assumed that it was performed prior and will be considered for derivation of baseline. If a scheduled pre-dose measurement actually occurred post-dose, then the corresponding measurement will be analyzed similar to an unscheduled post-dose measurement.

Absolute and percent changes from baseline are defined as:

- Absolute change = timepoint value – baseline value
- Percent change = 100 * (timepoint value – baseline value) / baseline value

9.3 Study Day / Study Treatment Day

In the safety run-in part, Day 1 is the day of start of study treatment, the day before is Day -1 (no Day 0 is defined).

In the randomized double-blind part, Day 1 is either the day of randomization or the day of start of study treatment depending of the analysis, as detailed in Table 5 5 of this document.

9.4 Definition of Duration and 'time since' Variables

Duration in days will be calculated by the difference of start and stop date + 1 if not otherwise specified. For example, duration of response (days) = date of PD/death/censoring – date of response + 1.

The time since an event (e.g. time since initial cancer diagnosis) will be calculated as reference date minus date of event. In general, the reference date will be study day 1.

The time to an event will be calculated by the difference between the time of event and the reference date + 1 if not otherwise specified. For example, survival time (days) = date of death - study day 1 + 1.

9.5 Conversion Factors

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

9.6 Date of Last Contact

The date of last contact will be derived for participants who are not known to have died at the analysis cut-off using the latest complete date prior to or at the data cut-off date from the following:

- All participant assessment dates (blood draws [laboratory, PK], vital signs, ECOG performance status, ECG, tumor assessments, quality of life assessments)
- Start and end dates of anticancer therapies administered after discontinuation of study treatment
- AE start and end dates
- Last known alive date in "Subject Status / Survival Follow-Up" eCRF page (do not use the follow-up date)
- Study treatment start and end dates
- Randomization date
- Date of discontinuation taken from the "Study Termination" eCRF page (do not use if reason for discontinuation is lost to follow-up or death)

Only dates associated with actual examinations of the participant reported in the eCRF 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 will not be used.

9.7 Time Window

Not applicable

9.8 Definition of On-treatment Period

The on-treatment period is defined as the time from the date of first study treatment administration to the date of last study treatment administration + 30 days or the cut-off date or death or the earliest date of subsequent anticancer therapies minus 1 day, whichever occurs first, unless otherwise stated. For participants with treatment ongoing at the cut-off date, all data from the date of first study treatment administration up to the cut-off date will be considered under the on-treatment period.

For immune-related AEs as listed in Section 15.2.5.2, an expanded on-treatment period will be used as a default for any analysis: Time from the first study intervention to the last study

intervention date + 90 days, death OR to the earliest date of subsequent anticancer therapy minus 1 day, whichever occurs first, unless otherwise stated.

Any systemic anticancer therapy, any anticancer surgery and any anticancer radiotherapy as documented in the "Anti-cancer treatment after discontinuation" eCRF page will be considered as subsequent anticancer therapy (palliative bone-directed radiotherapy and urgent palliative surgeries, diagnostic biopsies and stenting/percutaneous transhepatic biliary drainage for the purpose of releasing biliary tract obstruction occurring during the on-treatment period will be documented in the "Concurrent procedures" eCRF page).

9.9 Imputation of Missing Data

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

Missing statistics, e.g. when they cannot be calculated, will be presented as "nd" for "not determined. For example, if n=1, the measure of variability [e.g. StD] cannot be computed and will be presented as "nd".

Partial dates, which are not to be imputed according to the IAP, will be presented in the format like "_____YYYY". If values are imputed according to the IAP, imputed values will be presented in participant data listings and imputed information will be flagged.

Incomplete dates will be handled as specified in the following table:

Table 6 Imputation Rules for Incomplete Dates

Age Calculation	Incomplete dates (date of informed consent, date of birth) for the calculation of age will be imputed as follows: • In case of missing day for at least one date, but month and year available for both dates: the day of informed consent and the day of birth will be set to 1. • In case of missing month for at least one date, but year available for both dates, the day and the month of informed consent and the day and month of birth will be set to 1. • In all other cases, the incomplete dates will not be imputed.
Disease history	Incomplete dates for disease history (e.g. initial diagnosis date, date of documented, locally advanced, inoperable or metastatic disease diagnosis, date of response or progression in prior treatment) will be imputed as follows: • If the day is missing, it will be imputed to the 1st day of the month. • If both day and month are missing, the month and day will be imputed as January 1st • If the date is completely missing, no imputation will be performed.
Adverse events	Incomplete AE-related dates will be imputed as follows: • In case the onset date is missing completely or missing partially but the onset month and year, or the onset year are equal to the start of study treatment then the onset date will be imputed by the minimum of start of study treatment and AE resolution date (if not missing).

	• In all other cases, the missing onset day or missing onset month will be imputed by 1. • Incomplete stop dates will be imputed by the last day of the month (if day is missing only), if not resulting in a date later than the date of the participant's death. In the latter case, the date of death will be used to impute the incomplete stop date. • In all other cases, the incomplete stop date will not be imputed.
Previous and concomitant medication	Incomplete dates for previous and concomitant medications will be imputed as follows: • For start date of medication o If the day is missing, it will be imputed to the 1st day of the month. o If both day and month are missing, the month and day will be imputed as January 1st. o If the date is completely missing, no imputation will be performed. • For end date medication: o If the day is missing, it will be imputed to the last day of the month. o If both day and month are missing, the month and day will be imputed as December 31st o If the date is completely missing, no imputation will be performed. Note: In case the imputation results in a date later than the date of the participant's death, then the date of death will be used to impute the incomplete stop date.
Subsequent Anticancer Therapy	Incomplete dates for the start date of subsequent anticancer therapy (drug therapy, radiotherapy, surgery) will be imputed as follows and will be used for determining censoring dates for efficacy analyses and in the derivation of the end of the on-treatment period: • If only day is missing, it will be imputed as the last day of the month unless the end date of subsequent anticancer therapy is before that date. In that case, the incomplete anticancer therapy start date will be imputed as the end date of the anticancer therapy. • If both day and month are missing, no imputation will be performed. Incomplete subsequent anticancer therapy stop dates will not be imputed.
Death	For the purposes of survival analyses, partially missing death dates will be imputed as follows: • If only the day is missing, the death date will be imputed to the maximum of the (non-imputed) day after the date of last known alive date and the 15th day of the month • Otherwise it will not be imputed

Pharmacokinetic parameters will be calculated using the actual elapsed time since dosing. When the actual sampling time is missing, calculations will be performed using the scheduled time. Otherwise, there will be no further imputation of missing data. Concentrations will be set to missing in summary tables if the value is reported as no result.

9.10 Scoring of HRQOL Data

Unless otherwise specified, HRQOL questionnaires will be scored using their published administration and scoring manual. For items with missing responses, the response will be managed as per the scoring manual. See Section 16.3 for details.

9.11 Presentation of Continuous and Qualitative Variables

Continuous (non-PK) variables will be summarized using descriptive statistics i.e.,

- Number of participants (N), number of participants with missing values
- Mean, standard deviation (StD)
- Median, 25th percentile - 75th percentile (Q1-Q3)
- Minimum (Min), and maximum (Max)

If there are no missing values, the number of participants with missing values should be set to 0.

Mean, Median, Q1, Q3, Min, Max have the same precision as SDTM data (number of digits) for non-derived data. Statistics on derived data will be rounded to reasonable digits, whereas maximal digits will be available in ADaM datasets. SD is to be presented with one digit more than mean.

If not otherwise specified, p-value will be presented using the SAS format "pvalue6.4" using 4 decimals. HR and odds ratio will be displayed with 2 decimals.

Qualitative variables will be summarized by counts and percentages. Unless otherwise stated, the calculation of proportions will be based on the number of participants in the analysis set of interest. Therefore, counts of missing observations will be included in the denominator and presented as a separate category.

Percentages will be reported to one decimal place.

Descriptive statistics by nominal visit or time point, e.g. for laboratory measurements, will include only data from scheduled visits. Unscheduled visits will be included in the derivation of baseline or worst on-treatment values.

9.12 Significance Level

The overall significance level is 2.5% one-sided. Statistical tests are described in Section 14.

CCI [redacted]

9.13 Pooling of Centers

In order to provide overall estimates of treatment effects, data will be pooled across sites. The "site" factor will not be considered in statistical models or for subgroup analyses due to the high

number of participating sites in contrast to the anticipated small number of participants at each site.

9.14 Unscheduled Assessments

As per database definition, unscheduled safety assessments are always linked to a scheduled timepoint (each unscheduled assessment is linked to the previous scheduled timepoint). Safety data retrieved from an unscheduled timepoint (vital signs, electrocardiogram [ECG] and laboratory data) will be analyzed according to the following scenario:

- For shift tables, they will be taken into account in the definition of the worst assessment during study
- For description at each timepoint post-baseline, the first available result (in chronological order) per timepoint will be taken into account in the analysis in case of multiple values
- For description at baseline, the last available result before the first administration of study treatment will be taken into account in the analysis in case of multiple values (see Section 9.2)

For immunogenicity analysis, unscheduled visits will also be taken into account in the analysis following the same rules as detailed above.

For PK analysis, unscheduled samples will not be linked to a scheduled timepoint and will be excluded from summaries.

9.15 Study Treatment

In this study, bintrafusp alfa, placebo, and chemotherapies are considered as study treatments. The date of first study treatment administration will be defined as the earliest administration date of any treatments (bintrafusp alfa, placebo or chemotherapies). The date of last study treatment administration will be defined as the latest administration date of any treatments (bintrafusp alfa, placebo or chemotherapies).

9.16 Preferred Term for analysis of World Health Organization's Drug Dictionary (WHO-DD) coded data

For data coded according to WHO Drug B3 (e.g., concomitant medications), summaries will be done on the preferred term (PT) level where the PT corresponds to codes ending in 01001. With this approach, variations of salt forms of active ingredients will be analyzed under the same term, e.g. diphenhydramine and diphenhydramine hydrochloride will be analyzed as the same PT, diphenhydramine.

CCI [redacted]

9.18 Re-screened Participants

Re-screened participants will be only counted once in the screening analysis set, considering the latest screening (screening with latest informed consent).

9.19 Software

All statistical analyses will be performed using statistical software SAS® (Statistical Analysis System, SAS-Institute, Cary NC, USA, Windows Version 9.4 or higher) in the SAS Grid environment.

The computer programPPD Version 8.0, or higher PPD USA) will be used to derive PK parameters applying non-compartmental analysis.

10 Study Participants

The subsections in this section include specifications for reporting participant disposition and study treatment/study discontinuations. Additionally, procedures for reporting protocol deviations are provided.

10.1 Disposition of Participants and Discontinuations

Descriptive statistics will be used to summarize participant disposition based on the eCRF data. All participants within the screening analysis set will be considered.

The number and percentage of participants in each of the following disposition categories will be presented by cohort/arm and total. Percentages will be calculated based on the number of participants in the SAFRI / ITT analysis set, except for the number of screened and re screened participants where no percentage will be provided:

- Total number of participants screened (i.e., all participants who gave informed consent)
- Number of participants who discontinued from the study prior to randomization (or prior to first study treatment administration for safety run-in part) overall and grouped by the main reason (e.g. the specific failed inclusion or exclusion criteria, adverse event, lost to follow-up, death, progressive disease, withdrawal of consent and other)
- Number of re-screened participants
- Number of participants who received at least one dose of study treatment (bintrafusp alfa and/or chemotherapies) overall and by cohort for safety run-in part
- Number of randomized participants overall and by treatment arm for randomized, double-blind part

The end of treatment status will be summarized by:

- Number of randomized participants who did not receive any study treatment by treatment arm.

- Number of participants with bintrafusp alfa/placebo) ongoing at the data cut-off date overall and by cohort/treatment arm
- Number of participants with cisplatin ongoing at the data cut-off date overall and by cohort/treatment arm
- Number of participants with gemcitabine ongoing at the data cut-off date overall and by cohort/treatment arm
- Number of participants off bintrafusp alfa/placebo, grouped by main reason (treatment completed as per protocol, progressive disease (PD), death, adverse event, lost to follow-up, protocol non-compliance, withdrew consent, other)
- Number of participants off cisplatin, grouped by main reason (treatment completed as per protocol, progressive disease, death, adverse event, lost to follow-up, protocol non-compliance, withdrew consent, other)
- Number of participants off gemcitabine, grouped by main reason (treatment completed as per protocol, progressive disease, death, adverse event, lost to follow-up, protocol non-compliance, withdrew consent, other)

The end of study status will be summarized by:

- Number of participants with bintrafusp alfa/placebo ongoing at the end of study overall and by cohort/treatment arm
- Number of participants with cisplatin ongoing at the end of study overall and by cohort/treatment arm
- Number of participants with gemcitabine ongoing at the end of study overall and by cohort/treatment arm
- Number of participants who completed/discontinued study participation, with the associated main reason (study completed according to protocol, adverse event, lost to follow-up, protocol non-compliance, death, withdrew consent, other)

In addition, the number of participants screened and randomized in each analysis set as defined in Section 8.11 will be provided overall, by region, by country within region, and by site.

For the randomized, double-blind part, the results of the randomization algorithm (according to IXRS) will be summarized as follows:

- Number and percentage of randomized participants overall, by region, by country within region
- Number and percentage of randomized participants by randomization strata (IXRS)
- Number and percentage of randomized participants by randomization strata (CRF)
- Listing of participants for whom stratum by IXRS is different from stratum by CRF
- Cross tabulation: participants randomized vs. treated and corresponding listing

The listing of participant disposition will include all participants (i.e., including screening failures, but not re-screened participants (at their screen failure time) who will be listed in a separate listing). The listing will include the following information: cohort/planned arm (participants having actual arm different from planned arm will be flagged and displayed in a separate listing), participant identifier, date of informed consent, participant continue beyond screening yes/no (if not reason for not to continue), date of randomization (in randomized part), date of first/last study treatment (for bintrafusp alfa/placebo, cisplatin, and gemcitabine), date and reason off-treatment, date and reason off-study, population flags. When the reason such as reason off-treatment is categorized as "Other, specify" or "Withdrew consent from treatment, specify", the verbatim text as entered in the eCRF will be presented in the listing.

If any re-screened participants are observed, they will be presented in a specific listing that will include: cohort/planned arm, participant identifier (identifier at inclusion in the study), date of informed consent at inclusion, date of first study treatment, initial participant identifier (identifier at screen failure), date of informed consent at screen failure, date and reason of screen failure. Note if participants are screened several times, all screening attempts will be listed.

For the IDMC meetings, Kaplan-Meier's plots of time to study treatments discontinuation and time to study discontinuation will be provided (if more than 10 events will be observed).

In addition, for the assessment of COVID-19 impact on this study, an overview table will be presented by treatment arm with the following information:

- Participants potentially affected by COVID-19 (i.e., participants who started treatment after start of the COVID-19 pandemic, or who started treatment prior to start of the COVID-19 pandemic and are still ongoing after the start of the pandemic)
- Participants with at least one COVID-19 impact
- Participants with at least one COVID-19 impact in the following categories: adverse events, death, protocol deviations, missed drug administration, treatment administration modifications, missed tumor assessments, missed visits, tele visits replacing on-site visits, treatment discontinuation, study discontinuation
- Number of participants with missed tumor assessments, delayed tumor assessments, missed visits, tele-visits replacing on-site visits (1 / 2 / 3 / >3)

The frequency distribution of participants who started the treatment before, during or after the pandemic per country-specific start of COVID-19 study period (see Section 7.2 for details on the COVID-19 categorization by study period) will be also displayed.

Listings of COVID-19 impact will be provided:

- Participants affected by COVID-19 with following information: cohort/planned arm, participant identifier, first and last date of treatment administration, date of the event, visit, category, event, event description/reason

10.2 Protocol Deviations

Analysis Set: SAFRI / ITT

10.2.1 Important Protocol Deviations

Important protocol deviations (IPDs) are protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a participant's rights, safety, or well-being.

IPDs will be defined in a specific document (see Appendix 1). They could include, but are not limited to:

- Participants who are dosed on the study despite not satisfying the enrolment criteria
- Participants who develop withdrawal criteria whilst on the study but are not withdrawn
- Participants who receive the wrong study treatment or an incorrect dose
- Participants who receive an excluded concomitant medication
- Failure to collect data necessary to interpret primary endpoints
- Failure to collect necessary key safety data
- Deviation from Good Clinical Practice
- Any other protocol deviation that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a participant's rights, safety, or well-being

IPDs will be identified for all participants by either site monitoring, medical review processes or programming and confirmed prior to or at the Data Review Meeting at the latest.

All IPDs will be documented in SDTM datasets whether identified through site monitoring, medical review or programming.

Additional clinically important protocol deviations may be identified during the course of the study but will not require amendment to this IAP.

IPDs will be summarized for

- Deviations from the inclusion and exclusion criteria
- Deviations post-inclusion

The following summary table and listing will be provided:

- Frequency table per reason of IPDs, separated for pre-/post inclusion deviations
- Listing of all protocol deviations (including COVID-19 minor PD) which will include: cohort/planned arm, participant identifier, category of the deviation (e.g. inclusion/exclusion), a description of the deviation, if important or not (yes/no). The listing will be enriched by the variable "COVID-19 Related Protocol Deviations" flagging all protocol deviations due to COVID-19.

CCI [redacted]

10.2.2 Reasons Leading to the Exclusion from an Analysis Population

A frequency table per reason of exclusion from the DLT and PK analysis sets as well as a listing will be provided. A listing of participants included in the ITT analysis set but excluded from the SAF analysis set will also be provided.

11 Demographics and Other Baseline Characteristics

Analysis Set: SAFRI / ITT

11.1 Demographics

Demographic characteristics and physical measurements will be summarized descriptively using the following information from the Screening/Baseline Visit eCRF pages.

Demographic characteristics will be summarized using the following information:

- Sex: male, female, undifferentiated
- Ethnicity: Hispanic or Latino, not Hispanic or Latino; Japanese, not Japanese
- Race:
    - For participants reporting one race only: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Not collected at the site, Other.
    - For participants reporting multiple races, all combinations will be reported under 'More than one race" category.
- Age (years): summary statistics
- Age categories:

- < 65 years,
- ≥ 65 years
  - 65-74 years,
  - 75-84 years,
  - ≥ 85 years

- Geographic Region: North America, Latin America, Europe, Asia & Pacific
- Body Surface Area (BSA) (m^2) at Baseline
- Height (cm) at Baseline
- Weight (kg) at Baseline
- Body Mass Index (BMI) (kg/m^2) at Baseline
- ECOG Performance status:
  - 0: Fully active, able to carry on all pre-disease performance without restriction
  - 1: Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
  - 2: Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours
  - 3: Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours
  - 4: Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair
  - 5: Dead

Specifications for computation (at ADaM level):

- Age [years]: (date of given informed consent - date of birth + 1) / 365.25

  The integer part of the calculated age will be used for reporting purposes. Note for rescreened participants, their latest informed consent date will be used.
- BMI (kg/m^2) = weight(kg)/[height(cm)²/10000]
- BSA (m^2) = ([height (cm) x weight (kg)] / 3600) ½
- Investigator site codes will be used for the determination of the participant's geographic region.

Demographic characteristics including cohort/planned arm, participant identifier, sex, race (including all reported races in case of "multiple" races, and details in case of "other" race), ethnicity, geographic region, age, body mass index, body surface area, weight and height will be presented in a listing.

11.2 Medical History

Relevant past and ongoing medical conditions at Baseline will be summarized from the "Medical History Details" eCRF page, using the latest available version of the Medical Dictionary for Regulatory Activities (MedDRA), PT as event category and MedDRA system organ class (SOC) body term as Body System category.

Medical history will be displayed in frequency tables, ordered by primary SOC and PT in alphabetical order. Each participant will be counted only once within each PT or SOC.

Listing of medical history including cohort/planned arm, participant identifier, age, sex, race, preferred term, reported medical history term, start/end dates, related study condition, ongoing at screening and toxicity grade (when medical history is ongoing) will be presented.

11.3 Other Baseline Characteristics

11.3.1 Disease History

Information on disease characteristics collected on the "Disease History" eCRF page will be summarized as follows:

- Biliary tract cancer (BTC) subtype and tumor histology:
    - Intrahepatic CCA
        - Intrahepatic cholangiocarcinoma
        - Combined hepatocellular cholangiocarcinoma
        - Others [carcinosarcoma, other]
    - Extrahepatic CCA
        - Perihilar CCA
            - Adenocarcinoma [adenocarcinoma, adenocarcinoma (biliary type, intestinal type, gastric foveolar type), mucinous adenocarcinoma, clear cell adenocarcinoma, signet ring cell carcinoma]
            - Squamous [squamous cell carcinoma, adenosquamous carcinoma]
            - Others
        - Distal CCA
            - Adenocarcinoma [adenocarcinoma, adenocarcinoma (biliary type, intestinal type, gastric foveolar type), mucinous adenocarcinoma, clear cell adenocarcinoma, signet ring cell carcinoma]
            - Squamous [squamous cell carcinoma, adenosquamous carcinoma]
            - Others
    - Gallbladder cancer

- Adenocarcinoma [adenocarcinoma, adenocarcinoma (biliary type, intestinal type, gastric foveolar type), mucinous adenocarcinoma, clear cell adenocarcinoma, signet ring cell carcinoma]
- Squamous [squamous cell carcinoma, adenosquamous carcinoma]
- Others

- Ampulla of Vater's cancer
  - Adenocarcinoma [adenocarcinoma, adenocarcinoma (invasive intestinal type, pancreaticobiliary type), clear cell carcinoma, hepatoid adenocarcinoma, mucinous carcinoma, signet ring cell carcinoma]
  - Squamous [squamous cell carcinoma, adenosquamous carcinoma]
  - Others

- Histology grade (Well differentiated, Moderately differentiated, Poorly differentiated, Grade cannot be assessed)
- Time since initial cancer diagnosis (months)
- Time since documented, locally advanced or metastatic disease (months)
- Time since last progression of disease prior to study entry (months)
- TNM classification at initial diagnosis and at study entry: each T, N, M category will be described (TX, T0, N1, etc.)
- Molecular abnormalities:
  - FGFR1/ FGFR2/ FGFR3 (mutated, wild type)
  - IDH1/ IDH2 (mutated, wild type)
- MSI status: MSI high, MSI low, MSS, unknown

Dedicated listings will also be provided with the following information:

- Cohort/planned arm, participant identifier, age, sex, race, BTC classification, tumor histology, date of initial cancer diagnosis (months), date of documented, locally advanced or metastatic disease (months), date of last disease progression (months), TNM classification at initial diagnosis, TNM classification at study entry and MSI status
- Molecular abnormality (result for FGFR1 (mutated/wild type), FGFR2 (mutated/wild type), FGFR3 (mutated/wild type), IDH1 (mutated/wild type) and IDH2 (mutated/wild type), others (normal/abnormal))

11.3.2 Skin Status History

Skin status history is collected on the "Skin Status History" eCRF page and will be summarized by the frequency and percentage of participants having the following history of:

- Frequent sunburn (Yes, No, Unknown)

- Easy sunburn (Yes, No, Unknown)
- Skin cancer (Yes, No, Unknown)
- Significant UV exposure (Yes, No, Unknown)
- Photosensitivity due to skin disorder (Yes, No, Unknown)
- Photosensitivity due to medication (Yes, No, Unknown)
- Family history of skin cancer in first degree relative (i.e., parents, siblings and/or children) (Yes, No, Unknown)
- Number of participants having history of the skin conditions above (No condition, 1 condition, 2 conditions, 3 or more conditions)

A listing of skin status history will be provided.

11.4 Previous Anticancer Therapy

Previous anticancer treatments and procedures are collected under the "Prior Anti-Cancer Drug Therapies Details" and "Prior Anti-Cancer Surgeries Details" eCRF pages.

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

- Participants with at least one type of previous anticancer treatment or procedure (i.e., drug therapy, radiotherapy or surgery)
- Participants with at least one previous anticancer drug therapy

- Participants with at least one previous anticancer radiotherapy
- Participants with at least one previous anticancer surgery

The following details for previous systemic anticancer drug therapy will also be summarized:

- Participants with at least one previous adjuvant therapy
- Participants with at least one previous neoadjuvant therapy
- Participants with at least one previous therapy for metastatic or locally advanced disease
- Number of any previous anticancer therapy lines for metastatic or locally advanced disease (as categorical variable)
- Best response to last treatment regimen

Previous anticancer drugs, previous anticancer radiotherapy, and previous anticancer surgery will be presented in separate listings:

- The previous anticancer drug listing will contain cohort/planned arm, participant identifier, age, sex, race, regimen ID/name, intent of therapy, best response, date of progression, medication name, preferred term, start date, end date.
- The previous anticancer radiotherapy listing will contain cohort/planned arm, participant identifier, age, sex, race, start date, end date, location of radiotherapy.
- The previous anticancer surgery listing will contain cohort/planned arm, participant identifier, age, sex, race, name and location of surgery, date of surgery, curative intent of surgery (yes/no), and outcome of surgery.

12 Previous or Concomitant Medications/Procedures

Analysis set: SAFRI / ITT

12.1 Previous and Concomitant Medications

Concomitant and previous medications are reported on the "Concomitant Medications Details" eCRF page.

Concomitant treatments are medications, other than study treatment, which are taken by participants any time during the on-treatment period, see Section 9.8.

Previous medications are medications, other than study treatment and pre-medications for study treatments, which started before first administration of study treatments.

A medication may be classified as both concomitant and previous. The respective flags will be derived based on start and end date.

Concomitant and previous treatment each will be summarized by number and percentage of participants. ATC-2nd level and preferred term will be tabulated as given from the WHO-DD

dictionary most current version. If any previous or concomitant medication is classified into multiple ATC classes, the medication will be summarized separately under each of these ATC classes. The summary tables will be sorted by decreasing frequency of drug class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used. In case any specific medication does not have an ATC classification level 2 coded term, it will be summarized under "Unavailable ATC classification" category. Each participant will only be counted once within a given drug class and within a given drug name, even if he/she received the same medication at different times.

Specific rules will be used to define if a medication/procedure is considered as a previous, concomitant or both previous and concomitant medication/procedure as detailed in Appendix 2.

Previous and concomitant medications will be presented in listings which will include: cohort/planned arm, participant identifier, age, sex, race, preferred term, medication name as provided by the investigator, start date, end date, dose, dose units, frequency, route, reason for the medication.

A separate listing of concomitant COVID-19 vaccinations will be provided, this listing will include the same information as concomitant medications listing: country and standardized drug grouping (SDG) subgroup information will also be provided.

12.2 Premedications

Premedications are medications administered per protocol on the same day as, but prior to, the study treatment.

The number of participants receiving premedication for bintrafusp alfa/placebo will be summarized for each treatment visit based on "Study Treatment Premedication details" eCRF page. The number of participants receiving premedication for chemotherapy (cisplatin and/or gemcitabine) will be summarized for each treatment visit based on the "Chemotherapy Premedication details" eCRF page.

Percentages will be calculated on the number of participants who received a dose of bintrafusp alfa/placebo (or a chemotherapy dose, respectively) at the associated visit.

Listings of premedication for bintrafusp alfa/placebo and of premedication for chemotherapy will be provided and they will include: cohort/planned arm, participant identifier, age, sex, race, medication name, visit, date/time of study treatment, dose, dose units, and route.

12.3 Concurrent Procedures

Concurrent procedures are reported according to the "Concomitant Procedures Details" eCRF page.

Concurrent procedures will be presented in a listing which will include cohort/planned arm, participant identifier, age, sex, race, name of procedure (as provided by the investigator), start date,

end date, indication, reason for procedure and type of specimen collected. A flag will be displayed to identify each procedure as prior to treatment or on-treatment.

12.4 Subsequent Anticancer Therapy

Anticancer treatment after discontinuation of study drug will be summarized according to the "Anti-cancer Treatment After Discontinuation", "Radiotherapy After Discontinuation" and "Surgery After Discontinuation" for anticancer surgery eCRF pages.

The number of participants in each of the following anticancer treatment categories will be tabulated:

- Participants with at least one subsequent anticancer treatment or procedure (i.e., drug therapy, radiotherapy or surgery)
- Participants with at least one subsequent anticancer drug therapy
- Participants with at least one subsequent anticancer radiotherapy
- Participants with at least one subsequent anticancer surgery
- In addition, the number and percentage of participants who received subsequent immune checkpoint inhibitors (CPI) (avelumab, nivolumab, pembrolizumab, lambrolizumab, atezolizumab, durvalumab, tremelimumab or ipilimumab) and targeted therapy will be tabulated by treatment arms. The final list of subsequent CPI and targeted therapy will be provided upon medical review of all subsequent anticancer therapies and will be listed in Appendix 6.
- The number and percentage of participants with any anticancer drug therapy will also be presented by type (based on medical review results):
- Cytotoxic therapy
- Immune CPI
- Monoclonal antibodies therapy other than CPI
- Targeted therapy with small molecules
- Other

The best response to the first subsequent anticancer therapy will be described (CR/ PR/ SD /PD/ Non-CR/Non-PD/ Not Evaluable (NE)/ NA/ Unknown) as well as the best overall response among all anti-cancer drug therapies.

The subsequent anti-cancer drug therapies will also be summarized based on the number and percentage of participants by ATC-2nd level and preferred term. A participant will be counted only once within a given ATC class and within a preferred term, even if he/she received the same medication at different times. The summary will be sorted on decreasing frequency of ATC-2nd level and decreasing frequency of preferred term within a given ATC drug class. In case of equal frequency regarding ATC-2nd level (respectively preferred term), alphabetical order will be used.

Subsequent CPI and subsequent targeted therapy will be summarized. A participant will be counted only once within a CPI/targeted therapy, even if he/she received the same medication at different times. The summary will be sorted on decreasing frequency of CPI/targeted therapy. In case of equal frequency regarding CPI/targeted therapy level (respectively preferred term), alphabetical order will be used.

In addition, the anticancer treatment after discontinuation of study treatment will be provided in two listings:

- For drug therapy: cohort/planned arm, participant identifier, age, sex, race, preferred term / drug name, regimen name, start date, end date and best response
- For radiotherapy and surgery: cohort/planned arm, participant identifier, age, sex, race, type of therapy, start date, end date, radiotherapy site or name of surgery/location, if surgery outcome and was the surgery curative in intent (yes/no)

13 Study Treatment: Compliance and Exposure

Analysis set: SAFRI / SAF

Participants will be treated with bintrafusp alfa/placebo (bintrafusp alfa or placebo for the randomized part, bintrafusp alfa for the open-label part) in combination with cisplatin and gemcitabine. Bintrafusp alfa/placebo will be given at a flat dose of 2400 mg once every 3 weeks or corresponding placebo. Gemcitabine will be given at a dose of 1000 mg/m² and cisplatin at a dose of 25 mg/m² on Days 1 and 8 of every 21-day cycle for 8 cycles. If chemotherapy cannot be administered on Day 1 or Day 8 of a given cycle (for safety or tolerability reasons), the missed dose may be administered on Day 15 of that cycle. Administration details, including duration of treatment, of all study drugs are given in Section 6 of the study protocol.

All dosing calculations below and summaries will be based on the "Study Treatment Administration Details", "Gemcitabine Administration Details" and "Cisplatin Administration Details" eCRFs pages. Dosing calculations for the two chemotherapies will be based on the corresponding eCRF pages mentioned previously, and the planned dose (prepared by the unblinded pharmacist) will be captured in the drug preparation form.

Duration of Therapy and Number of Infusions for Bintrafusp alfa/Placebo and Chemotherapy

An infusion is regarded as being administered if the actual dose received is > 0.

The **overall duration** of each study treatment will be calculated as follows:

- Duration of bintrafusp alfa/placebo (weeks) = (date of last dose – date of first dose + 21) / 7
- Duration of cisplatin/gemcitabine (weeks) = (date of last dose – date of first dose + x) / 7 where x will be equal to 21 if participant only receive day 1 at the last cycle, x=14 if the last dose received is at day 8, and x=7 if the last dose received is at day 15

Note: Cycle X day 1, day 8 and day 15 are identified based on visit name as per the protocol scheduled activities (see Appendix 5)

Incomplete/Missing dates of last dose will be imputed as specified in Table 6 .

The following summary tables will be provided per cohort/arm:

- Duration of therapy (weeks) for each treatment separately (bintrafusp alfa/placebo, cisplatin, gemcitabine), by summary statistics and by categories (<3 weeks, 3 to 6 weeks, 6 to 9 weeks, up to ≥48 weeks)
- Number of infusions received for each treatment separately (bintrafusp alfa/placebo, cisplatin, gemcitabine), by summary statistics and by categories (1, 2, up to 16 doses)

Cumulative Dose, Dose Intensity, and Relative Dose Intensity

Bintrafusp alfa/placebo and chemotherapy calculations and summaries will be based on the "Administration Details" eCRF pages for each therapy.

Cumulative dose, dose intensity and relative dose intensity of chemotherapies will be calculated for each chemotherapy agent separately (cisplatin and gemcitabine).

The **cumulative dose** of bintrafusp alfa/placebo and chemotherapies per participant in a given time period will be the sum of the actual dose levels that the participant received within that period (i.e., total dose administered [mg or mg/m2]).

Dose Intensity (DI)

- For bintrafusp alfa/placebo:
  - Overall DI (mg/3 weeks) = [overall cumulative dose (mg)]/[overall duration (weeks)/3]
- For chemotherapies:
  - Overall DI (mg/m²/3 weeks) = [overall cumulative dose (mg)/BSA (m²)]/[overall duration (weeks)/3]

Note: in case the BSA is missing in the given time period, the latest previous BSA available will be used.

Relative dose intensity (RDI)

- For bintrafusp alfa/placebo and chemotherapies:
  - Overall RDI (%) = 100 x [overall DI]/ planned dose intensity

where the dose planned is 2400 mg for bintrafusp alfa/placebo, 2000 mg/m² for gemcitabine and 50 mg/m² for cisplatin.

The following summary table will be provided for bintrafusp alfa/placebo and chemotherapies by treatment arm:

- Duration of therapy (weeks)

- Total number of infusions received
- Cumulative dose (mg or mg/m²)
- DI (mg/3 weeks or mg/m²/3 weeks)
- RDI (%) as continue variable, and categorized as
    - < 80% / 80%-90% / >90% for bintrafusp alfa/placebo
    - <50% / 50%-75% / >75% for chemotherapy

Two listings will be presented:

- A listing of study treatment exposure data, which will provide: cohort/planned arm, participant identifier, age, sex, race, visit, infusion start date and time, infusion end date and time, infusion rate (mL/hr), volume infused (for bintrafusp alfa/placebo), actual dose (for chemotherapy), administration modification and reason for modification, change in administration detail, and treatment delay (days).
- An additional listing of treatment exposure and compliance which will include cohort/planned arm, participant identifier, age, sex, race, duration of therapy (weeks), total number of infusions received, cumulative dose of therapy, DI, and RDI (%).

Dose Modification

Per protocol, dose modification for bintrafusp alfa/placebo is not allowed. No summaries will be provided.

Chemotherapy dose reductions as recorded on the "Gemcitabine Administration Details" and "Cisplatin Administration Details" pages of the eCRF (i.e., answer to the question "Is there a change in dose?" = "Dose adjusted") will be used for analysis. The number of participants with at least one dose reduction, reasons for dose reduction (Adverse Event / Other), as well as a categorization of the number of infusions with dose reduced (1 / 2 / ≥ 3) will be summarized for each regimen by treatment arm.

Therapy Delays

Therapy delays will be derived for bintrafusp alfa/placebo as follows:

> Therapy delays for bintrafusp alfa/placebo = start date of current infusion – start date of the previous infusion – 21

Chemotherapy administration delayed as recorded on the "Gemcitabine Administration Details" and "Cisplatin Administration Details" pages of the eCRF (i.e., answer to the question "Was the Cisplatin/Gemcitabine administration delayed?" = "Yes") will be used for analysis.

The following will be summarized in a table for bintrafusp alfa/placebo and chemotherapy for each treatment arm:

- Number of participants with at least one delay

- Number of participants with no delay
- Number of delays per participant (1 delay, 2 delays, 3 delays, ≥ 4 delays)
- Longest delay per participant (1-2 days, 3-8 days, 9-15 days, ≥16 days) (only for bintrafusp alfa/placebo)

Temporary Infusion Interruptions

Temporary interruptions of bintrafusp alfa/placebo and chemotherapy infusion as recorded on the "Administration Details" pages of the eCRF will be used for analysis. The number of participants with at least one study drug temporary infusion interruption, reason for infusion temporary interruption (Adverse Event / Other), as well as a categorization of the number of temporary infusion interruptions (1 / 2 / ≥3) will be summarized by treatment arm for both bintrafusp alfa/placebo and chemotherapy.

Infusion Rate Reductions

Infusion rate reductions as recorded on the "Administration Details" eCRF pages will be used for analysis. The number of participants with at least one infusion rate reduction, reason for infusion rate reductions (Adverse Event / Other), as well as a categorization of the number of infusion rate reductions (1 / 2 / ≥3) will be summarized by treatment arm for both bintrafusp alfa/placebo and chemotherapy.

14 Efficacy Analyses

Analysis Sets: SAFRI / ITT

Depending on the decision to expand or not the study into Phase III sample size, all efficacy endpoints will be considered either as the primary analysis or a sensitivity analysis according to Table 7. Data related to investigator's tumor assessments will be collected on the "Assessment of disease based on imaging (according to RECIST 1.1)" eCRF page.

Table 7 Definition of Primary and Sensitivity Analysis of Efficacy Tumor-related Endpoints according to the Adaptation Decision

Adaptation decision	Primary Analysis	Sensitivity Analysis to the type of tumor assessment (Investigator vs IRC)
Expansion into Phase III (500 participants)	Efficacy endpoint as assessed by the investigator	Efficacy endpoint as assessed by the IRC in the first 150 randomized antibiotics naïve participants in Phase II
No expansion	Efficacy endpoint as assessed by the IRC	Efficacy endpoint as assessed by the investigator

Multiplicity adjustment for primary and key secondary endpoints

Out of the secondary endpoints, PFS as well as objective response rate are considered as key secondary endpoints.

If the study is not expanded into Phase III sample size, the statistical test for the primary endpoint will be performed at one-sided alpha of 0.025. Otherwise, two sequential tests (interim and primary analyses) will be performed for the primary and key secondary endpoints.

A hierarchical testing procedure will be applied to the confirmatory tests for the primary and key secondary endpoints. The hypothesis for the primary endpoint (OS) will be tested using a group sequential approach in an interim analysis and a final analysis at a significance level of 0.025 one-sided. If the null hypothesis for the primary endpoint is rejected, the hypothesis for the first key secondary endpoint (PFS) will be tested confirmatory using the same information fraction for a group sequential approach. Finally, if this last null hypothesis is rejected then the last key secondary endpoints (objective response rate) will be tested likewise.

The group sequential approach for testing OS, PFS and objective response will use alpha-spending function of Lan-DeMets approach according to the actual number of observed events regarding the total event number planned for the primary analysis for OS (334 events), i.e., the same information fraction with OS will be used for PFS and objective response. O'Brien-Fleming boundaries are used for OS and Pocock boundaries are used for PFS and objective response (Glimm 2010).

Pre-specifying the order of the hypotheses to be tested, the family wise error rate for this approach to sequentially rejective multiple test procedure (Maurer 1995) is controlled at 2.5% one-sided alpha level.

CCI [redacted]

14.1 Overall Survival (Primary Endpoint)

14.1.1 Primary Objective: Derivation and analysis of Overall Survival

Overall survival is defined as the time from study day 1 to the date of death due to any cause:

OS (months) = (date of event or censoring – date of study day 1 + 1)/30.4375.

The date of event / censoring is defined in Table 8.

Table 8 Survival Event / Censoring

Survival Status	Date of Event/censoring	Censoring
Participants who died before or at cut-off date	Date of death	No
Participants have not died before or at cut-off date	Date of last contact	Yes

The date of last contact will be defined following Section 9.6.

Kaplan-Meier Analysis

The analysis of OS will be performed with a Kaplan-Meier method. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median OS with two-sided 95% CIs. In particular, OS rates with their two-sided 95% CIs at 3, 6, 9, 12 and 24 months at least will be presented, as well as the number of participants at risk and failed. The CIs for the median will be calculated according to Brookmeyer and Crowley and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (conftype=loglog default option in SAS PROC LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of participants with an event (death) and censoring reasons will also be presented by treatment group. Censoring reasons are the following:

- Alive (administrative censoring)
- Withdrawal of consent
- Lost to follow-up

Lost to follow-up will include the following participants:

- Lost to follow-up status is collected on the eCRF page prior to the analysis cut-off
- Participants with the last contact date > 14 weeks prior to the analysis cut-off date (duration of 14 weeks is based on the assessment schedule of every 3 months for survival follow-up interval + 2-week window)

Alive will include censored participants who did not withdraw from consent and were not lost to follow-up based on the previous definition.

A participant listing will be provided including the following information: cohort/planned arm, participant identifier, age, sex, race, date of study day 1, date of event/censoring, event/censoring reason, time to event/censoring.

OS will also be presented graphically with Kaplan-Meier figures.

Follow-up Duration

A Kaplan-Meier analysis for OS follow-up duration will also be performed reversing the OS censoring and event indicators. This analysis will be based on the main Kaplan-Meier analysis of OS described above.

Stratified Log-rank Test

OS will be compared between the two treatment groups (bintrafusp alfa *vs* placebo) using a one-sided stratified log-rank test.

The following null hypothesis will be tested:

$$H_0^{OS}: \lambda_B^{OS}(t) = \theta\lambda_P^{OS}(t), \theta \geq 1 \text{ versus } H_1^{OS}: \lambda_B^{OS}(t) = \theta\lambda_P^{OS}(t), \theta < 1$$

Where $\lambda^{OS}(t)$ represents the hazard at time t and θ the unknown constant of proportionality of hazards in the treatment groups B (bintrafusp alfa) and P (placebo).

Cox's Regression Model

The treatment effect on OS will be estimated using a Cox's proportional hazards model stratified by the randomization data (BTC subtype, Asian/Non-Asian and disease status at initial diagnosis) through the HR computation. Randomization strata will be taken as specified and documented in the IXRS. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), i.e., for the i-th stratum the hazard function is expressed as: h(i; t) = h(i, 0; t) exp(xβ), where h(i, 0; t) defines the baseline hazard function for the i-th stratum and x defines the treatment group (0 = placebo, 1 = bintrafusp alfa) and β is the unknown regression parameter.

Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

A table including the Kaplan-Meier estimates and the Cox's proportional hazards model results will be provided.

Methods for evaluating the validity of model assumptions

Schoenfeld residuals for the stratified Cox proportional hazards model will be plotted including a LOESS curve to investigate graphically violations from the proportional hazards assumption; a non-zero slope is evidence of deviation from proportional hazards. Schoenfeld residuals will be computed in SAS using the PHREG procedure and using the OUTPUT statement and the keyword RESSCH. With proportional hazards the LOESS curve should be parallel to the x-axis.

The proportional hazards assumption will be formally tested using Schoenfeld's residual test (Schoenfeld, 1980; Therneau & Grambsch, 2000). Large departures from proportional hazards 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 proportional hazards assumption will be checked visually by plotting for each randomization stratum

log(-log(S(t))) versus log(t),

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

If these methods show large departures from proportional hazards, then OS will also be analyzed based on restricted mean survival time (RMST) differences (Zhang, 2013). See Section 14.1.2.

Multivariable Cox Regression

Multivariable Cox regression analysis will be carried out to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact. Relevant baseline factors will be defined as the subgroups defined in Section 8.2. The following categorizations will be considered for those subgroups with multiple definitions:

- Blood-based and/or tumor-based TMB at baseline, grouped in quartiles
    - TMB low (Q1+Q2) (reference level)
    - TMB high (Q3+Q4)
- PD-L1 expression on tumor cells (TC) at baseline
    - $< 1\%$ (reference level)
    - $\geq 1\%$
- PD-L1 expression on immune cells (IC) at baseline
    - $< 1\%$ (reference level)
    - $\geq 1\%$

A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure, the stepwise selection method in SAS (PROC PHREG) will be used. The level of significance for an explanatory variable to enter the model will be set to 0.15 and the significance level for removing to 0.40. The Cox's proportional hazard model is defined as: $h(t) = h(0; t) \exp(Xb)$, where $h(0; t)$ defines the baseline hazard function and X defines the vector of explanatory variables and b the unknown vector of regression parameters. The hazard ratios of all selected explanatory variables and of treatment effects will be reported including 2-sided 95% CIs. No interactions will be considered. Post-baseline factors will not be considered for the model.

14.1.2 Sensitivity Analyses of Overall Survival

To assess the robustness of the OS analysis, the following sensitivity analysis will be conducted:

- An unstratified analysis will be performed to compare OS between the two treatment groups and to estimate the treatment effect. This analysis will be performed as described in Section 14.1.1.
- The primary analysis will be repeated by censoring participants who received subsequent immune checkpoint inhibitor therapy such as avelumab, nivolumab, pembrolizumab, atezolizumab, durvalumab, tremelimumab or ipilimumab after discontinuing the study treatment at the date of the first dose of subsequent immune checkpoint inhibitor therapy minus 1 day. The final list of subsequent immune checkpoint inhibitor therapy will be provided upon medical review of all subsequent anti-cancer therapies.
- The measurement of treatment effect on OS stratified by the randomization data will be repeated using randomization strata as specified and documented in the eCRF.

If the analysis of OS based on Cox's regression model has been shown to not meet the proportional hazards assumption, then the following analysis will be performed.

Restricted Mean Survival Time (RMST)

The RMST methodology is applicable independently of the proportional hazards (PH) assumption and can be used, at a minimum, as a sensitivity analysis to explore the robustness of the primary analysis results.

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 then be interpreted as the expected survival time restricted to the common follow-up time τ among all patients. The selection of τ should ensure that the RMST evaluation will not go beyond the maximum time point where the evaluation can be performed while also taking into account a large period of time that is expected to provide a meaningful assessment of treatment effect. To avoid arbitrary selection of the common cut-off τ for both treatment arms, three sets of analyses will be performed:

- τ_1 = minimum of (largest observed survival time for bintrafusp alfa arm, largest observed survival time for placebo arm).
- τ_2 = minimum of (largest survival event time for bintrafusp alfa arm, largest survival event time for placebo arm).
- τ_3 = midpoint between τ_1 and τ_2.

In this section, 'survival' is meant to denote OS.

The treatment effect between the bintrafusp alfa arm and the placebo arm will be assessed based on the difference in RMST. The associated 95% CI for the difference in means and one-sided p-value will be generated. RMST as a function of τ and the associated treatment effect between the two treatment arms will be plotted against time τ (from τ = 0 to the largest observed survival time across both arms).

CCI

14.2 Progression-Free Survival According to RECIST 1.1 (Secondary Endpoint)

Progression-Free Survival time is defined as the time from study day 1 until the first documentation of progression of disease or death due to any cause, whichever occurs first.

PFS = (date of PD or death or censoring – date of study day 1 + 1)/ 30.4375 (months).

Censoring rules for primary and sensitivity analyses are summarized in Table 9.

Table 9 Censoring Rules for Primary and Sensitivity Analysis of PFS

Situation		Primary Analysis / Sensitivity Analysis to the type of tumor assessment	Sensitivity Analysis 2 (Primary Analysis PFS counting all events (PD and death) regardless of the start of a new anticancer therapy and ignoring the number of missing evaluable tumor assessments before progression or death)
No PD and no death	New anticancer therapy is not initiated	Censored at last tumor assessment*	Censored at last tumor assessment*
	New anticancer therapy is initiated	Censored at last tumor assessment* before new anticancer therapy	Censored at last tumor assessment*
No baseline assessment or no evaluable post-baseline assessment	No death or death >12 weeks after start of study treatment	Censored on study day 1	Censored on study day 1
	Death ≤12 weeks after start of study treatment	Progressed at date of death	Progressed at date of death
PD or death	After ≤1 missing tumor assessment[a]	Progressed at date of documented PD or death, whichever came first	Progressed at date of documented PD or death, whichever came first
	After ≥2 subsequent missing tumor assessments[a]	Censored at last tumor assessment* before the missing assessments.	Progressed at date of documented PD or death, whichever came first

Situation		Primary Analysis / Sensitivity Analysis to the type of tumor assessment	Sensitivity Analysis 2 (Primary Analysis PFS counting all events (PD and death) regardless of the start of a new anticancer therapy and ignoring the number of missing evaluable tumor assessments before progression or death)
	Before started new anticancer therapy[b]	Progressed at date of documented PD or death, whichever came first	Progressed at date of documented PD or death, whichever came first
	After started new anticancer therapy[b]	Censored at last tumor assessment* before new anticancer therapy	Progressed at date of documented PD or death, whichever came first

* with outcome CR, PR, SD or Non-CR/Non-PD. If no adequate tumor assessment, censored on study day 1.
[a] No assessments in 84 days during the first 9 months of follow-up or 168 days after the first 9 months of follow-up.
[b] Definition of new anticancer therapy to be considered is given in Section 9.8.

14.2.1 Main Analysis for PFS

Main analysis for PFS will be based on participants' tumor responses either as assessed by the investigator or by the IRC according to the rule defined in Table 7.

Kaplan-Meier Analysis

The analysis of PFS will be performed with a Kaplan-Meier method. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median PFS with two-sided 95% CIs. In particular, PFS rates with their two-sided 95% CIs at 3, 6, 9, 12, 18 and 24 months at least (depending on the actual data) will be presented, as well as the number of participants at risk and failed. The CIs for the median will be calculated according to Brookmeyer and Crowley and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (conftype=loglog default option in SAS PROC LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Frequency (number and percentage) of participants with and without an event, overall and per event type (PD or death) will be presented by treatment arm. Censoring reasons will be described by treatment arm, as follows:

- Administrative censoring (ongoing in the study without an event)
- No baseline assessment i.e no tumor scan recorded before randomization
- No evaluable post-baseline assessment
- Start of new anticancer therapy
- Event after 2 or more missing or non-evaluable post-baseline assessments
- Withdrawal of consent

- Lost to follow-up

Lost to follow-up will include the following participants:

- Lost to follow-up status is collected on the "Treatment Termination" eCRF page or the "Study Termination" eCRF page prior to the analysis cut-off
- Participants with censoring date > 14 weeks prior to the analysis cut-off date (duration of 14 weeks is based on the assessment schedule of every 3 months for survival follow-up interval + 2-week window) and participant ongoing without event
- Participants who discontinued/completed the study without an event

A participant listing will be provided including the following information: cohort/planned arm, participant identifier, age, sex, race, date of study day 1, date of event/censoring, event/censoring reason, time to event/censoring.

PFS will also be presented graphically with Kaplan-Meier figures.

Follow-up Duration

A plot will be generated to compare planned and actual relative day of tumor assessments by treatment arm. A Kaplan-Meier analysis for PFS follow-up duration will also be performed reversing the PFS censoring and event indicators. This analysis will be based on the main Kaplan-Meier analysis of PFS described above.

Stratified Log-rank Test

PFS will be compared between the two treatment groups (bintrafusp alfa *vs* placebo) using a one-sided stratified log-rank test.

The following null hypothesis will be tested:

$H_0^{PFS}: \lambda_M^{PFS}(t) = \theta\lambda_P^{PFS}(t), \theta \geq 1$ versus $H_1^{PFS}: \lambda_M^{PFS}(t) = \theta\lambda_P^{PFS}(t), \theta < 1$

Where $\lambda^{PFS}(t)$ represents the hazard at time t and θ the unknown constant of proportionality of hazards in the treatment groups M (bintrafusp alfa) and P (placebo).

Tests will be considered as confirmatory if conditions on multiplicity adjustments stated in the introduction of Section 14 are fulfilled.

Cox's Regression Model

The treatment effect on PFS will be estimated using a stratified Cox's proportional hazards model following the methodology described in Section 14.1.1.

Methods for evaluating the validity of model assumptions

Potential departure from the model assumptions will be assessed following the methodology described in Section 14.1.1.

Multivariable Cox Regression

Multivariable Cox regression analysis for PFS will be carried out to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact. It will be conducted in the same way as described in Section 14.1.1.

14.2.2 Sensitivity Analysis for PFS

The following two sensitivity analyses will be conducted:

- PFS counting all events (PD and death) regardless of the start of a new anticancer therapy and ignoring the number of missing evaluable tumor assessments before progression or death
- Main analysis described for the secondary endpoint PFS will also be performed on tumor responses based on investigator or IRC assessment according to Table 7. No formal test will be performed.
- The measurement of treatment effect on PFS stratified by the randomization data will be repeated using randomization strata as specified and documented in the eCRF.

These sensitivity analyses will be supported with Kaplan-Meier plots and participant-listings.

Progressive Disease by Investigator vs IRC

For all participants with both tumor assessment per investigator and IRC, a summary of the IRC assessment versus investigator assessment will be provided including numbers of concordant and discordant assessments as well as the number of cases where PD was assessed at different timepoints by IRC and investigator. Table 10 outlines the possible outcomes by investigator and IRC.

Table 10 Possible Outcomes for Investigator vs IRC

		IRC	
		PD	No PD
Investigator	PD	a = a1 + a2 + a3	b
	No PD	c	d

a1: number of agreements on timing and occurrence of PD
a2: number of agreements on PD but investigator declares PD later than IRC
a3: number of agreements on PD but investigator declares PD earlier than IRC
N= a+b+c+d.
The timing agreement of progression has to be defined as a window of 7 days

The following measures of agreement and discordance will be calculated for each treatment arm:

- Total event discrepancy rate: (b+c) / N

- Total event agreement rate: (a+d) / N
- Rate of unconfirmed investigator PDs: b / (a+b)
- 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 progression earlier than IRC within each arm as a proportion of the total number of investigator-assessed PDs.

The LDR quantifies the frequency with which the investigator declares progression later than IRC as a proportion of the total number of discrepancies within the arm.

Discordance metrics are calculated for each treatment arm and, for each metric, the difference in discordance between the bintrafusp alfa and placebo arms is used to evaluate potential bias. If the discordance is similar across the treatment arms, then this 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 bintrafusp alfa arm.

14.3 Objective Response Rate According to RECIST 1.1 (Secondary Endpoint)

Objective Response Rate (ORR)

ORR is defined as the proportion of participants having a best overall response of CR or PR. For confirmed ORR, confirmed CR or PR will be used.

Best overall response will be assessed based on overall responses reported at different evaluation time points from randomization until the first documented disease progression according to RECIST 1.1, taking requirements for confirmation into account as detailed below. Only tumor assessments performed before or at start of any subsequent anticancer therapies (see Section 9.8 for subsequent anticancer therapies to be considered) will be considered in the assessment of best overall response. Clinical deterioration will not be considered as documented disease progression.

Best overall response based on confirmed responses:

- CR = at least two determinations of CR at least 4 weeks apart and before progression

  Note: it is reasonable to consider CR-NE-CR or CR-PR-CR as CR as long as the second CR is at least 4 weeks apart the first time point.

- PR = at least two determinations of PR or better (PR followed by PR or PR followed by CR) at least 4 weeks apart and before progression (and not qualifying for a CR)

  Note: It is reasonable to consider PR-NE-PR or PR-SD-PR as PR provided the second PR is at least 4 weeks apart the first time point

- SD = at least one SD assessment (or better) $\geq$ 6 weeks after start date and before progression (and not qualifying for CR or PR)
- Non-CR/Non-PD = at least one non-CR/non-PD assessment (or better) $\geq$ 6 weeks after start date and before progression with no measurable disease at baseline (and not qualifying for CR or PR)

  Non-CR/non-PD is specific to IRC assessment, i.e., it does not apply for the Investigator assessment as investigator should include only participants with measurable disease at baseline as per inclusion/exclusion criteria.

- PD = progression $\leq$ 12 weeks after start date (and not qualifying for CR, PR, SD or non-CR/non-PD)
- ND = at least one ND assessment

  Note: Overall response is rated as "No Disease" (ND) when the IRC is not able to identify any disease at baseline (target or non-target lesions). ND is specific to IRC assessment, i.e., it does not apply for the Investigator assessment as investigator should include only participants with measurable disease at baseline as per inclusion/exclusion criteria.

- NE = all assessments are NE and/or no (or not evaluable) tumor assessments on-treatment, or participant does not meet any of the responses above

Best overall response based on unconfirmed responses:

- CR = at least one determination of CR documented before progression
- PR = at least one determination of PR documented before progression (and not qualifying for an CR)

- SD = at least one SD assessment ≥ 6 weeks after start date and before progression (and not qualifying for CR or PR)
- Non-CR/Non-PD= at least one non-CR/non-PD assessment ≥ 6 weeks after start date and before progression with no measurable disease (and not qualifying for CR or PR)

  Non-CR/non-PD is specific to IRC assessment, i.e., it does not apply for the Investigator assessment as investigator should include only participants with measurable disease at baseline as per inclusion/exclusion criteria.
- PD = progression ≤ 12 weeks after start date (and not qualifying for CR, PR, SD or non-CR/non-PD)
- ND = at least one ND assessment

  Note: Overall response is rated as "No Disease" (ND) when the IRC is not able to identify any disease at baseline (target or non-target lesions). ND is specific to IRC assessment, i.e., it does not apply for Investigator assessment as investigator should include only participants with measurable disease at baseline as per inclusion/exclusion criteria.
- NE = all assessments are NE or participant has a missing (or not evaluable) baseline tumor assessment and/or no (or not evaluable) tumor assessments on-treatment, or participant does not meet any of the responses above

Disease Control Rate (DCR)

DCR is defined as the proportion of participants having a best overall response of CR or PR, SD or Non-CR/Non-PD. For confirmed DCR, confirmed CR or PR or SD or Non-CR/Non-PD will be used.

14.3.1 Main Analysis for ORR

Main analysis for ORR will be based on participants' tumor responses either as assessed by the investigator or by the IRC according to the rule defined in Table 7.

Number and percentage of participants with best overall response of confirmed CR, PR, SD, non-CR/non-PD, PD, ND and NE as well as the ORR and the DCR will be tabulated by treatment arm. The ORR and DCR will be presented with a two-sided 95% CI using the Clopper-Pearson method (exact CI for a binomial proportion as computed by default by the SAS FREQ procedure using the EXACT option).

The association of treatment effect and OR will be tested by the general association statistic of the Cochran-Mantel-Haenszel test (CMH) with the randomization strata taken into account. The null hypothesis of no association in any of the randomization strata is tested against the alternative, which specifies that there is an association between treatment and tumor response at least in one randomization stratum. A one-sided CMH test will be performed at an overall alpha level of 0.025. If the study is expanded into Phase III, testing will be performed as explained in the introduction of Section 8.1.

The stratified odds ratio in terms of confirmed OR will also be estimated along with its 95% CI to assess the treatment effect. The odds ratio is defined as the odds of confirmed OR with test treatment divided by the odds of confirmed OR with control treatment. The Breslow-Day test will be used to check the homogeneity of the odds ratio across the randomization strata. It tests the null hypothesis that odds ratios in all strata are equal against the alternative hypothesis that the odds ratio is different in at least one stratum.

In case the null hypothesis of homogeneity of odds ratios across strata is not rejected at the alpha level of 5% two-sided, the common odds ratio will be determined using the Mantel-Haenszel estimate (by the FREQ procedure using CMH option in SAS); if the null hypothesis of homogeneity of odds ratio across all strata is rejected, the odds ratio per stratum will be calculated with the corresponding exact CI.

The individual percentage of change in the sum of diameter of all target lesions (longest diameter for non-nodal lesions and short axis for nodal target lesions) since baseline for each treatment arm will be displayed over time on a spider plot, together with the first occurrence of new lesion, the participant off-treatment. This spider plot will be produced for the whole safety run-in analysis set and it will be restricted to participants with PR or better at any time during the study for the randomized analyses. The percent change in the sum of diameters of all target lesions between baseline and the best post-baseline assessment (i.e., lowest change since baseline) for each treatment arm will also be displayed on a waterfall graph.

For spider plots and waterfall plots, the percent change from baseline in the sum of diameters of all target lesions will be displayed for valid timepoint assessments only. For the purpose of this analysis, a valid timepoint assessment is defined as a complete assessment of all target lesions reported at baseline. Further, split and coalesced lesions must be taken into account appropriately to determine if a timepoint assessment is valid to derive the percent change from baseline in the sum of diameters. All sum of diameters should be used, including the ones beyond first PD (for waterfall plots, the best post-baseline sum of diameters will be used even if occurring beyond PD). For waterfall plots, the percent change from baseline to the 6-week assessment, as well as the percent change from baseline to the best post-baseline sum of diameters, will be displayed for each participant.

Listings of tumor assessment will be provided with the following information: cohort/planned arm, participant identifier, age, sex, race, date of randomization (date of first study treatment administration for non-randomized part), date of start of subsequent anticancer therapy, date of death when death occurs, best overall response based on confirmed and unconfirmed responses, visit, date(s) of imaging, description of target lesions (size, site, type, method, split/merged, response), non-target lesions (status, site, type, method, response), and new lesions (site, type, method), sum of lesion diameters, percent change in target lesions from baseline, and overall response of participant. Note that response of target and non-target lesions is not applicable at baseline.

Multivariable Logistic Regression

Multivariable logistic regression analysis will be performed on confirmed best overall response to assess and adjust the treatment effect for relevant baseline factors of potential prognostic impact.

The subgroup variables defined in Section 8.2 will be included in the model. A stepwise selection procedure will serve to identify explanatory variables of potential prognostic values additional to the randomization strata which will be included in all models during the selection procedure. 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). Once this procedure stops, the factor 'treatment group' will be added to the last selected model in order to evaluate the effect of treatment on best overall response when adjusted for the selected explanatory variables. The odd ratios of all selected explanatory variables and of treatment effects will be reported including 2-sided 95% CI. No interactions will be considered. Post-baseline factors will not be considered for the model.

Non-evaluable best overall response

A summary table of the reasons for non-evaluable best overall response based on confirmed responses will be provided by randomized arm. The following reasons will be used in hierarchical order:

- No baseline assessment, i.e no tumor scan recorded before randomization
- No post-baseline assessments due to death within 6 weeks after start date
- No post-baseline assessments due to other reasons
- All post-baseline assessments have overall response 'Non-evaluable'
- New anticancer therapy started before first evaluable post-baseline assessment
- SD (or better) of insufficient duration (<6 weeks after start date without further evaluable tumor assessment)

  Note: special cases where best overall response is NE due to both early SD and late PD will be classified into this category
- Non-CR/non-PD of insufficient duration (<6 weeks after start date without further evaluable tumor assessment) (if appears in the data
- PD too late (>12 weeks after start date)
- "Not determined" categories may be added, if applicable

A listing of reasons for non-evaluable best overall response based on confirmed responses will also be created including: cohort/planned arm, participant identifier, age, sex, race, date of randomization (randomized phase only), date of first and last dose, date(s) of imaging, overall response and the reason for confirmed best overall response non-evaluable.

14.3.2 Sensitivity Analysis for ORR

- Number and percentage of participants with best overall response of unconfirmed CR, PR, SD, non-CR/non-PD, PD, ND and NE as well as the ORR and the DCR based on unconfirmed responses will be tabulated by treatment arm. The ORR and DCR will be presented with a two-

sided 95% CI using the Clopper-Pearson method (exact CI for a binomial proportion as computed by default by the SAS FREQ procedure using the EXACT option).

- Main analysis described for the secondary endpoint ORR based on confirmed responses will also be performed on tumor responses based on investigator or IRC assessment according to Table 7. No formal test will be performed.
    - In addition, a summary of the best overall response as adjudicated by IRC versus Investigator assessment including numbers of concordant and discordant assessments, and a listing of inconsistencies will be provided.
- The measurement of stratified odds ratio in terms of confirmed OR will be repeated using randomization strata as specified and documented in the eCRF.

14.4 Duration of Response According to RECIST 1.1 (Secondary Endpoint)

Duration of response (DOR) is defined for participants with a confirmed objective response, as the time from first documentation of confirmed objective response (CR or PR based on confirmed responses) to the date of first documentation of objective PD or death due to any cause whichever occurs first:

DOR = (date of PD or death or censoring – date of confirmed objective response + 1)/30.4375 (months).

The censoring rules for DOR are as described for PFS (primary definition in Table 9) in Section 14.2.

14.4.1 Main Analysis for DOR

Primary analysis for DOR will be based on participants' tumor responses either as assessed by the investigator or by the IRC according to the rule defined in Table 7.

DOR of CR/PR based on confirmed responses according to RECIST 1.1 will be described by treatment arm. As CR/PR based on confirmed responses involves having at least two objective responses, the date of occurrence of the first CR/PR will be used as the date of objective response.

The analysis of DOR will be performed with a Kaplan-Meier method. Kaplan-Meier estimates (product-limit estimates) will be presented by treatment group together with a summary of associated statistics including the median DOR with two-sided 95% CIs. In particular, DOR rates with their two-sided 95% CIs at 3, 6, 9, 12 and 24 months at least will be presented, as well as the number of participants at risk and failed. The CIs for the median will be calculated according to Brookmeyer and Crowley and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation according to Kalbfleisch and Prentice (conftype=loglog default option in SAS PROC LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

The time to and duration of response per participant having an objective response (including delayed response, see Section 14.10 below) based on confirmed responses will be displayed in swimmer graphs. Kaplan-Meier figures will also be provided.

Listings will be provided with the following information: planned arm, participant identifier, age, sex, race, date of randomization, date of first response, date of last tumor assessment, censored (yes/no), date of event/censoring, event/censoring reason and duration of response.

14.4.2 Sensitivity Analysis for DOR

- Main analysis described for the secondary endpoint DOR will also be performed from tumor responses based on investigator or IRC assessment according to Table 7. No formal test will be performed.

- Main analysis described for the secondary endpoint DOR will also be performed from tumor responses based on unconfirmed responses for both investigator and IRC.

14.5 Durable Response According to RECIST 1.1 (Secondary Endpoint)

Durable Response Rate (DRR) is defined as the number of participants having a DOR lasting > 6 months, out of the total number of participants.

14.5.1 Main Analysis for DRR

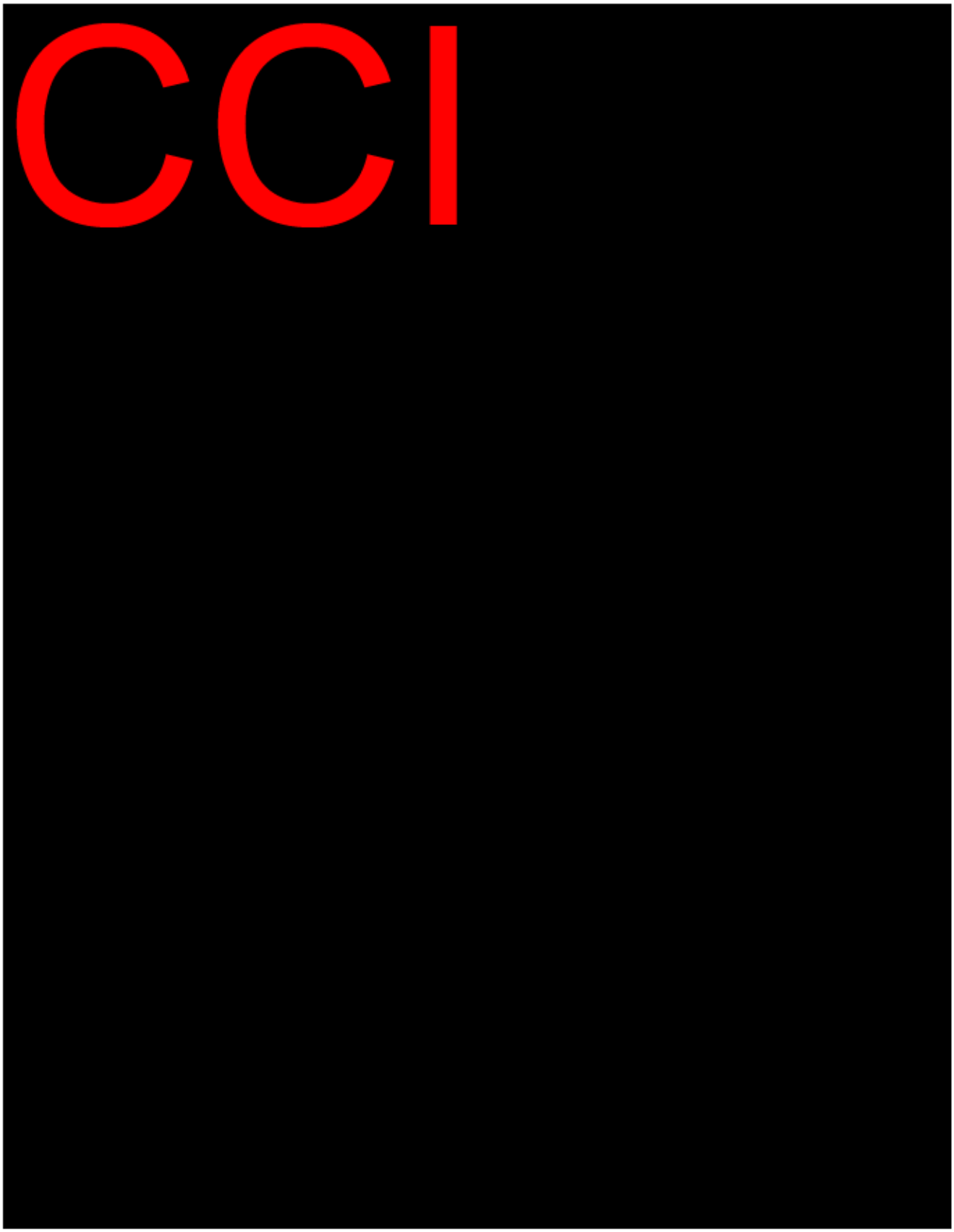
Main analysis for DRR will be based on participants' tumor responses either as assessed by the investigator or by the IRC according to the rule defined in Table 7.

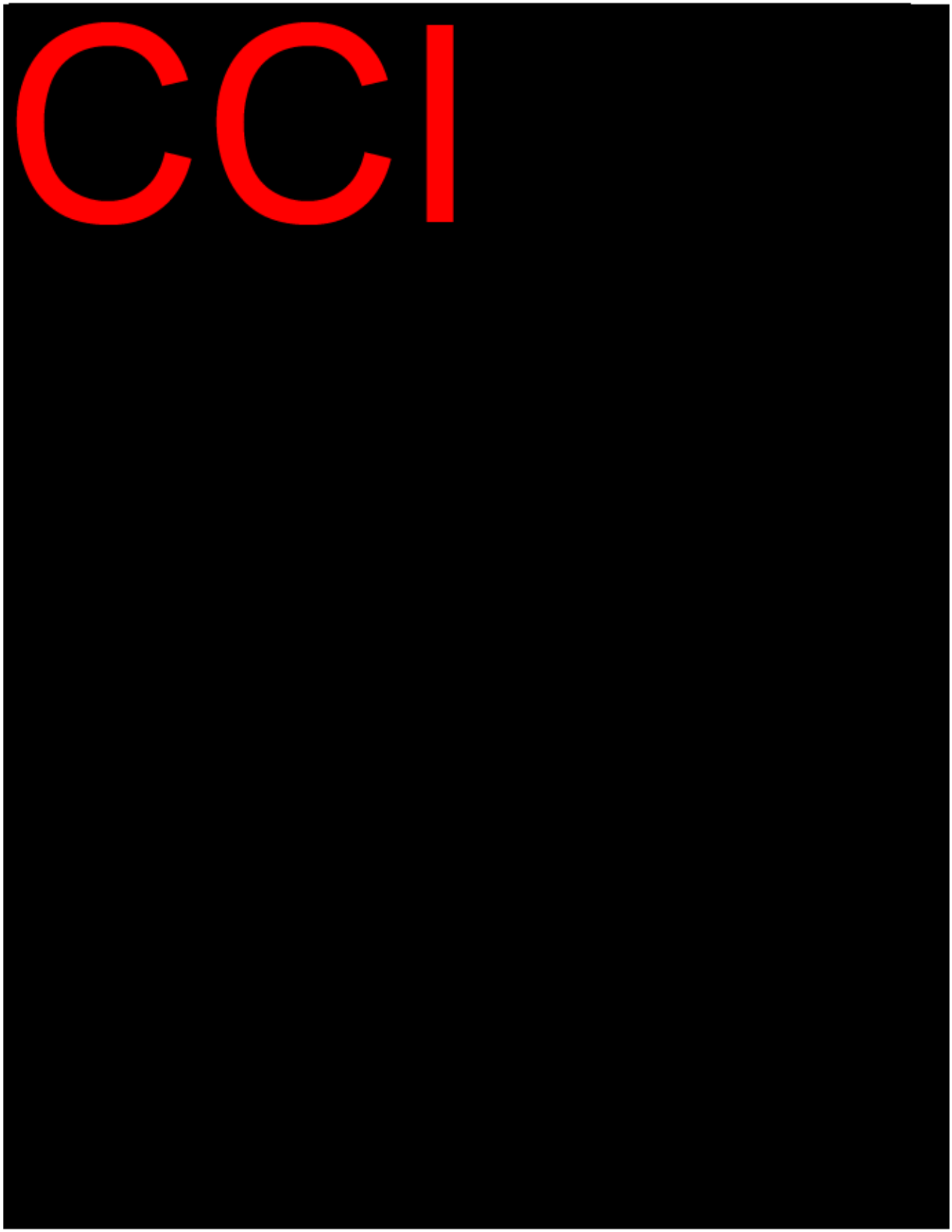
The DRR will be provided per treatment arm with a two-sided 95% CI using the Clopper-Pearson method (exact CI for a binomial proportion as computed by default by the FREQ procedure using the EXACT option).

The association between DRR and treatment as well as the magnitude of the treatment effect (stratified odds ratio obtained from CMH method) will be assessed using the same methodology as described in Section 14.3.1.

14.5.2 Sensitivity Analysis for DRR

Main analysis described for the secondary endpoint DRR will also be performed from tumor responses based on investigator or IRC assessment according to Table 7. No formal test will be performed.


CCI


CCI

15 Safety Analyses

Analysis set: SAFRI / SAF

The following subsections include specifications for summarizing safety endpoints that are common across clinical studies such as AE, laboratory tests and vital signs.

15.1 Adverse Events

Treatment-emergent adverse events are those events with onset dates occurring during the on-treatment period or if the worsening of an event is during the on-treatment period.

Changes in toxicity grade, seriousness or outcome of AEs are recorded as separate entries in the eCRF with associated end and start dates (start date equals end date of previous entry). Such entries reporting the same event in such immediately consecutive periods will be considered as one event in the analysis. These events will be kept as separate records in the database in order to maintain the full detailed history of the events. The start date of the initial record in the sequence is taken as the start date of the entire event, similarly the end date of the last event in the sequence is taken as the end date of the entire event. The overall outcome of the AE is the outcome of the last event in the sequence. Duration of the AE and the TEAE flag will be adjusted accordingly in the analysis.

All analyses described in this section will be based on TEAEs, unless otherwise specified, and will be described by treatment arm. The AE listings will include all AEs. AEs outside the on-treatment period (prior or after) will be flagged in listings.

Incomplete AE-related dates will be handled as stated in Section 9.9 (Table 6).

Dose-limiting Toxicity Adverse Events are those AEs reported on the "Adverse Event Details" eCRF page with the "Is this adverse event a dose limiting toxicity" or "Is this adverse event a dose limiting toxicity upon SMC review?" fields ticked "Yes".

Bintrafusp alfa/placebo Related Adverse Events are those AEs with relationship to bintrafusp alfa/placebo reported by the Investigator as related (i.e., answer to the question "Relationship with study treatment" = "Related" on "Adverse Event Details" eCRF page) and those of unknown relationship (i.e., no answer to the question "Relationship with M7824").

Chemotherapy Related Adverse Events are those AEs with relationship to gemcitabine or cisplatin defined as follows:

- For gemcitabine reported by the Investigator as related (i.e., answer to the question "Relationship with Gemcitabine" = "Related" or "missing" on "Adverse Event Details" eCRF page)
- For cisplatin reported by the Investigator as related (i.e., answer to the question "Relationship with Cisplatin" = "Related" or "missing" on "Adverse Event Details" eCRF page)

Treatment-related Adverse Events are bintrafusp alfa/placebo and/or chemotherapy related adverse events.

Serious Adverse Events are those AEs reported on the "Adverse Events Details" eCRF page, with the "Serious Adverse Event" field ticked "Yes".

Adverse Events Leading to Dose Reduction of Chemotherapy are those AEs leading to dose reduction of gemcitabine or cisplatin defined as follows:

- For gemcitabine consider answer to the question "Action(s) taken with Gemcitabine" = "Dose reduced" on "Adverse Event Details" eCRF page)
- For cisplatin consider answer to the question "Action(s) taken with Cisplatin" = "Dose reduced" on "Adverse Event Details" eCRF page)

Adverse Events Leading to Temporary Discontinuation of Bintrafusp alfa/placebo are those AEs leading to temporary discontinuation of bintrafusp alfa/placebo (answer to the question "Action(s) taken with Study treatment" = "Drug interrupted" on "Adverse Event Details" eCRF page).

Adverse Events Leading to Temporary Discontinuation of Chemotherapy are those AEs leading to temporary discontinuation of gemcitabine or cisplatin defined as follows:

- For gemcitabine consider answer to the question "Action(s) taken with Gemcitabine" = "Drug interrupted" on "Adverse Event Details" eCRF page
- For cisplatin consider answer to the question "Action(s) taken with Cisplatin" = "Drug interrupted" on "Adverse Event Details" eCRF page

Adverse Events Leading to Permanent Discontinuation of Bintrafusp alfa/placebo are those AEs leading to permanent discontinuation of bintrafusp alfa/placebo (i.e., answer to the question "Action(s) taken with Study treatment" = "Drug withdrawn" on "Adverse Event Details" eCRF page).

Adverse Events Leading to Permanent Discontinuation of Chemotherapy are those AEs leading to permanent discontinuation of gemcitabine or cisplatin defined as follows:

- For gemcitabine consider answer to the question "Action(s) taken with Gemcitabine" = "Drug withdrawn" on "Adverse Event Details" eCRF page
- For cisplatin consider answer to the question "Action(s) taken with Cisplatin" = "Drug withdrawn" on "Adverse Event Details" eCRF page

Adverse Events Leading to Death are those AEs leading to death (as recorded on the "Adverse Event Details" eCRF page, change in grade = "No" and outcome = "Fatal", or Grade = "Grade 5 or death related to AE" or Serious adverse event = "Yes" and seriousness criteria include "Results in death").

Adverse Events of Special Interest are those AEs identified according to a pre-specified search list of MedDRA Preferred Terms (PTs). Categories of AESI include:

- Infusion-related Reactions (IRRs)
- Immune-related Adverse Events (irAEs)
- Skin AEs possibly related to transforming growth factor beta (TGF-β) inhibition
- Anemia

  Anemia events are those AEs belonging to the MedDRA HLT Anaemias NEC, HLT Anaemias haemolytic immune, HLT Anaemias haemolytic NEC or PT = Haemoglobin decreased.

Bleeding events are those AEs belonging to the MedDRA SMQ Haemorrhage terms (excluding laboratory terms).

15.1.1 All Adverse Events

Adverse events will be summarized by worst severity (according to National Cancer Institute - Common Terminology Criteria for Adverse Events [NCI-CTCAE] version 5.0) per participant, using the latest version of MedDRA, PT as event category and MedDRA primary SOC body term as Body System category.

Unless otherwise stated, AEs will be displayed in terms of frequency tables: PTs and primary SOCs in alphabetical order.

Each participant will be counted only once within each PT or SOC. If a participant experiences more than one AE within a PT or SOC for the same recording period, only the AE with the strongest relationship or the worst severity, as appropriate, will be included in the summaries of relationship and severity.

If an adverse event is reported for a given participant more than once during treatment, the worst severity and the worst relationship to study treatment will be tabulated. In case a participant had events with missing and non-missing grades, the maximum of the non-missing grades will be displayed.

A table presenting the overall summary of AEs will be presented including the frequency (number and percentage) of participants within each of the following categories:

- TEAEs
- Treatment-related TEAEs
- Bintrafusp alfa/placebo related TEAEs
- Chemotherapy related TEAEs
- Serious TEAEs
- Treatment-related serious TEAEs
- Bintrafusp alfa/placebo related serious TEAEs
- Chemotherapy related serious TEAEs
- TEAEs NCI-CTCAE severity grade (≥3, ≥4)
- Treatment-related TEAEs NCI-CTCAE severity grade (≥3, ≥4)
- Bintrafusp alfa/placebo related TEAEs NCI-CTCAE severity grade (≥3, ≥4)
- Chemotherapy related TEAEs NCI-CTCAE severity grade (≥3, ≥4)
- TEAEs leading to death
- Treatment-related TEAEs leading to death

- Bintrafusp alfa/placebo related TEAEs leading to death
- Chemotherapy related TEAEs leading to death
- AESIs and bintrafusp alfa/placebo related AESI:
  - IRRs (see Section 15.2.5.1 for definition)
  - irAEs (see Section 15.2.5.2 for definition)
  - TGF-β inhibition mediated skin adverse events
  - Anemia
- TEAEs, bleeding events (see Section 15.1 for definition)
- Bintrafusp alfa/placebo related TEAEs, bleeding events
- Chemotherapy related TEAEs, bleeding events

Frequency tables for TEAEs corresponding to each category in the overview table above will be provided by:

- MedDRA primary SOC (ordered alphabetically) and PT (ordered alphabetically), (except for AESI)

TEAEs, treatment-related TEAEs, bintrafusp alfa/placebo-related TEAEs and chemotherapy-related TEAEs will also be summarized by worst grade, and the most frequent PT (at least 5%) will be presented graphically by worst grade and PT with bar chart figures.

Requirements for Clinicaltrials.gov and EudraCT

Summary table for non-serious TEAEs excluding SAEs applying frequency threshold of 5% will be provided.

Listings of adverse events will contain the following information: cohort/planned arm, participant identifier, age, sex, race, dates of first and last administrations of study treatment, PT, reported term for the AE, start date, end date, duration of AE (in days), day relative to the first infusion, day relative to the most recent infusion prior to AE onset, relationship to study treatment, toxicity grade, action(s) taken, outcome, seriousness (yes/no), AESI infusion-related (yes/no), AESI immune-related (yes/no), TGF-β inhibition mediated skin AESI (yes/no), anemia AESI (yes/no).

The following listings will be provided with the relevant information:

- Listing of all AEs (whether treatment-emergent or not). TEAEs will be flagged.
- Listing of TEAEs.
- Listing of non-TEAEs for AEs occurring after enrollment (date of first signature of informed consent/date of first signature of first informed consent) but prior to the first dose of study treatment.
- Listing of AEs with onset or worsening after the on-treatment period.

Evaluation of Potential Effect of ADA on Safety of Bintrafusp alfa

The table providing overall summary of AEs, as described above will also be provided by ADA status (ever positive, never positive).

Listings of all AEs and all IRRs for ever-positive ADA participants (pre-existing, transient treatment-emergent, persistent-treatment emergent) will be prepared including participant identifier and showing the date(s) of the positive ADA result together with the AEs or IRRs. For the AEs and IRRs, start and stop date will be shown along with grade. Adverse events recorded during the period of two weeks prior to the positive ADA value until two weeks after the positive ADA value will be flagged.

Evaluation of COVID-19 effects on AEs

The direct effect of COVID-19 for AEs will be assessed by means of:

- An overview table of treatment-emergent AEs associated to COVID-19 by MedDRA primary SOC, high level group terms (HLGT), high level terms (HLT) and PTs (ordered alphabetically).
- A listing of COVID-19 related AEs. The following information will be provided:
    - Participant ID, country, age, sex, race
    - Date of first, last treatment with study drug
    - COVID-19-associated AE start date (day), COVID-19 associated AE stop date (day)
    - AE preferred term, verbatim
    - Toxicity grade
    - Seriousness
    - Relationship to treatment
    - Action taken
    - Outcome

The COVID-19 related AEs will be identified as specified in the "COVID-19 Data and Reporting for Merck Ongoing Studies managed by IQVIA" document version 1.1.

The overview table will be provided if more than 3 treatment-emergent AEs will be observed in the analysis population.

15.1.2 Adverse Events Leading to Discontinuation of Study Treatment

Frequency tables summarizing the following actions taken with study treatment will be presented by PT and primary SOC in alphabetical order:

- TEAEs leading to temporary drug interruption (at least one of the study treatments)
- Bintrafusp alfa/placebo related TEAEs leading to temporary drug interruption

- Chemotherapy related TEAEs leading to temporary drug interruption
- TEAEs leading to permanent treatment discontinuation (of at least one of the study treatments)
- Bintrafusp alfa/placebo related TEAEs leading to permanent treatment discontinuation
- Chemotherapy related TEAEs leading to permanent treatment discontinuation
- TEAEs leading to infusion rate reduction (of at least one of the study treatments)
- Treatment related TEAEs leading to infusion rate reduction
- TEAEs leading to dose reductions (of at least one of the chemotherapy regimens)
- Chemotherapy related TEAEs leading to dose reductions

In addition, the incidences for above items will be summarized in an overview table.

A listing of TEAEs leading to permanent treatment discontinuation will also be provided with the relevant information.

15.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

15.2.1 Deaths

All deaths, deaths within 30 days after the last dose of study treatment, death within 60 days after the first dose of study treatment (for all participants, the first dose will be the first dose of the any study treatment) as well as the primary reason for death will be tabulated based on information from the "Death" eCRF pages.

The following summaries will be provided:

- Number of deaths
- Number of deaths within 30 days after last dose of study treatment
- Number of deaths within 60 days after first dose of study treatment
- Primary reason for death
    - Progressive disease and/or disease related condition
    - Event unrelated to study treatment
    - Event related to study treatment
    - Unknown

In addition, date and cause of death will be provided in an individual participant data listing together with the following information: cohort/randomized arm, participant identifier, analysis sets to which the participant belongs to, age, sex, race, dates of first/last doses of study treatment, number of infusions, day relative to the first and the last infusion, autopsy (yes/no/unknown), AEs with fatal outcome (list PTs of AEs with outcome=Fatal, as well as Grade 5 AEs or Serious AEs

resulting in death), flag for death within 30 days of last dose of study treatment and flag for death within 60 days of first dose of study treatment.

15.2.2 Serious Adverse Events

The number of participants with SAEs will be described by SOC and PT:

- SAEs
- Treatment-related SAEs
- Bintrafusp alfa/placebo-related SAEs
- Chemotherapy-related SAEs

Bar charts presenting SAEs, treatment-related SAEs, bintrafusp alfa/placebo-related SAEs and chemotherapy-related SAEs of the most frequent PT (at least 5%) will be displayed by worst grade and PT.

A listing of SAEs will also be provided (see Section 15.1.1).

15.2.3 Dose-limiting Toxicity (Primary Endpoint Safety Run-in)

Criteria to define DLT are specified in protocol section 6.6.2. DLT will only be defined in the open-label safety run-in phase.

A table of the occurrence of DLT AE (defined in Section 15.1) will be provided including:

- Number and percentage of participants who experienced a DLT during the DLT evaluation period per investigator and per SMC
- Number of DLTs per participant experienced during the DLT evaluation period (1 / 2 / ≥3) per investigator and per SMC
- DLT experienced during the DLT evaluation period per investigator and per SMC by SOC and PT based on the latest available MedDRA version.

A listing of DLTs will also be provided with the relevant information (see Section 15.1.1).

15.2.4 Bleeding Events

Bleeding events are defined in Section 15.1. Bleeding events and trial drug related bleeding events (treatment-related, bintrafusp alfa/placebo-related and chemotherapy-related) will be summarized in a frequency table presenting SOC and PT sorted by alphabetical order. The worst grade per participant, per SOC, and per PT will be reported:

- Any grade (including missing grade)
- Grade 1
- Grade 2

- Grade 3
- Grade 4
- Grade 5

15.2.5 Adverse Events of Special Interest

15.2.5.1 Infusion-Related Reaction including Immediate Hypersensitivity

IRRs are defined as adverse events with PTs according to a pre-specified MedDRA search list, and are divided into two subcategories: "Reactions" and "Signs and symptoms" based on the temporal relationship between the day of the infusion and onset of the event:

Reactions of IRRs: should be considered when the onset is on the day of infusion (during or after the infusion if timing related to the infusion is available) or the day after the infusion (irrespective of resolution date) for any infusion-related reaction, drug hypersensitivity, anaphylactic reaction, hypersensitivity and/or Type 1 hypersensitivity.

Signs and symptoms of IRRs and hypersensitivity/allergic reactions: should be considered when the onset is on the day of infusion (during or after the infusion if timing related to the infusion is available) and resolved completely, with the end date on the same day of the infusion or the day after for any of the following: pyrexia, chills, flushing, hypotension, dyspnea, wheezing, back pain, abdominal pain and urticaria.

Note: all infusions (i.e., of bintrafusp alfa/placebo and chemotherapies) have to be considered.

IRRs, overall and by subcategories, will be summarized by the following variables:

- Number of participants with at least one IRR by the worst NCI-CTCAE toxicity grade (grade 1/ grade 2/ grade 3/ grade 4/ grade 5/ missing grade)
- Number of participants with IRR leading to permanent treatment discontinuation of each study treatments
- Time related to first onset (infusion 1/ infusion 2/ infusion 3/ infusion 4 or later) of each study treatment. The events will be assigned to the actual drug infusions that the participant received, not to the planned dates. An IRR is assigned to a drug infusion if its onset is on the same date (but not before dosing when timing related to the infusion is available) or the day following the drug infusion.

A frequency table of IRR AEs by worst grade, SOC, and PT will also be provided.

A listing of IRRs will be provided with the relevant information (see Section 15.1.1). One additional listing will display the study drug administration details together with the IRR including administration date (day) /time, reason for modification, type of modification, modification start time, use of pre-medication, IRR AE PT, IRR AE grade, IRR AE start day / stop day, and IRR AE timing related to infusion (if available).

15.2.5.2 Immune-related Adverse Events

irAEs will be identified programmatically. AEs which satisfy all the following criteria will be flagged as immune-related:

- The AE PT matches a PT on the list of pre-selected MedDRA terms.
- The AE onset or worsening occurs after the first bintrafusp alfa/placebo infusion and no more than 90 days after last bintrafusp alfa/placebo dose, or to the earliest date of subsequent anticancer therapy minus 1 day, whichever occurs first, unless otherwise stated.
- On the "AE" eCRF page, the question "Were corticosteroids, Immunosuppressants, hormonal therapy (e.g. Thyroid) or Insulin applied?" has the answer "Yes" selected.
- On the "imAE SPECIFIC QUESTIONS" eCRF page, either:
  1. The question "Does any of the following provide a clear etiology for the event, other than immune related AE?" has the answer "No" selected, indicating that the AE is not attributable to underlying cancer disease/PD, prior or concomitant medications/procedures, nor another medical condition such as an infection or pre-existing disease.

  OR

  2. The "imAE SPECIFIC QUESTIONS" eCRF page indicates that a biopsy was performed and the question "Is the histopathology/biopsy consistent with an immune-mediated event?" has the answer "Yes" selected.

Where criteria (1) through (3) are met, and entries for condition (4) are missing, the following rules will be applied:

- If the answer to "Does any of the following provide a clear etiology for the event, other than immune related AE?" (4a) is missing, the event will be considered as irAE (irrespective of biopsy results).
- If the answer to "Is the histopathology/biopsy consistent with an immune-mediated event?" (4b) is missing, or if no biopsy was performed, and condition (4a) is not satisfied, i.e., "Yes" is selected (i.e., at least one (clear) etiology of the event is provided) as the answer to the question "Does any of the following provide a clear etiology for the event, other than immune related AE?", the event will be considered as a non-irAE.

PTs will be compiled into categories (subcategories): Immune-related rash, Immune-related colitis, Immune-related pneumonitis, Immune-related hepatitis, Immune-related nephritis and renal dysfunction, Immune-related endocrinopathies (adrenal insufficiency, hypogonadism, pituitary dysfunction, Type 1 diabetes mellitus, thyroid disorders), and Other immune-related adverse events (Anemia, Encephalitis, GVHD, Guillain-Barre Syndrome, Myasthenic syndrome, Myocarditis, Myositis, Neurologic events, Pancreatitis, Uveitis, Vasculitis, Other). PTs belonging to the category "Immune related endocrinopathies – Thyroid disorders" will also be compiled into sub-subcategories: Hyperthyroidism, Hypothyroidism, Thyroiditis.

irAEs will be summarized by the following variables:

- Any irAEs

- irAEs by the worst grade
- irAEs leading to permanent treatment discontinuation of each study treatments
- Serious irAEs

A frequency table of irAEs by worst grade, category, subcategory (for immune-related endocrinopathies and other immune-related adverse events), sub-subcategory (for immune-related endocrinopathies – thyroid disorders) and PT will also be provided.

A listing of irAEs will also be provided with the relevant information, including additional interventions for irAEs (e.g. biopsies, surgical procedures, medical procedures) (see Section 15.1.1).

15.2.5.3 TGF-β inhibition mediated skin adverse events

To identify potential skin AEs possibly related to TGF-β inhibition, MedDRA PT queries will be used to search for skin AEs of interest in the clinical database. A listing containing these pre-specified PT search terms will be generated. PTs will be compiled into two categories: Narrow definition, and Broad definition:

Narrow definition:

- Keratoacanthoma
- Squamous cell carcinoma of skin

Broad definition has additional PTs:

- Hyperkeratosis
- Actinic keratosis
- Basal cell carcinoma
- Lip squamous cell carcinoma
- Bowen´s disease

Further details (e.g. MedDRA PT queries) are regularly updated based on the current MedDRA version.

The overall summary of skin TEAE will include the following categories for narrow and broad definition:

- All skin TEAEs
- All skin TEAEs by worst grade
- Skin TEAEs leading to permanent treatment discontinuation of each study treatment
- Serious skin TEAEs

Frequency table for skin TEAEs by worst grade will be provided by MedDRA PTs (including both narrow and broad definition PTs).

A listing of skin AEs will also be provided. This listing will specifically include, for the AEs identified from the PT list, the number of lesions, if a biopsy or an excision was done and if it confirmed the diagnosis, and the location for each lesion coming from eCRF page "TGFβ MEDIATED SKIN REACTION" if available.

15.2.5.4 Anemia

Anemia events are defined in Section 15.1. Anemia events and trial drug related anemia events (treatment-related, bintrafusp alfa/placebo-related and chemotherapy-related) will be summarized in a frequency table presenting SOC and PT sorted by alphabetical order. The worst grade per participant, per SOC, and per PT will be reported. A table displaying the overall summary of Anemia AEs by treatment arm will be also presented.

A listing of anemia adverse event will be provided with the relevant information (see Section 15.1.1).

Dedicated listings may also be given:

- TEAE anemia ≥ grade 3
- Serious TEAE anemia with ≥ grade 3

15.2.6 Three-tier Approach to Summarizing and Analyzing Adverse Events

The 3-tier approach is a systematic way to summarize and analyze adverse events in clinical studies. AEs are categorized in different tiers and analyzed using different levels of statistical analysis. All analyses will be performed on TEAEs only.

The 3-tier approach will be performed only if the study is expanded in Phase III.

Risk Measures

The following statistics will be provided to summarize safety data:

- **Crude Rate:** is calculated as the number of participants with a specific AE out of the total number of participants at risk expressed as percentage.
- **Exposure Adjusted Incidence Rate (EAIR):** is defined as the number of participants with a specific AE out of the total exposure-time among the participants in the treatment arm at risk of the initial occurrence of the event. If a participant has multiple events, the exposure period of the first event will be used. For a participant with no event, the exposure period will be censored at the last follow-up time for the AE summarization period. EAIR will be measured in 100 participant years, unless a different provides a more reasonable number of digits for reporting purpose.

Tier 1, 2, and 3 Identifications

All AEs will be classified into Tier 2 or Tier 3 based on the Rule-of-4. If there are 4 or more participants with the reported preferred term in any treatment group that preferred term will be included in Tier 2. Otherwise it will be included in Tier 3.

Further to this, Tier 1 AEs will be identified from a list of pre-selected MedDRA terms (using the latest version of MedDRA available). In case these events fulfill the Rule-of-4, Tier 1 analysis methods will be used. Otherwise, analyses will be done according to Tier 3. The following list of pre-selected composite terms based on specified MedDRA queries will be used:

- Immune-related adverse events
    - Immune-related pneumonitis
    - Immune-related colitis
    - Immune-related hepatitis
    - Immune-related endocrinopathies
        - Thyroid disorders
            - Hypothyroidism
            - Hyperthyroidism
            - Thyroiditis
        - Adrenal insufficiency
        - Type 1 Diabetes Mellitus
        - Pituitary dysfunction
        - Hypogonadism
    - Immune-related nephritis and renal dysfunction
    - Immune-related rash
    - Other immune-related adverse events

- Infusion-related reactions
    - Reactions
    - Signs and Symptoms
- TGF-β inhibition mediated skin reactions
- Anemia, as defined in Section 15.2.5.4
- Bleeding adverse events
- Impaired wound healing

Summary statistics

The following summaries will be presented:

For the pre-specified terms

- Number of participants and percentage by pre-specified terms in each treatment arm accompanied by the difference in crude rate between treatment arms with 95% CI. The CI will be generated using the Miettinen & Nurminen (MN) method. CI will not be provided for terms not fulfilling the rule-of-4.
- Forest plot showing crude rate and crude rate difference for each pre-specified term in Tier 1 with 95% MN CI
- Number of participants, number of participant-years and EAIR with 95% CI using Poisson method by pre-specified terms in each treatment arm. In addition, the difference in EAIR between treatment arms along with 95% CI will be shown. The CI for EAIR difference will be estimated based on MN or equivalent statistical methods depending on their computational efficiency. CIs will not be provided for terms belonging to Tier 3.
- Forest plot showing EAIR and EAIR difference for each pre-specified term in Tier 1 with 95% CI

For the any TEAEs:

- Number of participants and percentage by PT in Tiers 2 and 3 and overall in each treatment arm accompanied by the difference in crude rate between treatment arms with 95% CI. CI will be generated using MN method. CI will not be provided for terms belonging to Tier 3.
- Forest plot showing crude rate and crude rate difference for each Tier 2 TEAEs with 95% MN CI
- Number of participants, number of participant-years and EAIR with 95% CI using Poisson method by PT in Tiers 2 and 3 in each treatment arm. In addition, the difference in EAIR between treatment arms along with 95% CI will be shown. The CI for EAIR difference will be estimated based on MN or equivalent statistical methods depending on their computational efficiency. CIs will not be provided for terms belonging to Tier 3.
- Forest plot showing EAIR and EAIR difference for each Tier 2 TEAEs with 95% CI

No multiplicity adjustment will be applied for the analysis of TEAEs.

For presentation of tier 2 and 3 events in the main body of the CSR text, a cut of 5% in any of the treatment arms will be applied, while the corresponding end-of-text table will provide all of tier 2 and 3 events.

15.3 Clinical Laboratory Evaluation

Baseline and on-treatment laboratory values (including corresponding normal ranges), converted into standard units, will be used for summary statistics and shift tables.

Laboratory results will be classified according to the NCI-CTCAE criteria version 5.0 and as specified in Appendix 3. Additional laboratory results that are not part of NCI-CTCAE will be

categorized as follows: below normal limits, within normal limits, and above normal limits (according to the original laboratory normal ranges).

All analyses described in this section will be described by treatment arm if not otherwise specified.

Quantitative data will be summarized using descriptive statistics (mean, StD, median, Q1, Q3, Min, and Max) of actual baseline values, on-treatment values, and changes from baseline to each on-treatment visit over time. Refer to Section 9.14 for further information about handling of unscheduled assessments. Qualitative data based on reference ranges will be described according to the categories (i.e., Abnormal, Normal).

Summary tables over time will present summary statistics for continuous and categorical variables by timepoint.

The following figures will be provided:

- Boxplots of the laboratory values by timepoint
- Boxplots of the change from baseline by timepoint

Boxplots for laboratory parameters where toxicity grades are defined based on the ratio of the parameter value and the upper limit of normal (ULN) will be displayed using the ratio of the measured value over ULN. This comprises alkaline phosphatase (ALP), alanine aminotransferase (ALT), aspartate aminotransferase (AST), Activated Partial Thromboplastin Time (aPTT), bilirubin, and creatinine.

Laboratory parameters with NCI-CTC grades available

Laboratory parameters with NCI-CTC grades available will be analyzed with their respective NCI-CTC name and direction of abnormality. For parameters that are graded with both low and high values, such as potassium (hypokalemia/hyperkalemia), the toxicities will be summarized separately. Low direction toxicity (e.g. hypokalemia) grades at baseline and post-baseline will be set to 0 when the variables are derived for summarizing high direction toxicity (e.g. hyperkalemia), and vice versa.

The following summaries will be displayed:

- Number and percentage of participants by worst on-treatment grade (>=1, >=3, >=4)
- Shift in toxicity grading from baseline to highest on-treatment toxicity grade

The definitions of the NCI-CTCAE toxicity grading version 5.0 are provided in Appendix 3.

Table 11 NCI-CTC Gradable Parameters

Parameter	Parameter Code	Name in NCI-CTC	Direction of Abnormality
Biochemistry			

Parameter	Parameter Code	Name in NCI-CTC	Direction of Abnormality
Alanine Aminotransferase	ALT	Alanine aminotransferase increased	High
Albumin	ALB	Hypoalbuminemia	Low
Alkaline Phosphatase	ALP	Alkaline phosphatase increased	High
Amylase	AMYLASE	Serum amylase increased	High
Aspartate Aminotransferase	AST	Aspartate aminotransferase increased	High
Bilirubin total	BILI	Blood bilirubin increased	High
Calcium[a]	CA	Hypercalcemia/Hypocalcemia[a]	High/Low
Creatinine	CREAT	Creatinine increased	High
Gamma Glutamyl Transferase	GGT	GGT increased	High
Glucose	GLUC	Hypoglycemia[b]	Low
Lactate Dehydrogenase	LDH	Blood lactate dehydrogenase increased	High
Lipase	LIPASET	Lipase increased	High
Magnesium	MG	Hypermagnesemia/Hypomagnesemia	High/Low
Potassium	K	Hyperkalemia/Hypokalemia	High/Low
Sodium	SODIUM	Hypernatremia/Hyponatremia	High/Low
Creatine Kinase	CK	CPK increased	High
Hematology			
Eosinophils	EOS	Eosinophilia	High
Lymphocytes	LYM	Lymphocyte count decreased/ Lymphocyte count increased	Low/High
Neutrophils	NEUT	Neutrophil count decreased	Low
Leukocytes	WBC	White blood cell decreased / Leukocytosis	Low/High
Hemoglobin	HGB	Anemia/Hemoglobin increased	Low/High
Platelet count	PLAT	Platelet count decreased	Low
Coagulation			
Activated Partial Thromboplastin Time[c]	APTT	Activated partial thromboplastin time prolonged[c]	High
Activated Partial Thromboplastin Time/standard thromboplastin time[c]	APTTSTND	Activated partial thromboplastin time prolonged/standard thromboplastin time[c]	High
Prothrombin International Normalized Ratio[c]	INR	INR increased[c]	High

[a] Based on corrected calcium (see Appendix 3).

[b] Hyperglycemia is defined as per NCI-CTCAE V5.0 but grades cannot be computed based on glucose values only.

[c] Reported on the "Coagulation" eCRF page.

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

- Corrected Calcium (mg/dL) = Calcium (mg/dL) – 0.8 [Albumin (g/dL)-4]

or

- Corrected calcium (mmol/L) = measured total calcium (mmol/L) + 0.02 (40 – serum albumin [g/L])

Laboratory parameters with NCI-CTC grades not available

Table 12 Non-NCI-CTC Gradable Parameters

Parameter (LBTEST)	
Biochemistry	
Bilirubin direct	BILIDIR
Bilirubin Indirect	BILINDIR
Chloride	CL
C-Reactive Protein	CRP
Phosphate[a]	PHOS
Total Protein	PROT
Urea Nitrogen	BUN
Creatinine Clearance	CREATCLR
Hematology	
Absolute Reticulocytes	RETI
Basophils	BASO
Basophils/Leukocytes	BASOLE
Eosinophils/Leukocytes	EOSLE
Erythrocytes (RBC)	RBC
Hematocrit	HCT
Lymphocytes/Leukocytes	LYMLE
Mean Corpuscular Hemoglobin	MCH
Mean Corpuscular HGB Concentration	MCHC
Mean Corpuscular Volume	MCV
Monocytes	MONO
Monocytes/Leukocytes	MONOLE
Neutrophils/Leukocytes	NEUTLE
Reticulocytes/Erythrocytes	RETIRBC
Coagulation	
Prothrombin Time[b]	PT

[a] Phosphate is a gradable parameter as per NCI-CTCAE V5.0 but grades cannot be computed based on phosphate results only.
[b] Reported on the "Coagulation" eCRF page.

For **WBC differential counts** (neutrophil, lymphocyte counts), the absolute value will be used when reported. When only percentages are available (this is mainly important for neutrophils and

lymphocytes, because the CTCAE grading is based on the absolute counts), the absolute value is derived as follows:

Derived differential absolute count = (WBC count) * (Differential %value / 100)

For all non-gradable parameters, the following summaries will be displayed:

- Number and percentage of participants by lowest on-treatment value (classified as normal, high, low)
- Number and percentage of participants by highest on-treatment value (classified as normal, high, low)
- Shift from baseline to highest/lowest on-treatment value (classified as normal, high, low)

Separate listings of hematology, biochemistry and coagulation will be created. Each listing will include: cohort/planned arm, participant identifier, age, sex, race, first dose date, last dose date, laboratory parameter (units), visit, date/time (study day), SI value, change in SI value from baseline, LLN, ULN, indicator of normal range (low, normal, high), toxicity grade according to NCI-CTCAE (when applicable) and highest/lowest on-treatment value flag. Baseline and post-baseline values after the on-treatment period will be flagged.

Liver function tests

ALT, AST, and total bilirubin will be used to assess possible drug induced liver toxicity. The ratios of test result over ULN will be calculated and classified for these three parameters during the on-treatment period.

The number and percentage of participants within each of the following liver function categories during the on-treatment period will be described:

- ALT <3×ULN, ALT ⩾ 3×ULN, ALT ⩾ 5×ULN, ALT ⩾ 10×ULN, ALT ⩾ 20×ULN
- AST <3×ULN, AST ⩾ 3×ULN, AST ⩾ 5×ULN, AST ⩾ 10×ULN, AST ⩾ 20×ULN
- (ALT and AST) <3×ULN, (ALT or AST) ≥ 3×ULN, (ALT or AST) ≥ 5×ULN, (ALT or AST) ≥ 10×ULN, (ALT or AST) ≥ 20×ULN
- BILI < 2×ULN, BILI ≥ 2×ULN
- Concurrent ALT ≥ 3×ULN and BILI ≥ 2×ULN
- Concurrent AST ≥ 3×ULN and BILI ≥ 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and BILI ≥ 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and BILI ≥ 2×ULN and ALP > 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and BILI ≥ 2×ULN and ALP ≤ 2×ULN or missing

Concurrent measurements are those occurring on the same date.

Categories will be cumulative, i.e., a participant with an elevation of AST ≥10×ULN will also appear in the categories ≥5×ULN and ≥3×ULN.

A plot of peak ALT versus peak total bilirubin, both relative to the ULN will be provided. This eDISH plot (evaluation of drug-induced serious hepatotoxicity) will be divided into 4 quadrants by the lines through ALT = 3×ULN and total bilirubin = 2×ULN. The left lower quadrant is then considered normal or insignificant elevations in liver chemistries, the upper left quadrants indicate participants with possible Gilbert's cholestasis; the right upper quadrant are the possible Hy's Law participants; the right lower quadrant shows possible Temple's Corollary (participants with ALT≥ 3 x ULN but not satisfying Hy's Law). The same plot will be provided for AST.

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

Urinalysis / urinalysis microscopic evaluation, anemia, hormone tests, serum, serology, cancer antigen 19.9

All test results for urinalysis /urinalysis microscopic evaluation, hormone tests, serums and serology parameters will be listed in dedicated listings:

- Urinalysis parameters:
  - Urinalysis full parameters: physical appearance, pH, specific gravity, protein, glucose, ketones, nitrite, leukocyte esterase, blood, urobilinogen, bilirubin, color
  - Urinalysis microscopic parameters: erythrocytes (RBC), leukocytes (WBC), epithelial cells, bacteria, crystals, casts
- Anemia parameters: red cell volume distribution width standard deviation (RDW-SD), red cell distribution coefficient of variation (RDW-CV), total iron binding capacity, transferrin, iron, ferritin, vitamin B12, serum folate, absolute reticulocytes
- Hormone parameters: thyroxine free (Free T4), thyrotropin (thyroid-stimulating hormone; TSH)
- Cancer antigen 19.9 (CA 19-9)
- Serum parameters: Krebs von den Lungen-6 (KL-6), surfactant protein A (SP-A), surfactant protein D (SP-D) (only for Japanese sites)
- Serology parameters: hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis C RNA

Pregnancy test

Results for pregnancy and post-menopausal status as collected on the "Pregnancy Test" eCRF page will also be listed:

- Pregnancy parameters (serum or highly sensitive urine human chorionic gonadotropin [hCG])
- Post-menopausal status

15.4 Vital Signs

The following summaries will be prepared for vital signs: body temperature, pulse rate, systolic blood pressure (SBP), diastolic blood pressure (DBP), respiration rate and oxygen saturation considering only participants with post-baseline values during on-treatment period:

- Summary statistics over time

The following potentially clinically significant abnormalities will be summarized:

- $\leq$ 95 mmHg and decrease from baseline $\geq$ 20 mmHg in systolic blood pressure
- $\geq$ 140 mmHg and increase from baseline $\geq$ 20 mmHg in systolic blood pressure
- $\leq$ 45 mmHg and decrease from baseline $\geq$ 20 mmHg in diastolic blood pressure
- $\geq$ 90 mmHg and increase from baseline $\geq$ 20 mmHg in diastolic blood pressure
- $\leq$ 50 beats/min and decrease from baseline $\geq$ 20 beats/min in heart rate
- $\geq$ 100 beats/min and increase from baseline $\geq$ 20 beats/min in heart rate
- $<$ 20 breaths/min and decrease from baseline $\geq$ 5 breaths/min in respiratory rate
- $\geq$ 20 breaths/min and increase from baseline $\geq$ 5 breaths/min in respiratory rate
- $\geq$ 10% weight increase
- $\geq$ 10% weight decrease

A listing of vital signs will be provided including cohort/planned arm, participant identifier, age, sex, race, vital sign parameter, visit, timepoint, date, time, value, unit, and change from baseline. Baseline values and post-baseline values collected after the on-treatment period will be flagged.

15.5 Other Safety or Tolerability Evaluations

ECG

Single 12-lead ECGs will be obtained as outlined in the Schedule of Activities (Table 1) in the protocol, and ECG parameters will be described from the data collected on the "Electrocardiogram" eCRF page. Shift table for ECG interpretation value at baseline to the value at end of treatment will be provided.

A listing of ECG values will be provided including cohort/planned arm, participant identifier, age, sex, race, ECG parameter and unit, visit/time (study day), ECG date, value, change from baseline. Baseline values and post-baseline values collected after the on-treatment period will be flagged. Qualitative ECG results will also be provided.

ECOG Performance Status

The ECOG Performance Status will be derived from the data collected on the "ECOG Performance Status" eCRF page.

The ECOG shift from baseline to the highest score during the on-treatment period will be summarized.

ECOG performance status will also be presented in a listing at each timepoint.

16 Analyses of Other Endpoints

16.1 Pharmacokinetics

The analyses described in this section will be performed by the Clinical PK/Pharmacodynamics group (CPK) of Translational Medicine, Merck Healthcare KGaA, Darmstadt, Germany, or by a Contract Research Organization selected by the Sponsor. Pharmacokinetic listings and individual data will be presented based on the Safety Analysis Set. Summaries and statistical analyses will be based on the PKAS. Integrated analyses across studies, such as the Population PK analysis, will be described separately from this IAP and presented separately from the main CSR.

Pharmacokinetic concentrations/PK parameters mentioned herein refer to those of bintrafusp alfa.

16.1.1 Missing/non-quantifiable/unscheduled PK Data Handling

Concentrations Below the Lower Limit of Assay Quantification

Pharmacokinetic concentrations below the lower limit of quantification (BLQ) will be set to zero for calculating parameters and descriptive statistics.

Pharmacokinetic concentrations that are BLQ and occur before the last quantifiable data point will be taken as zero for calculating PK parameters of single dose profiles. Concentrations that are BLQ and occur after the last quantifiable data point will not be considered in the calculation of the terminal first order rate constant (λz).

Unscheduled Samples

For PK analysis, unscheduled samples will not be linked to a scheduled timepoint and will be excluded from summaries.

Deviations, Missing Concentrations, and Anomalous Values

Pharmacokinetic parameters will be calculated using the actual elapsed time since dosing. When the actual sampling time is missing, calculations will be performed using the scheduled time. Otherwise, there will be no further imputation of missing data. Concentrations will be set to missing in summary tables if the value is reported as no result.

Pharmacokinetic concentrations which are erroneous due to a protocol violation (as defined in the clinical study protocol and/or Section 8.1), documented handling error, or analytical error (as documented in the protocol deviation log, bioanalytical data, and/or bioanalytical report) may be excluded from the PK analysis if agreed upon prior to performing the PK analysis. In this case the rationale for exclusion will be provided in the CSR. Any other PK concentrations that appear

implausible to the Pharmacokineticist/PK[redacted] Data Analyst will not be excluded from the analysis. Any implausible data will be documented in the CSR.

Exclusions for concentration data descriptive statistics:

- Positive pre-dose values on Day 1
- Concentration observed at the end of infusion (C_{EOI}) < lower limit of quantification (LLOQ)
- In case of missed dose, exclude all concentrations until intended dosing is resumed
- Concentration observed at the end of the dosing interval (C_{trough}) values in case samples are taken at least 7 days late or early

Exclusions for Non-Compartmental Analysis (NCA):

- Positive pre-dose values on Day 1
- C_{EOI} < LLOQ
- In case of missed dose, exclude all concentrations until intended dosing is resumed

Any other PK concentrations that appear implausible to the Pharmacokineticist/PK[redacted] Data Analyst will not be excluded from the analysis. Any implausible data will be documented in the CSR.

If a PK parameter cannot be derived from a participant's concentration data, the parameter will be coded as NC (not calculated) (Note that NC values will not be generated beyond the day that a participant discontinues the treatment). For statistical analyses (i.e., descriptive statistics), PK parameters coded as NC will be set to missing.

If an individual participant has a known biased estimate of a PK parameter (due for example to a deviation from the assigned dose level), this participant/value will be excluded from the descriptive statistics based on the PK analysis set and instead the result will be listed only.

Relevant decisions on participant inclusion in the PK analysis set will be made in a two-step process:

3. Before database lock in a blinded data review meeting.

4. After database lock and unblinding, the PK analysis set may be updated based upon unblinded data relevant to PK.

16.1.2 Descriptive PK Analysis

Presentation of PK Concentration Data

Listings

Individual PK sample times, time deviations, and concentration data will be listed by part, participant, regional cohort (Asian or non-Asian), study day, and nominal time. Concentration listings will be based on the safety run-in analysis set for the safety run-in part and the SAF for the randomized, double-blind part. Pharmacokinetic concentrations will be reported with the same

precision as the source data provided by the bioanalytical laboratory. Actual elapsed sample collection times will be rounded to two decimal places with units of hours for reporting purposes in listings.

Tables

Pharmacokinetic concentration data for the first dose in the safety run-in part will be presented in tables and descriptively summarized for the PKAS by regional cohort, overall (both cohorts combined), and nominal time using: number of non-missing observations (n), arithmetic mean (Mean), geometric mean (GeoMean), StD, StD of log-transformed data (logStD), coefficient of variation (CV%), geometric coefficient of variation (GeoCV%), Min, median (Median), and Max. Summaries of pre-dose and end-of-infusion concentrations (other than first dosing interval) for the safety run-in part, as well as all pre-dose and end-of-infusion concentrations in the randomized, double-blind part, will be covered by the parameter summaries only.

Additional table(s) will summarize PK concentrations with further stratification by ethnicity (e.g., Japanese, non-Japanese Asian, all Asian, non-Asian, Chinese mainland, non-Chinese mainland Asian), based on the PKAS. Additional table(s) will summarize PK concentrations with further stratification by ADA subsets ever positive and never positive, and additionally "Treatment emergent (ALL)" versus "ADA Non-Treatment-emergent", based on the PKADA. Additional table(s) will summarize PK concentrations with further sub-stratification by nAb subsets ever positive and never positive, based on the PKNAB. For nAb subsets, summaries will be done for ever positive in either of 2 nAb assays (PD-L1 and TGFß receptor neutralization; see Section 16.4. Additional table(s) may summarize PK concentration at the time points with ADA measurement with stratification by the ADA results (Negative or Positive, see Section 16.4) at each time point based on the PKADA. Additional table(s) may summarize PK concentrations with further stratification by antibiotics status subgroups. Only subgroup sample size with a minimum of 3 participants will be displayed.

Descriptive statistics of PK concentration data will be calculated using values with the same precision as the source data provided by the bioanalytical laboratory and rounded for reporting purposes only. The following conventions will be applied when reporting descriptive statistics of PK concentration data:

Mean, Min, Median, Max, GeoMean:	3 significant digits
StD, logStD:	4 significant digits
CV%%, GeoCV%:	1 decimal place

Figures

Individual PK concentration-time profiles showing all participants by regional cohort will be created for serial sampling (i.e., safety run-in part, first dose) using the actual time points and the numeric concentration data. Arithmetic mean, GeoMean and Median concentration-time profiles by regional cohort and overall (both cohorts) for serial sampling (i.e., safety run-in part, first dose) will be provided using scheduled (nominal) time points and the numeric concentration data. All concentration-time plots for PK data will be presented both on a linear and on a semi-logarithmic

scale. Mean and GeoMean PK plots will include StD or logStD error bars when plotted on a linear scale. Plots of individual data will be presented based on the safety run-in analysis set, and summary plots will be based on the PKAS.. Figures of pre-dose and end-of-infusion concentrations other than first dosing interval will be covered by the parameter summaries only.

Additional figures will present Mean, GeoMean, and Median PK concentration-time profiles with further stratification by ethnicity (e.g., Japanese, non-Japanese Asian, all Asian, non-Asian, Chinese mainland, non-Chinese mainland Asian), based on the PKAS. Additional figures will present Mean, GeoMean, and Median PK concentration-time profiles with further stratification by ADA subsets ever positive and never positive, and additionally "Treatment emergent (ALL)" and "ADA Non-Treatment-emergent", based on the PKADA. Additional box-whisker plots may be presented of PK concentration at the time points with ADA measurement with stratification by the ADA results (Negative or Positive, see Section 16.4) at each time point based on the PKADA. Additional figures will present Mean, GeoMean and Median PK concentration-time profiles with further sub-stratification by nAb subsets ever positive and never positive (see Section 16.4 based on the PKNAB. Additional figure(s) of individual data overlaid will be presented with further stratification by ADA subgroups (ever positive and never positive). Only subgroup sample size with a minimum of 3 participants will be displayed.

16.1.3 Pharmacokinetic Non-Compartmental Analysis

The PK parameters listed below will be calculated for bintrafusp alfa (if estimable) using the actual time elapsed from dosing (or using scheduled time if actual time is not available, e.g., any analyses prior to final analysis) by non-compartmental methods. Pharmacokinetic concentrations will be analyzed with the same precision as the source data provided by the bioanalytical laboratory. Actual elapsed sample collection times will be analyzed unrounded (maximum of 14 significant digits). The pre-dose samples will be considered as if they had been taken simultaneously with the start of infusion (i.e., time zero).

C_{EOI}	The concentration observed immediately at the end of infusion. This will be taken directly from the observed bintrafusp alfa concentration-time data.
C_{trough}	The concentration observed immediately before next dosing (corresponding to pre-dose or trough concentration for multiple dosing). This will be taken directly from the observed bintrafusp alfa concentration-time data.
AUC_{0-t}	Area under the concentration-time curve (AUC) from time zero (i.e., infusion start time) to the last sampling time (t_{last}) at which the concentration is at or above the lower limit of quantification. (SAFRI part only; first dose only).
$AUC_{0-\infty}$	The AUC from time zero (infusion start time) extrapolated to infinity, based on the predicted value for the concentration at t_{last} ($C_{last\ pred}$), as estimated using the linear regression from λ_Z

	determination. $AUC_{0-\infty} = AUC_{0-t}+C_{last\ pred}/\ \lambda_Z$. (SAFRI part only; first dose only).
$t_{1/2}$	Apparent terminal (elimination) half-life, calculated by ln2/ λ_Z. (SAFRI part only; first dose only).
λ_Z	Terminal first order (elimination) rate constant, determined from the terminal slope of the log-transformed concentration-time curve using linear regression on terminal data points of the curve. (SAFRI part only; first dose only).
C_{max}	Maximum observed concentration. (SAFRI part only; first dose only).
t_{max}	The time to reach the maximum observed concentration collected during a dosing interval (unless otherwise defined, take the 1st occurrence in case of multiple/identical C_{max} values). (SAFRI part only; first dose only).
CL	Total body clearance. $CL = Dose/AUC_{0-\infty}$. (SAFRI part only; first dose only).
V_z	Volume of distribution during terminal phase following intravenous dosing. $V_z = Dose/(AUC_{0-\infty}*\lambda_Z)$. (SAFRI part only; first dose only).

The following PK parameters (SAFRI part only; first dose only) will be calculated for diagnostic purposes and listed, but will not be summarized:

- The time interval (h) of the log-linear regression to determine λ_Z ($t_{1/2}$, Interval).
- The starting time point (h) of the time interval of the log-linear regression to determine λ_z (Lambda_z_low).
- The ending time point (h) of the time interval of the log-linear regression to determine λ_z (Lambda_z_upp).
- Number of data points ($t_{1/2}$, N) included in the log-linear regression analysis to determine λ_Z.
- Goodness-of-fit statistic (adjusted Rsq) for calculation of λ_Z.
- Percentage of $AUC_{0-\infty}$ obtained by extrapolation ($AUC_{extra\%}$), calculated by $(1 - [AUC_{0-t}/AUC_{0-\infty}])\times 100$.

The regression analysis (determination of λ_z) should contain as many data points as possible (but excluding C_{max}) and has to include concentration data from at least 3 different time points, consistent with the assessment of a straight line (the terminal elimination phase) on the log-transformed scale. The observation period over which the regression line is estimated should be at least two-fold the resulting $t_{1/2}$ itself. If $AUC_{extra\%}$ is >20.0% and/or adjusted Rsq of λ_z is <0.800, and/or $t_{1/2}$, Interval is less than two-fold the resulting $t_{1/2}$, then λ_z and all derived parameters

(e.g. $t_{1/2}$, $AUC_{0-\infty}$, CL, etc.) will be listed, flagged, and discussed/reviewed with the Merck Clinical PK/CCI Scientist on a case-by-case basis for potential exclusion from descriptive statistics.

The calculation of the AUCs will be performed using the mixed log-linear trapezoidal method (linear up, log down). Extrapolated areas will always be computed using the predicted last concentration that is estimated using the linear regression from terminal rate constant determination.

Presentation of PK Parameter Data

Listings

Individual PK parameter data will be listed by part, participant, regional cohort, and study day. Parameter listings will be based on the SAFRI analysis set for the safety run-in part and the SAF for the randomized, double-blind part. Pharmacokinetic parameters will be reported to 3 significant digits in listings.

The Phoenix WinNonlin non-compartmental Core Output will be provided in a separate listing.

Tables

Pharmacokinetic parameter data will be presented in tables and descriptively summarized for the PKAS by part, regional cohort, overall by part (both cohorts within each part), overall (all parts/cohorts), and day using: n, Mean, StD, CV%, Min, Median, Max, GeoMean, logStD, GeoCV%, and the 95% CI for the GeoMean. Summaries will be based on the PKAS. Additional table(s) will summarize select parameters with further stratification by ADA subgroups ever positive and never positive.

Additional table(s) will summarize C_{trough} and C_{EOI} with further stratification by ethnicity (e.g., Japanese, non-Japanese Asian, all Asian, non-Asian, Chinese mainland, non-Chinese mainland Asian), based on the PKAS. Additional table(s) will summarize C_{trough} and C_{EOI} with further stratification by ADA subsets ever positive and never positive, and may additionally summarize C_{trough} and C_{EOI} at the time points with ADA measurement with stratification by the ADA results (Negative or Positive, see Section 16.4) at each time point, based on the PKADA. Additional table(s) will summarize C_{trough} and C_{EOI} with further sub-stratification by nAb subsets ever positive and never positive, based on the PKNAB. For nAb subsets, summaries will be done for ever positive in either of 2 nAb assays (PDL1 and TGFß receptor neutralization; PDL1+/TGFß+, PDL1+/TGFß−, PDL1−/TGFß+) versus never positive (PDL1−/TGFß−). Additional table(s) may summarize PK parameters with further stratification by antibiotics status subgroups.

Additional table(s) will summarize C_{trough} and C_{EOI} with further sub-stratification by ADA subsets ever positive and never positive and, nested within, regional cohort based on the PKADA. Additional table(s) will summarize C_{trough} of ADA ever positive participants with further sub-stratification by ADA subgroups "Treatment emergent (ALL)" and "ADA Non-Treatment-emergent", based on the PKADA. For nAb ever-positive participants, serum bintrafusp alpha C_{trough} will be descriptively summarized in additional table(s) for nAb status subgroups (positive in any of 2 nAb assays), based on the PKNAB. Additional table(s) will summarize C_{trough} of ADA

ever-positive participants with further sub-stratification by ADA subgroups and, nested within, regional cohort (and may be further stratified by ethnicity as above), based on the PKADA. For nAb ever-positive participants, serum bintrafusp alpha C_{trough} will be descriptively summarized in additional table(s) for nAb status subgroups (positive in any of 2 nAb assays) and, nested within, regional cohort (and may be further stratified by ethnicity as above), based on the PKNAB. Additional table(s) will summarize C_{trough} of ADA Treatment-emergent participants and nAb Treatment-emergent participants by PK day relative to seroconversion, for all participants and by regional cohort (and may be further stratified by ethnicity as above) based on the PKADA and PKNAB, respectively. Only subgroup sample size with a minimal 3 participants will be displayed for any of the above.

In export datasets, as well as in the SDTM PP domain, PK parameters will be provided with full precision and will not be rounded. Descriptive statistics of PK parameter data will be calculated using full precision values and rounded for reporting purposes only.

The following conventions will be applied when reporting descriptive statistics of PK parameter data:

Mean, Min, Median, Max, GeoMean, 95% CI:	3 significant digits
StD, logStD:	4 significant digits
CV%, GeoCV%:	1 decimal place

Figures

Individual PK C_{trough} and C_{EOI} values will be plotted against actual study day on a linear scale, for all participants by part and regional cohort. Plots of individual data will be presented based on the safety run-in analysis set for the safety run-in part and the SAF for the randomized, double-blind part.

Arithmetic mean (±StD), GeoMean (±logStD), and Median C_{trough} and C_{EOI} will be plotted versus nominal study day by part, regional cohort, overall by part (both cohorts within each part), and overall (all parts/cohorts) on a linear scale. Summaries will be based on the PKAS. Additional figure(s) will be presented of Mean (±StD), GeoMean (±logStD), and Median C_{trough} and C_{EOI} with further stratification by ethnicity (e.g., Japanese, non-Japanese Asian, all Asian, non-Asian, Chinese mainland, non-Chinese mainland Asian), based on the PKAS. Additional figure(s) will be presented with further stratification by ADA subgroups (ever positive and never positive), and additionally "Treatment emergent (ALL)" and "ADA Non-Treatment-emergent", based on the PKADA. Additional figure(s) will present Mean, GeoMean, Median, C_{trough} and C_{EOI} with further sub-stratification by nAb subsets ever positive and never positive, based on the PKNAB. Summary figures of Mean, GeoMean, or Median will contain a footnote denoting n per time point. All Mean, GeoMean, and Median plots will only display time points with n>2, which will also be added as footnote.

For ADA treatment-emergent (induced) participants with at least one C_{trough} measurement before and after ADA seroconversion, individual C_{trough} will be plotted versus PK day relative to seroconversion (for readability, split further into groups of 10 participants or fewer as needed), for

all participants and by regional cohort, based on the SAF. Box plots will be prepared for C_{trough} versus PK day relative to seroconversion, for all participants and by regional cohort, based on the PKADA.

For nAb treatment-emergent participants with at least one C_{trough} measurement before and after nAb seroconversion (earliest of 2 assays if positive in both), individual C_{trough} will be plotted versus PK day relative to seroconversion (for readability, split further into groups of 10 participants or fewer as needed), for all participants and by regional cohort, based on the SAF. Box plots will be prepared for C_{trough} versus PK day relative to seroconversion, for all participants and by regional cohort, based on the PKNAB.

CCI

CCI

Time to onset of response for treatment-emergent ALL

CCI

CCI

CCI


CCI

CCI

[redacted]

CCI

CCI

CCI

CCI

CCI

17 References

Chua W, Charles KA, Baracos VE, Clarke SJ. Neutrophil/lymphocyte ratio predicts chemotherapy outcomes in patients with advanced colorectal cancer. Br J Cancer. 2011 Apr 12;104(8):1288-95.

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; 45:228-47.

Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Stat Med. 2010 Jan 30;29(2):219-28.

Kao SC, Pavlakis N, Harvie R, et al. High blood neutrophil-to-lymphocyte ratio is an indicator of poor prognosis in malignant mesothelioma patients undergoing systemic therapy. Clin Cancer Res. 2010 Dec 1;16(23):5805-13.

Maurer W, Hothorn L, Lehmacher W. Multiple comparisons in drug clinical trials and preclinical assays: a priori ordered hypotheses. In Biometrie in der Chemisch-Pharmazeutischen Industrie, Vol. 6, Vollmar, J. (ed.), 3–18. Stuttgart: Fischer Verlag, 1995.

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.

European Centre for Disease Prevention and Control. Data on COVID-19 geographic distribution by country. https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide, accessed and downloaded on 26 June 2020.

18 Appendices

18.1.1 Appendix 1 - Definition of Important Protocol Deviations

See document: ctp-ms200647-0055-iap-appendix-ipd-v2.docx

18.1.2 Appendix 2 - Rules for Identification of Previous or Concomitant Medications/Procedures

See document: ctp-ms200647-0055-iap-appendix-identification-previous-conc-medications-v1.docx

18.1.3 Appendix 3 - Definition of NCI-CTCAE Grading

See document: ctp-ms200647-0055-iap-appendix-nci-ctcae-v2.xlsx.

18.1.4 Appendix 4 – Adaptation Plan

See documents: ctp-ms200647-0055-iap-appendix-adaptation-plan-v1.docx and ctp-ms200647-0055-iap-appendix-simulation-report-v1.docx.

Updated simulation report following protocol version 5.0: ctp-ms200647-0055-iap-appendix-simulation-report-v2.docx.

18.1.5 Appendix 5 – Cycle X Day 1/Day 8/Day 15 determination

Cycle X Day 1 determination:

Visit Name	Cycle X Day 1
WEEK 1 DAY 1	Cycle 1 Day 1
WEEK 4 DAY 22	Cycle 2 Day 1
WEEK 7 DAY 43	Cycle 3 Day 1
WEEK 10 DAY 64	Cycle 4 Day 1
WEEK 13 DAY 85	Cycle 5 Day 1
WEEK 16 DAY 106	Cycle 6 Day 1
WEEK 19 DAY 127	Cycle 7 Day 1
WEEK 22 DAY 148	Cycle 8 Day 1
WEEK 25 DAY 169	Cycle 9 Day 1
WEEK 28 DAY 190	Cycle 10 Day 1
WEEK 31 DAY 211	Cycle 11 Day 1
WEEK 34 DAY 232	Cycle 12 Day 1
WEEK 37 DAY 253	Cycle 13 Day 1
WEEK 40 DAY 274	Cycle 14 Day 1
WEEK 43 DAY 295	Cycle 15 Day 1
WEEK 46 DAY 316	Cycle 16 Day 1

Cycle X Day 8 determination

Visit Name	Cycle X Day 8
WEEK 2 DAY 8	Cycle 1 Day 8
WEEK 5 DAY 29	Cycle 2 Day 8
WEEK 8 DAY 50	Cycle 3 Day 8
WEEK 11 DAY 71	Cycle 4 Day 8
WEEK 14 DAY 92	Cycle 5 Day 8
WEEK 17 DAY 113	Cycle 6 Day 8
WEEK 20 DAY 134	Cycle 7 Day 8
WEEK 23 DAY 155	Cycle 8 Day 8
WEEK 26 DAY 176	Cycle 9 Day 8
WEEK 29 DAY 197	Cycle 10 Day 8
WEEK 32 DAY 218	Cycle 11 Day 8
WEEK 35 DAY 239	Cycle 12 Day 8
WEEK 38 DAY 260	Cycle 13 Day 8
WEEK 41 DAY 281	Cycle 14 Day 8
WEEK 44 DAY 302	Cycle 15 Day 8
WEEK 47 DAY 323	Cycle 16 Day 8

Cycle X Day 15 determination

Visit Name	Cycle X Day 15
WEEK 3 DAY 15	Cycle 1 Day 15
WEEK 6 DAY 36	Cycle 2 Day 15
WEEK 9 DAY 57	Cycle 3 Day 15
WEEK 12 DAY 78	Cycle 4 Day 15
WEEK 15 DAY 99	Cycle 5 Day 15
WEEK 18 DAY 120	Cycle 6 Day 15
WEEK 21 DAY 141	Cycle 7 Day 15
WEEK 24 DAY 162	Cycle 8 Day 15
WEEK 27 DAY 183	Cycle 9 Day 15
WEEK 32 DAY 204	Cycle 10 Day 15
WEEK 35 DAY 225	Cycle 11 Day 15
WEEK 38 DAY 246	Cycle 12 Day 15
WEEK 41 DAY 267	Cycle 13 Day 15
WEEK 44 DAY 288	Cycle 14 Day 15
WEEK 47 DAY 309	Cycle 15 Day 15
WEEK 50 DAY 330	Cycle 16 Day 15

18.1.6 Appendix 6 – Subsequent Immune Checkpoint Inhibitors

As of 30th July 2021, the list of subsequent immune check point inhibitors included in the analysis is the following:

- Avelumab
- Nivolumab
- Pembrolizumab
- Lambrolizumab
- Atezolizumab
- Durvalumab
- Tremelimumab
- Ipilimumab

Signature Page for CCI [redacted] v1.0

Approval	PPD Clinical 03-Feb-2022 16:51:45 GMT+0000

Signature Page for CCI [redacted] v1.0