# Integrated Analysis Plan for CSR Addendum (May 2020 cut-off)

| | |
|---|---|
| **Clinical Study Protocol Identification No.** | MS200647-0008 |
| **Title** | A Phase I, open-label, multiple-ascending dose trial to investigate the safety, tolerability, pharmacokinetics, biological and clinical activity of MSB0011359C in subjects with metastatic or locally advanced solid tumors with expansion to selected indications in Asia |
| **Study Phase** | Phase I |
| **Investigational Medicinal Product(s)** | Bintrafusp alfa |
| **Clinical Study Protocol Version** | 03 March 2020 / Version 7 |
| **Integrated Analysis Plan Author** | **Coordinating Authors**<br>PPD PPD<br>PPD PPD |
| **Integrated Analysis Plan Date and Version** | 11 April 2022, Version 2 |
| **Integrated Analysis Plan Reviewers** | Biostatistics Senior, Merck Healthcare PPD<br>PPD, Merck Healthcare PPD<br>PPD PPD<br>PPD PPD<br>Medical Lead, EMD Serono PPD<br>PPD EMD Serono PPD<br>PPD Merck Healthcare PPD<br>PPD, EMD Serono PPD<br>PPD PPD<br>PPD PPD<br>Medical writer PPD |

# Approval Page

**Integrated Analysis Plan:** MS 200647-0008

A Phase I, open-label, multiple-ascending dose trial to investigate the safety, tolerability, pharmacokinetics, biological and clinical activity of MSB0011359C in subjects with metastatic or locally advanced solid tumors with expansion to selected indications in Asia

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 analysis 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 |
|---|---|
| ACTH | Adrenocorticotropic hormone |
| ADA | Antidrug antibody |
| ADaM | Analysis Data Model |
| AE | Adverse Event |
| AESI | Adverse Events of Special Interest |
| ALT | Alanine aminotransferase |
| ALP | Albumin Alkaline Phosphatase |
| ANA | Antinuclear antibody |
| ANC | Absolute neutrophil count |
| aPTT | Activated partial thromboplastin time |
| AST | Aspartate Aminotransferase |
| ATP | Adenosine Triphosphate |
| AUC | Area Under the Curve |
| $AUC_{0-t}$ | Area under the concentration-time curve from time of dosing to the time of the last observation |
| $AUC_{0-\infty}$ | Area under the concentration-time curve from time of dosing extrapolated to infinity |
| $AUC_{0-336}$ | Area Under the Serum Concentration-Time Curve from time zero to 336 hours |
| $AUC_{extra\%}$ | Percentage of $AUC_{0-\infty}$ obtained by extrapolation |
| BOR | Best overall response |
| BTC | Biliary Tract Cancer |
| cBOR | Confirmed best overall response |
| CCA | Cholangio cell carcinoma |
| CDISC | Clinical Data Interchange Standards Consortium |
| $C_{EOI}$ | Serum concentration at end of infusion |
| CI | Confidence Interval |
| $C_{max}$ | Maximum serum concentration observed post-dose |
| $C_{min}$ | Minimum serum concentration observed post-dose |
| CR | Complete Response |

| | |
|---|---|
| CSR | Clinical Study Report |
| CTCAE | Common Terminology Criteria for Adverse Events |
| $C_{trough}$ | Serum trough concentration |
| CV% | Coefficient of Variation (%) |
| DBP | Diastolic Blood Pressure |
| DCR | Disease control rate |
| DI | Dose Intensity |
| DR | Duration of Response |
| ECG | Electrocardiogram |
| ECOG | Eastern Cooperative Oncology Group |
| ECOG PS | Eastern Cooperative Oncology Group performance status |
| eCRF | Electronic Case Report Form |
| eDISH | Evaluation of Drug-Induced Serious Hepatotoxicity |
| EORTC | European Organisation for Research and Treatment of Cancer |
| EOT | End of Treatment |
| ESCC | Esophageal Squamous Cell Cancer |
| GC | Gastric Cancer |
| GCP | Good Clinical Practice |
| GGT | Gamma-Glutamyl Transpeptidase |
| GeoCV% | Geometric Coefficient of Variation |
| GeoMean | Geometric Mean |
| HBV | Hepatitis B Virus |
| HCC | Hepatocellular carcinoma |
| HCV | Hepatitis C virus |
| HDV | Hepatitis D virus |
| HGB | Hemoglobin |
| HIV | Human Immunodeficiency Virus |
| IAP | Integrated Analysis Plan |
| ICF | Informed Consent Form |
| ICH | International Council for Harmonization |
| IMP | Investigational Medicinal Product |
| irAE | Immune-Related Adverse Event |

| | |
|---|---|
| IRC | Independent Review Committee |
| IRR | Infusion-Related Reaction |
| irTEAE | Immune-related Treatment-Emergent Adverse Event |
| LDH | Lactate dehydrogenase |
| LLN | Lower Limit of Normal |
| LLQ/LLOQ | Lower Limit of Quantification |
| logStD | Standard deviation of log-transformed data |
| Max | Maximum |
| MCH | Mean corpuscular hemoglobin |
| MCHC | Mean corpuscular hemoglobin concentration |
| MCV | Mean Corpuscular Volume |
| Mean | Arithmetic Mean |
| MedDRA | Medical Dictionary For Regulatory Activities |
| Min | Minimum |
| MSI | Microsatellite Instability |
| MSI-H | Microsatellite Instability-high |
| MSS | Microsatellite Stable |
| n | Number of non-missing values |
| NA | Not Applicable |
| nAb | Neutralizing Antibody |
| NC | Not calculated |
| NCA | Non-compartmental Analysis |
| NCI-CTCAE | National Cancer Institute – Common Terminology Criteria for Adverse Events |
| OR | Objective Response |
| ORR | Objective Response Rate |
| OS | Overall Survival |
| PD | Progressive Disease |
| PD-L1 | Programmed death ligand 1 |
| CCI | CCI |
| PFS | Progression Free Survival |
| PK | Pharmacokinetic |

| | |
|---|---|
| PKADA | Pharmacokinetic ADA Analysis Set |
| PKAS | PK Analysis Set |
| PKNAB | Pharmacokinetic nAb Analysis Set |
| PR | Partial Response |
| PT | Prothrombin Time/Preferred Term |
| Q1 | First Quartile |
| Q3 | Third Quartile |
| RBC | Red Blood Cell |
| RDI | Relative Dose Intensity |
| RECIST | Response Evaluation Criteria in Solid Tumors |
| RF | Rheumatoid Factor |
| RNA | Ribonucleic Acid |
| RR | RR interval |
| SAE | Serious Adverse Event |
| SAF | Safety Analysis Set |
| SBP | Systolic Blood Pressure |
| SD | Stable Disease |
| SDTM | Study Data Tabulation Model |
| SOC | System Organ Class |
| StD | Standard Deviation |
| $t_{1/2}$ | Terminal Half-Life |
| T4 | Thyroxine |
| TBILI | Total Bilirubin |
| TC | Tumor Cells |
| TEAE | Treatment-Emergent Adverse Event |
| TGFβ | Transforming Growth Factor-Beta |
| $t_{max}$ | Time to Maximum Concentration |
| TME | Tumor Microenvironment |
| TSH | Thyroid-stimulating hormone |
| ULN | Upper Limit of Normal |
| WBC | White Blood Cells |
| WHO-DD | WHO Drug Dictionary |

| | |
|---|---|
| WTU | Whole tumor |
| λz | Terminal Phase Rate Constant |

# 3 Modification History

| Unique Identifier for Version | Date of IAP Version | Author | Changes from the Previous Version |
|---|---|---|---|
| 1.0 | 29SEP2020 | PPD | NA |
| 2.0 | 11APR2022 | PPD | Cancellation of the final analysis after the database lock (IAP update performed after the CSR and CSR addendum deliveries in order to remove the protocol defined final data cut-off) |

# 4 Purpose of the Integrated Analysis Plan

The purpose of this integrated analysis plan (IAP) is to document technical and detailed specifications for the analyses of data collected for bintrafusp alfa in MS200647-0008 study for CSR addendum .

Results of the analyses described in this IAP will be included in the Clinical Study Report (CSR) addendum with May 2020 cut-off. 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 addendum but not identified in this prospective IAP will be clearly identified in the CSR addendum.

The current analyses purpose is to provide additional data update using 15May2020 data cutoff. The analyses for the CSR were based on 24Aug2018 data cutoff. Since the last participant started on 24th August 2017, the following data will be unchanged from CSR and analyses will not be repeated:

- Demographics, medical history, others baseline characteristics and previous anticancer therapies
- Confirmed and unconfirmed Best Overall Response (BOR) by Investigator (BOR by Independent Review Committee (IRC) will be described as changes for gastric cohort have been observed) and related subgroup analysis

Following analysis will not be run using the new cutoff due to the exploratory nature of the non-safety endpoints:

- All irRECIST and mRECIST analysis
- Efficacy subgroup analysis for gastric cancer histology and PD-L1 expression in combined positive score
- Patient-reported Outcomes (PRO) analysis
- Cytokines analysis

But in the opposite, new analysis will be added and described in this IAP such as:

- Neutralizing antibody data (nAb) analysis
- ADA analysis (duration of response, time to onset, ADA titer analysis)
- Efficacy subgroup analysis by PD-L1 whole tumor and PD-L1 tumor microenvironment
- Bleeding events and anemia description
- COVID-19 analysis

# 5 Objectives and Endpoints

Refer to protocol version 7.

# 6 Overview of Planned Analyses

This IAP will address analyses used to support CSR addendum.

Statistical analyses will be performed using cleaned eCRF data as well as external data including tumor assessment results by the IRC. All data will be included up to a clinical cut-off date.

The analyses will be performed separately for:

1. All dose levels in Dose Escalation (3, 10 and 20 mg/kg) and HCC cohorts.
2. All cohorts of Dose Expansion (BTC, ESCC, GC)

Important Protocol Deviations, Previous and Concomitant Medications and Safety analyses will be presented by cohort and all cohorts combined (i.e. all Dose Escalation and HCC combined in one set of outputs, all Dose Expansion combined in another set of outputs)

Treatment exposure, and efficacy analyses will be presented by cohort.

## 6.1 Decision after discontinuation of NSCLC and BTC trials

The data review outcome from the 3 randomized controlled studies in NSCLC and BTC (MS200647-0005, MS200647-0037, MS200647_0055) appears to indicate, consistently across 2 indications, either poorer observed hazard ratios for PFS and OS in the experimental arms with bintrafusp alfa or low likelihood for bintrafusp alfa to add benefits compared to standard of care. Based on this, the decision was made to not perform any final analysis after the final database lock. The study results have been described in the CSR approved on 31 March 2020 (cut-off date: 24$^{th}$ August 2018) and on CSR addendum approved on 25 March 2021 (cut-off date: 15$^{th}$ May 2020).

# 7 Changes to the Planned Analyses in the Clinical Study Protocol

There will be no final analysis after the final database lock.

## 7.1 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 will be generated to assess potential impacts of COVID-19 to this study including:

- Number of participant in pre/during COVID-19 study period
- Number of COVID-19 related protocol deviations,
- Listing of participants with any missedtumor assessments, samples/procedures or missed visits due to COVID-19,
- Number of participants with Adverse Events (AEs) in pre/during COVID-19 study period, and
- Listing of AEs related to COVID-19.

# 8 Analysis Populations and Subgroups

## 8.1 Definition of Analysis Populations

**Screening Analysis Set (SCR):** All participants who signed the informed consent form.

**Full Analysis Set (FAS) / Safety Analysis Set (SAF):** All participants who received at least 1 dose of trial treatment (> 0 mg and/or duration of infusion > 0). The SAF terminology will be used for the safety analysis and the FAS terminology will be used for efficacy analysis.

**PK Analysis Set (PKAS)**: All participants who complete at least 1 infusion of trial treatment, and who provide at least one post-dose sample with measurable concentration of bintrafusp alfa.

**PKADA** is defined as a subpopulation of the PK Analysis Set and restricted to participants who have in addition at least one valid result of antidrug antibody (ADA) at any time point.

**PKNAB** is defined as a subpopulation of the PK Analysis Set and restricted to participants who have in addition at least one valid result of nAb at any time point.

**Immunogenicity Analysis Set (IMM):** All participants who received at least 1 dose of trial drug and who have at least one valid result of ADA (i.e. ADA is negative, or positive with or without the titer result available) at any time point.

The analysis of efficacy and safety by nAb will be performed if we have nAb data for at least 5 participants.

## 8.2 Subgroup Definition and Parameterization

Subgroup analyses will be performed as specified in this IAP (see Section **Error! Reference source not found.**). All subgroup analyses will be exploratory, no adjustment for multiplicity will be performed.

For the definition of subgroup level, data as documented in the electronic case report form (eCRF) will be taken. The category "missing" will not be included in any subgroup analysis.

CCI [redacted]

The following subgroups will be defined:

- BTC classification
  - Intrahepatic Cholangio Cell Carcinoma (CCA) / Extrahepatic CCA / Carcinoma of Vater's ampulla / Gallbladder cancer
- PD-L1 expression on tumor cells (TC), tumor microenvironment (TME) and whole tumor (WTU) at baseline
  - $< 1\%, \geq 1\%$
- ADA status
  - Ever positive
  - Never positive
- nAb status
  - Ever positive
  - Never positive

# 9 General Specifications for Data Analyses

## 9.1 Data Handling After Cut-off Date

Data after cut-off do not undergo the cleaning process.

Data obtained after a cut-off will not be displayed in any listings or used for summary statistics, e.g. laboratory values of samples taken after data cut-off, AE with onset date after data cut-off, etc. will not be included in any analysis or listing.

Stop dates are not affected by this rule, e.g. a stop date of AEs, which starts prior to the cut-off, but stops after the date of cut-off, will not be changed.

## 9.2 Definition of Baseline and Change from Baseline

The last available assessment prior to the start of study treatment is defined as the "baseline" value or "baseline" assessment for safety and efficacy analyses. If an assessment is planned to be

performed prior to the first dose of study treatment in the protocol and the assessment is performed on the same day as the first dose of study treatment, it will be assumed that it was performed prior to study treatment administration, if assessment time point is not collected or is missing. If assessment time points are collected, the observed time point will be used to determine pre-dose on study day 1 for baseline calculation. Unscheduled assessments will be used in the determination of baseline. However, if time is missing, an unscheduled assessment on study day 1 will be considered to have been obtained after study treatment administration.

Patients who start treatment and discontinue from the study on the same day may have two different sets of data collected on study day 1 (one during study and one in the End of Treatment (EOT) visit. Data reported at the EOT visit are not eligible for baseline selection.

If a scheduled pre-dose measurement actually occurred post-dose, then the corresponding measurement will be treated and analyzed similar to an unscheduled post-dose measurement.

## 9.3 Treatment Day

Treatment day is defined relative to the start of trial treatment. Treatment day 1 is defined as the date of first administration of trial treatment. The day before the first administration of trial treatment is defined as Treatment day -1 (there is no Treatment day 0).

## 9.4 Definition of Duration and 'time since' Variables

Duration (days) will be calculated by the difference of start and stop date + 1 day, if not otherwise specified. For example, survival time (days) = date of death – date of first dose of trial treatment + 1.

The time since an event (e.g. time since first diagnosis) will be calculated as reference date minus date of event. In general, the reference date will be the date of first dose of trial treatment.

## 9.5 Conversion Factors

**Conversion of days to months or years:**

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

## 9.6 Time Window

There will be no difference between scheduled and unscheduled visits except for by-visit analyses of safety. The assignments of visit windows are described in the tables below for the purpose of by-visit analyses.

- No visit windowing will be performed at discontinuation, end of treatment, or safety follow-up visits for laboratory, vital sign, and ECG data, and post-dose assessment on Week 1 Day 1, Week 3 Day 15 visits for ECG data. Instead, the earliest non-missing observation among the unscheduled or scheduled assessments for each visit (discontinuation, end of

treatment, or safety follow-up) will be used for the analysis. For post-dose assessments on Week 1 Day 1, Week 3 Day 15 visits for ECG data, the earliest non-missing observation on Week 1 Day 1, Week 3 Day 15, respectively, will be used for the analysis.

- Scheduled and unscheduled assessments are included for visit windowing.
- If there are multiple assessments for any specified visit and some of them are from scheduled visits, the assessment from the scheduled visit with the closest distance to the planned study day will be used for analysis.
- If there are multiple assessments for any specified visit and none of them are from scheduled visits, the assessment with the closest distance to the planned study day will be used for analysis.
- If there are two or more unscheduled assessments with the same distance to the planned study day such as (-1/+1 day), the assessment prior to the planned study day such as -1 day will be used for windowing.
- There is no difference for visit windowing between tests from core serum chemistry panel and tests from the full serum chemistry panel, as those two panels will be analyzed together.
- For ECG assessment associated with study drug dose, only assessments where time points (prior to infusion or after infusion) are not missing will be considered for the analysis.

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

**Table 1 Visit Window Definition for Vital Signs, Hematology, Hemostaseology, and Chemistry Assessments**

| From (AWLO) | To (AWHI) | Planned Study Day (AWTARGET) | Analysis Visit (N), (AVISITN) | Analysis Visit (AVISIT) |
|---|---|---|---|---|
| ~ | <1 | | 1 | Baseline |
| 1 | 1 | 1 | 2 | Week 1 Day 1 |
| 2 | 2 | 2 | 3 | Week 1 Day 2* |
| 3 | 11 | 8 | 4 | Week 2 Day 8* |
| 12 | 18 | 15 | 5 | Week 3 Day 15 |
| 19 | 25 | 22 | 6 | Week 4 Day 22 |
| 26 | 36 | 29 | 7 | Week 5 Day 29 |
| 37 | 50 | 43 | 8 | Week 7 Day 43 |
| 51 | 64 | 57 | 9 | Week 9 Day 57 |
| 65 | 78 | 71 | 10 | Week 11 Day 71 |
| 79 | 92 | 85 | 11 | Week 13 Day 85 |
| 93 | 106 | 99 | 12 | Week 15 Day 99 |
| 107 | 120 | 113 | 13 | Week 17 Day 113 |
| 121 | 134 | 127 | 14 | Week 19 Day 127 |

| | | | | |
|---|---|---|---|---|
| 135 | 148 | 141 | 15 | Week 21 Day 141 |
| 149# | 162# | 155# | 16# | Week 23 Day 155# |

* For participants in dose expansion cohorts, there are no visits for vital signs at Week 1 Day 2 and Week 2 Day 8 and hence no visit windows for vital signs at Week 1 Day 2 and Week 2 Day 8 and the AWLO for vital sign at Week 3 Day 15 will be 2.
# Time windowing should continue every 2 weeks to cover the whole data collected

## 9.7 Definition of On-treatment Period

On-treatment period is defined as the time from the first trial drug administration to the last trial drug administration date + 30 days OR the earliest date of subsequent anti-cancer therapy (including drugs, surgeries, radiotherapies) minus 1 day, whichever occurs first, unless otherwise stated.

For immune-related AEs as listed in Section 16.1.3.2, 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.

## 9.8 Imputation of Missing Data

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.

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 trial treatment then the onset date will be replaced by the minimum of start of trial treatment and AE resolution date (if not missing).
- In all other cases the missing onset day or missing onset month will be replaced by 1.
- Incomplete stop dates will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of participant's death or cutoff date. In the latter case the date of death or cutoff date will be used to impute the incomplete stop date.
- In all other cases the incomplete stop date will not be imputed.

Incomplete dates for previous, concomitant and subsequent medications will be imputed as

follows:

- For start date of medication
  - 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.
- For end date of previous, concomitant and subsequent medications
  - If the day is missing, it will be imputed to the last day of the month.
  - If both day and month are missing, the month and day will be imputed as December 31st
  - 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 patient's death, then the date of death will be used to impute the incomplete stop date.

Missing or partial death dates will be imputed based on the last known date to be alive (refer to Section 14.4)

- If the date is missing it will be imputed as day after the last known date to be alive
- If the day or month is missing, death will be imputed as the maximum between the last known date to be alive and the imputed date of death where:
  - Missing day: 1st day of the month and year of death
  - Missing day and month: January 1st of the year of death

Missing or partial tumor assessments dates will be imputed based as follow:

- If there are multiple scan dates associated with an evaluation, the earliest of the scan dates associated with the evaluation will be used as the date of assessment.
- If one or more investigation dates for an evaluation are incomplete but other investigation dates are available, the incomplete date(s) are not considered for calculation of the assessment date and assessment date is calculated as the earliest of all investigation dates (e.g. X-ray, CT-scan).
- If all measurement dates for an evaluation have no day recorded, the 1st of the month is used.
- If the month is not completed, for any of the investigations for an evaluation, the respective assessment will be considered to be at the date which is exactly between the previous and the following assessment. If both a previous and following assessments are not available, this assessment will not be used for any calculations.

## 9.9 Pooling of centers

Data will be pooled across centers. The "center" factor will not be considered in statistical models due to the high number of participating centers in contrast to the anticipated small number of participants per center.

## 9.10 Significance level

No statistical tests will be performed for any of the study endpoints. For descriptive purposes, 95% (CIs) or 90% CIs will be calculated where indicated.

## 9.11 Presentation of continuous and qualitative variables

Continuous variables will be summarized using descriptive statistics i.e. number of non-missing values and number of missing values [i.e. n (missing)], mean, median, standard deviation (StD), Q1, Q3, minimum (min), and maximum (max).

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

In case the analysis refers only to certain visits, percentages will be based on the number of participants still present in the trial at that visit, unless otherwise specified.

Unless otherwise specified, time to event endpoints will be presented in months. Summary statistics will be reported with 1 digit.

## 9.12 Reporting conventions

Mean, median, Q1, Q3, Min, Max will have the same precision as SDTM data (number of digits) for non-derived data, and StD should be displayed to one digit more than the mean. Statistics on derived data will be rounded to reasonable digits, whereas maximal digits should be available in CDISC ADaM data sets. Percentages will be reported to one decimal place. The rounding will be performed to the closest integer / first decimal using the common mid-point between the two consecutive values. For example, 5.1 to 5.4 will be rounded to an integer of 5, and 5.5 to 5.9 will be rounded to an integer of 6.

## 9.13 Preferred term for analysis of WHO-DD coded data

For data coded according to WHO-DD (e.g., concomitant medications), summaries will be done on the preferred term level where the preferred term is corresponding 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 preferred term diphenhydramine.

## 9.14 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).

## 9.15 Data collected after re-initiated treatment

No re-initiation of treatment has been observed in this study at the time of cut-off date.

## 9.16 Categorization of participants for COVID-19 impact assessment

For the assessment of COVID-19 impact on this study, participants will be categorized as being affected by COVID-19 (either due to infection or due to circumstances of social distancing affecting the capabilities of sites/hospitals etc.) based on the COVID-19 study period defined as:

- The start of COVID-19 study period will be defined by country as the minimum of 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 26 June 2020 and 11 March 2020 (WHO-start of world-wide pandemic).
- Post-pandemic could be defined as date (1) vaccination is released, (2) WHO declares COVID-19 pandemic over, (3) region-specific calls are made to end social distancing measures with no relevant rise in cases thereafter. As study treatment for the last participant started on 24 August 2017, no participant will be grouped into the post COVID-19 study period for this study and no post-pandemic date will be defined.

Participants will be categorized into subgroups as follows:

- Pre: Participant started treatment prior to COVID-19 study period
- During: Participant started treatment during the COVID-19 study period
- Post: Participant started treatment post the COVID-19 study period

## 9.17 Software

All statistical analyses of efficacy and safety will be performed using SAS® Grid. The computer program Phoenix® WinNonlin® Version 6.4 or higher (Certara, L.P., Princeton, New Jersey, USA) will be used for non-compartmental analysis of PK data, CCI [redacted]
The same version of SAS as above will be used for PK/CCI [redacted] data statistical analyses and outputs.

# 10 Study Participants

## 10.1 Disposition of Participants and Discontinuations

*Analysis sets: Screening analysis set*

The following parameters will be summarized by all cohorts combined (where applicable) and by cohort. This summary will be presented for all participants.

- Total number of participants screened overall
- Total number of re-screened participants
- Number of participants who discontinued from the trial prior to first dose of treatment
- Reason for discontinuation of trial prior to first dose of treatment (did not meet eligibility criteria, withdraw of consent, other)
- Number of participants who received no treatment
- Number of participants who received at least 1 dose of trial treatment (Safety Analysis Set)
- Number of participants with treatment on-going
- Number of participants who completed the treatment
- Number of participants who discontinued the treatment
- Number of re-initiated participants
- Reasons for treatment discontinuation
    - Adverse event (AE)
    - Lost to follow-up
    - Protocol non-compliance
    - Death
    - Progressive disease
    - Withdrew consent
    - Other
- Number of participants who discontinued the treatment but are still in study for follow-up
- Number of participants who discontinued from the study
- Reasons for study discontinuation
    - Lost to follow-up
    - Death
    - Withdrew consent
    - Other

In addition, the number of participants in each analysis set defined in Section 8.1 will be summarized. The percentage of participants will be calculated based on the number of participants in the SAF population.

The follow-up time will be defined as the time between the date of first dose and the last known alive date of the subject (or cut-off date). Descriptive statistics will be provided. In addition, a Kaplan-Meier analysis will be performed using overall survival data by reverting the censoring flag (subjects still alive are counted as events, dead subjects are counted as censored).

## 10.2 Protocol Deviations

*Analysis set: Full Analysis Set*

### 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.

The following, but not limited to, are defined as important protocol deviations:

- Subjects that are dosed on the study despite not satisfying the inclusion criteria or meeting the exclusion criteria
- Subjects that develop withdrawal criteria whilst on the study but are not withdrawn
- Subjects that receive an incorrect dose
- Subjects that receive an excluded concomitant medication
- Deviation from Good Clinical Practice

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.

Frequency table per reason of important protocol deviations will be presented.

Important protocol deviations will also be presented in a data listing. This listing will include dose level/cohort, the participant identifier, age, sex, a category of the deviation (e.g. inclusion/exclusion), and a description of the deviation.

Important protocol deviations are specified in Appendix 1.

The number of participants with important protocol deviations overall as well as important protocol deviations due to COVID-19 will be tabulated. A listing of all COVID-19 Related Protocol Deviations (minor or important) will be provided. A listing of participants with any missed tumor assessments, missed samples/procedures and missed visits due to COVID-19 will

also be provided. The listing will be based on the COVID-19 protocol deviations included in CTMS, as well as the information recorded in the exposure eCRF page, if any.

# 11 Demographics and Other Baseline Characteristics

Demographics and other baseline characteristics will not be analyzed as there no change compared to CSR is expected.

# 12 Previous or Concomitant Medications/Procedures

*Analysis set: Full Analysis Set*

All Previous and Concomitant medications descriptions will be provided per dose level/cohort and overall (dose escalation overall, HCC overall, dose expansion overall [excluding HCC]).

## 12.1 Previous and concomitant medications

**Pre-ICF medications** are medications, other than study medications, which were taken within 30 days prior to signing the informed consent and were stopped prior to the signing of the informed consent.

**Previous medications** are medications, other than study medications and pre-medications for study drug, which are started before first dose date of study treatment. Pre-ICF medications which continue past the time of signing the informed consent will be categorized as previous medications. In case the date values will not allow to unequivocally allocate a medication to previous medication the medication will be considered as previous medication.

**Concomitant medications** are medications, other than study medications and pre-medications for study drug, which are taken by participants any time on-trial (on or after the first day of trial drug treatment for each participant) or within 30 days after last dose of trial drug OR the earliest date of subsequent anti-cancer drug therapies minus 1 day, whichever occurs first. In case the date values will not allow to unequivocally allocate a medication to concomitant medication the medication will be considered as concomitant medication.

**Post follow-up medications** are medications, other than study medications and pre-medications for study drug, which are started on or 30 days after last dose of trial drug OR on or after the earliest date of subsequent anti-cancer drug therapies, whichever occurs first.

Previous and Concomitant treatments will be summarized separately from the "Concomitant Medications Details" eCRF page. These summaries will present the number and percentage of participants by drug class and preferred term, overall and by dose level/cohort. Drug class will be derived as ATC classification Level 2, and the preferred term will be taken as the preferred drug name, based on the most recent available version of the WHO-DD dictionary. If multiple ATCs are assigned to a drug, all ATCs for that drug will be reported. A participant will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. The summary will be sorted by ATC-2nd Level and Preferred Term in alphabetical order.

Pre-ICF, previous, and concomitant medications will be presented in separate listings. Pre-ICF medications will come from the "Relevant Previous Medications" eCRF page. Each listing will include dose level/cohort, participant identifier, age, sex, preferred term, medication name, start date, end date, dose, dose units, frequency, route, reason for the medication, and adverse event (if reason for medication was an adverse event). The drug class and the preferred term use for Pre-ICF medications will follow the same rules for previous and concomitant medications described in previous paragraph.

## 12.2 Premedications

Premedications for study drug are medications e.g. diphenhydramine or acetaminophen, which are administrated the same day as, but prior to the study drug administration to mitigate potential infusion-related reactions.

The number of participants receiving pre-medication will be summarized for each treatment visit based on "Premedication details" eCRF page (participants for whom the question "Has the participant received any pre-medications before MSB0011359C infusion?" is answered "Yes" at the corresponding visit).

A listing will be reported with the relevant information collected on the "Premedication Details" eCRF page. The listing will include dose level/cohort, participant identifier, age, sex, preferred term, reported medication name, date/time of administration, dose, dose units, and route.

## 12.3 Concurrent procedures

All concurrent procedures, which were undertaken any time during the on-treatment period, will be presented in a listing with information collected from the "Concomitant Procedures Details" eCRF page. Concurrent procedures will be classified by medical review. Procedures will be coded using current version of Medical Dictionary for Regulatory Activities (MedDRA) and reported using preferred term (PT).

Number of participants with concurrent procedures (Prior to first dose of trial treatment, and during on-treatment period) will be described.

Concurrent procedures will be presented in a listing will include dose level/cohort, participant identifier, age, sex, name of procedure, preferred term, start date, end date, and reason for procedure.

## 12.4 Subsequent Anticancer Therapies

Anti-cancer treatment after discontinuation of bintrafusp alfa will be provided in a data listing with data retrieved from "Anti-Cancer Treatment After Discontinuation Details", "Radiotherapy After Discontinuation Details", and "Surgery After Discontinuation Details" eCRF pages.

The number of participants in each of the following anti-cancer treatment categories will be tabulated, all dose level/cohorts combined (where applicable), and by dose level/cohort:

- Participants with at least one subsequent anti-cancer treatment (i.e. therapy, radiotherapy or surgery)
- Participants with at least one subsequent anti-cancer drug therapy
- Participants with at least one subsequent anti-cancer radiotherapy
- Participants with at least one subsequent anti-cancer surgery

The type of subsequent anti-cancer drug therapy as provided in the e-CRF (i.e. anti-PD-1/anti-PD-L1, cytotoxic therapy, endocrine therapy, monoclonal antibodies therapy, small molecules, immunotherapy except anti-PD-1/anti-PD-L1, other) will be described, as well as the best response across all post trial treatments (complete response, partial response, stable disease, progressive disease, non-complete response/non-progressive disease, not assessable, unknown, not applicable).

In addition, the anti-cancer treatment after discontinuation of bintrafusp alfa will be provided in a data listing including with relevant information.

The earliest date of start of new anti-cancer therapy after discontinuation of bintrafusp alfa will be used for the definition of the on-treatment period and censoring for efficacy analyses

# 13 Study Treatment: Compliance and Exposure

*Analysis set: Safety Analysis Set*

The following analyses will be performed based on the safety analysis set by dose level/cohort. All dosing calculations and summaries will be based on "MSB0011359C Administration Details" eCRFs page. A listing of study drug administration will be created with the information collected on the "MSB0011359C Administration Details" eCRF page.

Subjects will receive an IV infusion of bintrafusp alfa over 1 hour (-10 minutes / +20 minutes, that is, over 50 to 80 minutes) once every 2 weeks as detailed in the Schedule of Assessments in the study protocol.

After confirmation of tolerability at 20 mg/kg in the dose escalation part, the bintrafusp alfa dose for further investigation is 1200 mg for GC, ESCC, and BTC participants. Summary statistics for GC, ESCC, and BTC cohorts will be derived based on the flat dose.

The duration of bintrafusp alfa treatment (in weeks) during the study for a patient is defined as:

$$\text{duration}=\left(\frac{\text{date of last dose-date of first dose+14}}{7}\right)$$

The cumulative dose of bintrafusp alfa per patient in a time period is the sum of the actual dose levels that the patient received within that period (i.e., total dose administered). The cumulative dose is expressed in mg/kg for dose escalation and HCC participants, and in mg for dose expansion participants excluding HCC participants.

The dose intensity (DI) and the relative dose intensity (RDI) will be calculated for each patient across all cycles. The dose intensity per cycle (mg/2 weeks) is defined as

$$DI=\left(\frac{\text{Cumulative dose}}{\text{treatment duration (weeks)/2}}\right)$$

The dose intensity is expressed in mg/kg/2 weeks for dose escalation and HCC participants, and in mg/2 weeks for dose expansion participants excluding HCC participants.

The relative dose intensity (RDI) is defined as the actual dose intensity divided by the planned dose intensity per cycle and expressed in percentage.

$$RDI\ (\%)=100\times\left(\frac{DI}{\text{planned dose level}}\right)$$

- The summary of treatment exposure and compliance for bintrafusp alfa will include:
  - Duration of therapy (weeks)
  - Total number of infusions received
  - Cumulative dose (mg/kg or mg)
  - Dose intensity (mg/kg/2 weeks or mg/2 weeks)
  - Relative dose intensity (%) as continue variable, also categorized as
    - < 80%
    - 80%-90%
    - >90%

All descriptions will be provided per dose level/cohort and overall (dose escalation overall, HCC overall, dose expansion overall (excluding HCC)).

- Two listings will be presented for treatment exposure and compliance:
  - A listing of study drug administration will be created with the information collected on the "MSB0011359C Administration Details" eCRF page. This listing will include assigned dose level/cohort, participant identifier, age, sex, infusion start date/time, infusion end date/time, infusion rate (mL/hr), actual dose (mg or mg/kg), treatment administration modification (Y/N), modification details (infusion rate reduction/temporary interruption and time infusion resumed/discontinued at this visit), reason for administration modification (AE/other), and treatment delay (days). Body weight will also be displayed for participants from dose escalation and HCC subjects
  - A listing of treatment exposure and compliance will include assigned dose level/cohort, participant identifier, age, sex, duration of therapy (weeks), total number of infusions received, cumulative dose (mg or mg/kg), dose intensity (mg/2 weeks or mg/kg/2weeks), and relative dose intensity (%)

### Dose Reductions

Dose reduction is not allowed per protocol and will not be summarized.

### Dose Delays

Delays will be derived based on infusion date and will be based on the deviation of the actual to the planned treatment administration day (relative to the previous treatment administration date). A delay is defined as 1 day or more of delay between the actual and the planned treatment administration day. For example, if one participant receives bintrafusp alfa on day 1, then the next administration date will be on day 15; however, if the participant receives trial treatment on day 17, this will be considered as a delay of 2 days.

The following will be summarized in a table, overall and by dose level/cohort:

- Number of participants with at least one delay
- Number of delays per participant, categorized as 0, 1, 2, 3, ≥4
- Number of cumulative delay days per participant, categorized as 0, 1-2, 3-6, ≥7 days

### Infusion Rate Reductions

Infusion rate reductions as recorded on the "MSB0011359C Administration Details" eCRF page will be used for analysis. Number of participants with at least one infusion rate reduction, number of participants with at least one infusion rate reduction due to adverse event, as well as a categorization of infusion rate reductions (1 / 2 / ≥3) will be summarized.

### Study Drug Temporary Interruptions

Study drug temporarily interrupted as recorded on the "MSB0011359C Administration Details" page of the eCRF will be used for analysis. Number of participants with at least one study drug interruption, number of participants with at least one study drug interruption due to adverse event, as well as a categorization of study drug interruptions (1 / 2 / ≥3) will be summarized.

## 14 Efficacy Analyses

*Analysis sets: Full Analysis set*

The endpoint evaluation will be performed separately for each cohort.

Table 2 provides an overview of the tumor response endpoints per criteria.

**Table 2 Overview of Tumor Response Endpoints**

| Endpoints | Escalation part | Expansion part (HCC Cohort) | Expansion part (GC, ESCC, BTC Cohorts) |
|---|---|---|---|
| BOR, Duration of Response (DR), PFS per | ✓, by investigator | ✓, by IRC (primary secondary) | ✓, by IRC (primary secondary) |

| Endpoints | Escalation part | Expansion part (HCC Cohort) | Expansion part (GC, ESCC, BTC Cohorts) |
|---|---|---|---|
| RECIST 1.1 | (secondary) | ✓, by investigator (secondary) | ✓, by investigator (secondary) |
| OS | ✓, (secondary) | ✓, (secondary) | ✓, (secondary) |

Of note, some endpoints in the Table 2 might not be analyzed using investigator assessment, but listings and/or graphs for investigator assessment will be provided.

## 14.1 Best Overall Response

**Best overall response (BOR)** will be assessed based on the tumor response at different evaluation time points from baseline until the first documented disease progression. The unconfirmed BOR will be defined by the best overall response reached by the participant during this period, while the confirmed BOR will have to meet additional rules of confirmation as defined in the following sections. Only tumor assessments performed before the start of any further anticancer treatment will be considered in the assessment of BOR. Clinical deterioration will not be considered as documented disease progression.

**Objective Response (OR)** is defined as complete response (CR) or partial response (PR) according to evaluation criteria from start date until documented disease progression. Patients who do not have an on-treatment radiographic tumor assessment due to early progression, who receive anti-tumor treatments other than the study treatments prior to reaching a CR or PR, or who die, progress, or drop out for any reason prior to reaching a CR or PR will be counted as non-responders in the assessment of OR. Each patient will have an objective response status (0: no OR; 1: OR). OR rate (ORR) is the proportion of patients with OR in the analysis set.

**Disease Control Rate** (DCR) is defined as the proportion of participants with BOR according to evaluation criteria of CR, PR, or SD. For RECIST 1.1 as adjudicated by IRC, non-CR/non-PD will also be considered in DCR evaluation.

### 14.1.1 Best Overall Response According to RECIST 1.1 as Assessed by Investigator

The confirmed BOR according to RECIST 1.1 as assessed by investigator will be based on reported overall responses at different evaluation time points from the start date until documented disease progression, according to the following rules:

- CR = at least two determinations of CR at least 4 weeks apart and before progression
- PR = at least two determinations of PR at least 4 weeks apart and before progression (and not qualifying for a CR)
- SD = at least one SD assessment (or better) ≥ 6 weeks after start date and before progression (and not qualifying for CR or PR).
- PD = progression ≤12 weeks after start date (and not qualifying for CR, PR or SD).

The following rule will additionally be applied regarding derivation of BOR:

- Participants who miss the first two post-baseline tumor evaluations and subsequently are observed to have PD will be assigned a not-evaluable BOR (NE) (i.e. tumor assessment of PD is >12 weeks after start date and there is no tumor assessment in between).
- When evaluating OR, both CR and PR must be confirmed by repeat assessments performed no less than 4 weeks after the criteria for response are first met. It is reasonable to consider a participant with time point response of PR-SD-PR, PR-NE-PR or CR-NE-CR as a confirmed response as long as the second CR or PR is more than 28 days away from the first time point.

The individual percentage of change in the sum of diameter since baseline will be displayed over time per dose level/cohort on a spider plot, together with the first occurrence of new lesion and participant off treatment.

A listing will present the tumor assessment and overall response per RECIST 1.1 as assessed by the investigator including: dose level/cohort, participant identifier, age, sex, date of start of subsequent therapy, MSI status, PD-L1 expression, date of death when death occurs, confirmed BOR, visit, date(s) of imaging, description of target lesions (size, site, type, method, response), non-target lesions (status, site, type, method, response), new lesions (site, type, method), sum of lesion diameters, percent change in target lesions from baseline, and overall response, sorted by dose level/cohort, participant identifier, and visit.

In addition, a listing presenting responses (CR, PR) occurring after a first progression will also be provided for dose expansion cohorts including: dose level/cohort, subject identifier, age, sex, MSI status, PD-L1 expression, unconfirmed and confirmed BOR, date of start of subsequent anti-cancer therapy, visit, date(s) of imaging, sum of lesion diameters, percent change in target lesions from baseline, overall response, and flag for response after progression, sorted by dose level/cohort, subject identifier, and visit.

### 14.1.2 Best Overall Response According to RECIST 1.1 as Adjudicated by the IRC

The confirmed BOR according to RECIST 1.1 as adjudicated by IRC will be assessed based on reported overall responses at different evaluation time points from the start date until documented disease progression, according to the following rules:

- CR = at least two determinations of CR at least 4 weeks apart and before progression
- PR = at least two determinations of PR at least 4 weeks apart and before progression (and not qualifying for a CR)
- SD = at least one SD assessment (or better) ≥ 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 does not meet criteria of CR
- PD = progression ≤12 weeks after start date (and not qualifying for CR, PR or SD)

- ND = at least one ND assessment

The following rule will additionally be applied regarding derivation of BOR:

- Participants who miss the first two post-baseline tumor evaluations and subsequently are observed to have PD will be assigned a not-evaluable BOR (NE) (i.e. tumor assessment of PD is >12 weeks after start date and there is no tumor assessment in between).

When evaluating OR, both CR and PR must be confirmed by repeat assessments performed no less than 4 weeks after the criteria for response are first met. It is reasonable to consider a participant with time point response of PR-SD-PR, PR-NE-PR or CR-NE-CR as a confirmed response as long as the second CR or PR is more than 28 days away from the first time point.

The confirmed and unconfirmed objective response rate (ORR) will be calculated 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).

As tumor lesions are evaluated by the IRC, it may happen that the independent reviewer disagrees with the investigator and does not assess any tumors as "measurable" at screening. In that case, the BOR could be rated as "non-CR/non-PD" (if no "CR" or "PD" are previously reported).

If the IRC is not able to identify any disease at baseline (target or non-target lesions), the BOR may be rated as "No Disease" (ND).

The following analyses will be performed:

- Number and percentage of participants with confirmed and unconfirmed BOR of CR, PR, SD, PD, and NE
- Number and percentage of participants with confirmed and unconfirmed ORR with a two-sided 95% CI
- Number and percentage of participants with confirmed and unconfirmed DCR with a two-sided 95% CI
- Spider graph of individual percentage of change in the sum of diameter since baseline
- Waterfall graph of the change in the sum of diameters between baseline and the best post-baseline assessment (i.e. minimum change since baseline)

Listing of tumor assessment: participant identifier, age, sex, MSI status, PD-L1 expression, date of death when death occurs, unconfirmed and confirmed BOR, visit, date(s) of imaging, description of target lesions (size, site, type, method, response), non-target lesions (status, site, type, method, response), and new lesions (site, type, method), sum of lesion diameters, percent change in target lesions at baseline, and overall response, sorted by dose level/cohort, participant identifier, and visit.

In addition, a listing presenting responses (CR, PR) occurring after a first progression will also be provided including: participant identifier, age, sex, MSI status, PD-L1 expression, date of start of subsequent therapy, unconfirmed and confirmed BOR, visit, date(s) of imaging, sum of lesion

diameters, percent change in target lesions from baseline, overall response and flag for response after progression, sorted by dose level/cohort, participant identifier, and visit.

When BOR as adjudicated by IRC and investigator assessment are both available, a listing of inconsistencies will be provided.

Listings of efficacy assessments for both Investigator and IRC assessment will also be provided including: cohort, participant identifier, age, sex, MSI status, PD-L1 expression, date of first study treatment, treatment duration, cBOR, PFS, time to response, DR and OS. Efficacy assessments will be given for both investigator and IRC. Subjects with treatment ongoing or response ongoing will be flagged.

## 14.2 Duration of Response

Duration of Response (DR) measured from the time measurement criteria are first met for CR/PR until the first date of PD or death due to any cause within 84 days (12 weeks) or last tumor assessment if death occurred during the treatment period, or within 168 days (24 weeks) of last tumor assessment if death occurred during the follow-up period.

The analysis of DR will be performed using confirmed CR/PR according to RECIST 1.1 as measured by the investigator. For dose expansion participants (excluding HCC as there is no responder), analysis of DR according to RECIST 1.1 as measured by the IRC will also be performed.

DR will be censored in the following scenarios:

- Participants who do not experience an event (PD or death) will be right-censored on the date of their last evaluable (non-missing and non “Not Evaluable”) tumor assessment.
- Participants who start new anti-cancer treatment prior to an event will be censored on the date of the last evaluable tumor assessment before anti-cancer therapy is given.
- Participants with an event after two or more subsequent missing response assessments (i.e. 2 times the scheduled time interval between two subsequent response assessments) will be censored on the date of the last evaluable (non-missing and non “Not Evaluable”) tumor assessment.

  Note: 2 times the scheduled time interval means 84 days during the treatment period and 168 days during the follow-up period (for participants who ended the treatment not for progression).
- Participants who do not have a baseline tumor assessment or who do not have any post-baseline tumor assessments will be censored on the date of first dose unless death occurred on or before the time of the second planned tumor assessment in which case the death will be considered an event.

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

The analysis of DR will be performed with a Kaplan-Meier method (product-limit estimates) and a summary of associated statistics will be presented including corresponding two-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and CIs for the survival function estimates at month 3, 6, 12, 18, 24, 36, and 48 will be derived using the log-log transformation according to Kalbfleisch and Prentice (1980) (CONFTYPE=loglog default option in SAS PROC LIFETEST). The estimate of standard error will be computed using Greenwood's formula. DR rates with their CI at 3, 6, 9, 12, 18, 24, 36, and 48 months will be presented, as well as the number of participants at risk and failed.

The time and duration of response per participant will be displayed in a swimmer graph.

Listing will be provided with the following information: dose level/cohort, participant identifier, age, sex, MSI status, PD-L1 expression, date of first response, date of first PD or death, duration of response, and censoring reason.

## 14.3 Progression-Free Survival

Progression-Free Survival (PFS) time is defined as the time (in weeks) from first administration of trial treatment until the first date of PD or death due to any cause within 84 days (12 weeks) of last tumor assessment if death occurred during the treatment period, or within 168 days (24 weeks) of last tumor assessment if death occurred during the follow-up period.

PFS will be censored in the following scenarios:

- Participants who do not experience an event (PD or death) will be right-censored on the date of the last evaluable (non-missing and non "Not Evaluable") tumor assessment.
- Participants who start new anti-cancer treatment prior to an event will be censored on the date of the last evaluable (non-missing and non "Not Evaluable") tumor assessment before anti-cancer therapy is given.
- Participants with an event after two or more subsequent missing response assessments (i.e. 2 times the scheduled time interval between two subsequent response assessments) will be censored on the date of the last evaluable (non-missing and non "Not Evaluable") tumor assessment.
- Note: 2 times the scheduled time interval means 84 days during the treatment period and 168 days during the follow-up period (for participants who ended the treatment not for progression)
- Participants who do not have a baseline tumor assessment or who do not have any post-baseline tumor assessments will be censored on the date of first dose unless death occurred on or before the time of the second planned tumor assessment in which case the death will be considered an event.

PFS = (date of PD or death/censoring – date of the first dose + 1)/ 30.4375 (months).

PFS event / censoring are defined in Table 3:

**Table 3 Progression-free Survival Event / Censoring**

| PFS Event Status | | Censoring | Date of event / censoring |
|---|---|---|---|
| Progressed or died | Within two subsequent scheduled tumor assessments after last response assessment of CR, PR or SD | Event | Minimum (Date of PD, Date of death) |
| | Otherwise | Censored | Date of last tumor assessment with outcome CR, PR or SD or date of first dose, whatever is later |
| Neither progressed nor died | | Censored | Date of last tumor assessment with outcome CR, PR or SD or date of first dose, whatever is later |

CR = complete response; PD = progressive disease; PFS= progression-free survival ; PR = partial response; SD = stable disease.

The analysis of PFS will be performed according to the RECIST 1.1 per Investigator. For dose expansion participants, analysis of PFS according to RECIST 1.1 as measured by the IRC will also be performed.

The analysis of PFS time will be performed with a Kaplan-Meier method with the same approach as for DR described in Section 14.2. Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including corresponding two-sided 95% CIs. PFS rates with their CI at 3, 6, 9, 12, 18, 24, 36, and 48 months will be presented, as well as the number of participants at risk and failed.

Listing of PFS will be provided with the following information: dose level/cohort, participant identifier, age, sex, MSI status, PD-L1 expression, date of first administration, date of last tumor assessment, date of event/censoring, event/censoring reason, time to event.

## 14.4 Overall Survival

The overall survival (OS) time is defined as the date from first dose to death due to any cause:

OS = (date of event/censoring – date of the first dose + 1)/30.4375 (months).

For patients alive at the time of data cut-off date or who are lost to follow up, OS will be censored at the last date known to be alive.

The date of event / censoring is defined in Table .

**Table 4 Survival Event / Censoring**

| | Date of event / censoring | Censoring |
|---|---|---|
| Participants alive or lost to follow-up before or at cut-off date | Last date known to be alive | Yes |
| Participants who died before or at cut-off date | Date of death | No |

The following dates will be considered to determine the last date known to be alive. Only the ones among them that are before or at data cut-off shall be used in the derivation. Dates past the data cut-off, e.g.: stop dates after cut-off, will be ignored by the derivation:

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

The last known to be alive date will be completed by the date of death for dead participants.

The analysis of OS will be performed with a Kaplan-Meier method (product-limit estimates) and a summary of associated statistics will be presented including corresponding two-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and CIs for the survival function estimates at Month 3, 6, and 12, 18, 24, 36, and 48 months (depending on actual data) will be derived using the log-log transformation according to Kalbfleisch and Prentice (1980) (CONFTYPE=loglog default option in SAS PROC LIFETEST). The estimate of standard error will be computed using Greenwood's formula. OS rates with their CI at 3, 6, 12, 18, 24, 36 and 48 months will be presented, as well as the number of participants at risk and failed.

A participant listing will provide the following information: dose level/cohort, participant identifier, age, sex, MSI status, PD-L1 expression, date of first administration, date of last tumor assessment, date of event/censoring, event/censoring reason, time to event.

## 14.5 Subgroup Analysis

BOR, DR, PFS according to RECIST 1.1 by Investigator and IRC, as well as OS analyses will be performed on subgroups as shown below:

**Table 4 Efficacy assessments by subgroups**

| | BOR | | PFS | | DR[a] | | OS |
|---|---|---|---|---|---|---|---|
| | Inv | IRC | Inv | IRC | Inv | IRC | |
| BTC classification | | | ✓ | ✓ | | | ✓ |
| PD-L1 TC[b] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| PD-L1 TME[b] | | ✓ | | ✓ | | ✓ | ✓ |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| PD-L1 WTU[b] | | ✓ | | ✓ | | ✓ | ✓ |
| MSI | | ✓ | | | | | |
| ADA | ✓ | ✓ | ✓ | ✓ | | | ✓ |
| nAb | ✓ | ✓ | ✓ | ✓ | | | ✓ |

[a] Only for dose expansion participants (excluding HCC)
[b] 1% cutoff
ADA = antidrug antibody; BOR = best overall response; BTC = biliary tract cancer; DR = duration of response; HCC = hepatocellular carcinoma; Inv = Investigator; IRC = Independent Review Committee; MSI: Microsatellite Instability; nAb = neutralizing antibody; OS = overall survival; PD-L1 = programmed death ligand 1; PFS = progression-free survival.

Corresponding Kaplan-Meier figures will be provided for PFS, DR and OS. For BOR, corresponding spider plot will be displayed.

Patients with missing subgroup will not be included in analyses.

# 15 Anti-Drug Antibody and Neutralizing Antibody

## 15.1 Anti-Drug Antibody

*Analysis Sets: Immunogenicity analysis set*

Anti-drug antibody (ADA) will be assessed before the trial treatment start, and on Days 15, 29, 43, 57, 71, and 85 after the start of infusion, 6-weekly up to/including Week 25, every 12 weeks after Week 25, and during safety follow-up (the term HAHA is used in the eCRF). Samples collected after the on-treatment period (e.g. safety follow-up) will be included in the analysis as well. If the sample is positive for ADA, it will be re-analyzed to determine the titer. The ADA results will be derived based on the algorithm in Table 5.

**Table 5 Algorithm for the Derivation of ADA Results**

| Sample Screen Result | Confirmatory | Titer | ADA Result |
|---|---|---|---|
| Negative | NA | NA | Negative |
| NR | NA | NA | NR |
| Positive | Negative | NA | Negative |
| Positive | NR | NA | NR |
| Positive | Positive | Number | Number |
| Positive | Positive | NR | Positive-TNR |

NR = no result, NA = not applicable, TNR = titer no result.

Negative, number, or positive-TNR are valid results while number and positive-TNR are considered as positive. Participants will be characterized into different categories based on the criteria in

Table 6 below.

**Table 6 Participants Characterized based on ADA Results**

| Category | Definition | Participant at Risk (Denominator for Incidence) |
|---|---|---|
| Never positive | No positive results at any time point | Number of participants with at least one valid result at any time point |
| Ever positive | At least one positive result at any time point, including baseline | Number of participants with at least one valid result at any time point |
| Pre-existing | A positive ADA result prior to treatment with bintrafusp alfa | Number of participants with valid baseline result |
| Treatment boosted | A positive ADA result prior to treatment with bintrafusp alfa and the titer ≥ 8*baseline titer at least one post-baseline value | Number of participants with valid baseline result and at least one valid post-baseline result |
| Treatment emergent | Not positive prior to treatment with bintrafusp alfa and with at least one positive post-baseline result | Number of participants with at least one valid post-baseline result and without positive baseline result (including NR) |
| Transient positive | If treatment emergent participants have (a single positive evaluation, or duration between first and last positive result <16 weeks) and last assessment not positive.”. | Number of participants with at least one valid post-baseline result and without positive baseline result (including NR) |
| Persistent positive | If treatment emergent participants have duration between first and last positive result ≥16 weeks or a positive evaluation at the last assessment | Number of participants with at least one valid post-baseline result and without positive baseline result (including NR) |

**Start of ADA Immunogenicity Response (Seroconversion)**

For participants with any positive ADA response, the date of the first assessment with positive ADA result will be considered as start date of ADA response.

Time to onset (weeks) of ADA response will be calculated as:

(Date of first positive ADA assessment – start date of bintrafusp alfa treatment + 1) / 7

Note: If the first positive is prior to the start of treatment, the formula is revised to:

(Date of first positive ADA assessment – start date of bintrafusp alfa treatment) / 7

**Duration of ADA Immunogenicity Response**

Duration of ADA immunogenicity response (weeks) is defined as:

(Date of last positive ADA assessment – date of first positive ADA assessment + 1) / 7

For participants with pre-existing positive, duration will be calculated from start date of bintrafusp alfa treatment rather than from date of first positive assessment. Participants still on treatment at the data cut-off and positive assessment at their last assessment before cut-off will be censored at the date of last assessment.

For ADA results, the following analysis will be described in each cohort and in all cohorts combined:

- The frequency and percentage of each ADA category will be tabulated

- The ADA titer value by timepoint will be summarized
- The maximum observed ADA titer per participant will be tabulated for each ADA status group. For each discrete titer value, percentages will be calculated using the total number of participants in each ADA status group as the denominator.
- The time to first ADA positive response will be summarized
- The duration of ADA immunogenicity response will be summarized

These analyses will be performed overall and by subgroups:

- Ethnicity (Japanese, Non-Japanese) for dose expansion participants (excluding HCC)
- BTC classification (Carcinoma of Vater's ampulla, Extrahepatic cholangio cell carcinoma, Gallbladder cancer, Intrahepatic cholangio cell carcinoma)

A listing of all individual ADA results from ever positive participants will be prepared by time point. The listing will include cohort, participant identifier, age, gender, race, ADA categories, date of assessment and results of screening, confirmatory and titer values.

A further listing will be provided including: cohort, participant identifier, age, sex, date of first ADA positive result, responder per IRC/investigator, date of response and timing of response related to the date of first ADA positive result. Responders will be defined as participants meeting confirmed CR or PR and non-responders as all other participants.

## 15.2 Neutralizing Antibody

*Analysis Sets: Immunogenicity analysis set*

Samples with a reportable ADA titer will also be tested in the two nAb assays, PD-L1 and TGFβ.

nAb results are positive or negative in a single assay and only derived when not performed because ADA was negative (see **Error! Reference source not found.**Table ). Participants will be characterized in the two assays in different nAb categories based on the criteria in **Error! Reference source not found.**

Table .

**Table 7 Algorithm for the Derivation of nAb Results**

| ADA Confirmatory Result | nAb Result | Derived nAb Result |
|---|---|---|
| Negative | NA | Negative |
| NR | NA | NR |
| NA (screen NR) | NA | NA |
| NA (screen negative) | NA | Negative |
| Positive | NR | NR |
| Positive | Positive | Positive |

| | | |
|---|---|---|
| Positive | Negative | Negative |

ADA = antidrug antibody, NA = not applicable, nAb = neutralizing antibody, NR = no result.

**Table 8 Participants Characterized based on nAb Results**

| Category | Definition | Participant at Risk (Denominator for Incidence) |
|---|---|---|
| Never positive | No nAb positive results at any time point | Number of participants with at least one valid ADA result at any time point |
| Ever positive | At least one nAb positive result at any time point | Number of participants with at least one valid ADA result at any time point |
| Pre-existing | A positive nAb result prior to treatment with bintrafusp alfa | Number of participants with valid ADA baseline result |
| Treatment emergent | Not nAb positive prior to treatment with bintrafusp alfa and with at least one nAb positive post-baseline result | Number of participants with at least one ADA valid post-baseline result and without nAb positive baseline results (including missing, NR) |
| Transient positive | If treatment emergent participants have (a single nAb positive evaluation, or duration between first and last nAb positive result <16 weeks) and last ADA assessment not nAb positive. | Number of participants with at least one ADA valid post-baseline result and without nAb positive baseline results (including missing, NR) |
| Persistent positive | If treatment emergent participants have duration between first and last nAb positive result ≥16 weeks or a nAb positive evaluation at the last ADA assessment | Number of participants with at least one ADA valid post-baseline result and without nAb positive baseline results (including missing, NR) |

ADA = antidrug antibody, nAb = neutralizing antibody, NR = no result.

**Start of nAb Immunogenicity Response (Seroconversion)**

For participants with any positive nAb response, the date of the first assessment with positive nAb result will be considered as start date of nAb response.

Time to onset (weeks) of nAb response will be calculated as:

(Date of first positive nAb assessment – start date of bintrafusp alfa treatment + 1) / 7

Note: If the first positive is prior to the start of treatment, the formula is revised to:

(Date of first positive nAb assessment – start date of bintrafusp alfa treatment) / 7

**Duration of nAb Immunogenicity Response**

Duration of nAb immunogenicity response (weeks) is defined as:

(Date of last positive nAb assessment – date of first positive nAb assessment + 1) / 7

For participants with pre-existing positive, duration will be calculated from start date of bintrafusp alfa treatment rather than from date of first positive assessment. Participants still on treatment at the data cut-off and positive assessment at their last assessment before cut-off will be censored at the date of last assessment.

For nAb results, the following analysis will be described in each cohort and in all cohorts combined:

- The frequency and percentage of each nAb category will be tabulated for each assay individually and combined.
- The nAb titer value by timepoint will be summarized
- The maximum observed nAb titer per participant will be tabulated for each nAb status group. For each discrete titer value, percentages will be calculated using the total number of participants in each nAb status group as the denominator.
- The time to first nAb positive response will be summarized
- The duration of nAb immunogenicity response will be summarized

These analyses will be performed overall and by subgroups:

- Ethnicity (Japanese, Non-Japanese) for dose expansion participants (excluding HCC)
- BTC classification (Carcinoma of Vater's ampulla, Extrahepatic cholangio cell carcinoma, Gallbladder cancer, Intrahepatic cholangio cell carcinoma)

Listings of nAb results from ever positive participants (in either assay) will be provided with the following: cohort, participant identifier, age, sex, race, assay (PD-L1 or TGFβ), nAb categories status, visit, date of assessment and results of screening, and titer values.

A further listing will be provided including: cohort, participant identifier, age, sex, date of first nAb positive result, responder per IRC/investigator, date of response and timing of response related to the date of first nAb positive result. Responders will be defined as participants meeting confirmed CR or PR and non-responders as all other participants.

## 15.3 Evaluation of Potential Effect of ADA/nAb on Bintrafusp alfa Efficacy

*Analysis Sets: Immunogenicity analysis set*

Following analysis will be described by subgroup of ADA (never positive/ever positive) and nAb (never positive [ie: no PD-L1 positive assay and no TGFβ positive assay at any timepoint including baseline]/ever positive [ie either PD-L1 or TGFβ positive assay at any timepoint including baseline]):

- Confirmed BOR per Investigator as described in Section 14.1.1 for dose escalation cohorts

- Confirmed BOR per Investigator and IRC as described in Section 14.1.1 and Section 14.1.2 for dose expansion cohorts

- PFS per Investigator as described in Section 14.3 for dose escalation cohorts and respective Kaplan-Meier curve

- PFS per Investigator and IRC as described in Section 14.3 for dose expansion cohorts and respective Kaplan-Meier curve

- OS as described in Section 14.4 and respective Kaplan-Meier curve

- Listings of efficacy assessments and immunology parameters for ADA positive subjects will include: cohort, participant identifier, age, sex, ADA and nAb categories, date of first study treatment, treatment duration, cBOR, PFS, time to response, DR and OS. Subjects with treatment ongoing or response ongoing will be flagged.

## 15.4 Evaluation of Potential Effect of ADA/nAb on Bintrafusp alfa Safety

*Analysis Sets: Immunogenicity analysis set*

The following analyses frequency and percentage of AEs by ADA subgroups and nAb (never positive [i.e. no PD-L1 positive assay and no TGFβ positive assay at any timepoint including baseline]/ever positive [i.e. either PD-L1 or TGFβ positive assay at any timepoint including baseline]) will be performed in each cohort (refer to Section 16):

- TEAEs
- TEAEs, grade ≥ 3
- TEAEs leading to permanent treatment discontinuation
- TEAEs excluding IRRs leading to drug interruptions
- Serious TEAEs
- TEAEs leading to death
- irAEs
- IRRs
- Skin AESI
- Anemia (using the search preferred term ´Anaemia´)

## 15.5 Evaluation of Potential Effect of ADA on Bintrafusp alfa PK

Bintrafusp alfa concentration in serum will be descriptively summarized as described in Section 16.3 to evaluate the potential effect of ADA on PK.

# 16 Safety Evaluation

*Analysis set: Safety Analysis Set*

All safety description will be provided per dose level/cohort and overall (dose escalation overall, HCC overall, dose expansion overall (excluding HCC)).

## 16.1 Adverse Events

Treatment emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period for the first time, or if the worsening of an event is during the on-treatment period as defined in Section 9.7. Adverse events will be coded according to the most current version of MedDRA. Severity of AEs will be graded using the NCI-CTCAE v4.03 toxicity grading scale.

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 in case of a change in toxicity grade. 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 start date of the entire event. 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 if not otherwise specified. The AE listings will include all AEs (whether treatment-emergent or not).

- **Related Adverse Events**: adverse events with relationship to study treatment (as recorded on the "Adverse Events Details" eCRF page, Relationship with MSB0011359C= Related) reported by the investigator and those of unknown relationship (i.e. no answer to the question "Relationship with MSB0011359C").
- **Serious Adverse Events (SAE)**: serious adverse events (as recorded on the "Adverse Events Details" eCRF page, Serious Adverse Event = Yes).
- **Adverse Events leading to Treatment Discontinuation**: adverse events leading to permanent discontinuation of study treatment (as recorded on the "Adverse Events Details" eCRF page, Action taken with MSB0011359C = Drug withdrawn).
- **Adverse Events leading to Death**: adverse event leading to death (as recorded on the "Adverse Events Details" eCRF page, Outcome = Fatal, as well as AEs of Grade 5).
- **Adverse Events of Special Interest (AESI):** adverse events of special interest (as identified according to a pre-specified search list of MedDRA Preferred Terms). Categories of AESI include:

- Immune-Related Adverse Events (irAE)
- Infusion-Related Reactions (IRR) including hypersensitivity
- Skin AE possibly related to TGFβ 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).

### 16.1.1 All Adverse Events

Adverse events will be summarized by worst severity (according to NCI-CTCAE version 4.03) per patient, using the latest version of MedDRA preferred term (PT) as event category and MedDRA primary system organ class (SOC) body term as Body System category.

Each patient will be counted only once within each PT or SOC. If a participant experience more than on 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.

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 in case of an improvement in toxicity grade. These events will be kept as separate records in the database in order to maintain the full detailed history of the event. The start date of the initial record in the sequence is taken as start date of the event of entire event. Duration of the AE and the TEAE flag will be adjusted accordingly in the analysis. Of note if a record has an outcome resolved (with or without sequelae), this can't be an 'intermediate' record.

Unless otherwise stated adverse events will be displayed in terms of frequency tables: PT and primary SOC in alphabetical order.

- If an adverse event is reported for a given participant more than once during treatment, the worst severity and the worst relationship to trial 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.
- The overall summary of AEs table will include the frequency (number and percentage) of patients with each of the following:
  - TEAEs
  - Related TEAEs
  - Serious TEAEs

- Related Serious TEAEs
- TEAEs, Grade ≥3
- Related TEAEs, Grade ≥3
- TEAEs leading to death
- Related TEAEs leading to death
- TEAEs and related TEAE of special interest:
  - Infusion-related reactions (IRRs)
  - Immune-related AEs (irAEs)
  - Potential TGF-β-mediated skin AEs
  - Anemia
- TEAE, bleeding events
- Related TEAEs bleeding events
- TEAEs leading to study termination
- Related TEAEs leading to study termination

The following incidences will be summarized by SOC and PT

- TEAEs
- TEAEs by worst grade
- Related TEAEs
- Related TEAEs by worst grade
- TEAEs leading to death
- Related TEAEs leading to death

Further TEAE tables will be presented by PT in descending order of frequency overall:

- TEAEs
- TEAEs, grade ≥ 3
- Related TEAEs
- Related TEAEs, grade ≥ 3
- Serious TEAEs
- Related serious TEAEs
- TEAEs leading to death

TEAEs and related TEAEs by worst grade will also be summarized, and the most frequent PT (at least 5%) will be presenting graphically by worst grade and PT with bar chart figures.

A listing of all adverse events (whether treatment-emergent or not) and a listing of TEAEs will be created separately with the relevant information.

Listing of adverse events will contain the following information: dose level/cohort, participant identifier, age, sex, preferred term, reported term for the adverse event, 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, TEAE (Y/N), seriousness (Y/N), AE before on-treatment period (Y/N), AE onset or worsening after on-treatment period (Y/N), AESI infusion-related (Y/N), AESI immune-related (Y/N), medication administered. The listing will be sorted by dose level/cohort, participant identifier, and start date.

**Evaluation of COVID-19 effects on AEs**

The direct effect of COVID-19 for AEs will be assessed via listings of COVID-19 related AEs. The following listings will be generated using the 'COVID-19 related terms MedDRA 23.0 update Spreadsheet' (https://www.meddra.org/covid-19-related-terms-meddra-230-update-spreadsheet, last accessed on 28 May 2020) as available from Maintenance and Support Services Organization (MSSO), considering all 'search terms for COVID-19-related' ='Y'. The following information will be provided:

- Subject 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 indirect effect of COVID-19 for AEs will not be assessed due to the low number of participants still on treatment at the onset of COVID-19.

### 16.1.2 Adverse Events Leading to Treatment Discontinuation/ Interruption / Modification

An overview table will display all dose level/cohorts combined (where applicable) and by dose level/cohort with the following information:

- TEAEs leading to temporary drug interruption
- Related TEAEs leading to temporary drug interruption

- TEAEs leading to permanent treatment discontinuation
- Related TEAEs leading to permanent treatment discontinuation
- TEAEs leading to infusion rate reduction
- Related TEAEs leading to infusion rate reduction

In addition, the incidences for above items will be summarized by SOC and PT.

The listing of all TEAEs leading to permanent treatment discontinuation will also be provided with the relevant information (see description of listing in Section 16.1.1).

## 16.2 Deaths, Serious Adverse Events, and Adverse Events of Special Interest

### 16.2.1 Deaths

All deaths, deaths within 30 days after last dose of study drug, death within 60 days after first dose as well as the primary reason for death will be tabulated based on information from the "Death" and "Survival Follow-Up" 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
    - Disease Progression
    - Event unrelated to study treatment
    - Event related to study treatment
    - Other
    - Unknown

In addition, date and cause of death will be provided in individual patient data listing together with selected dosing information (study treatment received, date of first / last administration, number of infusions, day relative to the first infusion, day relative to the most recent infusion) and will include the following information:

- AEs with fatal outcome (preferred terms of AEs with outcome=Fatal, as well as AEs of Grade 5),
- Flag for death within 30 days of last dose of study treatment.
- Flag for death within 60 days of first dose of study treatment.

### 16.2.2 Serious Adverse Events

The frequency (number and percentage) of patients with each of the following will be presented for treatment emergent SAEs by SOC and PT:

- SAEs
- Related SAEs

The listings of SAEs will also be provided with the relevant information (see description of listing in Section 16.1.1).

### 16.2.3 Adverse Events of Special Interest

#### 16.2.3.1 Infusion-Related Reaction

Infusion-Related Reactions (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 criteria on the timely relationship as detailed below:

**Reactions of IRR:** should be considered when onset is on the day of bintrafusp alfa infusion (during or after the infusion) 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 onset is on the day of bintrafusp alfa infusion (during or after the infusion) and resolved completely with the end date on the same day of the infusion or the day after.

IRR, 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
- Number of participants with IRR leading to infusion rate reduction
- Time related to first onset (infusion 1/ infusion 2/ infusion 3/ infusion 4 or later). The events should 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 at the same date (but not before dosing when time is recorded) or the following day of drug infusion.

The incidence of IRR will be described all dose level/cohorts combined (where applicable) and by dose level/cohort, separately for overall IRR and IRR subcategories:

- Incidence of IRR by Worst Grade, SOC, and PT
- Incidence of related IRR by Worst Grade, SOC, and PT

The listing of IRRs will be provided with the relevant information. Two listings for IRR subcategories will also be provided.

One additional listing will display the study drug administration details together with the infusion related adverse event. This listing will include administration date (day) /time, reason for modification, type of modification, modification start time, use of pre-medication, IRR AE Preferred Term, IRR AE grade, IRR AE start day /stop day, IRR AE time related to infusion.

### 16.2.3.2 Immune-Related Adverse Events

Immune-related adverse events (irAEs) will be identified programmatically. AEs which satisfy all of the following criteria will be flagged as immune-related:

1) The AE preferred term matches a preferred term on the list of pre-selected MedDRA terms.

2) The AE onset or worsening occurs after the first study drug administration and no more than 90 days after last dose, death or the earliest date of subsequent anticancer therapy minus 1 day, whichever occurs first.

3) On the AE eCRF page, the question "Were Corticosteroids, Immunosuppresants, or hormonal therapy (e.g. Thyroid) applied?" has the answer "Yes" selected.

4) On the imAE eCRF page, either:

a. The question "Does any of the following provide a clear etiology for the event?" 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

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

In the case that 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?" (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 as the answer to the question "Does any of the following provide a clear etiology for the event?"), the event will be considered as a non-irAE.

PTs will be compiled into categories: Immune-mediated rash, Immune-mediated colitis, Immune-mediated pneumonitis, Immune-related hepatitis, Immune-related nephritis and renal

dysfunction, Immune-mediated endocrinopathies (Adrenal insufficiency, Hypogonadism, Pituitary dysfunction, Type 1 Diabetes Mellitus, Thyroid disorders), Other immune-related adverse events (GVHD, Myasthenic syndrome, Myocarditis, Myositis, Neurologic events, Other, Pancreatitis, Uveitis, Vasculitis).

The overall summary of immune-related AEs (irTEAE) will include the following categories:

- Any irAEs
- irAEs by the worst grade
- Related irAEs
- Related irAEs by the worst grade
- irAEs leading to permanent treatment discontinuation
- Related irAEs leading to permanent treatment discontinuation
- Serious irAEs
- Related serious irAEs

The incidence of irAE will be described for all dose level/cohorts combined (where applicable) and by dose level/cohort:

- Incidence of irAE by Worst Grade, cluster and PT
- Incidence of related irAE by Worst Grade, cluster and PT
- irAEs leading to permanent treatment discontinuation by cluster and PT
- Related irAEs leading to permanent treatment discontinuation by cluster and PT

The listing of irTEAE will also be provided with the relevant information, including additional interventions for irAE (e.g. biopsies, surgical procedures, medical procedures) with a flag for irAEs with onset outside of the on-treatment period. A separate listing of irAEs with onset after the on-treatment period will also be provided.

### 16.2.3.3 Potential TGF-β-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 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

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

- All skin TEAE
- All skin TEAE by worst grade
- Skin TEAE leading to permanent treatment discontinuation
- Serious skin TEAEs

Tables for skin TEAEs frequency will be provided by MedDRA PTs (including both narrow and broad definition PTs). A listing of skin AEs will also be provided, containing participant identifier, age, sex, race, first and last date of study intervention, preferred term, 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 (Y/N).

#### 16.2.3.4 Anemia Adverse Events

The incidence of anemia will be described all dose level/cohorts combined (where applicable) and by dose level/cohort:

- Incidence by Worst Grade, SOC, and PT
- Incidence of related anemia by Worst Grade, SOC, and PT

### 16.2.4 Bleeding Events

Bleeding events of interest are the preferred terms belonging to the MedDRA SMQ Haemorrhage terms (excluding laboratory terms).

Bleeding events and study drug related bleeding events 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

## 16.3 Patient Narratives

An excel file will be provided for all patients falling into at least one of the following defined narrative categories:

1. A treatment-emergent fatal case with reason other than Progression Disease i.e. all subjects with TEAE with "Outcome" = Fatal in the "Adverse event details" page and AEDECOD not in ("Disease progression", "Neoplasm progression"), with reason other than "Disease progression" in the "Report subject's death" page (where Death date equal to AE end date), during the on-treatment period.
2. A treatment-related life-threatening case (with reason other than progression of disease, i.e. AEDECOD not in ("Disease progression", "Neoplasm progression")) during the on-treatment period and after the on-treatment period.
3. A bintrafusp alfa treatment related SAE during the on-treatment period and after the on-treatment period i.e. subjects with AE with "Relationship with M7824 treatment" = Related and "Serious adverse event" = Yes in the "Adverse event details" page.
4. Grade >3 treatment related AESI during the on-treatment period and after the on-treatment period.
5. A treatment-emergent AE leading to withdrawal from the treatment (only those AEs as primary reason for withdrawal), and not due to progression of disease (i.e. AEDECOD not in ("Disease progression", "Neoplasm progression")) during the on-treatment period.
6. Cases of significant medical information during the treatment-period and after the on-treatment period are determined on a case-by-case basis.

## 16.4 Clinical Laboratory Evaluation

### 16.4.1 Hematology and Chemistry Parameters

Baseline Laboratory Assessments are defined as the last non-missing observation prior to first dose of trial treatment, while Treatment Emergent Laboratory Assessments are any sample collected after 1st drug administration and within the on-treatment period. Laboratory values (including corresponding normal ranges) converted in standard unit will be used for summary statistics and shift tables.

Laboratory results will be classified according to the NCI-CTCAE criteria version 4.03. Non-numerical qualifiers (with exception of fasting flags) will not be taken into consideration in the derivation of CTCAE criteria (e.g., hypokalemia Grade 1 and Grade 2 are only distinguished by a non-numerical qualifier and therefore Grade 2 will not be derived, similar for Grade 1 and Grade 3 hyperuricemia where Grade 3 will not be derived). 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). Only subjects with post-baseline laboratory values will be included in the summaries.

Quantitative data will be summarized all dose level/cohorts combined (where applicable) and by dose level/cohort using descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline over time. Unscheduled assessments will be taken into account using the time-windows described in Table 1. End of Treatment visit will be summarized separately. Qualitative data based on reference ranges will be described according to the categories (i.e. Low, Normal, and High).

**Gradable parameters**

Gradable parameters to analyze with their respective NCI-CTC name and direction of abnormality are provided in Table 9. For parameters which are graded with two toxicities 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.

**Table 9 NCI-CTC Gradable parameters**

| Parameter (LBTEST) | Parameter code (LBTESTCD) | Name in NCI-CTC | Direction of abnormality |
|---|---|---|---|
| **Biochemistry** | | | |
| 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 | BILI | Blood bilirubin increased | High |
| Calcium | CA | Hypercalcemia/Hypocalcemia | High/Low |
| Cholesterol | CHOL | Cholesterol high | High |
| Creatinine | CREAT | Creatinine increased | High |
| Gamma Glutamyl Transferase | GGT | GGT increased | High |
| Glucose | GLUC | Hyperglycemia/Hypoglycemia | High/Low |
| Lipase | LIPASET | Lipase increased | High |
| Magnesium | MG | Hypermagnesemia/Hypomagnesemia | High/Low |
| Phosphate | PHOS | Hypophosphatemia | Low |
| Potassium | K | Hyperkalemia/Hypokalemia | High/Low |
| Sodium | SODIUM | Hypernatremia/Hyponatremia | High/Low |
| Triglycerides | TRIG | Hypertriglyceridemia | High |
| Urate | URATE | Hyperuricemia | High |
| **Hematology** | | | |
| Activated Partial Thromboplastin Time | APTT | Activated partial thromboplastin time prolonged | High |
| Haptoglobin[1] | HAPTOG | Haptoglobin decreased | Low |
| Hemoglobin | HGB | Anemia/Hemoglobin increased | Low/High |
| Leukocytes | WBC | Leukocytosis/White blood cell decreased | High/Low |
| Lymphocytes | LYM | Lymphocyte count decreased/Lymphocyte count increased | High/Low |
| Neutrophils | NEUT | Neutrophil count decreased | Low |
| Platelets | PLAT | Platelet count decreased | Low |
| Prothrombin Intl. Normalized Ratio | INR | INR increased | High |

[1] Haptoglobin is collected only in case of episode of anemia.

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)

If the range for the differential absolute count is not available (only range for value in % is available) then Grade 1 will be attributed as follows:

- Lymphocyte count decreased:
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % Lower Limit of Normal (LLN) value, and
  - derived absolute count ≥ 800/mm$^3$
- Neutrophil count decreased:
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥ 1500/mm$^3$

For **calcium**, CTCAE grading is based on Corrected Calcium and Ionized Calcium (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].

**Non-gradable parameters**

Non-gradable parameters are provided in Table 10.

**Table 10 Non-NCI-CTC Gradable Parameters**

| Parameter (LBTEST) | Parameter code (LBTESTCD) |
|---|---|
| **Biochemistry** | |
| Blood Urea Nitrogen | BUN |
| C Reactive Protein | CRP |
| Chloride | CL |
| Creatine Kinase | CK |
| Creatinine Clearance[1] | CREATCLR |
| Indirect Bilirubin[2] | BILIND |
| Lactate Dehydrogenase | LDH |
| Protein | PROT |
| Urea | UREA |
| **Hematology** | |
| Basophils/Leukocytes | BASOLE |
| Eosinophils/Leukocytes | EOSLE |
| Ery. Mean Corpuscular HGB Concentration | MCHC |
| Ery. Mean Corpuscular Hemoglobin | MCH |
| Ery. Mean Corpuscular Volume | MCV |
| Erythrocytes | RBC |
| Hematocrit | HCT |
| Monocytes/Leukocytes | MONOLE |

| Parameter (LBTEST) | Parameter code (LBTESTCD) |
|---|---|
| Prothrombin Time | PT |
| Reticulocytes/Erythrocytes | RETIRBC |

[1] Creatinine clearance is only collected at screening to check inclusion criteria. It will be presented in listings only.
[2] Indirect bilirubin is not measured but obtained by site calculation (=Total bilirubin – Direct bilirubin). It will be presented in listings only.

For all non-gradable parameters, the following summaries will be displayed per dose level/cohort and all dose level/cohorts combined:

- Shift from baseline to highest/lowest on-treatment value

For gradable parameters, the following summaries will be displayed per cohort and all cohorts combined:

- Number and percentage of subjects by worst on-treatment values (>=G1, >=G3, >=G4)
- Shift in toxicity grading from baseline to highest post-baseline toxicity

The following figures will also be provided for each dose level/cohort:

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

**Liver function tests**: Alanine aminotransferase (ALT), aspartate aminotransferase (AST), and Total Bilirubin are used to assess possible drug induced liver toxicity. The ratios of test result over upper limit of normal (ULN) will be calculated and classified for these three parameters during the on-treatment period.

Summary of liver function tests will include the following categories. The number and percentage of patients with each of the following during the on-treatment period will be summarized for all dose level/cohorts combined (where applicable) and by dose level/cohort:

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

Concurrent measurements are those occurring on the same date.

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

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will also be created by graphically displaying peak serum ALT(/ULN) vs. peak total bilirubin (/ULN) including reference lines at ALT >3×ULN and total bilirubin >2×ULN. The display 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 quadrants indicate patients with possible Gilbert's cholestasis; the right upper quadrant are the possible Hy's Law patients; the right lower quadrant is possible Temple's Corollary (patients with ALT ≥ 3 x ULN but not satisfying Hy's Law).

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

Separate listings of Hematology and Chemistry will be created. Each listing will include assigned dose level/cohort, subject identifier, age, sex, first dose date, last dose date, dose, number of doses, laboratory parameter (units), visit, date, SI value, LLN, ULN, Indicator of Normal Range (Low, Normal, High), and toxicity grade (if applicable). These listings will be sorted by assigned dose level/cohort, subject identifier, and laboratory measurement date.

### 16.4.2 Viral Status

Listing of viral status parameters (HIV, HBV, HCV and HDV) will be provided.

### 16.4.3 Other Laboratory Parameters

All other parameters collected on the eCRF will be listed in dedicated listings presenting all data collected in the eCRF:

- Basic urinalysis (pH, Specific gravity, Glucose, Bilirubin, Ketones, Blood, Color, Protein, Nitrite, Leukocyte esterase, Urobilinogen, Macroscopic appearance) and Urinalysis microscopic evaluation (Red blood cell count, White blood cell count, Epithelial cells, Bacteria, Crystals, Casts, Mucus)
- Serum electrophoresis (Total protein, Albumin) – Only for dose escalation
- Hormonal test (ACTH, ANA, RF, T4, TSH)
- Pregnancy test

Those listings will be sorted by assigned dose level/cohort, subject identifier, lab category, and parameters within each lab category, date.

## 16.5 Vital Signs

Quantitative data will be summarized all dose level/cohorts combined (where applicable) and by dose level/cohort using simple descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline over time.

The maximum changes of vital sign measurements from the baseline to maximum on-treatment change will be grouped as follows:

**Table 11 Categories of Maximum Change from Baseline in Vital Signs**

| Vital Sign Parameter | Increase/ decrease | Baseline category | Change from baseline category |
|---|---|---|---|
| Temperature | Increase | <37 °C,<br>37 - <38 °C<br>38 - <39 °C<br>39 - <40 °C<br>≥ 40 °C | < 1°C,<br>1-<2°C,<br>2-<3°C,<br>≥ 3 °C |
| Weight | Increase | Any | <10%,<br>≥10% |
| | Decrease | Any | <10%,<br>≥10% |
| Heart rate | Increase | <100 beats/min;<br>≥ 100 beats/min | ≤20 beats/min,<br>>20 – 40 beats/min,<br>>40 beats/min |
| | Decrease | <50 beats/min;<br>≥ 50 beats/min | ≤20 beats/min,<br>>20 – 40 beats/min,<br>>40 beats/min |
| SBP | Increase | <140 mmHg;<br>≥ 140 mmHg | ≤20 mmHg,<br>>20 – 40 mmHg,<br>>40 mmHg |
| | Decrease | <95 mmHg;<br>≥ 95 mmHg, | ≤20 mmHg,<br>>20 – 40 mmHg,<br>>40 mmHg |
| DBP | Increase | <90 mmHg;<br>≥ 90 mmHg | ≤20 mmHg,<br>>20 – 40 mmHg,<br>>40 mmHg |
| | Decrease | <45 mmHg;<br>≥ 45 mmHg | ≤20 mmHg,<br>>20 – 40 mmHg,<br>>40 mmHg |
| Respiration rate | Increase | <20 breaths/min; | ≤5 breaths/min, |

| Vital Sign Parameter | Increase/ decrease | Baseline category | Change from baseline category |
|---|---|---|---|
| | | ≥ 20 breaths/min/min | >5 – 10 breaths/min,<br>>10 breaths/min |
| | Decrease | <20 breaths/min;<br>≥ 20 breaths/min | ≤5 breaths/min,<br>>5 – 10 breaths/min,<br>>10 breaths/min |

For each patient, the worst on-treatment value will be calculated. Missing values will define a separate category.

The following summaries will be prepared for vital sign parameters as grouped above, considering only subjects with post baseline values:

- Maximal Shifts (changes in categories)
- Listing of highest change per subject

A table of potentially clinically meaningful changes will also be displayed:

- ≤ 95 mmHg and decrease from baseline ≥ 20 mmHg in systolic blood pressure
- ≥ 140 mmHg and increase from baseline ≥ 20 mmHg in systolic blood pressure
- ≤ 45 mmHg and decrease from baseline ≥ 20 mmHg in diastolic blood pressure
- ≥ 90 mmHg and increase from baseline ≥ 20 mmHg in diastolic blood pressure
- ≤ 50 beats/min and decrease from baseline ≥ 20 beats/min in heart rate
- ≥ 100 beats/min and increase from baseline ≥ 20 beats/min in heart rate
- ≥ 10% weight increase
- ≥ 10% weight decrease

A listing of vital sings will be provided. The Listing will include assigned dose level/cohort, subject identifier, age, sex, vital sign parameter (units), visit, date, value, baseline value, change from baseline. The listings will be sorted by assigned dose level/cohort, subject identifier, vital sign parameter, and vital sign measurement date.

## 16.6 Other Safety or Tolerability Evaluations

### 16.6.1 ECG

The 12-lead Electrocardiogram (ECG) assessments will be performed according to the Schedule of Assessments in the Clinical Trial Protocol.

ECG summaries will include all ECG assessments at baseline and during the on-treatment period. All ECG assessments will be listed, and those collected outside the on-treatment period will be flagged in the listing.

The following analyses will be performed all dose level/cohorts combined (where applicable) and by dose level/cohort, during the on-treatment period.

- Shift table from normal baseline values to abnormal post-baseline values.

A listing of ECG values will be provided. The listing will include dose level/cohort, subject identifier, age, sex, ECG parameter, visit, ECG date, value, unit, baseline value, change from baseline, any abnormal findings. Qualitative ECG results will also be provided in the listing. The listing will be sorted by dose level/cohort, subject identifier, ECG parameter, and ECG date.

### 16.6.2 ECOG Performance Status

The ECOG shift from baseline to the highest score during the on-treatment period will be summarized for all dose level/cohorts combined (where applicable) and by dose level/cohort. ECOG performance status will also be presented in a listing at each time point.

The outcomes of ECOG Performance Status are defined as below:

- 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.

### 16.6.3 Oxygen Saturation

By-visit summaries of oxygen saturation and change from baseline will be provided.

Oxygen saturation will be presented in a listing at each time point with changed values from baseline.

# 17 Pharmacokinetics and Pharmacodynamics

The analyses described in this section will be performed by the Clinical PK/CCI 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 SAF. Summaries and statistical analyses will be based on the PKAS. CCI

Pharmacokinetic concentrations/parameters refer to bintrafusp alfa concentrations/PK parameters. Pharmacodynamic concentrations / PDyn parameters refer to "CD3+ PD-L1 RO%" (also referred to PD-L1 target occupancy), "TGFß-1", "TGFß-2", and "TGFß-3".

## 17.1 Missing PK/CCI Data

**Concentrations below the lower limit of assay quantification**

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

Pharmacokinetic concentrations <LLOQ, which are before the last quantifiable data point, will be taken as zero for calculating the area under the serum concentration-time curve (AUC) of single dose profiles. Concentration below LLOQ, which occur after the last quantifiable data point will not be considered in the calculation of the terminal first order rate constant ($\lambda_z$).

**Deviations, missing concentrations, and anomalous values**

There will be no imputation of missing data. Concentrations will be set to missing in summary tables if the value is reported as no result. Pharmacokinetic/CCI concentrations which are erroneous due to a protocol violation (as defined in the clinical study protocol), documented handling error, or analytical error (as documented in the bioanalytical report) may be excluded from the PK/CCI analysis if agreed upon prior to performing a statistical analysis. In this case the rationale for exclusion must be provided in the relevant listing/table.

Exclusions for concentration data descriptive statistics

- Positive pre-dose values on Day 1
- Concentration observed at the end of infusion ($C_{EOI}$) < 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/CCI concentrations that appear implausible to the Pharmacokineticist/PK/CCI Data Analyst must not be excluded from the analysis. Any implausible data will be documented in the relevant listing/table.

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, 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, and instead the result will be listed in a separate table.

## 17.2 Descriptive PK and CCI Analysis

**Presentation of PK/PD Concentration Data**

A by-participant listing will present PK CCI sample times, time deviations, and concentrations based on the SAF. Concentrations will be reported with the same precision as the source data.

**Tables**

Pharmacokinetic/CCI concentration or data will be presented in tables and descriptively summarized by treatment and cohort (as appropriate), day and nominal time using: number of non-missing observations (n), arithmetic mean (Mean), StD, Min, median (Median), and Max. Pharmacokinetic summaries will additionally present coefficient of variation (CV%), geometric mean (GeoMean), StD of log-transformed data (logStD), and geometric coefficient of variation (GeoCV%). Pharmacodynamic data summaries will additionally present number and percentage of missing observations, Q1, and Q3. Summaries will be based on the PKAS.

Additional table(s) will summarize PK concentrations with further stratification by ADA subsets ever positive and never positive, 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 15). Additional table(s) will summarize PK concentrations with stratification by Japanese, andnon-Japanese as applicable, based on the PKAS. CCI

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

| | |
|---|---|
| Mean, Min, Median, Max, Q1, Q3, GeoMean: | 3 significant digits |
| StD, logStD: | 4 significant digits |

CV%, GeoCV%: 1 decimal place

**Figures**

Individual PK concentration-time profiles showing all participants by treatment and cohort will be created using the actual time points and the numeric concentration data. Plots of individual data will be based on the SAF. Median, GeoMean and Mean concentration-time profiles by treatment and cohort 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. Summary (Mean/Median) PK plots will be 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, 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 15), based on the PKNAB. Additional figures will present Mean and Median PK concentration-time profiles with stratification by Japanese and non-Japanese as applicable, based on the PKAS. CCI

CCI A line displaying the LLOQ of the corresponding assay will be added. Plots will be presented on a linear scale only, and data will be based on the SAF.

CCI A line displaying the LLOQ of the corresponding assay will be added. Data will be based on the PKAS.

Pharmacodynamic metrics ("TGFß" concentrations) will be plotted against PK concentrations (bintrafusp alfa), for all participants both on a linear and on a semi-logarithmic scale. Only PK CCI samples, both taken at the same time points following intravenous (IV) infusion on Day 1 are part of these plots. These plots will be based on the PKAS.

## 17.3 Pharmacokinetic Non-Compartmental Analysis

The PK parameters listed below will be calculated for bintrafusp alfa using the actual time elapsed from dosing (or using scheduled time if actual time is not available).

| | |
|---|---|
| $C_{max}$ | Maximum observed concentration in serum |
| $t_{max}$ | Time to reach $C_{max}$ |
| $AUC_{0-t}$ | Area under the serum concentration-time curve from time zero to the last quantifiable concentration |

| | |
|---|---|
| $C_{trough}$ | The concentration observed at the end of the dosing interval, before next dosing (serum trough concentration). This PK parameter will be taken directly from the observed bintrafusp alfa concentration-time data. |
| $C_{EOI}$ | The concentration observed at the end of infusion. This PK parameter will be taken directly from the observed bintrafusp alfa concentration-time data. |

When applicable the following parameters will also be calculated:

| | |
|---|---|
| $AUC_{0-\infty}$ | Area under the serum concentration-time curve from time zero extrapolated to infinity. $AUC_{0-\infty} = AUC_{0-t} + AUC_{extra}$, where $AUC_{extra} = C_{lastpred}/ \lambda_z$. |
| $AUC_{extra\%}$ | Percentage of $AUC_{0-\infty}$ obtained by extrapolation, calculated by $(1 - [AUC_{0-t}/AUC_{0-\infty}])\times100$. If $AUC_{extra\%}$ is greater than 20.0%, $AUC_{0-\infty}$ and $\lambda_z$ and parameters derived from them (e.g. $t_{1/2}$, CL, and $V_z$ will be included in the Phoenix® WinNonlin® parameter outputs, summaries, and inferential statistics, but will be flagged. |
| $AUC_{0-336}$ | Area under the serum concentration-time curve from time zero to 336 h. This parameter will be calculated using nominal time at 336 hours, by extrapolation/interpolation as necessary. |
| $t_{1/2}$ | Elimination half-life, calculated by $\ln2/ \lambda_z$ |
| $\lambda_z$ | Terminal 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 |
| CL | Total body clearance of drug from serum. CL = Dose IV / $AUC_{0-\infty}$ |
| Vz | Volume of distribution during terminal phase. $V_Z$ = Dose/($AUC_{0-\infty}$ * $\lambda_z$) |
| $AUC_{0-t}$/Dose | Dose-Normalized $AUC_{0-t}$. Two dose normalizations are performed:<br>1) Dose normalization using actual Dose in mg/kg<br>2) Dose normalization using actual Dose in mg. |
| $C_{max}$/Dose | Dose-Normalized $C_{max}$. Two dose normalizations are performed:<br>1) Dose normalization using actual Dose in mg/kg<br>2) Dose normalization using actual Dose in mg |
| $AUC_{0-\infty}$/Dose | Dose-Normalized $AUC_{0-\infty}$. Two dose normalizations are performed:<br>1) Dose normalization using actual Dose in mg/kg<br>2) Dose normalization using actual Dose in mg. |
| $AUC_{0-336}$/Dose | Dose-Normalized $AUC_{0-336}$. Two dose normalizations are performed:<br>1) Dose normalization using actual Dose in mg/kg<br>2) Dose normalization using actual Dose in mg |

The following PK parameters will be calculated for diagnostic purposes and listed, but will not be summarized:

- The time interval (h) of the log-linear regression to determine $\lambda_z$ ($t_{½}$, Interval).
- Number of data points ($t_{½}$, N) included in the log-linear regression analysis to determine $\lambda_z$.
- Goodness of fit statistic (Rsq) for calculation of $\lambda_z$. If Rsq is <0.800, $\lambda_z$ and parameters derived from it (e.g. $t_{1/2}$, $AUC_{0-\infty}$, CL, and $V_z$) will be included in the Phoenix® WinNonlin® parameter outputs, summaries, and inferential statistics, but will be flagged.

The regression analysis (determination of $\lambda_z$) 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 twofold the resulting $t_{½}$ itself. Phoenix® WinNonlin® best fit methodology will be used as standard. The last quantifiable concentration should always be included in the regression analysis, while the concentration at $t_{max}$ and any concentrations <LLOQ which occur after the last quantifiable data point should not be used.

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. The pre-dose samples will be considered as if they had been taken simultaneously with the administration.

The analysis of dose proportionality of PK parameters of M7824 will be quantified as part of an exploratory analysis using the power model on the parameters ($\ln[PKparameter] = \alpha + \beta \times \ln[dose]$) (mg/kg dosing only, e.g., dose escalation and HCC). This analysis will be based on $AUC_{0-\infty}$, $AUC_{0-t}$, $AUC_{0-336}$, and $C_{max}$. The intercept $\alpha$ and the slope $\beta$ together with 90% CIs will be estimated and presented. A p-value testing whether $\beta = 1$ will also be presented.

**Presentation of PK Parameter Data**

Individual PK parameters will be listed by nominal study day based on the SAF. Individual PK parameters will be reported with the same precision as the source data.

Pharmacokinetic parameters will be reported to 3 significant figures. In export datasets, as well as in the SDTM PP/XD 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.

Pharmacokinetic parameter data will be listed in tables and descriptively summarized by treatment (dose, e.g. 1200 mg) and cohort separately, and day using: n, Mean, StD, CV%, Min, Median, Max, GeoMean, logStD, GeoCV%, and the 95% CI for the GeoMean (LCI 95% GM, UCI 95% GM). Summaries will be based on the PKAS.

Additional table(s) will summarize NCA parameters following first infusion, $C_{trough}$ and $C_{EOI}$ with further stratification by ADA subsets ever positive and never positive, based on the PKADA. Additional table(s) will summarize NCA parameters following first infusion, $C_{trough}$

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 (PD-L1 and TGFß receptor neutralizationversus never positive (see Section 15). Additional table(s) will summarize NCA parameters following first infusion, $C_{trough}$, and $C_{EOI}$ with stratification by Japanese andnon-Japanese as applicable, based on the PKAS. Additional table(s) will summarize NCA parameters following first infusion, $C_{trough}$, and $C_{EOI}$ with further sub-stratification by ADA subsets ever positive and never positive and, nested within, Japanese and non-Japanese as applicable based on the PKADA. Additional table(s) will summarize $C_{trough}$ of ADA ever positive participants with further sub-stratification by ADA subgroups (e.g. Pre-existing, Treatment boosted, Treatment emergent, Transient positive, Persistent positive), based on the PKADA. For nAb ever-positive participants, serum M7824 $C_{trough}$ 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 Treatment-emergent participants and nAb Treatment-emergent participants by PK day relative to seroconversion, based on the PKADA and PKNAB, respectively. CCI [redacted] All above will be summarized by treatment (dose, e.g. 1200 mg across cohorts) and cohort separately.

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 |

Individual $C_{trough}$ and $C_{EOI}$ values will be plotted against actual time points on a linear scale, for all participants by treatment and cohort. Plots of individual data will be based on the SAF.

Arithmetic mean $C_{trough}$ ±StD and $C_{EOI}$ ±StD and GeoMean $C_{trough}$ ±logStD and $C_{EOI}$ ±logStD will be plotted versus nominal day on a linear scale. Median $C_{trough}$ and $C_{EOI}$ will also be plotted versus nominal day on a linear scale. Summary plots will be based on the PKAS. Additional figures will present Mean, GeoMean, and Median $C_{trough}$ and $C_{EOI}$ with further stratification by ADA subsets ever positive and never positive, based on the PKADA. Additional figures will present Mean, GeoMean and Median $C_{trough}$ 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 versus never positive (see Section 15). Additional figures will present Mean, GeoMean and Median $C_{trough}$ and $C_{EOI}$ with stratification by Japanese and non-Japanese, based on the PKAS. CCI [redacted]

For ADA treatment-emergent subjects 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 each cohort (for readability, split further into groups of 10 subjects or fewer as needed). Box plots will be prepared for $C_{trough}$ versus PK day relative to seroconversion.

For nAb treatment-emergent subjects 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 each cohort (for readability, split further into groups of 10 subjects or fewer as needed). Box plots will be prepared for $C_{trough}$ versus PK day relative to seroconversion.

All above will be summarized by treatment (dose, e.g. 1200 mg across cohorts) and cohort separately.

Dose proportionality will be evaluated for using data from participants with a full PK profile, and will be presented graphically as follows:

Regression plots for individual $AUC_{0\text{-}t}$, $AUC_{0\text{-}336}$, $AUC_{0\text{-}\infty}$ and $C_{max}$ values (where applicable) and dose normalized $AUC_{0\text{-}t}$, $AUC_{0\text{-}336}$, $AUC_{0\text{-}\infty}$ and $C_{max}$ versus dose on a linear scale. All the above dose proportionality plots will be repeated with further substratification by Japanese and non-Japanese as applicable. Additional plots showing fit of power model described above will be presented for $AUC_0$-t and Cmax. These plots will be based on the PKAS.

The Phoenix WinNonlin NCA Core Output will be provided in a separate listing.

# 18 References

Brookmeyer, R. and Crowley, J. (1982), A Confidence Interval for the Median Survival Time, Biometrics, 38, 29–41.

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

Eisenhauer EA, Therasse P, Bogaerts J, et al. (2009), New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer; 45:228-47.

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

Kaplan EL, Meier P (1958). Nonparametric estimation from incomplete observations. J Am Stat Assoc. 53: 457-81.

# 19 Appendices

## 19.1 Appendix 1: Important Protocol Deviations

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| **Inclusion criteria:**<br>**For the subject to be eligible for inclusion, each criterion must be checked 'YES':** | | | | | | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Signed written informed consent | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation # 1<br>Expansion # 1 | PDEV01 | 5.3.1.1 for dose escalation; 5.3.1.2 for expansion cohorts | Programmed<br>Medical Review. | |
| Male or female subjects aged ≥ 20 years | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation # 2<br>Expansion # 2 | PDEV02 | 5.3.1.1 for dose escalation; 5.3.1.2 for expansion cohorts | Programmed. | |
| Histologically or cytologically provencondition as described in inclusion criterion #3 of protocol | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation: # 3<br>Expansion #3 | PDEV03 | 5.3.1.1 for dose escalation; 5.3.1.2 for expansion cohorts | Medical Review | |
| Availability of tumor (primary or metastatic) archival material or fresh biopsies within 28 days before first administration of IMP is mandatory. If no archival material is available and only one lesion is amenable for biopsy and it is the only target lesion, the Medical Monitor should be consulted for subject eligibility). Tumor biopsies and tumor archival material must be suitable for biomarker assessment. | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation NA<br>Expansion #4 | PDEV04 | 5.3.1.2 for expansion cohorts | Programmed | |
| Disease must be measurable with at least 1 unidimensionally measurable lesion by RECIST 1.1. | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation NA<br>Expansion #5 | PDEV05 | 5.3.1.2 for expansion cohorts | Programmed | |
| ECOG performance status of 0 to 1 at trial entry | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation #4<br>Expansion #6 | PDEV06 | 5.3.1.1 for dose escalation; 5.3.1.2 for expansion cohorts | Programmed | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Life expectancy ≥12 weeks as judged by the Investigator. | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation #5<br>Expansion #7 | PDEV07 | 5.3.1.1 for dose escalation; 5.3.1.2 for expansion cohorts | Medical Review | |
| Adequate hematological function defined by White Blood Cell (WBC) count ≥ 3 x $10^9$/L with absolute neutrophil count (ANC) ≥ 1.5 x $10^9$/L, lymphocyte count ≥ 0.5 x $10^9$/L, platelet count ≥ 120 x $10^9$/L, and Hgb ≥ 9 g/dL (in absence of blood transfusion) | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation #6<br>Expansion #8 excluding HCC cohorts | PDEV08 | 5.3.1.1 for dose escalation; 5.3.1.2 for expansion cohorts | Programmed | |
| Adequate hepatic function defined by a total bilirubin level ≤ 1.5 x the upper limit of normal range (ULN), an AST level ≤ 2.5 x ULN, and an ALT level ≤ 2.5 x ULN | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Expansion #9 excluding HCC cohort | PDEV09 | 5.3.1.2 for expansion cohorts | Programmed | |
| Adequate renal function defined by an estimated creatinine clearance > 50 mL/min according to the Cockcroft-Gault formula or by measure of creatinine clearance from 24 hour urine collection | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation #8<br>Expansion #10 | PDEV10 | 5.3.1.1 for dose escalation; 5.3.1.2 for expansion cohorts | Programmed | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Highly effective contraception (that is, methods with a failure rate of less than 1% per year) for both male and female subjects if the risk of conception exists Highly effective contraception must be used 30 days prior to first trial treatment administration, for the duration of trial treatment, and at least for 4 months after stopping trial treatment. Should a woman become pregnant or suspect she is pregnant while she or her partner is participating in this trial, the treating physician should be informed immediately. | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation #9<br>Expansion #11 | PDEV11 | 5.3.1.1 for dose escalation; 5.3.1.2 for expansion cohorts | Medical Review | |
| Woman of childbearing potential must have a negative serum pregnancy test at screening visit and a negative serum or urine pregnancy test at Day 1 before dosing, if applicable. | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation #10<br>Expansion #12 | PDEV12 | 5.3.1.1 for dose escalation; 5.3.1.2 for expansion cohorts | Medical Review | |
| Adequate hepatic function defined by a total bilirubin level ≤ 1.5 × ULN, an AST level ≤ 2.5 × ULN, and an ALT level ≤ 2.5 × ULN. For subjects with liver involvement in their tumor, AST ≤ 5.0 × ULN, ALT ≤ 5.0 × ULN, and bilirubin ≤ 3.0 is acceptable. | Eligibility and Entry Criteria | Subject did not meet Inclusion Criteria:<br>Escalation #7 | PDEV13 | 5.3.1.1 for dose escalation | Medical Review | |
| **Exclusion criteria:**<br>**For the subject to be eligible for inclusion, each criterion must be checked 'NO':** | | | | | | |
| Concurrent treatment with non-permitted drugs | Eligibility and Entry Criteria | Subject met Exclusion Criteria 1 | PDEV14 | 5.3.2, 6.5.2 | Medical Review | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Prior therapy with any antibody/drug targeting T cell co-regulatory proteins (immune checkpoints) such as anti-PD-1, anti-PD-L1, anti-CTLA-4 antibody (consult Medical Monitor if necessary), or anti-4-1BB antibody is not allowed, inclusive of intrahepatic, localized administration of such agents. | Eligibility and Entry Criteria | Subject met Exclusion Criterion 2 | PDEV15 | 5.3.2 | Medical Review | |
| Prior therapy with any antibody/drug targeting TGFβ/TGFβ receptor. | Eligibility and Entry Criteria | Subject met Exclusion Criterion 3 | PDEV16 | 5.3.2 | Medical Review | |
| Anticancer treatment within 21 days before the start of trial treatment, eg, cytoreductive therapy, radiotherapy involving more than 30% of the bone marrow (with the exception of palliative bone directed radiotherapy), immune therapy, or cytokine therapy | Eligibility and Entry Criteria | Subject met Exclusion Criterion 4 | PDEV17 | 5.3.2 | Medical Review | |
| Anticancer treatment with antibody within 28 days before the start of trial treatment. | Eligibility and Entry Criteria | Subject met Exclusion Criterion 5 | PDEV18 | 5.3.2 | Medical Review | |
| Major surgery within 28 days before the start of trial treatment (excluding prior diagnostic biopsy); | Eligibility and Entry Criteria | Subject met Exclusion Criterion 6 | PDEV19 | 5.3.2 | Medical Review | |
| Systemic therapy with immunosuppressive agents within 7 days before the start of the trial treatment; or use of any investigational drug within 28 days before the start of trial treatment. | Eligibility and Entry Criteria | Subject met Exclusion Criterion 7 | PDEV20 | 5.3.2 | Medical Review | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Previous malignant disease other than the target malignancy to be investigated in this trial with the exception of cervical carcinoma in situ and superficial or non-invasive bladder cancer (treated with curative intent) within the last 5 years or basal cell or squamous cell carcinoma in situ within the last 3 years. | Eligibility and Entry Criteria | Subject met Exclusion Criteria 8 | PDEV21 | 5.3.2 | Medical Review | |
| Rapidly progressive disease which, in the opinion of the Investigator, may predispose to inability to tolerate treatment or trial procedures. | Eligibility and Entry Criteria | Subject met Exclusion Criteria 9 | PDEV22 | 5.3.2 | Medical Review | |
| Subjects with active central nervous system (CNS) metastases causing clinical symptoms or metastases that require therapeutic intervention are excluded. Subjects with a history of treated CNS metastases (by surgery or radiation therapy) are not eligible unless they have fully recovered from treatment, demonstrated no progression for at least 2 months, and do not require continued steroid therapy. Subjects with CNS metastases incidentally detected during Screening which do not cause clinical symptoms and where no therapeutic intervention is needed should be discussed with the Sponsor. | Eligibility and Entry Criteria | Subject met Exclusion Criteria 10 | PDEV23 | 5.3.2 | Medical Review | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Receipt of any organ transplantation, including allogenic stem-cell transplantation, but with the exception of transplants that do not require immunosuppression (eg, corneal transplant, hair transplant) | Eligibility and Entry Criteria | Subject met Exclusion Criteria 11 | PDEV24 | 5.3.2 | Medical Review | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Significant acute or chronic infections including, among others:<br>• Positive test for human immunodeficiency virus (HIV) or known acquired immunodeficiency syndrome<br>• Except for the HCC cohort, HBV infection (HBV surface antigen positive or HBV core antibody positive with reflex to positive HBV DNA) or HCV infection (positive HCV antibody with reflex to positive HCV RNA)<br>• Subjects with active tuberculosis (history of exposure or history of positive tuberculosis test; plus presence of clinical symptoms, physical or radiographic findings) | Eligibility and Entry Criteria | Subject met Exclusion Criteria 12 | PDEV25 | 5.3.2 | Medical Review | |
| Active autoimmune disease that might deteriorate when receiving an immunostimulatory agent (exceptions noted in protocol) | Eligibility and Entry Criteria | Subject met Exclusion Criterion 13 | PDEV26 | 5.3.2 | Medical Review | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Interstitial lung disease or its history. | Eligibility and Entry Criteria | Subject met Exclusion Criterion 14 | PDEV27 | 5.3.2 | Medical Review | |
| Known history of hypersensitivity reactions to M7824 or its products, or known severe hypersensitivity reactions to monoclonal antibodies (Grade ≥3 NCI-CTCAE v4.03), any history of anaphylaxis, or recent, within 5 months, history of uncontrolled asthma. | Eligibility and Entry Criteria | Subject met Exclusion Criterion 15 | PDEV28 | 5.3.2 | Medical Review | |
| Persisting toxicity (except alopecia and vitiligo) related to prior therapy Grade > 1 NCI-CTCAE v4.03, however sensory neuropathy Grade ≤ 2 is acceptable | Eligibility and Entry Criteria | Subject met Exclusion Criterion 16 | PDEV29 | 5.3.2 | Medical Review | |
| Pregnancy or currently in lactation | Eligibility and Entry Criteria | Subject met Exclusion Criterion 17 | PDEV30 | 5.3.2 | Medical Review | |
| Known alcohol or drug abuse | Eligibility and Entry Criteria | Subject met Exclusion Criterion 18 | PDEV31 | 5.3.2 | Medical Review | |
| Clinically significant cardiovascular / cerebrovascular disease as follows: cerebral vascular accident / stroke (< 6 months prior to enrollment), myocardial infarction (< 6 months prior to enrollment), unstable angina, congestive heart failure (New York Heart Association Classification Class ≥ II), or serious cardiac arrhythmia | Eligibility and Entry Criteria | Subject met Exclusion Criterion 19 | PDEV32 | 5.3.2 | Medical Review | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Clinically relevant diseases (eg, inflammatory bowel disease) and/or uncontrolled medical conditions, which, in the opinion of the Investigator, might impair the subject's tolerance or ability to participate in the trial. | Eligibility and Entry Criteria | Subject met Exclusion Criterion 20 | PDEV33 | 5.3.2 | Medical Review | |
| Any psychiatric condition that would prohibit the understanding or rendering of informed consent | Eligibility and Entry Criteria | Subject met Exclusion Criterion 21 | PDEV34 | 5.3.2 | Medical Review | |
| Legal incapacity or limited legal capacity | Eligibility and Entry Criteria | Subject met Exclusion Criterion 22 | PDEV35 | 5.3.2 | Medical Review | |
| Vaccine Administration within 4 weeks of IMP administration. Vaccination with live vaccines while on trial is prohibited. Administration of inactivated vaccines is allowed (for example, inactivated influenza vaccines). | Eligibility and Entry Criteria | Subject met Exclusion Criterion 23 | PDEV36 | 5.3.2 | Medical Review | |
| **Additional exclusion criteria for subjects in the HCC cohort** | | | | | | |
| Any prior or current ascites that requires paracentesis for control; or history of variceal bleeding; or history of hepatic encephalopathy; or history of obstructive jaundice. | Eligibility and Entry Criteria | Subject met Exclusion Criterion 24 | PDEV37 | 5.3.2 | Medical Review | |
| Hepatitis D virus (HDV) co-infection with hepatitis B virus (HBV; if HBV surface antigen or HBV DNA positivity at Screening then must check for HDV status) | Eligibility and Entry Criteria | Subject met Exclusion Criterion 25 | PDEV38 | 5.3.2 | Medical Review | |
| Chemoembolization or radioembolization within 28 days prior to IMP administration | Eligibility and Entry Criteria | Subject met Exclusion Criterion 26 | PDEV39 | 5.3.2 | Medical Review | |
| **Concomitant Medication** | | | | | | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Non-permitted concomitant medication during the study | Concomitant Medication | Subjects took immunotherapy, immunosuppressive drugs (for example, chemotherapy or systemic corticosteroids except for short term treatment of allergic reactions, endocrine replacement therapy at low dose prednisone [≤ 10 mg daily] or equivalent, or for the treatment of irAEs or other appropriate short term steroid use), or other experimental pharmaceutical products. Short term administration of systemic steroid or other immunosuppressant such as infliximab or mycophenolate (that is, for allergic reactions or the management of irAEs) is allowed. Steroids with no or minimal systemic effect (topical, inhalation) are allowed. **Note:** for subjects with glioblastoma, steroid use is allowed. | PDEV40 | 6.5.2 | Medical Review | |
| Non-permitted concomitant medication during the study | Concomitant Medication | Adefovir | PDEV41 | 6.5.2 | Medical Review | |

| Protocol Reference Text | Protocol Deviation Category | Description of Protocol Deviation | Deviation Code | Protocol Section | Medical review/programmed* | |
|---|---|---|---|---|---|---|
| Non-permitted concomitant medication during the study | Concomitant Medication | Prophylactic corticosteroids for infusion related reactions were administered to the subject. | PDEV42 | 6.5.2 | Medical Review | |
| Non-permitted concomitant medication during the study | Concomitant Medication | Any live vaccine therapies for the prevention of infectious disease were administered to the subject. | PDEV43 | 6.5.2 | Medical Review | |
| Non-permitted concomitant medication during the study | Concomitant Medication | Blood transfusions and erythroid growth factors were administered to a patient during the 21 days DLT window during the dose escalation phase. | PDEV44 | 6.5.2 | Medical Review | |
| **Withdrawal from treatment/trial** | | | | | | |
| Subjects that developed withdrawal criteria whilst on study but were not withdrawn; | Other Criteria | Subject became pregnant, but continued on study. | PDEV45 | 5.5.1 | Medical Review Required. Programmed | |
| Subjects that developed withdrawal criteria whilst on study but were not withdrawn; | Other Criteria | Subject developed any Grade ≥ 3 ADRs or repetitive Grade 2 ADRs, but continued on study. | PDEV46 | 5.5.1 | Medical Review | |
| Subjects that developed withdrawal criteria while on the study but were not withdrawn; | Other Criteria | Subjects were not withdrawn despite therapeutic failure requiring urgent additional drug. | PDEV47 | 5.5.1 | Medical Review | |
| **Other Criteria** | | | | | | |
| Subjects dosing error (≥ +/- 20% assigned dose) | IP Compliance | Subject had dosing error. | PDEV48 | 6.2 | Programmed | |
| NA | Other Criteria | Other protocol deviation | PDEV99 | Medical defined | Medical Review | |

* Programmed by Data Management team.

Signature Page for VV-CLIN-253237 v4.0

| | |
|---|---|
| Approval | PPD<br>-2022 23:39:57 GMT+0000 |

Signature Page for VV-CLIN-253237 v4.0