# Integrated Analysis Plan

| | |
|---|---|
| **Clinical Study Protocol Identification No.** | MS200647_0047 |
| **Title:** | A Phase II, Multicenter, Open-label Study to Investigate the Clinical Efficacy of M7824 Monotherapy in Participants With Locally Advanced or Metastatic Biliary Tract Cancer Who Fail or are Intolerant to First-line Platinum-Based Chemotherapy |
| **Study Phase** | Phase II |
| **Investigational Medicinal Product(s)** | Bintrafusp Alfa (M7824) |
| **Clinical Study Protocol Version** | 10 October 2019/Version 3.0 |
| **Integrated Analysis Plan Author** | **Coordinating Authors**<br>PPD, PPD<br>PPD, PPD<br>PPD |
| **Integrated Analysis Plan Date and Version** | 26 August 2020, Version 2 |
| **Integrated Analysis Plan Reviewers** | **Function**<br>PPD, Merck Healthcare<br>PPD, Merck Healthcare<br>PPD, PPD<br>PPD, PPD<br>PPD, Merck Healthcare<br>PPD, Merck Healthcare<br>PPD, PPD<br>PPD, PPD<br>PPD, EMD Serono<br>PPD, Merck Healthcare<br>PPD Merck Healthcare<br>PPD EMD Serono<br>PPD, PPD<br>PPD, PPD<br>PPD, PPD<br>PPD, EMD Serono<br>PPD, Merck Healthcare<br>PPD, EMD Serono<br>PPD, EMD Serono<br>PPD |

# Approval Page

**Integrated Analysis Plan:** MS200647_0047

A Phase II, Multicenter, Open-label Study to Investigate the Clinical Efficacy of M7824 Monotherapy in Participants With Locally Advanced or Metastatic Biliary Tract Cancer Who Fail or are Intolerant to First-line Platinum-Based Chemotherapy

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

# 1 Table of Contents

# 2 List of Abbreviations and Definition of Terms

| Abbreviation | Definition |
|---|---|
| 2L | Second-line |
| ADA | Antidrug antibody |
| ADaM | Analysis Data Model |
| AE | Adverse Event |
| AESI | Adverse Events of Special Interest |
| ALT | Alanine aminotransferase |
| ALP | Alkaline Phosphatase |
| aPTT | Activated Partial Thromboplastin Time |
| AST | Aspartate Aminotransferase |
| BMI | Body Mass Index |
| BOR | Best overall response |
| BTC | Biliary Tract Cancer |
| cBOR | Confirmed best overall response |
| $C_{EOI}$ | The concentration observed immediately at the end of infusion |
| CI | Confidence Interval |
| COVID-19 | Coronavirus disease 2019 |
| CR | Complete Response |
| CSR | Clinical Study Report |
| $C_{trough}$ | The concentration observed immediately before next dosing (corresponding to pre-dose or trough concentration for multiple dosing) |
| CV% | Coefficient of Variation (%) |
| DBP | Diastolic Blood Pressure |
| DCR | Disease control rate |
| DI | Dose Intensity |
| DNA | Deoxyribonucleic Acid |
| DOR | Duration of Response |
| ECG | Electrocardiogram |
| ECOG PS | Eastern Cooperative Oncology Group performance status |
| eCRF | Electronic Case Report Form |

| | |
|---|---|
| eDISH | Evaluation of Drug-Induced Serious Hepatotoxicity |
| CCI [redacted] | [redacted] |
| CCI [redacted] | [redacted] |
| CCI [redacted] | [redacted] |
| CCI [redacted] | [redacted] |
| GeoCV% | Geometric Coefficient of Variation (%) |
| GeoMean | Geometric mean |
| HBV | Hepatitis B Virus |
| HCV | Hepatitis C virus |
| HGB | Hemoglobin |
| HLT | High level term |
| CCI [redacted] | [redacted] |
| IAP | Integrated Analysis Plan |
| IC | Immune Cells |
| ICH | International Council for Harmonization |
| IDMC | Independent Data Monitoring Committee |
| irAE | Immune-Related Adverse Event |
| IRC | Independent Review Committee |
| CCI [redacted] | |
| IRR | Infusion-Related Reaction |
| ITT | Intent-to-Treat |
| LLN | Lower Limit of Normal |
| LLOQ | lower limit of quantification |
| logStD | StD of log-transformed data |
| Max | Maximum |
| MCH | Mean corpuscular hemoglobin |
| MCHC | Mean corpuscular hemoglobin concentration |
| MCV | Mean Corpuscular Volume |

| | |
|---|---|
| Mean | Arithmetic mean |
| Med | Median |
| MedDRA | Medical Dictionary for Regulatory Activities |
| Min | Minimum |
| CCI | [redacted] |
| MSS | Microsatellite stable |
| CCI | [redacted] |
| NC | Not calculated |
| NCI-CTCAE | National Cancer Institute – Common Terminology Criteria for Adverse Events |
| nd | Not determined |
| ND | No Disease |
| OR | Objective Response |
| ORR | Objective Response Rate |
| OS | Overall Survival |
| PA | Primary Analysis |
| PD | Progressive Disease |
| CCI | [redacted] |
| PFS | Progression-Free Survival |
| CCI | [redacted] |
| PK | Pharmacokinetic(s) |
| PR | Partial Response |
| CCI | [redacted] |
| PT | Prothrombin Time/Preferred Term |
| Q1 | First Quartile |
| Q3 | Third Quartile |
| RBC | Red Blood Cell |
| RDI | Relative Dose Intensity |
| RECIST 1.1 | Response Evaluation Criteria in Solid Tumors Version 1.1 |
| RNA | Ribonucleic Acid |
| SAE | Serious Adverse Event |
| SAF | Safety Analysis Set |
| SAP | Statistical Analysis Plan |

| | |
|---|---|
| SBP | Systolic Blood Pressure |
| SCR | Screening (analysis population) |
| SD | Stable Disease |
| SDTM | Study Data Tabulation Model |
| SI | International System of Units |
| SOC | System Organ Class |
| StD | Standard Deviation |
| T4 | Thyroxine |
| TBILI | Total Bilirubin |
| TC | Tumor Cells |
| TEAE | Treatment-Emergent Adverse Event |
| TGF-β | Transforming Growth Factor-Beta |
| CCI | [redacted] |
| TMTB | Total measured tumor burden |
| TNM | Tumor, Lymph Nodes, Metastasis |
| TSH | Thyroid-stimulating hormone |
| ULN | Upper Limit of Normal |
| VAS | Visual analog scale |
| WBC | White Blood Cells |
| WHO | World Health Organization |

# 3 Modification History

| Unique Identifier for IAP Version | Date of IAP Version | Author | Changes from the Previous Version |
|---|---|---|---|
| 1 | 05JUL2019 | PPD | Not Applicable |
| 2 | 26AUG2020 | PPD | Updated to be in line with updated protocols, including addition of CCI per biliary tract cancer (BTC) subtype, Section 6.2.<br>Replaced 'M7824' with 'bintrafusp alfa' in IAP specific text.<br>Section 7.1 added (COVID-19 impact), with analysis details added in sections 10.1, 10.2 and 15.1.1.<br>Section 8.2: Added genetic profiling at baseline and nAb status subgroups.<br>Section 9: The on-treatment period definition was updated to include the treatment reinitiation period. Added categorization of participants for COVID-19 impact assessment.<br>Section 12.4: Added "Number of previous lines of therapy for metastatic/locally advanced disease" to the previous systemic anticancer drug therapy summary.<br>Section 13: Updated to include the treatment re-initiation period if a participant re-initiates treatment.<br>Section 14.1: Updated definition of PD for unconfirmed best overall response from "PD = at least one PD assessment (and does not meet criteria of CR, PR, SD or Non-CR/Non-PD)" to "PD = progression ⩽16 weeks after first date of study intervention (and does not meet criteria of CR, PR, SD or Non-CR/Non-PD)";<br>Sections 14.1.1, 14.4.2, 14.8.1: Added listings by ADA and nAb status.<br>Section 14.3: Added the following to the definition of PD "PD is also to be confirmed if the participant:<br>- dies within 12 weeks after the initial observation of PD, or<br>- discontinues treatment due to disease progression prior to or within 12 weeks after the assessment of PD."<br>Section 15.2.4.3: Added the narrow and broad definitions for potential TGF-β-mediated skin adverse events.<br>Section 15.5: Added listing description for non-protocol Related Hospital Visits.<br>Section 161.2: Added PK summaries by ADA and nAb subgroups.<br>Section 16.3: Referenced the CCI SAP.<br>Section 16.4.1: Added definitions for start of immunogenicity response and duration of immunogenicity response and analyses by race.<br>Section 16.4.2 added (neutralizing anti-drug antibody).<br>Section 16.5.3: Added pooled derivation details since data from multiple sources.<br>CCI ).<br>Added Protocol Deviation and Definition of NCI-CTCAE Grading appendices. |

# 4 Purpose of the Integrated Analysis Plan

The purpose of this integrated analysis plan (IAP) is to document technical and detailed specifications for primary, second and BTC subtype analyses of data collected for protocol MS200647_0047. Independent Data Monitoring Committee (IDMC) analysis will be developed in a separate statistical analysis plan (SAP).

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

The IAP is based upon Section 9 (Statistical Considerations) of the study protocol and is prepared in compliance with International Council for Harmonization (ICH) Guideline E9. It describes analyses planned in the protocol.

# 5 Objectives and Endpoints

## Table 1 Study Objectives and Endpoints

| Objectives | Endpoints (Outcome Measures) | IAP section |
|---|---|---|
| Primary | | |
| To evaluate clinical efficacy of M7824 based on ORR | • Confirmed objective response (OR) according to Response Evaluation Criteria in Solid Tumors Version 1.1 (RECIST 1.1) assessed by an Independent Review Committee (IRC) | 14.1 |
| Secondary | | |
| To evaluate clinical efficacy of M7824 based on duration of response (DOR) | • DOR assessed from complete response (CR) or partial response (PR) until progression of disease (PD), death, or last tumor assessment assessed by an IRC | 14.5 |
| To evaluate clinical efficacy of M7824 based on durable response rate (DRR) | • Durable response of at least 6 months according to RECIST 1.1 assessed by IRC | 14.6 |
| To evaluate clinical safety of M7824 | • Occurrence of treatment-emergent adverse events (TEAEs) and treatment-related AEs, including adverse events of special interest (AESIs) | 15.1, 15.2 |
| To evaluate clinical efficacy based on progression-free survival (PFS) | • PFS according to RECIST 1.1 assessed by IRC | 14.4 |
| To evaluate ORR, DOR, DRR, and PFS by Investigator read | • OR, DOR, DRR, and PFS according to RECIST 1.1 assessed by Investigator read | 14.2, 14.4, 14.5, 14.6 |
| To evaluate clinical efficacy based on overall survival | • OS | 14.8 |
| To characterize the pharmacokinetic (PK) profile of M7824 | • The concentration observed immediately at the end of infusion ($C_{EOI}$) of M7824<br>• The concentration observed immediately before next dosing (corresponding to pre-dose or trough concentration [$C_{trough}$] for multiple dosing) of M7824 | 16.1 |
| To characterize the immunogenicity of M7824 | • Immunogenicity of M7824 as measured by antidrug antibody (ADA) assay, from Screening through 12 weeks (± 2 weeks) after last treatment | 16.4 |
| To evaluate clinical efficacy of M7824 based on ORR, DOR and DRR according to programmed death ligand 1 (PD-L1) expression and microsatellite instability (MSI) status retrospectively | • Confirmed OR according to RECIST 1.1 assessed by an IRC according to PD-L1 expression and MSI status<br>• DOR and durable response of at least 6 months according to RECIST 1.1 assessed by an IRC according to PD-L1 expression and MSI status | 14.1.1, 14.5.1, 14.6.1 |
| CCI [redacted] | | |

| Objectives | Endpoints (Outcome Measures) | IAP section |
|---|---|---|
| CCI | | |

* Items noted above with an * symbol are not collected in sites in China.

# 6 Overview of Planned Analyses

This IAP addresses the primary and second analyses for the global study as well as BTC subtype efficacy analysis. Additional statistical analysis plans for Independent Data Monitoring Committee (IDMC) will be developed separately.

All statistical analyses will be performed using cleaned eCRF data as well as external data including tumor assessment measured by the Independent Review Committee (IRC). All data used in the analysis will be included up to a data cut-off point.

## 6.1 Primary and Second Analysis

Table 2 displays an overview of the primary and secondary analyses to be provided for this study.

**Table 2 Overview of Analyses**

| Analysis | Data cut-off point | Endpoints |
|---|---|---|
| **Primary analysis (PA) at 9 Months** | 9 months after the first dose of the last of 141 planned participants | Full evaluation of all efficacy and safety endpoints. Plus, assessment of impact of COVID-19. |
| **Second analysis at 15 Months** | 15 months after the first dose of the last of 141 planned participants | Full evaluation of all efficacy and safety endpoints |

The timing of conduct of the primary and secondary analyses will be triggered by the data cut-off points displayed in Table 2 above and will evaluate all criteria described in this IAP.


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

The statistical methods as described in the protocol were adopted. There are no changes to the planned analyses.

## 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 (summary tables, listings and figures) will be generated to assess potential impacts of COVID-19 to this study including:

- number of participants in pre/during COVID-19 study period,
- number of COVID-19 related protocol deviations,
- number of participants in pre/during COVID-19 study period with AEs,
- listing of participants with any missed treatment administrations, tumor assessments or missed visits due to COVID-19, and
- listing of AEs related to COVID-19.

Details of the categorization of participants for COVID-19 impact assessment is provided in Section 9 and details of the analyses is provided in Sections 10.1, 10.2, and 15.1.1.

# 8 Protocol Deviations and Analysis Populations

## 8.1 Definition of Protocol Deviations

Important protocol deviations 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.

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

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

Important protocol deviations will be identified and confirmed prior to or at Data Review Meeting and will include:

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

All important protocol deviations will be documented in Study Data Tabulation Model (SDTM) datasets whether identified through site monitoring, medical review or data-management programming.

**Further considerations for PK analysis:**

Examples of protocol deviations or important events for PK analysis and PK result interpretation may include, but may not be limited to, the following:

- Dose delayed outside the allowed window
- Actual dosing time not recorded
- Dose change or missed dose
- Pre-dose sample collected after the actual start of infusion
- End-of-infusion sample collected before the actual end of infusion
- Sample processing errors that may lead to inaccurate bioanalytical results

For the above protocol deviations or important events for PK, the relevant PK data will be excluded from summaries based on the PK analysis set.

Refer to Section 16.1 for more details of protocol deviations and handling relevant to PK.

## 8.2 Definition of Analysis Populations and Subgroups

**Screening Analysis Set (SCR)**

All participants who signed the informed consent.

**Intention-to-Treat Set (ITT)/Safety Analysis Set (SAF)**

The Intention-To-Treat analysis set as well as the Safety Analysis Set include all participants who received at least 1 dose of bintrafusp alfa.

The ITT analysis set will be used for all analyses of demographics and baseline characteristics, efficacy, CCI and CCI. The SAF analysis set will be used for all exposure and safety analyses.

**Pharmacokinetic Analysis Set (PK)**

All participants who completed at least 1 dose of bintrafusp alfa, and who provided at least 1 sample with a measurable concentration of bintrafusp alfa, without important protocol deviations or events deemed to affect PK evaluation.

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

**Pharmacokinetic ADA Analysis Set (PKADA)**

A subpopulation of the PK Analysis Set restricted to subjects who have in addition at least one valid result of ADA at any time point.

**Pharmacokinetic nAb Analysis Set (PKNAB)**

A subpopulation of the PK Analysis Set restricted to subjects who have in addition at least one valid result of neutralizing anti-drug antibodies (nAb) at any time point.

**Immunogenicity analysis set (IMM)**

All participants who received at least 1 dose of bintrafusp alfa and have at least one valid ADA result.

All immunogenicity analyses will be based on this analysis set.

Table 3 displays the use of the analysis sets in the different analyses:

**Table 3 Overview of the Analysis Set Used in the Analyses**

| Analyses | SCR | SAF | ITT | PK | PKADA | PKNAB | IMM |
|---|---|---|---|---|---|---|---|
| Disposition | ✓ | | | | | | |
| Demographics | | | ✓ | | | | |
| Baseline Assessments | | | ✓ | | | | |
| Previous and Concomitant Therapies | | | ✓ | | | | |
| Compliance and Exposure | | ✓ | | | | | |
| Efficacy | | | ✓ | | | | |
| Safety and Tolerability | | ✓ | | | | | |
| Pharmacokinetics | | | | ✓ | ✓ | ✓ | |
| CCI | | | | | | | |
| Immunogenicity | | | | | | | ✓ |
| CCI | | | | | | | |

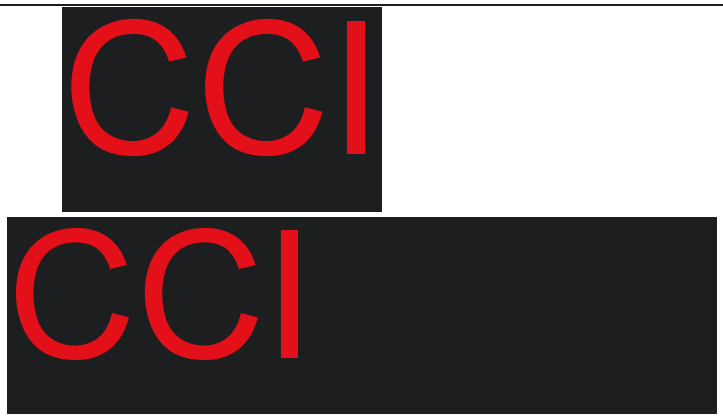
**Additional Subgroup Analysis Sets**

Analysis of primary, key secondary efficacy and safety endpoints may be performed on subgroups of interest as specified in this IAP Section 14 and Section 15.1 respectively. Since the study is not powered for any subgroup analysis, all the subgroup analyses are CCI in nature.

The following major subgroups will be used for analysis:

- BTC subtype
  - Intrahepatic cholangiocarcinoma
  - Extrahepatic cholangiocarcinoma
  - Gallbladder cancer
- Age group
  - Age < 65 years
  - Age ≥ 65 years
- ECOG PS at baseline
  - ECOG PS 0
  - ECOG PS 1
- Sex
  - Male
  - Female

- Race
  - White
  - Black or African American
  - Asian
  - American Indian or Alaska Native
  - Native Hawaiian or other Pacific Islander
  - More than one race
  - Not collected at the site
  - Other
- Geographic Region
  - Asia
  - US
  - Europe
- Time since completion of 1L platinum-based treatment
  - <6 months
  - ≥6 months
- Disease status at initial diagnosis (based on TNM classification)
  - Locally advanced (M0)
  - Metastatic disease (M1)
  - Unknown (MX)
- PD-L1 expression on tumor cells (TC) and on immune cells (IC) at baseline
  - <1%, ≥1%; <5%, ≥5%; <25%, ≥25%; <50%, ≥50% for TC
  - <1%, ≥1%; <5%, ≥5%; <25%, ≥25%; <50%, ≥50% for IC
  - <1%, 1-<50%, ≥50% for TC
  - <1%, 1-<50%, ≥50% for IC




- ADA status
  - Ever positive
  - Never positive
- nAb status
  - Ever positive (for either assay TGF-β or PD-L1)
  - Never positive

For Extrahepatic cholangiocarcinoma participants, following subgroups may also be used for analysis:

- Extrahepatic cholangiocarcinoma subtype
  - Perihilar cholangiocarcinoma
  - Distal cholangiocarcinoma

For ADA and nAb status subgroups, the subgroup analyses will be performed if there are at least 3 ever positive participants.

For all other subgroup variables with more than two categories, any subgroup category with less than 7 participants (approximately 5% of the population) will not be reported in the analysis.

# 9 General Specifications for Data Analyses

Refer to Section 16.1 for PK data handling/analysis details.

**Data handling after cut-off date:**

Data after cut-off will not undergo the cleaning process, and will be cut at SDTM level, i.e. before the process of ADAM creation. Data after 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.

Stop dates will not be affected by this rule, e.g. a stop date of an AEs, with start date prior to the cut-off, but stop date after the date of cut-off, will be used for analysis as reported.

**Pooling of centers:**

Because of the high number of participating centers and the anticipated small number of participants treated in each center, data will be pooled across centers, and the factor center will not be considered in subgroup analysis.

**Significance level:**

The overall significance level is 2.5% one-sided. Statistical tests will be described in Section 14. If confidence intervals are to be calculated, these will be two-sided with a confidence probability of 95%, unless otherwise specified in this IAP.

**Presentation of continuous and qualitative variables:**

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

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

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

Pharmacokinetic variables (concentrations and parameters) will be summarized as described in Section 16.1.2.

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 study at that visit, unless otherwise specified.

**Definition of baseline:**

The last available/non-missing measurement prior to the first study intervention will serve as the baseline measurement for safety and efficacy analyses.

If an assessment is planned to be performed prior to the first study intervention in the protocol and the assessment is performed on the same day as the first study intervention, it will be assumed that it was performed prior to study intervention, if assessment time is not collected or is missing. If assessment time is collected, the observed time as well as time of first dose 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 intervention.

**Definition of change from baseline**

Change from baseline = timepoint value – baseline value

Percent Change from Baseline = 100 * (timepoint value – baseline value) / baseline value

**Definition of on-treatment period:**

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

For immune-related AEs as listed in Section 15.2.4.2, an expanded on-treatment period will be used as a default for any analysis:

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

Any systemic anticancer therapy, any anticancer surgery and any anticancer radiotherapy as documented in the ”Anti-cancer treatment after discontinuation” CRF page will be considered as subsequent anticancer therapy.

The on-treatment period will include the initial treatment period as well as the reinitiation of treatment period, as applicable. Whether a participant reinitiates treatment (following the rules as outlined in the protocol) or not, the on-treatment period is defined as the time from the first study intervention administration to the last study intervention administration date + 30 days or the earliest date of subsequent anticancer drug (anticancer therapy, anticancer surgery and anticancer radiotherapy) therapy minus 1 day, whichever occurs first, unless otherwise stated.

**Unscheduled assessments:**

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

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

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

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

**Definition of duration:**

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

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

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

**Common calculations:**

For quantitative measurements, change from baseline will be calculated as:

- Test Value at timepoint X – Baseline Value

**Conversion factors:**

The following conversion factors will be used to convert days into months or years:

- 1 month = 30.4375 days
- 1 year = 365.25 days

**Handling of missing data:**

Unless otherwise specified in this IAP, missing data will not be replaced.

Missing statistics, e.g. when they cannot be calculated, should be presented as “nd” for “not determined. For example, if n=1, the measure of variability [e.g. standard deviation (StD)] cannot be computed and should be presented as “nd”.

**Handling of incomplete dates:**

- **Missing data handling rules for age calculation**

  Incomplete dates (date of informed consent, date of birth) for the calculation of age will be imputed as follows:

  - In case of missing day for at least one date, but month and year available for both dates: the day of informed consent and the day of birth will be set to 1.

- In case of missing month for at least one date, but year available for both dates, the day and the month of informed consent and the day and month of birth will be set to 1.
- In all other cases, the incomplete dates will not be imputed.

- **Missing data handling rules for disease history**

Incomplete dates for disease history (date of initial diagnosis date, date of documented, locally advanced, inoperable or metastatic disease diagnosis, date of response or progression in prior treatment) will be imputed as follows:

- If the day is missing, it will be imputed to the 1st day of the month.
- If both day and month are missing, the month and day will be imputed as January 1st
- If the date is completely missing, no imputation will be performed.

- **Missing data handling rules for adverse events**

Incomplete AE-related dates will be imputed as follows:

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

- **Missing data handling rules for previous and concomitant medications**

Incomplete dates for previous and concomitant 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 medication:

- 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 participant's death, then the date of death will be used to impute the incomplete stop date.

- **Missing data handling rules for subsequent anticancer therapy**

Incomplete dates for start date of subsequent anticancer therapy (drug therapy, radiotherapy, surgery) will be imputed as follows and will be used for determining censoring dates for efficacy analyses and in the derivation of the end of on-treatment period:

  - If only day is missing, it will be imputed as the last day of the month unless the end date of subsequent anticancer therapy is before that date. In that case, the incomplete anticancer therapy start date will be imputed as the end date of the anticancer therapy.
  - Otherwise it will not be imputed.

Incomplete subsequent anticancer therapy stop dates will not be imputed.

- **Missing data handling rules for death date**

For the purpose of survival analyses, partially missing death dates will be imputed as follows:

  - If only the day is missing, the death date will be imputed to the maximum of the (non-imputed) day after the date of last known alive date and the 15$^{th}$ day of the month
  - Otherwise it will not be imputed

- **Handling rules for tumor assessments**
  - 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

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

**Treatment day**

Day 1 is the day of start of study intervention, the day before is Day -1 (no Day 0 is defined). Study day/Study intervention day is defined relative to Day 1.

**Preferred term for analysis of WHO-DD coded data**

For data coded according to WHO Drug B3 (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.

**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).

**Data collected after re-initiated treatment:**

Data collected after re-initiation of treatment will be included in the summary statistics. A data listing will include AE data for all reinitiated participants.

**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 26th 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 3rd January 2020, 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 according to the methods shown in Table 4 depending on the purpose of analysis.

**Table 4 Types of categorization for COVID-19 assessment**

| Type of categorization | Subgroup | Definition |
|---|---|---|
| Categorization for assessment on efficacy | Pre | Participant started treatment prior to and had progressive disease or died or withdrew from tumor assessments prior to COVID-19 study period |
| | During | Participant started treatment prior to or during COVID-19 study period and had any tumor assessments during the COVID-19 study period |
| | Post* | Participant started treatment post the COVID-19 study period |
| Categorization for assessment on safety without time window | 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 |
| Categorization for assessment on safety with time window | Pre | Participant started treatment 3 months prior to COVID-19 study period |
| | During | Participant started treatment on or after the date 3 months prior to COVID-19 study period |
| | Post* | Participant started treatment post the COVID-19 study period |

* No participants will be categorized into the 'post COVID-19 study period' or the 'during COVID-19 study period for safety without time window' for this study.

**Software:**

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

The computer program Phoenix® WinNonlin® Version 8.0, or higher (Certara, L.P., Princeton, New Jersey, USA) could be used for PK data.

# 10 Study Participants

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

## 10.1 Disposition of Participants and Discontinuations

Descriptive statistics will be used to summarize participant disposition based on the electronic case report form (eCRF) data.

The following information will be reported:

- Total number of participants screened (i.e. participants who gave informed consent)
- Number of participants who discontinued from the study prior to study intervention overall and grouped by the main reason (e.g. the failed specific inclusion or exclusion criteria, adverse event, lost to follow-up, death, progressive disease, withdrawal of consent and other)

- Number of re-screened participants
- Number of participants who received at least one dose of study intervention (ITT/SAF analysis set)
- Number of participants with treatment ongoing at the data cutoff date
- Number of participants off study treatment, grouped by main reason (treatment completed as per protocol, progressive disease, death, adverse event, lost to follow-up, protocol non-compliance, withdrew consent, other)
- Number of participants who re-initiated the study treatment and number of participants who discontinued the study intervention after re-initiation
- Number of participants with treatment ongoing
- Number of participants who completed/discontinued the study participation, with the associated main primary reason (study completed according to protocol, adverse event, lost to follow-up, protocol non-compliance, death, withdrew consent, other)

Percentages will be calculated based on the number of participants in the ITT/SAF population, except for the number of screened and re-screened participants where no percentage will be provided.

In addition, the number of participants in each analysis set defined in Section 8.2 will be summarized, overall and by region (Europe, North America, Asia), country within region and center.

The listing of participant disposition will include all participants (i.e. including screening failures, but not re-screened participants [at their screen failure time] which will be listed in a specific listing). The listing will include the following information: participant identifier, date of informed consent, included in the study (if not reason for exclusion), first/last study intervention date, date and reason off-treatment, date and reason off-study, population flags. When the reason such as reason off-treatment will be categorized as "Other, specify" or "Withdrew consent from treatment, specify", the verbatim text as entered in the eCRF will be presented in the listing.

In addition, a listing of participants for which bintrafusp alfa has been reinitiated will be provided with the following information: participant identifier, date of first study intervention, date of last study intervention, reason for treatment termination, first and last re-initiation bintrafusp alfa administration date and status at end of treatment re-initiation including reason for treatment discontinuation, as applicable.

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

In addition, for the assessment of COVID-19 impact on this study, number of participants in subgroup of pre/during/post COVID-19 study period according to each categorization method described in Table 4 (Section 9) will be provided.

## 10.2 Protocol Deviations

*Analysis Set: ITT*

The following summary tables and listings of important protocol deviations will be provided:

- Frequency table per reason of important protocol deviations
- Listing of important protocol deviations which will include participant identifier, category of the deviation (e.g. inclusion/exclusion), and a description of the deviation.

Potential impact of COVID-19 pandemic in MS200647-0047 will be evaluated by an analysis of protocol deviations. The number of participants with important protocol deviations overall as well as all protocol deviations due to COVID-19 will be tabulated. A subject listing of participants with any missed treatment administrations, missed tumor assessments, and missed visits due to COVID-19 will also be provided.

# 11 Demographics and Other Baseline Characteristics

*Analysis Set: ITT*

## 11.1 Demographics

Demographic characteristics will be summarized using the following information:

- Sex: male, female
- Ethnicity: Hispanic or Latino, not Hispanic or Latino; Japanese, not Japanese
- Race:
  - For participants reporting one race only: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Not collected at the site, Other.
  - For participants reporting several races, all combinations will be reported under ‘More than one race” category.
- Age (years): summary statistics
- Age categories:
  - < 65 years, ≥ 65 years
  - 65-74 years, 75-84 years, ≥ 85 years

- Geographic Region: North America, Europe, Asia

Specifications for computation:

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

    The integer part of the calculated age will be used for reporting purposes.
  - Investigator site codes will be used for the determination of the participant's geographic region.

Demographic characteristics including participant identifier, sex, race (including all reported races in case of "multiple" races, and details in case of "other" race), ethnicity, geographic region, age (years) will be presented in a listing.

## 11.2 Medical History

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

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

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

## 11.3 Other Baseline Characteristics

### 11.3.1 Disease History

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

- Biliary Tract Cancer (BTC) subtype classification: Intrahepatic cholangiocarcinoma, Extrahepatic cholangiocarcinoma (perihilar vs distal cholangiocarcinoma) and Gallbladder cancer
- Tumor histology: Adenocarcinoma well differentiated, Adenocarcinoma moderate differentiated, Adenocarcinoma poorly differentiated, Adenocarcinoma unknown, Other
- Time since initial cancer diagnosis (months)
- Time since documented, locally advanced or metastatic disease (months)
- Time since last progression of disease prior to study entry (months)
- TNM classification at initial diagnosis and at study entry: each T, N, M category will be described (TX, T0, N1, etc.)


### 11.3.2 Vital signs at Baseline

The following vital signs at baseline will be collected from the "Vital signs" eCRF page and will be summarized:

- Height (cm)
- Weight at baseline (kg)
- Body Mass Index (BMI) ($kg/m^2$)

Specifications for computation:

- BMI ($kg/m^2$) = weight(kg)/[height(m)]$^2$

Heigth, weight and BMI will be listed in the demographics listing (see Section 11.1).

### 11.3.3 ECOG Performance Status at Baseline

The ECOG Performance Status will be described from the data collected on the "ECOG Performance Status" eCRF page. It will be described at baseline by the frequency and percentage of participants in each category:

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 housework, 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.

### 11.3.4 Skin status history

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

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

A listing of skin status history will be provided.

### 11.3.5 Tumor Biopsy

Tumor Biopsy is collected on the "Tumor Biopsy" eCRF page, a listing of tumor biopsy will be provided including: participant identifier, age, sex, race, visit, collection date, sample type (fresh biopsy or archival tissue sample), sample identifier, sample type for tumor block and for tumor tissue the number of slides collected.

### 11.3.6 PD-L1 test

PD-L1 expression on tumor cells or on immune cells as collected at screening will be descripted using the following categories:

- <1%, ≥1%; <5%, ≥5%; <25%, ≥25%; <50%, ≥50% for TC
- <1%, 1-50, ≥50% for TC
- <1%, ≥1%; <5%, ≥5%; <25%, ≥25%; <50%, ≥50% for IC
- <1%, 1-50, ≥50% for IC

Listing will also be provided and include: participant identifier, age, sex, race, unique sample identifier, visit, collection date and PD-L1 expression.

## 12 Previous or Concomitant Medications/Procedures

*Analysis set: ITT*

## 12.1 Previous and concomitant medications

**Previous medications** are medications, other than study medications, which started before first study intervention. 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, which are taken by participants any time on-treatment (on or after the first day of study intervention for each participant) or within 30 days after last dose of study drug OR the earliest date of subsequent anticancer 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. All medications starting the same day as study drug will be considered as concomitant. Medications starting the 30$^{th}$ days after the last dose of study treatment will also be considered as concomitant.

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

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

Summaries of previous and concomitant medications will present the number and percentage of participants by drug class and preferred term according to the WHO-Drug B3 dictionary. Drug class will be derived as ATC classification Level 2, and the preferred term will be taken as the preferred drug name. 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.

Previous and concomitant medications will be presented in listing and include: participant identifier, age, sex, race, preferred term, medication name as provided by the Investigator, start date, end date, dose, dose units, frequency, route, reason for the medication.

## 12.2 Premedications for bintrafusp alfa

Premedications are medications administered per protocol on the same day as, but prior to, the study intervention to mitigate potential infusion-related reactions.

As per protocol, the first 15 participants will not receive premedications (except if requested by Investigator) before first and second bintrafusp alfa infusions to establish whether mandatory use of premedication is required. If two or more Grade 2 infusion reactions are seen during the first 2 infusions for the first 15 participants (without medication), IDMC will determine if mandatory premedication is needed. Study enrollment will continue in parallel to this review and sites will be notified accordingly if premedication becomes required. If IDMC determines mandatory premedication is justified, premedication prior to each dose of bintrafusp alfa will be mandatory

for the first 2 infusions. After 2 infusions, premedication will be optional and at the discretion of the Investigator.

The number of participants receiving premedication will be summarized for each treatment visit based on "Premedication details" eCRF page. If the IDMC determines that mandatory premedication is required, then the table will be split into two columns, participants enrolled before vs after the review of the first 15 participants.

Percentages will be calculated on the number of participants who actually received an infusion at the associated visit.

Listing will also be provided including: participant identifier, age, sex, race, medication name, visit, date/time of study intervention, dose, dose units, and route.

## 12.3 Concurrent procedures

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

Concurrent procedures will be presented in a listing which will include participant identifier, age, sex, race, name of procedure (as provided by the Investigator), start date, end date, indication, reason for procedure and type of specimen collected. A flag will be displayed to identify each procedure as prior to treatment and on-treatment.

## 12.4 Previous anticancer treatments and procedures

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

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

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

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

- Number of any previous anticancer therapy regimens: 0 / 1 / 2 / 3 / ≥ 4
- Number of previous lines of therapy for metastatic/locally advanced disease: 0 / 1 / 2 / 3 / ≥ 4 (categories will be data dependent)
- Intent of therapy: Neoadjuvant / Adjuvant / Metastatic or Locally advanced

- Best response of last treatment regimen: Complete Response (CR) / Partial Response (PR) / Stable Disease (SD) / Progressive Disease (PD) / Non-Complete Response/Non-Progressive Disease (non-CR/non-PD) / Not Assessable / Unknown / Not applicable.
- The time since completion of first-line (1L) of platinum-based treatment will also be summarized. Platinum-based treatment will be identified for participants who received at least one platinum-based drug (cisplatin, oxaliplatin or carboplatin) and at least one of the following drugs in the same regimen: gemcitabine, fluorouracil or capecitabine. The time since completion will be measured using the maximum stop date of any of platinum-based drug/gemcitabine/fluorouracil/capecitabine treatment (maximum start date will be used in case the stop date is not available).
- Number of participants per 1L regimen, e.g., cisplatin/gemcitabine, cisplatin/fluorouracil.

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

- The previous anticancer drug listing will contain participant identifier, age, sex, race, regimen ID, preferred term, medication name, start date, end date, intent of therapy, best response, date of progression.
- The previous anticancer radiotherapy listing will contain participant identifier, age, sex, race, start date, end date, was prior radiotherapy to bone for palliation only, location of radiotherapy, total dose, number of fractions.
- The previous anticancer surgery listing will contain participant identifier, age, sex, race, date of surgery, name and location of surgery, curative intent of surgery (Y/N), and outcome of surgery.

## 12.5 Anticancer Treatment after Discontinuation

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

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

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

The type of subsequent anticancer drug therapy as provided in the e-CRF (i.e. Cytotoxic therapy/ Monoclonal antibodies therapy/ Small molecules/ Immunotherapy/ Other) will be described.

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

- For medication: participant identifier, age, sex, race, preferred term/medication name, medication type, regimen name, intent of therapy, start date and end date.
- For radiotherapy and surgery: participant identifier, age, sex, race, type of therapy, start date, end date, radiotherapy site or name of surgery/location, if radiotherapy total dose and number of fractions, if surgery outcome and was the surgery curative in intent (Y/N)

# 13 Study intervention Compliance and Exposure

*Analysis set: SAF*

Participants will be treated with bintrafusp alfa at a dose of 1200 mg once every 2 weeks, until confirmed progression of disease (PD), death, unacceptable toxicity, study withdrawal, or up to 24 months. This dose will be administered as "flat" doses independent of the body weight.

All dosing calculations below and summaries will be based on "Study Treatment Administration Details" eCRFs page. Data collected during the treatment re-initiation phase will be included in the summary statistics and flagged in the listings.

For the analysis of exposure, a dose is regarded to be administered if the actual dose received is > 0 mg.

Whether the participant re-initiates the treatment or not, the **duration** of treatment of bintrafusp alfa (in weeks) during the study is defined as:

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

The **cumulative dose** of bintrafusp alfa per participant in a time period is the sum of the actual dose that the participants received within that period (i.e., total dose administered (mg)).

The **dose intensity** of treatment (DI) (mg/cycle) of bintrafusp alfa per 2-week/cycle is defined as

$$\text{DI } of \text{ treatment} = \left(\frac{\text{Cumulative dose of treatment (mg)}}{\text{Duration of } bintrafusp \text{ alfa of treatment(in weeks)/2}}\right)$$

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

$$\text{RDI of treatment (\%)} = 100 \times \left(\frac{\text{DI of treatment (mg/cycle)}}{\text{planned dose level (mg/cycle)}}\right)$$

The following summary tables will be provided:

- Duration of therapy (weeks)

- Total number of infusions received
- Cumulative dose (mg)
- Dose intensity (mg/cycle)
- Relative dose intensity (%) as continue variable, and categorized as
  - < 80%
  - 80%-90%
  - >90%

Two listings will be presented:

- A listing of study intervention which will provide: participant identifier, age, sex, race, visit, infusion start date and time, infusion end date and time, infusion rate (mL/hr), actual dose (mg), route, administration modification and reason for modification, change in administration detail, treatment delay (days). Data collected during the treatment re-initiation phase will be flagged.
- An additional listing of treatment exposure and compliance which will include participant identifier, age, sex, race, duration of therapy (weeks), total number of infusions received, cumulative dose of therapy (mg), dose intensity (mg/cycle), and relative dose intensity (%). For data collected during the treatment re-initiation phase, only the duration of therapy as well as the number of infusions during the re-initiation phase will be summarized.

**Dose Modification**

Dose modification is not allowed per protocol. No summaries will be provided.

**Therapy Delays**

Delays of therapy will be derived for each infusion as the number of days since last infusion – 14:

Therapy Delays = start date of current infusion – start date of the previous infusion – 14

If the result is >0 day, then this will be classed as a delay. A participant may have more than one treatment delay throughout the course of treatment.

The following will be summarized in a table:

- Number of participants with at least one delay
- Number of participants with no delay
- Number of delays per participant (0 delay, 1 delay, 2 delays, 3 delays, ≥ 4 delays)
- Longest delay per participant (no delay, 1-2 days, 3-8 days, 9-15 days, ≥16 days)

### Infusion Temporary Interruptions

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

### Infusion Rate Reductions

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

# 14 Efficacy Analyses

*Analysis Sets: ITT*

## 14.1 Best Overall Response according to RECIST 1.1 as adjudicated by IRC (Primary endpoint)

| Analysis (Analysis Population) | Derivation | Statistical Analysis Methods | Missing data handling |
|---|---|---|---|
| **Primary endpoint: confirmed Objective Response (OR)** | | | |
| Primary analysis (ITT) | Confirmed Objective Response Rate (ORR) according to RECIST 1.1 as adjudicated by the IRC is defined as the number of participants having a confirmed objective response assessment of Complete Response (CR) or Partial Response (PR) (at least two determinations of CR/PR at least 4 weeks apart and before progression), out of the total number of participants. | ORR will be provided with a two-sided 95% Confidence Interval (CI) using the Clopper-Pearson method (exact CI for a binomial proportion as computed by default by the FREQ procedure using the EXACT option).<br><br>Exact binomial test (1 sample) against the null hypothesis for the ORR will be calculated to determine whether the null hypothesis of an ORR ≤ 10% can be rejected. | Participants with missing post-baseline data are considered as having no objective response |

### Best overall response (BOR)

Best overall response will be assessed based on the tumor response at different evaluation timepoints from the first study intervention until the first documented disease progression for RECIST criterion. 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.

BOR can be defined as confirmed or unconfirmed, both definitions are provided below.

**Confirmed best overall response (cBOR)**

cBOR according to RECIST 1.1 as adjudicated by IRC will be assessed according to the following rules:

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

  Note: It is reasonable to consider CR-NE-CR or CR-PR-CR as CR as long as the second CR is more than 28 days away from the first timepoint.

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

  Note: It is reasonable to consider PR-NE-PR or PR-SD-PR as PR as long as the second PR is more than 28 days away from the first timepoint

- SD = at least one SD assessment (or better) ≥ 6 weeks after first date of study intervention 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 first date of study intervention and before progression with no measurable disease and does not meet criteria of CR

  Note: 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. However, as per inclusion criteria, measurable disease must be confirmed by IRC at inclusion, so such case is not expected to occur. But in case of Non-CR/Non-PD, the overall response is rated as "non-CR/non-PD" (if no "CR" or "PD" are previously reported) by the IRC. Non-CR/Non-PD is specific to IRC assessment (it does not apply for Investigator assessment)

- PD = progression ≤16 weeks after first date of study intervention (and not qualifying for CR, PR or SD)

- ND = at least one ND assessment

  Note: Overall response is rated as "No Disease" (ND) when the IRC is not able to identify any disease at baseline (target or non-target lesions). However, as per inclusion criteria, measurable disease must be confirmed by IRC at inclusion, so such case is not expected to occur. ND is specific to IRC assessment (it does not apply for Investigator assessment)

- NE = all assessments are NE or participant has a missing (or not evaluable) baseline tumor assessment and/or no (or not evaluable) tumor assessments on-treatment, or participant does not complete any of the following response above

**Unconfirmed best Overall Response (uBOR)**

uBOR according to RECIST 1.1 as adjudicated by IRC will be assessed according to the following rules:

- CR = at least one determination of CR

- PR = at least one determination of PR (and does not meet criteria of CR)
- SD = at least one SD assessment ≥ 6 weeks after start date and before progression (and does not meet criteria of 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)

  Note: 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. However, as per inclusion criteria, measurable disease must be confirmed by IRC at inclusion, so such case is not expected to occur. But in case of Non-CR/Non-PD, the overall response is rated as "non-CR/non-PD" (if no "CR" or "PD" are previously reported) by the IRC. Non-CR/Non-PD is specific to IRC assessment (it does not apply for Investigator assessment)
- PD = progression ⩽16 weeks after first date of study intervention (and does not meet criteria of CR, PR, SD or Non-CR/Non-PD)
- ND = at least one ND assessment

  Note: Overall response is rated as "No Disease" (ND) when the IRC is not able to identify any disease at baseline (target or non-target lesions). However, as per inclusion criteria, measurable disease must be confirmed by IRC at inclusion, so such case is not expected to occur. ND is specific to IRC assessment (it does not apply for Investigator assessment)
- NE = all assessments are NE or participant has a missing (or not evaluable) baseline tumor assessment and/or no (or not evaluable) tumor assessments on-treatment, or participant does not complete any of the following response above

**Objective Response Rate (ORR)**

The Objective Response Rate (ORR) is defined as the number of participants having an OR assessment of Complete Response (CR) or Partial Response (PR), out of the total number of participants.

Confirmed ORR will be determined as the proportion of participants with a confirmed OR of PR or CR (having at least two determinations of CR/PR at least 4 weeks apart and before progression).

The confirmed ORR according to RECIST 1.1 as assessed by the IRC is the primary endpoint of the study.

**Disease Control Rate (DCR)**

Disease Control Rate (DCR) is defined as the proportion of participants with BOR according to evaluation criteria of CR, PR, or SD out of the total number of participants.

Separate tables will be provided for confirmed and unconfirmed BOR. Each one will provide the number and percentage of participants with cBOR (respectively uBOR) of CR, PR, SD, Non-CR/Non-PD, PD, ND and NE as well as the ORR and the DCR. The ORR and DCR will be provided 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). Exact binomial test (1 sample) against the null hypothesis for the ORR will be calculated for confirmed BOR to determine whether the null hypothesis of an ORR ≤ 10% can be rejected.

The individual percentage of change in the sum of diameter since baseline will be displayed over time on a spider plot, together with the first occurrence of new lesion and the participant off treatment. The change in the sum of diameters between baseline and the best post-baseline assessment (i.e. minimum change since baseline) will be displayed on a waterfall graph. Note the sum of diameters includes all target lesions (longest diameter for non-nodal lesions and short axis for nodal target lesions)

The best percent change in sum of target lesion diameters will also be presented in a waterfall plot with confirmed BOR per IRC and investigator, PD-L1 in tumor cells and immune cells, CCI, CCI and CCI status.

For spider plots and waterfall plots, the percent change from baseline in the sum of diameters will be displayed for valid timepoint assessments, only. For the purpose of this analysis, a valid timepoint assessment is defined as a complete assessment of all target lesions reported at baseline. Further, split and coalesced lesions have to be taken into account appropriately to determine if a timepoint assessment is valid to derive the percent change from baseline in sum of diameters. All sum of diameters should be used, including the ones beyond first PD (for waterfall plot, the best post-baseline sum of diameters will be used even if occurring beyond PD). For waterfall plot, the percent change from baseline to 8-weeks assessment, as well as the percent change from baseline to the best post-baseline sum of diameters will be displayed for each participant.

Listing of tumor assessment will be provided with the following information: participant identifier, age, sex, race, date of start of subsequent anticancer therapy, 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 from baseline, percent change in target lesions from nadir and overall response of participant. Note that response of target and non-target lesions and change from nadir is not applicable at baseline.

In addition, a summary table of the reasons for non-evaluable confirmed BOR will be provided, the following reasons will be detailed:

- No baseline assessment (if applicable)
- No post-baseline assessments due to death within 8 weeks after the start of study treatment
- No post-baseline assessments due to other reasons
- All post-baseline assessments have overall response 'Non-evaluable'
- New anticancer therapy started before first evaluable post-baseline assessment
- SD of insufficient duration (<6 weeks after the start of study treatment)

Note: Special cases where BOR is NE due to both early SD and late PD will be classified into this category

- Non-CR/Non-PD of insufficient duration (<6 weeks after the start of study treatment) (if required by the data)
- No evaluable tumor assessment >16 weeks followed by PD (i.e. tumor assessment of PD was >16 weeks after start of study treatment and there was no evaluable tumor assessment in between)
- No IRC review and Not determined categories may also be added if applicable

A listing of reasons for non-evaluable confirmed BOR will also be created including: participant identifier, age, sex, race, date of first and last dose, date(s) of imaging, overall response and the reason for confirmed BOR non-evaluable.

### 14.1.1 Subgroup analysis for BOR

Confirmed BOR, ORR and DCR according to RECIST 1.1 as adjudicated by IRC will also be evaluated on all subgroups as defined in Section 8.2. This will be repeated for Japanese participants including only the ADA and nAb status subgroup.

For ADA and nAb subgroups a listing will present the tumor assessment and overall response per RECIST 1.1 as assessed by the IRC including: participant ID, age, sex, race, ADA Status, nAb PD-L1 Assay Status, nAb TGF-β Assay Status, visit, date(s) of imaging, description of target lesions (size, site, status, type, method, response), non target lesions (status, site, method, response), and new lesions (site, status, method), sum of lesion diameters, percent change in target lesions at baseline, and overall response. Two listings will be created: for the ADA positive participants and for nAb positive in either away (PD-L1 and TGF-β) participants.

Participants with missing subgroup category will not be included in the related subgroup analysis. Those subgroups analyses will be supported by forest plots.

## 14.2 Best Overall Response according to RECIST 1.1 as assessed by Investigator (Secondary endpoint)

| Analysis (Analysis Population) | Derivation | Statistical Analysis Methods | Missing data handling |
|---|---|---|---|
| **Secondary endpoint: confirmed Objective Response (OR)** | | | |
| Secondary analysis (ITT) | Confirmed Objective Response Rate (ORR) according to RECIST 1.1 as assessed by the Investigator is defined as the number of participants having a confirmed objective response assessment of Complete Response (CR) or Partial Response (PR) (at least two determinations of CR/PR at least 4 weeks apart and before progression), out of the total number of participants. | ORR will be provided 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). | Participants with missing post-baseline data are considered as having a non-evaluable response |

Data collected in the eCRF for RECIST 1.1 response criteria as assessed by the Investigator ("Assessment of disease based on imaging (according to RECIST 1.1)" eCRF page) will be treated in the same way as data adjudicated by the IRC. Analyses described in Section 14.1 will be repeated based on Investigator assessment, i.e.:

- Number and percentage of participants with confirmed BOR of CR, PR, SD, PD and NE and corresponding listing
- ORR and DCR with their two-sided 95% CI
- Spider and Waterfall plots
- A summary table of the reasons for non-evaluable confirmed BOR and related listing, with the following reasons:
  - No baseline assessment (if applicable)
  - No post-baseline assessments due to death within 8 weeks after the start of study treatment
  - No post-baseline assessments due to other reasons
  - All post-baseline assessments have overall response 'Non-evaluable'
  - New anticancer therapy started before first evaluable post-baseline assessment
  - SD of insufficient duration (<6 weeks after the start of study treatment)
  - No evaluable tumor assessment >16 weeks followed by PD (i.e. tumor assessment of PD was >16 weeks after start of study treatment and there was no evaluable tumor assessment in between)
  - Not determined category may also be added if applicable

All above analyses will be repeated for unconfirmed BOR.

In addition, a summary of the BOR (confirmed and unconfirmed) as adjudicated by IRC versus Investigator assessment will be provided included numbers of concordant and discordant assessment, and a listing of inconsistencies will be provided.

CCI


CCI

CCI

## 14.4 Progression-Free Survival (Secondary Endpoint)

| Analysis (Analysis Population) | Derivation | Statistical Analysis Methods | Missing data handling |
|---|---|---|---|
| **Secondary endpoint: Progression-Free Survival (PFS)** | | | |
| Secondary analysis (ITT) | Time from first administration of study treatment until the first documentation of progression of disease (PD) or death due to any cause, whichever occur first, per RECIST 1.1 as measured by the IRC. | Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics (median time, 3-, 6-, 9-, 12-, 24-month rate estimates) including the corresponding two-sided 95% confidence intervals. The confidence intervals for the median will be calculated according to Brookmeyer and Crowley (1982) and confidence intervals for the survival function estimates at above defined timepoints will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST). The estimate of the standard error will be computed using Greenwood's formula. | PFS will be censored at the date of last progression-free tumor assessment or date of first study treatment, whatever is later.<br><br>Censoring rules (details provided in body text):<br>- No event (PD or death)<br>- No baseline and/or post-baseline assessment<br>- Start of new anticancer treatment<br>- Event after two or more subsequent missing response assessments |
| Sensitivity analysis 1 (ITT) | | | PFS will be censored at the date of last tumor assessment with outcome CR, PR or SD or date of first study treatment, whatever is later.<br><br>Censoring rules (details provided in body text):<br>- No event (PD or death)<br>- No baseline and/or post-baseline assessment |

| | | | |
|---|---|---|---|
| Sensitivity analysis 2 (ITT) | Time from first administration of study treatment until the first documentation of progression of disease (PD) or death due to any cause, whichever occur first, per RECIST 1.1 as measured by the Investigator. | | PFS will be censored at the date of last tumor assessment with outcome CR, PR or SD or date of first study treatment, whatever is later.<br><br>Censoring rules (details provided in body text):<br>- No event (PD or death)<br>- No baseline and/or post-baseline assessment<br>- Start of new anticancer treatment<br>- Event after two or more subsequent missing response assessments |

Sensitivity Analysis 1: PFS counting all events (PD and death) regardless of the start of a new anticancer therapy and ignoring the number of missing evaluable tumor assessments before progression or death
Sensitivity Analysis 2: PFS as assessed by the Investigator

**Progression-Free Survival (PFS)** time is defined as the time from first administration of study treatment until the first documentation of progression of disease (PD) or death due to any cause, whichever occur first.

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

The following censoring rules will also be applied for the PFS computation:

- Participants with no event (PD or death) will be censored on the date of the last adequate tumor assessment
- Participants who do not have a baseline tumor assessment or who do not have any evaluable post-baseline tumor assessments will be censored at the date of first administration of study treatment unless death occurred on or before the time of the second planned tumor assessment (i.e. 16 weeks) in which case the death will be considered an event.
- Participants who start new anticancer treatment prior to an event will be censored on the date of the last evaluable tumor assessment before anticancer therapy is given.

  Note: any systemic anticancer therapy, any anticancer surgery and any anticancer radiotherapy with exception of palliative bone radiotherapy will be considered as new/subsequent anticancer therapy and will lead to censoring.
- Participants with an event after two or more subsequent missing response assessments (i.e. no assessments in 112 days during the first 12 months of follow-up or 168 days after the first 12 months of follow-up) will be censored on the date of the last evaluable tumor assessment.

The last tumor assessment date is defined as the last available and evaluable tumor assessment performed prior to the cut-off date (or prior to end of study, i.e. participants lost to follow-up or who withdraw consent) or prior to subsequent anticancer therapy. If no evaluable tumor assessment is available, this date will be the date of first administration of study treatment.

Censoring rules are also summarized in Table 5.

**Table 5 Censoring Rules for Primary and Sensitivity Analysis of PFS**

| Situation | | Primary Analysis | Sensitivity Analysis 1 | Sensitivity Analysis 2 |
|---|---|---|---|---|
| No PD and no death | New anticancer therapy is not initiated | Censored at last tumor assessment* | Censored at last tumor assessment* | Censored at last tumor assessment* |
| | New anticancer therapy is initiated | Censored at last tumor assessment* before new anticancer therapy | Censored at last tumor assessment* | Censored at last tumor assessment* before new anticancer therapy |
| No baseline assessment or no evaluable post-baseline assessment | No death or death >16 weeks after start of study treatment | Censored at date of first administration of study treatment | Censored at date of first administration of study treatment | Censored at date of first administration of study treatment |
| | Death ≤16 weeks after start of study treatment | Progressed at date of death | Progressed at date of death | Progressed at date of death |
| PD or death | After ≤1 subsequent missing response assessment[a] | Progressed at date of document PD or death, whichever came first | Progressed at date of documented PD or death, whichever came first | Progressed at date of documented PD or death, whichever came first |
| | After ≥2 subsequent missing response assessment[a] | Censored at last tumor assessment* before missing assessments. | Progressed at date of documented PD or death, whichever came first | Censored at last tumor assessment* before missing assessments |
| | Before started new anticancer therapy | Progressed at date of document PD or death, whichever came first | Progressed at date of documented PD or death, whichever came first | Progressed at date of document PD or death, whichever came first |
| | After started new anticancer therapy | Censored at last tumor assessment* before new anticancer therapy | Progressed at date of documented PD or death, whichever came first | Censored at last tumor assessment* before new anticancer therapy |

Sensitivity Analysis 1: PFS counting all events (PD and death) regardless of the start of a new anticancer therapy and ignoring the number of missing evaluable tumor assessments before progression or death
Sensitivity Analysis 2: PFS as assessed by the Investigator
* with outcome CR, PR or SD. If no adequate tumor assessment, censored at date of first administration of study treatment
[a] No assessments in 112 days during the first 12 months of follow-up or 168 days after the first 12 months of follow-up

PFS according to RECIST 1.1 as measured by the IRC will be described (secondary endpoints).

CCI

The analysis of PFS 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 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.

Frequency (number and percentage) of participants with and without event and within each event type (PD or death) will be presented as well as the PFS median, min and max. Number of participants at risk, failed and PFS rates with their CI at 6, 12 and 24 months will also be provided. Censoring reasons will also be described. Censoring reasons are as follows:

- Administrative censoring (ongoing in the study without an event)
- No baseline assessment
- No evaluable post-baseline assessment
- Start of new anticancer therapy
- Event after 2 or more missing or non-evaluable post-baseline assessments
- Withdrawal of consent
- Lost to follow-up

Lost to follow-up will include the following participants:

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

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

Separate listings will be provided for the different assessment methods with the following information: participant identifier, age, sex, race, date of first administration of study treatment, date of last tumor assessment, date of event/censoring, event/censoring reason, time to event.

### 14.4.1 Sensitivity analysis for PFS

The two following sensitivity analyses will be conducted:

1. PFS counting all events (PD and death) regardless of the start of a new anticancer therapy and ignoring the number of missing evaluable tumor assessments before progression or death

2. PFS as assessed by the Investigator (described above in Section 14.4)

Those sensitive analyses will be supported with Kaplan-Meir plots. Listing of progression-free survival by the Investigator will also be provided.

### 14.4.2 Subgroup analysis for PFS

PFS according to RECIST 1.1 as adjudicated by IRC will also be evaluated on all subgroups as defined in Section 8.2. This will be repeated for Japanese participants including only the ADA and nAb status subgroup.

For the ADA and nAb subgroups, participant listings will provide the following information: participant ID, age, sex, race, ADA status, nAb PD-L1 assay status, nAb TGF-β assay status, date of first administration, date of last tumor assessment, date of event/censoring, event/censoring reason, time to event. Two listings will be created: for the ADA positive participants and for nAb positive in either assay (PD-L1 and TGF-β) participants.

Participants with missing subgroup category will not be included in the related subgroup analysis. Those subgroups analyses will be supported by Kaplan-Meir and forest plots (based on PFS rates at 6-month and 12-month).

## 14.5 Duration of Response (Secondary Endpoint)

| Analysis (Analysis Population) | Derivation | Statistical Analysis Methods | Missing data handling |
|---|---|---|---|
| **Secondary endpoint: Duration of Response (DOR)** | | | |
| Secondary analysis (ITT) | The duration of overall response is measured from the time from first documentation of confirmed objective response (CR or PR) to the date of first documentation of objective progression of disease (PD) or death due to any cause, whichever occurs first This definition applies only to participants who experienced confirmed objective response. It will be measured according to RECIST 1.1 as measured by the IRC and the Investigator | Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics (median time, 3-, 6-, 9-, 12-, 24-month rate estimates) including the corresponding two-sided 95% confidence intervals. The confidence intervals for the median will be calculated according to Brookmeyer and Crowley (1982) and confidence intervals for the survival function estimates at above defined timepoints will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST). The estimate of the standard error will be computed using Greenwood's formula. | Same censoring rules as for PFS. |

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

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

The censoring rules for DOR are as described above for PFS (primary definition in Table 5) in Section 14.4.

DOR of confirmed CR/PR according to RECIST 1.1 as measured by the IRC and the Investigator will be described. Considering confirmed CR/PR involve having at least two objective responses, the date of occurrence of the first CR/PR will be used as date of objective response.

The analysis of DOR 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 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.

DOR rates with their CI at 3, 6, 9, 12 and 24 months will be presented, as well as the number of participants at risk and failed.

The time to and duration of response per participant having a confirmed objective response (including delayed response, see Section 14.7 below) will be displayed in swimmer graphs. Kaplan-Meier figures will also be provided.

Listings will be provided with the following information: participant identifier, age, sex, race, date of first study treatment administration, date of first response, date of last tumor assessment, censored (Y/N), date of event/censoring, event/censoring reason and duration of response.

### 14.5.1 Subgroup analysis for DOR

DOR according to RECIST 1.1 as adjudicated by IRC will also be evaluated on all subgroups as defined in Section 8.2. This will be repeated for Japanese participants including only the ADA and nAb status subgroup.

Participants with missing subgroup category will not be included in the related subgroup analysis. Those subgroups analyses will be supported by forest plots (based on DOR rates at 6-month and 12-month).

## 14.6 Durable Response (Secondary Endpoint)

| Analysis (Analysis Population) | Derivation | Statistical Analysis Methods | Missing data handling |
|---|---|---|---|
| **Secondary endpoint: Durable Response Rate (DRR)** | | | |
| Secondary analysis (ITT) | Number of participants having a DOR of at least 6 months, out of the total number of participants | DRR will be provided 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). | Participants with missing baseline and post-baseline tumor assessments will be defined as non-evaluable, and considered as having no response. |

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

The DRR will be provided 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).

DRR will be described according to RECIST 1.1 as measured by the IRC and the Investigator.

### 14.6.1 Subgroup analysis for DRR

DRR according to RECIST 1.1 as adjudicated by IRC will also be evaluated on all subgroups as defined in Section 8.2. This will be repeated for Japanese participants including only the ADA and nAb status subgroup.

Participants with missing subgroup category will not be included in the related subgroup analysis. Those subgroups analyses will be supported by forest plots.

## 14.7 Delayed Response

**Delayed response** will be defined as a documented objective response [CR or PR] that occurred after the initial progression of disease. For those participants, the duration of delayed response (DOdR) will be defined as:

DOdR = (date of end of objective response – date of objective response + 1)/30.4375 (months).

with end of response being the earliest date between PD occurring after response, death, start of new anticancer treatment or cut-off date.

Two listings for participants having a delayed response according to RECIST 1.1 as measured by IRC will be described will be provided:

- For tumor assessment including: participant identifier, age, sex, race, date of first and last study intervention, date(s) of imaging, sum of lesion diameters, percent change in target lesions from baseline, and overall response.

- For duration of response including: participant identifier, age, sex, race, date of first and last study intervention, date/study day of first PD, date/study day of first response (CR or PR), ongoing response at cutoff date (Y/N), date/study day of end of response (with reason for end of response being PD, death, new anticancer treatment), duration of delayed response.

## 14.8 Overall survival

| Analysis (Analysis Population) | Derivation | Statistical Analysis Methods | Missing data handling |
|---|---|---|---|
| **Secondary endpoint: Overall Response (OS)** | | | |
| Secondary analysis (ITT) | Time from first administration of study treatment to the date of death due to any cause | Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics (median time, 3-, 6-, 9-, 12-, 24-month rate estimates) including the corresponding two-sided 95% confidence intervals. The confidence intervals for the median will be calculated according to Brookmeyer and Crowley (1982) and confidence intervals for the survival function estimates at above defined timepoints will be derived using the log-log transformation according to Kalbfleisch and Prentice (2011) (conftype=loglog default option in SAS Proc LIFETEST). The estimate of the standard error will be computed using Greenwood's formula. | For alive participants, OS will be censored at last date known to be alive<br><br>The following dates will be considered to determine the last date known to be alive (dates past the data cut-off will be ignored by the derivation; details provided in body text):<br>• All participant assessment dates<br>• Start and end dates of anticancer therapies administered after study treatment discontinuation.<br>• AE start and end dates<br>• Last known alive date collected on the 'Subject Status / Survival Follow-Up' eCRF page (do not use follow up date)<br>• Study drug start and end dates<br>Date of discontinuation from the "Study Termination" eCRF page (do not use if reason for discontinuation is lost to follow-up or death) |

**Overall survival (OS)** is defined as the time from first administration of study treatment to the date of death due to any cause:

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

For participants 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 6.

**Table 6 Survival Event / Censoring**

| Survival Status | 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 complete 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 and which are not imputed shall be used in the derivation:

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

Only dates associated with actual examinations of the participant reported in the eCRF will be used in the derivation. Dates associated with a technical operation unrelated to participant status such as the date a blood sample was processed will not be used.

Data collected after re-initiated treatment will be considered in the derivation of the last known to be alive date.

The analysis of OS time will be performed with a Kaplan-Meier method with the same approach as for PFS described in Section 14.4. Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including corresponding two-sided 95% CIs. OS rates with their CI at 3, 6, 9,12 and 24 months will be presented, as well as the number of participants at risk and failed. Censoring reasons will also be described. Censoring reasons are as follows:

- Alive at cut-off date
- Withdrawal of consent
- Lost to follow-up

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

A participant listing will provide the following information: participant identifier, age, sex, race, date of first study intervention, date of event/censoring, event/censoring reason, time to event.

### 14.8.1 Subgroup analysis for OS

Overall survival will also be evaluated on all subgroups as defined in Section 8.2. This will be repeated for Japanese participants including only the ADA and nAb status subgroup.

For the ADA and nAb subgroups, participant listings will provide the following information: participant ID, age, sex, race, ADA status, nAb PD-L1 assay status, nAb TGF-β assay status, date of first administration, date of last tumor assessment, date of event/censoring, event/censoring reason, time to event. Two listings will be created: for the ADA positive participants and for nAb positive in either assay (PD-L1 and TGF-β) participants.

Participants with missing subgroup category will not be included in the related subgroup analysis. Those subgroups analyses will be supported by Kaplan-Meier and forest plots (based on OS rates at 6-month and 12-month).

## 14.9 Follow-up time (PFS)

Kaplan-Meier analysis will be performed to estimate median time of duration of follow-up for PFS as adjudicated by IRC using the reverse censoring indicator of the PFS analysis (see Section 14.4).

Kaplan-Meier survival curves will also be presented with median duration of follow-up an its two-sided 95% CI. In particular, the follow-up rates at 3, 6, 9, 12 and 24 months will be estimated with their 95% CI.

## 14.10 Follow-up time (OS)

Kaplan-Meier analysis will be performed to estimate median time of duration of follow-up for OS using the reverse censoring indicator of the OS analysis (see Section 14.8).

Kaplan-Meier survival curves will also be presented with median duration of follow-up an its two-sided 95% CI. In particular, the follow-up rates at 3, 6, 9, 12 and 24 months will be estimated with their 95% CI.

# 15 Safety Analyses

*Analysis set: SAF*

The subsections in this section include specifications for summarizing safety endpoints that are common across clinical studies such as adverse events, laboratory tests and vital signs.

## 15.1 Adverse Events

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

Changes in toxicity grade, seriousness or outcome of AEs are recorded as separate entries in the eCRF with associated end and start dates (start date equals end date of previous entry). Such entries reporting the same event in such immediately consecutive periods will be considered as one event in the analysis. These events will be kept as separate records in the database in order to maintain

the full detailed history of the events. The start date of the initial record in the sequence is taken as start date of the entire event, similarly the end date of the last event in the sequence is taken as end date of the entire event. The overall outcome of the adverse event is the outcome of the last event in the sequence. Duration of the AE and the TEAE flag will be adjusted accordingly in the analysis.

All analyses described in this section will be based on TEAEs if not otherwise specified. The AE listings will include all AEs. AEs outside the on-treatment period (prior or after) will be flagged in listings. AEs occurring during the re-initiation phase will be considered in the summary tables and will be flagged in listings.

Incomplete AE-related dates will be handled as stated in Section 9.

**Related Adverse Events** are those events with relationship to study treatment (as recorded on the "Adverse Events Details" eCRF page, Relationship with M7824= Related) reported by the Investigator and those of unknown relationship (i.e. no answer to the question "Relationship with M7824").

**Serious Adverse Events (SAE)** are those events reported on the "Adverse Events Details" eCRF page, with the "Serious Adverse Event" field ticked "Yes".

**Adverse Events Leading to Temporary Discontinuation** are those events leading to temporary discontinuation of study treatment (answer to the question "Action(s) taken with M7824" = "Drug interrupted" on "Adverse Event Details" eCRF page).

**Adverse Events leading to Permanent Treatment Discontinuation** are those events leading to permanent discontinuation of study treatment (answer to the question "Action(s) taken with M7824" = "Drug withdrawn" on "Adverse Event Details" eCRF page).

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

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

- Infusion-Related Reactions (IRRs)
- Immune-Related Adverse Events (irAEs)
- Skin AEs possibly related to TGF-β inhibition
- Anemia

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

### 15.1.1 All Adverse Events

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

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

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

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

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

- TEAEs
- Related TEAEs
- Serious TEAEs
- Related Serious TEAEs
- TEAEs NCI-CTCAE severity grade (≥3, ≥4)
- Related TEAEs NCI-CTCAE severity grade (≥3, ≥4)
- TEAEs leading to death
- Related TEAEs leading to death
- AEs and related AE 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

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

- MedDRA primary SOC (ordered alphabetically) and PT (ordered alphabetically), (except for AE of special interest)

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.

**Clinicaltrials.gov and EudraCT -requirements**

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

Listings of adverse events will contain the following information: 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), AESI infusion-related (Y/N), AESI immune-related (Y/N), potential TGF-β-mediated skin AESI (Y/N), anemia AESI (Y/N). TEAEs and AEs occurring during the re-initiation period will be flagged.

Following listings will be provided with the relevant information:

- Listing of all AE (whether treatment-emergent or not) (TEAEs will be flagged) (AE occurring during re-initiation phase will be flagged)
- Listing of TEAEs
- Listing of non-TEAE for AEs occurring after enrollment (date of first signature of informed consent/date of first signature of first informed consent) but prior to the first dose of study intervention.
- Listing of AE with onset or worsening after the on-treatment period (AE occurring during re-initiation phase will be flagged)
- Listing of AEs for participants who reinitiated the treatment

**Evaluation of Potential Effect of ADA and nAb on bintrafusp alfa Safety**
The following analyses frequency and percentage of AEs by ADA status (ever positive, never positive), and by nAb status (ever positive for either assay, never positive) will be performed:

- TEAEs
- TEAEs, grade ≥3
- TEAEs leading to permanent treatment discontinuation
- TEAEs excluding IRRs leading to drug interruptions
- Serious TEAEs
- TEAEs leading to death
- irAEs (see definition in Section 15.2.4.2)
- IRRs (see definition in Section 15.2.4.1)

This will be repeated for Japanese participants.

Listings of all AEs and all IRRs for ever-positive ADA participants (pre-existing, transient treatment-emergent, persistent-treatment emergent) will be prepared including participant identifier and showing the date(s) of the positive ADA result together with the AEs or IRRs. For the AEs and IRRs, start and stop date will be shown along with grade. Adverse events recorded during the period of 2-weeks prior to the positive ADA value till two weeks after the positive ADA value will be flagged. This will be repeated for ever-positive nAb participants in either assay (PD-L1 and TGF-β). These analyses for ever positive ADA and ever positive nAb subjects in either assay will be repeated for Japanese participants.

**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 be assessed based on the difference in incidence rate between subgroups of pre and during COVID-19 period according to categorization method with time window:

- For common TEAEs applying frequency threshold of 10% in either subgroup of pre/during COVID-19 period, butterfly plots displaying incidence rates of TEAEs NCI-CTCAE severity grade ≤2 and ≥3 will be provided.
- For each AESIs defined in Section 15.2.4 occurred on 4 or more participants in any subgroup (Rule-of-4), forest plots displaying incidence rates, differences in incidence rates as well as the CI for the difference will be provided.

### 15.1.2 Adverse Events Leading to Study Intervention Discontinuation

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

- TEAEs leading to temporary drug interruption
- 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 in an overview table.

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

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

### 15.2.1 Deaths

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

The following summaries will be provided:

- Number of deaths (including deaths during re-initiation phase)
- Number of deaths within 30 days after last dose of study treatment
- Number of deaths within 60 days after first dose of study treatment
- Primary Reason for Death
    - Progressive disease and/or disease related condition
    - Event unrelated to study treatment
    - Event related to study treatment
    - Unknown

In addition, date and cause of death will be provided in an individual participant data listing together with following dosing information: participant identifier, age, sex, race, date of first/last

study intervention, number of infusions, day relative to the first and the last infusion, autopsy (Y/N/U), AEs with fatal outcome (list preferred terms of AEs with outcome=Fatal, as well as Grade 5 or Serious resulting in death), flag for death within 30 days of last dose of study treatment and flag for death within 60 days of first dose of study treatment.

### 15.2.2 Serious Adverse Events

The number of participants with serious AEs (SAEs) will be described by SOC and PT:

- SAEs
- Related SAEs

Bar charts presenting SAEs and related SAEs of the most frequent PT (at least 5%) will be displayed by worst grade and PT.

Listing of SAEs will also be provided (see description of listing in Section 15.1.1).

### 15.2.3 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

### 15.2.4 Adverse Events of Special Interest

#### 15.2.4.1 Infusion-Related Reaction including Immediate Hypersensitivity

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 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 infusion (during or after the infusion) and resolved completely with the end date on the same day of the infusion or the day after for any of the following: pyrexia, chills, flushing, hypotension, dyspnea, wheezing, back pain, abdominal pain and urticaria.

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.

When at least 10 participants will have received premedication for the first infusion and respectively at least 10 participants will have received the first study medication without premedication, results will also be split by premedication (Y/N) subgroups.

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

The listing of IRRs will be provided with the relevant information (see description of listing in Section 15.1.1). One additional listing will display the study drug administration details together with the infusion-related adverse event including 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.

### 15.2.4.2 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.

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

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 (i.e. at least one (clear) etiology of the event is provided) as the answer to the question "Does any of the following provide a clear etiology for the event?", the event will be considered as a non-irAE.

All non-pre-treatment AE records will be flagged for an event if at least one non-pre-treatment record satisfies the criteria.

PTs will be compiled into categories: Immune-mediated rash, Immune-mediated colitis, Immune-mediated pneumonitis, Immune-mediated hepatitis, Immune-mediated nephritis and renal dysfunction, Immune-mediated endocrinopathies (Adrenal insufficiency, Hypogonadism, Pituitary dysfunction, Type 1 Diabetes Mellitus, Thyroid disorders), Other immune-mediated adverse events (myositis, myocarditis, pancreatitis, neurologic events, other).

Immune-related adverse events (irAEs) will be summarized by the following variables:

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

The frequency table of immune-related AEs by worst grade, category, subcategory (for Immune-mediated endocrinopathies), sub-subcategory (for Immune-mediated endocrinopathies – thyroid disorders) and PT will also be provided.

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

### 15.2.4.3 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

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

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

- All skin 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). Plus, from the "TGFβ Mediated Skin Reaction" eCRF page, the number of lesions, confirmation of the diagnosis (Y/N) and lesion location.

### 15.2.4.4 Anemia

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

The following high level terms (HLT) and PTs will be used to select the anemia AEs to be included in the listing:

- Anaemias NEC (HLT)
- Anaemias haemolytic immune (HLT)
- Anaemias haemolytic NEC (HLT)
- Haemoglobin decreased (PT)

## 15.3 Clinical Laboratory Evaluation

Baseline and on-treatment 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 5.0 and as specified in Appendix 3. Additional laboratory results that are not part of NCI-CTCAE will be categorized as follows: below normal limits, within normal limits, and above normal limits (according to the original laboratory normal ranges).

Quantitative data will be summarized using descriptive statistics (mean, standard deviation, median, Q1, Q3, minimum, and maximum) of actual baseline values, on-treatment values and changes from baseline to each on-treatment visit over time. Refer to Section 9 to the consideration of Unscheduled assessments. Qualitative data based on reference ranges will be described according to the categories (i.e. Abnormal, Normal).

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

The following figures will be provided:

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

Boxplots for laboratory parameters where toxicity grades are defined based on the ratio of the parameter values and the upper limit of normal (ULN) will not be displayed using the unit of measurement but instead using the ratio of the measured value over ULN. This comprises alkaline phosphatase (ALP), alanine aminotransferase (ALT), aspartate aminotransferase (AST), Activated Partial Thromboplastin Time (aPTT), bilirubin, and creatinine.

**Laboratory parameters with NCI-CTC grades available**

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

The following summaries will be displayed:

- Number and percentage of participants by worst on-treatment grade (>=1, >=3, >=4)

- Shift in toxicity grading from baseline to highest on-treatment toxicity grade

The definitions of the NCI-CTCAE toxicity grading version 5.0 for each parameter are provided in Appendix 3 of this IAP.

**Table 7 NCI-CTC Gradable parameters**

| Parameter | Parameter code | 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 total | BILI | Blood bilirubin increased | High |
| Calcium[a] | CA | Hypercalcemia/Hypocalcemia[a] | High/Low |
| Creatinine | CREAT | Creatinine increased | High |
| Glucose | GLUC | Hypoglycemia | Low |
| Lipase | LIPASET | Lipase increased | High |
| Potassium | K | Hyperkalemia/Hypokalemia | High/Low |
| Sodium | SODIUM | Hypernatremia/Hyponatremia | High/Low |
| **Hematology** | | | |
| Absolute eosinophils | EOS | Eosinophilia | High |
| Absolute lymphocyte | LYM | Lymphocyte count decreased/Lymphocyte count increased | High/Low |
| Absolute neutrophils | NEUT | Neutrophil count decreased | Low |
| Hemoglobin | HGB | Anemia/Hemoglobin increased | Low/High |
| Leukocytes (WBC) | WBC | Leukocytosis/White blood cell decreased | High/Low |
| Platelets count | PLAT | Platelet count decreased | Low |
| **Coagulation** | | | |
| Activated Partial Thromboplastin Time[b] | APTT | Activated partial thromboplastin time prolonged | High |
| Activated PTT/Standard[b] | APTTSTND | Activated partial thromboplastin time prolonged | High |
| Prothrombin International Normalized Ratio[b] | INR | INR increased | High |

[a] based on corrected calcium (see Appendix 3)
[b] reported on the "Coagulation" eCRF page

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

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

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

For calcium, CTCAE grading is based on corrected calcium. Corrected calcium is calculated from albumin and calcium as follows based on the International System of Units (SI):
Corrected calcium (mmol/L) = measured total calcium (mmol/L) + 0.02 (40 – serum albumin [g/L]).

**Laboratory parameters with NCI-CTC grades not available**

**Table 8 Non-NCI-CTC Gradable Parameters**

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

* reported on the "Coagulation" eCRF page

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

- Number and percentage of participants by lowest on-treatment value (classified as normal, high, low)

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

Separate listings of hematology, biochemistry and coagulation will be created. Each listing will include: participant identifier, age, sex, race, first dose date, last dose date, laboratory parameter (units), visit, date, International System of Units (SI) value, lower limit of normal (LLN), upper limit of normal (ULN), indicator of normal range (low, normal, high), toxicity grade according to NCI-CTCAE (when applicable) and highest/lowest on treatment value flag. Baseline and post-baseline values after the on-treatment period will be flagged. These listings will be sorted by participant identifier, parameter and laboratory measurement date.

**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.

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

- ALT <3×ULN, ALT ⩾ 3×ULN, ALT ⩾ 5×ULN, ALT ⩾ 10×ULN, ALT ⩾ 20×ULN
- AST <3×ULN, AST ⩾ 3×ULN, AST ⩾ 5×ULN, AST ⩾ 10×ULN, AST ⩾ 20×ULN
- (ALT and AST) <3×ULN, (ALT or AST) ⩾ 3×ULN, (ALT or AST) ⩾ 5×ULN, (ALT or AST) ⩾ 10×ULN, (ALT or AST) ⩾ 20×ULN
- 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 participant with an elevation of AST ⩾10×ULN will also appear in the categories ⩾5×ULN and ⩾3×ULN.

A plot of peak ALT versus peak total bilirubin, both relative to the ULN will be provided. This eDISH plot (evaluation of drug-induced serious hepatotoxicity) will be divided into 4 quadrants by the lines through ALT ≥ 3×ULN and total bilirubin ≥2×ULN. The left lower quadrant is then considered normal or insignificant elevations in liver chemistries, the upper quadrants indicate

participants with possible Gilbert's cholestasis; the right upper quadrant are the possible Hy's Law participants; the right lower quadrant is possible Temple's Corollary (participants with ALT≥ 3 x ULN but not satisfying Hy's Law). Same plot will be provided for AST.

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

**Urinalysis / urinalysis microscopic evaluation, hormonal tests, serum, serology, cancer antigen 19.9**

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

- Urinalysis parameters:
  - Urinalysis full parameters: physical appearance, pH, specific gravity, protein, glucose, ketones, nitrite, leukocyte esterase, blood, urobilinogen, bilirubin, color
  - Urinalysis microscopic parameters: erythrocytes (RBC), leukocytes (WBC), epithelial cells, bacteria, crystals, casts, mucus
- Hormonal parameters: thyroxine free (Free T4), thyrotropin (Thyroid-Stimulating Hormone; TSH)
- Serum parameters: Krebs von den Lungen-6 (KL-6), surfactant protein A (SP-A), surfactant protein D (SP-D) (only for Japanese sites)
- Cancer antigen 19.9
- Serology parameters: hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis B DNA, hepatitis C RNA

**Pregnancy test**

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

- Pregnancy parameters (serum or highly sensitive urine human chorionic gonadotropin (hCG))
- Post-menopausal status: FSH and estradiol parameters

## 15.4 Vital Signs

Summary table over time will present summary statistics for vital signs variables by timepoint.

The following potentially clinically significant abnormalities will be summarized:

- ≤ 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
- ≤ 20 breaths/min and decrease from baseline ≥ 5 breaths/min in respiratory rate
- ≥ 20 breaths/min and increase from baseline ≥ 5 breaths/min in respiratory rate
- ≥ 10% weight increase
- ≥ 10% weight decrease

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

**Oxygen Saturation**

Results for oxygen saturation as collected on the "Oxygen Saturation" eCRF page will be listed.

## 15.5 Other Safety or Tolerability Evaluations

**ECG**

Single 12-lead ECGs will be obtained as outlined in the protocol using an ECG machine that automatically calculates the heart rate and measures PR, QRS, QT, and QTc Intervals. ECG parameters will be derived from the data collected on the "Electrocardiogram" eCRF page. Shift table for ECG interpretation value at baseline and value at end of treatment will be provided.

A listing of ECG values will be provided including participant identifier, age, sex, race, ECG parameter and unit, visit, ECG date, value, change from baseline. Baseline values and post-baseline values collected after the on-treatment period will be flagged. Qualitative ECG results will also be provided in the listing.

**ECOG Performance Status**

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

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

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

**Non-protocol Related Hospital Visits**

Participant's non-protocol related hospital visits will be listed from data collected on the "Non Protocol Related Hospital Visit" eCRF page. The listing will include participant identifier, age, sex, race, total number of pre-planned ambulant (outpatient) hospital visits since the last study

visit, total number of unplanned ambulant (outpatient) hospital visits since the last study visit, total number of nights spent in the hospital for pre-planned overnight (inpatient) hospital visits since the last study visit and total number of nights spent in the hospital for unplanned overnight (inpatient) hospital visits since the last study visit.

# 16 Analyses of Other Endpoints

## 16.1 Pharmacokinetics

The analyses described in this section will be performed by the Clinical PK/PD group (CPK) of Translational Medicine, Merck Healthcare KGaA, Darmstadt, Germany, or by a Contract Research Organization selected by the Sponsor. Pharmacokinetic listings and individual data will be presented based on the Safety Analysis Set. Summaries and statistical analyses will be based on the PK Analysis Set. Only subgroup sample size with a minimal 3 subjects will be displayed.

Pharmacokinetic concentrations/parameters refer to bintrafusp alfa concentrations/PK parameters.

### 16.1.1 Missing/non-quantifiable PK Data Handling

**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.

**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 concentrations which are erroneous due to a protocol deviation (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 analysis if agreed upon prior to performing a statistical analysis. In this case the rationale for exclusion must be provided in the CSR.

**Exclusions for concentration data (and $C_{EOI}$/$C_{trough}$) 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

Any other PK concentrations that appear implausible to the Pharmacokineticist/PK/Pharmacodynamic 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.

## 16.1.2 Descriptive PK Analysis

**Presentation of PK Concentration Data**

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

**Presentation of PK Parameter Data**

The PK parameters listed below will be taken directly from the observed bintrafusp alfa concentration-time data.

| | |
|---|---|
| $C_{trough}$ | The concentration observed immediately before next dosing (corresponding to pre-dose or trough concentration for multiple dosing). |
| $C_{EOI}$ | The concentration observed immediately at the end of infusion. |

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

Pharmacokinetic parameter data will be presented in tables and descriptively summarized by nominal study day using: number of non-missing observations (n), arithmetic mean (Mean), standard deviation (StD), coefficient of variation (CV%), minimum (Min), median (Med), maximum (Max), geometric mean (GeoMean), StD of log-transformed data (logStD), the geometric coefficient of variation (GeoCV%) and the 95% CI for the GeoMean (LCI 95% GM, UCI 95% GM). Summaries will be based on the Pharmacokinetic Analysis Set. Subgroup analysis is done with at least 3 subjects

Additional table(s) will summarize $C_{trough}$ and $C_{EOI}$ with further stratification by ADA subsets ever positive and never positive, based on the PKADA. Additional table(s) will summarize $C_{trough}$ and $C_{EOI}$ with further sub-stratification by nAb subsets ever positive and never positive, based on the PKNAB. For nAb subsets, summaries will be done for ever positive in either of 2 nAb assays (PDL1 and TGFß receptor neutralization; PDL1+/TGFß+, PDL1+/TGFß−, PDL1−/TGFß+) versus never positive (PDL1−/TGFß−). Additional table(s) will summarize $C_{trough}$ and $C_{EOI}$ with stratification by Japanese, non-Japanese Asian, all Asian and all non-Asian participants based on the PKAS.

Additional table(s) will summarize $C_{trough}$ and $C_{EOI}$ with further sub-stratification by ADA subsets ever positive and never positive and, nested within, Japanese, non-Japanese Asian, all Asian and all non-Asian participants based on the PKADA. Additional table(s) will summarize $C_{trough}$ of ADA ever positive subjects 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 subjects, serum bintrafusp alpha $C_{trough}$ will be descriptively summarized in additional table(s) for nAb status subgroups (positive in any of 2 nAb assays), based on the PKNAB. Additional table(s) will summarize $C_{trough}$ of ADA ever-positive subjects with further sub-stratification by ADA subgroups and, nested within, Japanese, non-Japanese Asian, all Asian, and all non-Asian participants, based on the PKADA. For nAb ever-positive subjects, serum bintrafusp alpha $C_{trough}$ will be descriptively summarized in additional table(s) for nAb status subgroups (positive in any of 2 nAb assays) and, nested within, Japanese, non-Japanese Asian, all Asian, and all non-Asian participants, based on the PKNAB. Additional table(s) will summarize $C_{trough}$ of ADA Treatment-emergent subjects and nAb Treatment-emergent subjects by PK day relative to seroconversion, for all subjects, Japanese, non-Japanese Asian, all Asian and non-Asian participants based on the PKADA and PKNAB, respectively. Only subgroup sample size with a minimal 3 subjects will be displayed.

Pharmacokinetic parameters will be reported to 3 significant figures. In export datasets, as well as in the SDTM PP domain, PK parameters will be provided with full precision and will not be rounded. Descriptive statistics of PK parameter data will be calculated using full precision values and rounded for reporting purposes only.

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

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

Individual PK $C_{trough}$ and $C_{EOI}$ values will be plotted versus actual study day on a linear scale, for all participants. Individual data will be presented based on the Safety Analysis Set.

Arithmetic mean (±SD), GeoMean (±logStD), and Med $C_{trough}$ and $C_{EOI}$ will be plotted versus nominal study day on a linear scale. Summaries will be based on the Pharmacokinetic Analysis Set.

Additional figure(s) will present Mean, GeoMean, and Med $C_{trough}$ and $C_{EOI}$ with with further stratification by ADA subsets ever positive and never positive, based on the PKADA, as long as both subsets consist of 3 or more participants. Additional figure(s) will present Mean, GeoMean, and Med $C_{trough}$ and $C_{EOI}$ with further sub-stratification by nAb subsets ever positive and never positive, based on the PKNAB, as long as both subsets consist of 3 or more participants. Additional figure(s) will present Mean, GeoMean, and Med $C_{trough}$ and $C_{EOI}$ with stratification by Japanese, non-Japanese Asian, all Asian and all non-Asian participants, based on the PKAS. Only subgroup sample size with a minimal 3 subjects will be displayed.

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 readability, split further into groups of 10 subjects or fewer as needed), for all subjects, Japanese, non-Japanese Asian, all Asian, and all non-Asian participants, based on the SAF. Box plots will

be prepared for $C_{trough}$ versus PK day relative to seroconversion, for all subjects, Japanese, non-Japanese Asian, all Asian, and all non-Asian participants, based on the PKADA.

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 readability, split further into groups of 10 subjects or fewer as needed), for all subjects, Japanese, non-Japanese Asian, all Asian, and all non-Asian participants, based on the SAF. Box plots will be prepared for $C_{trough}$ versus PK day relative to seroconversion, for all subjects, Japanese, non-Japanese Asian, Asian, and non-Asian, based on the PKNAB.

CCI

CCI

## 16.4 Immunogenicity

*Analysis Sets: Immunogenicity analysis set*

### 16.4.1 Antidrug Antibody

The ADA results will be derived for each visit based on the algorithm in Table 9.

**Table 9 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.

Note that samples collected during and after the on-treatment period (e.g. safety follow-up) including the ones collected during the re-initiation period will be included in the analysis. 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 10.

**Table 10 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 missing, NR) |
| Transient positive | If treatment emergent participants have<br>- a single positive evaluation<br>or<br>- duration between first and last positive result <16 weeks<br>and last assessment not positive | Number of participants with at least one valid post-baseline result and without positive baseline result (including missing, 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 missing, NR) |

**Start of Immunogenicity Response**

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 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.

The following analysis will be described:

- 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 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.

The ADA category, titer, maximum titer, time to first ADA positive response and duration of response analyses will also be performed by subgroup:

- Japanese, non Japanese Asian, all Asian, non-Asian.

The following further analyses will also be described:

- Evaluation of potential effect of ADA on bintrafusp alfa safety (see section 15.1.1)
- Evaluation of potential effect of ADA on bintrafusp alfa efficacy
    - for best overall response according to RECIST 1.1 as adjudicated by IRC (see section 14.1.1)
    - for progression-free survival according to RECIST 1.1 as adjudicated by IRC (see section 14.4.2)
    - for duration of confirmed response according to RECIST 1.1 as adjudicated by IRC (see section 14.5.1)
    - for durable confirmed response according to RECIST 1.1 as adjudicated by IRC (see section 14.6.1)
    - for overall survival (see section 14.8.1)
- Evaluation of potential effect of ADA on bintrafusp alfa Pharmacokinetic (see section 16.1.2)

Potential effect of ADA on safety, efficacy and PK will be evaluated on ADA positive status (ever positive, never positive).

Listings of ADA results from ever positive participants will be provided with the following: participant ID, age, sex, race, ADA categories status, visit, date of assessment and results of screening, confirmatory and titer values. A further listing will contain: participant ID, age, sex, race, 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.

## 16.4.2 Neutralizing Anti-drug Antibody

*Analysis Sets: Immunogenicity analysis set*

Samples with a reportable ADA titer will also be tested in two nAb assays, PD-L1 and TGF-β. nAb results are positive or negative in a single assay and only derived when not performed because ADA was negative (see Table 11). Subjects will be characterized into different nAb categories for each assay based on the criteria in Table 12. Treatment boosted is not defined for nAb.

**Table 11 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 12 Participants Characterized based on nAb Results**

| Category | Definition | Subject at Risk (Denominator for Incidence) |
|---|---|---|
| Never positive | No nAb positive results at any time point | Number of subjects 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 subjects 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 subjects 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 subjects with at least one valid ADA post-baseline result and without nAb positive baseline results (including missing, NR) |
| Transient positive | If treatment emergent subjects 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 subjects with at least one valid ADA post-baseline result and without nAb positive baseline results (including missing, NR) |
| Persistent positive | If treatment emergent subjects have duration between first and last nAb positive result ≥16 weeks or a nAb positive evaluation at the last ADA assessment | Number of subjects with at least one valid ADA post-baseline result and without nAb positive baseline results (including missing, NR) |

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

The following analysis will be described (the titer summaries will be provided if the titer data is available):

- The frequency and percentage of each nAb category will be tabulated for each assay individually and combined (either assay).
- The nAb titer value by timepoint will be summarized, per assay
- The maximum observed nAb titer per participant per assay will be tabulated. 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, per assay
- The duration of nAb response will be summarized, per assay.

See Section 16.4.1 for the derivation of time to first positive response and duration of response.

The nAb category, titer, maximum titer, time to first nAb positive response and duration of response analyses will also be performed by subgroup:

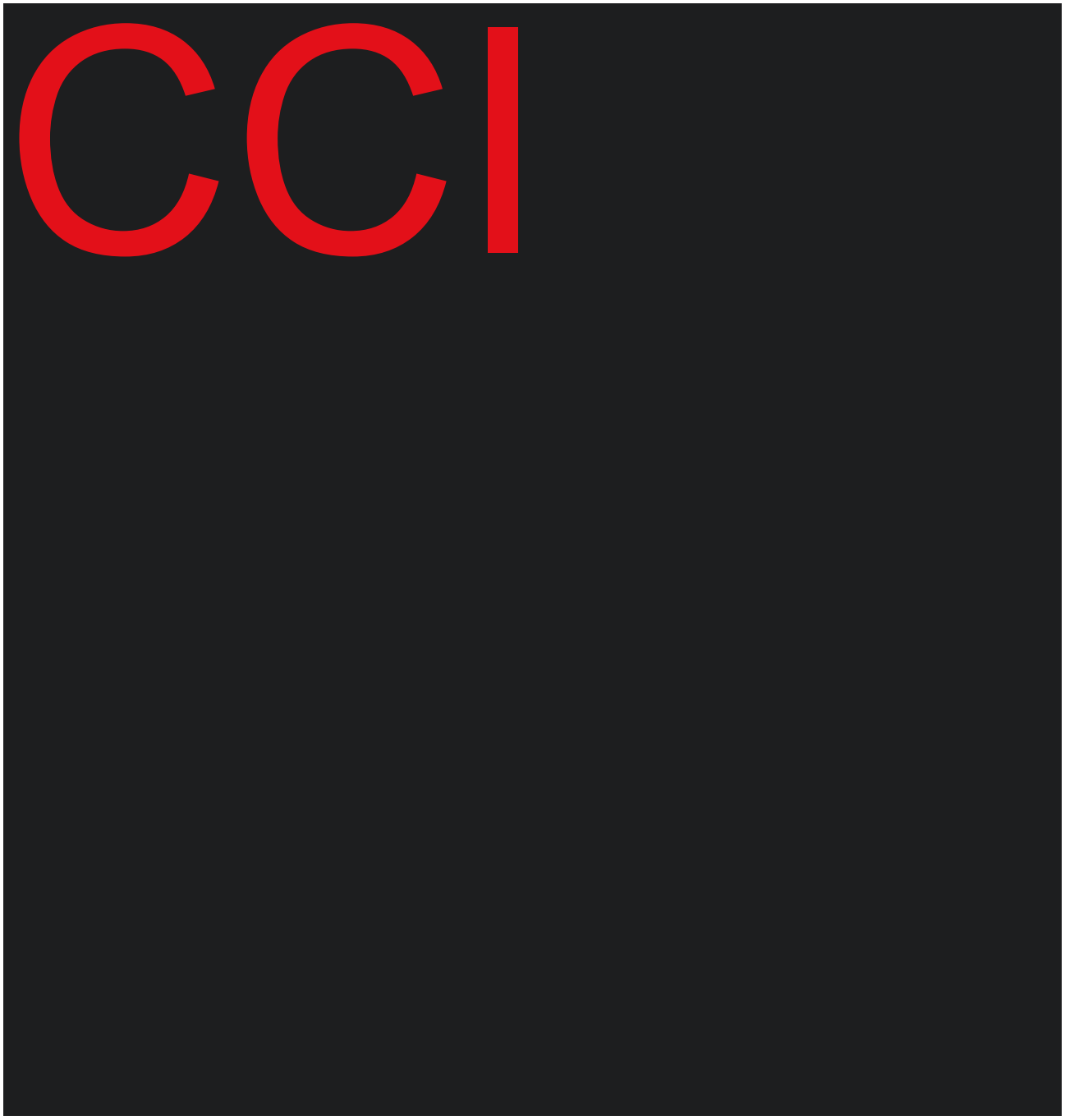
- Japanese, non Japanese Asian, all Asian, non-Asian.

A listing of nAb results from ever positive participants (in either assay) will be provided with the following: participant ID, age, sex, race, nAb categories status, visit, date of assessment and results of screening and titer values. A further listing will contain: participant ID, age, sex, race, 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.

The following further analyses will also be described as follows:

- Evaluation of potential effect of nAb on bintrafusp alfa safety (see section 15.1.1)
- Evaluation of potential effect of nAb on bintrafusp alfa efficacy
    - for best overall response according to RECIST 1.1 as adjudicated by IRC (see section 14.1.1)
    - for progression-free survival according to RECIST 1.1 as adjudicated by IRC (see section 14.4.2)
    - for duration of confirmed response according to RECIST 1.1 as adjudicated by IRC (see section 14.5.1)
    - for durable confirmed response according to RECIST 1.1 as adjudicated by IRC (see section 14.6.1)
    - for overall survival (see section 14.8.1)
- Evaluation of potential effect of nAb on bintrafusp alfa Pharmacokinetic (see section 16.1.2)

Potential effect of nAb on safety, efficacy and PK will be evaluated on nAb positive status (ever positive in either assay, never positive).




## 17 References

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

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

# 18 Appendices

## 18.1.1 Appendix 1: Definition of important protocol deviations

See document: *ctp-ms200647-0047-iap-v1-appendix-1.docx*

## 18.1.2 Appendix 2: Rules for identification of previous or concomitant medications/procedures

See document: *ctp-ms200647-0047-iap-v1-appendix-2.docx*

## 18.1.3 Appendix 3: Definition of NCI-CTCAE grading

See document: *CTC V5.0 guidance.xlsx*

# ELECTRONIC SIGNATURES

## Document: ctp-ms200647-0047-iap-v2

| Signed By | Event Name | Meaning of Signature | Server Date (dd-MMM-yyyy HH:mm 'GMT'Z) |
|---|---|---|---|
| PPD | Task Completed (Approval eSign): Approved | Technical Approval | PPD |
| PPD | Task Completed (Approval eSign): Approved | Technical Approval | PPD |
| PPD | Task Completed (Approval eSign): Approved | Business Approval | PPD |