# Integrated Analysis Plan of the Cervical Cancer analysis for the CSR addendum (cutoff date: 14May2021)

| | |
|---|---|
| **Clinical Study Protocol Identification No.** | EMR 200647-001 |
| **Title** | A Phase I, open-label, multiple-ascending dose trial to investigate the safety, tolerability, pharmacokinetics, biological and clinical activity of M7824 (MSB0011359C) in subjects with metastatic or locally advanced solid tumors and expansion to selected indications |
| **Study Phase** | Phase I |
| **Investigational Medicinal Product(s)** | M7824 (MSB0011359C) |
| **Clinical Study Protocol Version** | 31 January 2019 / Version 7.0 |
| **Integrated Analysis Plan Author** | Coordinating Author<br>PPD |
| **Integrated Analysis Plan Date and Version** | 05 Apr 2022 / Version 2.0 |

**Integrated Analysis Plan Reviewers**

| Function | Name |
|---|---|
| PPD | PPD |
| PPD | PPD |
| Director, Biostatistics, Merck Healthcare KGaA | PPD |
| Study Biostatistician, EMD Serono | PPD |
| Biostatistician, Merck Healthcare KGaA | PPD |
| PPD EMD Serono | PPD |
| PPD Merck Healthcare KGaA | PPD |
| PPD Merck Healthcare KGaA | PPD |

# Approval Page

**Integrated Analysis Plan:** EMR 200647-001

A Phase I, open-label, multiple-ascending dose trial to investigate the safety, tolerability, pharmacokinetics, biological and clinical activity of M7824 (MSB0011359C) in subjects with metastatic or locally advanced solid tumors and expansion to selected indications.

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

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

# 1 Table of Contents

# 2 List of Abbreviations and Definition of Terms

| | |
|---|---|
| ADA | anti-drug antibody |
| ADaM | analysis data model |
| AE | adverse event |
| AESI | adverse event of special interest |
| BOR | best overall response |
| cBOR | confirmed best overall response |
| CDISC | clinical data interchange standards consortium |
| $C_{EOI}$ | Concentration observed at the end of infusion |
| CI | confidence interval |
| CR | complete response |
| CSR | clinical study report |
| CTCAE | Common Terminology Criteria for Adverse Events |
| $C_{trough}$ | Concentration observed at the end of the dosing interval |
| CV% | coefficient of variation |
| DR | duration of response |
| ECG | electrocardiogram |
| eCRF | electronic case report form |
| EOT | end of treatment |
| GeoMean | geometric mean |
| GeoCV% | geometric coefficient of variation |
| HAHA | human antihuman antibody |
| HCC | hepatocellular carcinoma |
| IMP | investigational medicinal product |
| IAP | Integrated Analysis Plan |
| irAE | immune-related adverse event |
| IRR | Infusion-Related Reactions |
| LLN | lower limit of normal |
| logStD | standard deviation of log-transformed data |
| Max | miniimum |
| MedDRA | medical dictionary for regulatory activities |
| Min | maximum |
| NA | not applicable |
| nAb | neutralizing antibody |
| NC | not calculated |

| | |
|---|---|
| NCI | national cancer institute |
| NE | not evaluable |
| NR | no result |
| OS | overall survival |
| PD | progressive disease |
| PD-L1 | programmed death ligand 1 |
| PFS | progression-free survival |
| PK | pharmacokinetic(s) |
| PKADA | Pharmacokinetic ADA Analysis Set |
| PKADACC | Pharmacokinetic ADA Analysis Set for Cervical Cancer |
| PKAS | Pharmacokinetic Analysis Set |
| PKASCC | Pharmacokinetic Analysis Set for Cervical Cancer |
| PKNAB | Pharmacokinetic nAb Analysis Set |
| PKNABCC | Pharmacokinetic nAb Analysis Set for Cervical Cancer |
| PR | partial response |
| PT | Preferred term |
| RECIST | response evaluation criteria in solid tumors |
| SAF | safety analysis set |
| SAP | statistical analysis plan |
| SD | stable disease |
| SDTM | study data tabulation model |
| SI | system international |
| SMC | safety monitoring committee |
| SOC | system organ class |
| StD | standard deviation |
| TEAE | treatment-emergent adverse event |
| TGFβ | transforming growth factor-beta |
| TNR | titer no result |
| TTP | time to progression |
| ULN | upper limit of normal |
| WBC | white blood cell |
| WHO-DD | WHO Drug Dictionary |

# 3 Modification History

| Unique Identifier for Version | Date of IAP Version | Author | Changes from the Previous Version |
|---|---|---|---|
| 1.0 | 28MAY2021 | PPD | Not applicable |
| 2.0 | 05APR2022 | PPD | - Cancellation of the CSR addendum and the corresponding statistical analyses<br>- Cancellation of the final analysis after the database lock |

# 4 Purpose of the Integrated Analysis Plan

The purpose of this integrated analysis plan (IAP) is to document technical and detailed specifications for the analyses of data collected for bintrafusp alfa in the dose escalation and expansion phases under protocol EMR 200647-001 to get updated results for a Clinical Study Report (CSR) addendum and to get results more particularly in the participants with cervical cancer.

The laboratory code MSB0011359C was used in EMR200647-001 protocol, M7824 is the company code of the investigational anti-PD-L1/TGFβ Trap (anti-PD-L1/TGFβRII) drug and bintrafusp alfa is the International Nonproprietary Name for M7824. Results of analyses described in this IAP will be included in an addendum to the CSR. The data cutoff date for analyses will be 14 May 2021. Additionally, the planned analyses identified in this IAP will be included in regulatory submissions or future manuscripts. Any post-hoc, or unplanned analyses performed to provide results for inclusion in a CSR or an addendum, but not identified in this prospective IAP, will be clearly identified in the CSR addendum.

# 5 Objectives and Endpoints

This IAP will only address analyses for the purpose of the CSR addendum. The objectives and corresponding endpoints according to the protocol have already been described in the IAP version 1.0 (15 September 2020) for the CSR (02 March 2021).

# 6 Overview of Planned Analyses

This IAP will only address analyses for the purpose of the CSR addendum. Additional statistical analysis plans for Safety Monitoring Committee (SMC) analyses, for the interim and primary analyses by cohort, and for the CSR have been developed separately.

The analyses will be performed separately in 3 sets of outputs for:

1. Participants with cervical cancer will be defined according to the following criteria (CC set of outputs):
    - Participants in Dose Escalation having "CERVIX UTERI" as a tumor type according to the disease history
    - Participants in Dose Expansion being part of the Cervical cohort.
2. All dose levels in dose escalation where participants received other than 1200 mg or 500 mg flat dose, i.e. Dose Escalation cohorts (0.3 mg/kg ->10 mg/kg, 1 mg/kg, 1 mg/kg ->1200 mg, 3 mg/kg, 10 mg/kg, 20 mg/kg, 30 mg/kg, 2400 mg) and HCC 3 mg/kg cohort (DE set of outputs).
3. All cohorts of Dose Expansion receiving 1200 mg or 500 mg flat dose: HCC Dose Ascending Cohort -1200 mg, HCC 2L, Melanoma PD-L1 Fail, NSCLC Biomarker, NSCLC PD-L1 Fail, Pancreatic Adenocarcinoma, Esophageal Adenocarcinoma, Colorectal Carcinoma, Triple Negative Breast Cancer, Glioblastoma, Squamous Cell Carcinoma of Head and Neck, Cervical, NSCLC-2L 500 mg, NSCLC-2L 1200 mg, NSCLC-2L cohorts combined (XP set of outputs).

All the analyses will be performed in participants with cervical cancer (CC set of outputs). This first set of tables and figures will include 3 categories: Cervical Cancer - Dose Escalation, Cervical Cancer - Dose Expansion, Cervical Cancer - Pooled. In the listings, an additional column will be added to specify the information: Cervical Cancer - Dose Escalation, Cervical Cancer - Dose Expansion.

In addition, subject disposition, exposure, ADA/nAb categories and Safety analyses will be presented by cohort and all cohorts combined in the DE set of outputs and in the XP set of outputs.

A data review meeting will be held prior to any database lock. In addition, no database can be locked until this IAP has been approved (except for emergency).

**The analyses described above will be performed based on 14th May 2021 cutoff date.**

## 6.1 Decision after discontinuation of NSCLC and BTC trials

The data review outcome from the 3 randomized controlled studies in NSCLC and BTC (MS200647-0005, MS200647-0037, MS200647_0055) appears to indicate, consistently across 2 indications, either poorer observed hazard ratios for PFS and OS in the experimental arms with bintrafusp alfa or low likelihood for bintrafusp alfa to add benefits compared to standard of care. Based on this, the decision was made to cancel the CSR addendum and the corresponding statistical analyses described in this iAP. The study results have been described in CSR approved on 02Mar2021 based on data cutoff date of 15May2020.

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

There will be no final analysis after the database lock.

# 8 Analysis Populations and Subgroups

## 8.1 Definition of Analysis Populations

The following analysis sets are defined:

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

**Safety Analysis Set (SAF):** All participants who receive at least 1 dose of trial treatment.

**Full Analysis Set (FAS):** All participants who receive at least 1 dose of trial treatment.

**Pharmacokinetic (PK) Analysis Set (PKAS):** All participants who complete at least 1 infusion of investigational medicinal product (IMP), and who provide at least one post-dose sample with measurable concentration of bintrafusp alfa. The Pharmacokinetic ADA Analysis Set (PKADA) is defined as a subpopulation of the PK Analysis Set and restricted to participants who have in addition at least one valid result of ADA at any time point, and the Pharmacokinetic nAb Analysis Set (PKNAB) is defined as a subpopulation of the PKAS and restricted to participants who have in addition at least one valid result of nAb at any time point.

**PK Analysis Set for Cervical Cancer (PKASCC):** All participants diagnosed with cervical cancer who complete at least 1 infusion of IMP, and who provide at least one post-dose sample with measurable concentration of bintrafusp alfa. The Pharmacokinetic ADA Analysis Set for Cervical Cancer (PKADACC) is defined as a subpopulation of the PKASCC and restricted to participants who have in addition at least one valid result of ADA at any time point, and the Pharmacokinetic nAb Analysis Set for Cervical Cancer (PKNABCC) is defined as a subpopulation of the PKASCC and restricted to participants who have in addition at least one valid result of nAb at any time point.

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

The definition of the SAF and the FAS are identical in the non-randomized part of this study; the SAF terminology will be used for the safety analysis and the FAS terminology will be used for efficacy analysis. In the NSCLC second-line cohort where participants are randomized in 500 mg or 1200 mg arms, participants will be analyzed as randomized for efficacy analysis (FAS) and as actually treated for safety analysis (SAF). If a participant receives more than one dose level, the participant will be classified according to the first dose level received.

The following table summarizes the use of the analysis sets in the different analyses.

| | Analysis Population | | |
|---|---|---|---|
| Analyses | Full Analysis Set | Safety Analysis Set | Immunogenicity Set |
| Baseline Characteristics | ✓ | | |
| Compliance and Exposure | | ✓ | |
| Efficacy | ✓ | | |
| [redacted] | | | [redacted] |
| Safety and Tolerability | | ✓ | |

## 8.2 Subgroup Definition and Parameterization

Efficacy will be analyzed by subgroup categories (the category "missing" will not be included in any subgroup analysis):

- Age
  - Age ≥ 18 - < 50 years
  - Age ≥ 50 - < 65 years
  - Age ≥ 65 years
- Histology
  - Squamous cell carcinoma
  - Adenocarcinoma
  - Adenosquamous cell carcinoma
  - Other
- Number of prior treatment lines (metastatic/locally advanced prior anti-cancer drug therapies (without taking into account RT lines alone))
  - 0
  - 1
  - 2
  - 3
  - ≥4
- Prior Bevacizumab use
  - Yes
  - No

- Prior anti-cancer radiotherapy (any radiotherapy will be taken into account)
  - Yes
  - No
- ECOG PS at baseline
  - ECOG PS 0
  - ECOG PS 1
- Race
  - White
  - Black or African American
  - Asian
  - American Indian or Alaska Native
  - Native Hawaiian or Other Pacific Islander
  - Not collected at this site
  - Other
- Pooled Region
  - North America
  - Europe
  - Asia & Pacific
- ADA status
  - Ever positive
  - Never positive

# 9 General Specifications for Data Analyses

In order to provide overall estimates of treatment effects, data will be pooled across sites. The "site" factor will not be considered in statistical models or for subgroup analyses due to the high number of participating sites in contrast to the anticipated small number of participants treated at each site.

All analyses will be performed using SAS® Software version 9.2 or higher.

## 9.1 Data Handling After Cut-off Date

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

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

## 9.2 Definition of Baseline and Change from Baseline

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

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

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

## 9.3 Study Day / Study Treatment Day

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

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

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

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

## 9.5 Conversion Factors

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

1 week = 7 days, 1 month = 30.4375 days, 1 year = 365.25 days.

## 9.6 Date of Last Contact

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

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

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

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

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

## 9.7 Time Window

Not Applicable.

## 9.8 Definition of On-treatment Period

The on-treatment period 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 trial drug administration to the last trial drug 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.

For immune-related AEs as listed in Section 15.2.3.2, an expanded on-treatment period will be used as a default for any analysis: Time from the first trial drug administration to the last trial drug administration date + 90 days, death OR to the earliest date of subsequent anticancer therapy minus 1 day, whichever occurs first, unless otherwise stated.

## 9.9 Imputation of Missing Data

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

In all participant data listings, imputed values will be presented and imputed information will be flagged.

Missing statistics, e.g. when they cannot be calculated, should be presented as "nd". For example, if n=1, the measure of variability (SD) cannot be computed and should be presented as "nd".

| | |
|---|---|
| Age | Incomplete dates (date of informed consent, date of birth) for the calculation of age will be imputed as follows:<br>• 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.<br>• 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. |
| Death | Missing or partial dates of death will be imputed as follows:<br>• If the date is missing, it will be imputed as the last known date to be alive<br>• If the day or month is missing, death will be imputed as the maximum between the last known date to be alive and the imputed date of death where:<br>  • Missing day: 1st day of the month and year of death<br>  • Missing day and month: January 1st of the year of death |
| Adverse events | Incomplete AE-related dates will be imputed as follows:<br>• In case the onset date is missing completely or missing partially but the onset month and year, or the onset year are equal to the start of trial treatment then the onset date will be replaced by the minimum of start of trial treatment and AE resolution date (if not missing).<br>• In all other cases the missing onset day or missing onset month will be replaced by 1.<br>• Incomplete stop dates will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of participant's death. In the later 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. |

## 9.10 Significance Level

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

## 9.11 Presentation of Continuous and Qualitative Variables

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

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

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

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

## 9.12 Reporting Conventions

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

## 9.13 Preferred Term for Analysis of WHO-DD Coded Data

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

## 9.14 Data collected after Reinitiated Treatment

Data collected after reinitiation of treatment will be included in the summary statistics and will be shown in listings.

# 10 Study Participants

The subsections in this section include specifications for reporting participant disposition and study treatment/study discontinuations.

## 10.1 Disposition of Participants and Discontinuations

*Analysis sets: Screening analysis set - 3 different sets of outputs:*

- *Participants included in Cervical Cancer cohorts*
- *Participants included in Dose Escalation/HCC 3 mg/kg cohorts*
- *Participants included in Dose Expansion cohorts*

The following parameters will be summarized for all participants.

- Total number of participants screened overall
- Number of participants who discontinued from the trial prior to first dose of treatment
- Reason for discontinuation of trial prior to first dose of treatment (did not meet eligibility criteria, withdraw of consent, other)
- Number of participants who received no treatment
- Number of participants who received at least 1 dose of trial treatment (Safety Analysis Set)
- Number of participants with treatment on-going
- Number of participants who discontinued the treatment
- Number of reinitiated participants
- Reasons for treatment discontinuation – for reinitiated participants, the reason of treatment discontinuation during the first period will be described
    - Adverse event
    - Lost to follow-up
    - Protocol non-compliance
    - Death
    - Progressive disease
    - Withdrew consent
    - Other

- Number of participants who discontinued the treatment but are still in study for follow-up – this will exclude participants with re-initiated treatment ongoing
- Number of participants who discontinued from the study
- Reasons for study discontinuation
  - Lost to follow-up
  - Death
  - Withdrew consent
  - Other

*Analysis sets: Full analysis set*

The follow-up time till cut-off date will be defined as the time from the date of first dose to the cut-off date. Descriptive statistics will be provided.

## 10.2 Protocol Deviations / Exclusion from Analysis Populations

### 10.2.1 Important Protocol Deviations

Will not be reported.

### 10.2.2 Reasons Leading to the Exclusion from an Analysis Population

Will not be reported.

## 11 Demographics and Other Baseline Characteristics

The analysis of demographic and baseline characteristics of the study population is critical for describing the homogeneity or heterogeneity of the study population between study treatment groups. In general, summaries of the key demographics by study treatment group for the analyzed population are to be provided. If randomization strata or subgroups exist then a summary of the demographics for the subgroup or randomized strata may be relevant to fully characterize the study population.

*Analysis sets: Full analysis set – 1 set of outputs:*

- *Participants included in Cervical Cancer cohorts*

## 11.1 Demographics

Demographic and baseline characteristics will be summarized using the following information:

- Sex: Male, Female
- Race: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Other, Unknown
- Ethnicity: Hispanic or Latino, Not Hispanic or Latino
- Age (years): summary statistics
- Age categories:
  - < 65 years,
  - ≥ 65 years
    - 65-74,
    - 75-84,
    - ≥85 years
- Pooled Region:
  - North America
  - Europe
  - Asia
  - Rest of the World

Demographics will be extracted from the "Demographics" eCRF page.

Specifications for computation:

- Age (years) = (date of informed consent - date of birth + 1) / 365.25
- Investigator site codes will be used for the determination of the participant's geographic region.

## 11.2 Medical History

Will not be reported.

## 11.3 Other Baseline Characteristics

The following characteristics will be tabulated based on data collected at screening:

- PD-L1 expression in tumor cells: >=1%, >=5%, >=10%, >=25%, >=50%, >=80%,
- PD-L1 expression in tumor microenvironment: >=1%, >=5%, >=10%, >=25%, >=50%, >=80%,

- PD-L1 expression in whole tumor: >=1%, >=5%, >=10%, >=25%, >=50%, >=80%.

A listing of the PD-L1 expression at screening will also be provided with the following information: cohort, participant identifier, age, gender, race, confirmed Best Overall Response per RECIST 1.1 as assessed by Investigator, unique sample identifier, visit, collection date, PD-L1 in tumor cells, PD-L1 in tumor microenvironment, PD-L1 in whole tumor.

## 11.4 Prior Anti-cancer Therapy

Will not be reported.

# 12 Previous or Concomitant Medications/Procedures

Will not be reported.

# 13 Study Treatment: Compliance and Exposure

*Analysis Sets: Safety analysis set - 3 different sets of outputs:*

- *Participants included in Cervical Cancer cohorts*
- *Participants included in Dose Escalation/HCC 3 mg/kg cohorts*
- *Participants included in Dose Expansion cohorts*

Participants will receive an IV infusion of bintrafusp alfa over once every 2 weeks until PD has been confirmed by a scan, unacceptable toxicity, or occurrence of any criterion for withdrawal from the trial or the IMP as outlined in the protocol.

Each cycle of bintrafusp alfa is defined by a 2-week period.

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

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

For participants who reinitiate the treatment, the duration of therapy will be the overall treatment duration:

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

A summary table of the duration of therapy (weeks) will be provided, together with the associated listing

# 14 Efficacy Analyses

*Analysis Sets: Full analysis set – 1 set of outputs:*

- *Participants included in Cervical Cancer cohorts*

## 14.1 Best Overall Response

**Best overall response (BOR)** will be assessed based on the tumor response at different evaluation time points from baseline until the first documented disease progression for RECIST. Only tumor assessments performed before the start of any further anticancer treatment will be considered in the assessment of BOR. Clinical deterioration will not be considered as documented disease progression.

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

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

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

The confirmed BOR according to data collected in the eCRF for RECIST 1.1 response criteria as assessed by the investigator will be defined according to the following rules:

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

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

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

When evaluating OR, both CR and PR must be confirmed by repeated assessments performed no less than 4 weeks after the criteria for response are first met.

The confirmed ORR will be calculated for each cohort (CC dose escalation and CC expansion) and overall (all CC pooled together) 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).

In addition, the number and percentage of participants with confirmed BOR of CR, PR, SD, PD, and NE will be tabulated.

The number and percentage of participants with confirmed DCR will be tabulated together with a two-sided 95% CI.

A listing will present the tumor assessment and overall response per RECIST 1.1 as assessed by the investigator including: cohort, participant identifier, age, sex, race, unconfirmed and confirmed BOR, 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, sorted by cohort, participant identifier, and visit.

The individual percentage of change in the sum of diameters since baseline will be displayed over time per cohort on a spider plot, together with the first occurrence of new lesion and 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 for each participant per cohort on a waterfall graph. The sum of diameters includes all target lesions (longest diameter for non-nodal lesions and short axis for nodal target lesions). 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.

An additional waterfall plot will be created in which the PD-L1 expression at Screening (% in Tumor Cells, % in Tumor Microenvironment, % in Whole Tumor) will be presented; a different color will be used to display each of the values for the confirmed BOR as assessed by the investigator.

### 14.1.2 Subgroup Analysis for Best Overall Response According to RECIST 1.1 as Assessed by Investigator

Best Overall Response based on confirmed responses, ORR and DCR according to RECIST 1.1 as adjudicated by Investigator will also be evaluated in on all subgroups defined in 8.2. They will be calculated for each cohort (CC dose escalation and CC expansion) and overall (all CC pooled together) with a two-sided 95% CI.

## 14.2 Duration of Response

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

DR will be performed among participants of the Full Analysis Set who met confirmed CR/PR according to RECIST 1.1 as assessed by the Investigator

DR will be censored in the following scenarios:

- Participants who do not experience an event (PD or death) will be right-censored on the date of their last evaluable (non-missing and non "Not Evaluable") tumor assessment.
- If death without previously documented PD is observed after more than 84 days (12 weeks) during the first year of follow-up, or after more than 168 days (24 weeks) during the following years of follow-up of last tumor assessment, the patient will be right-censored at the date of the last evaluable tumor assessment.
- Participants who start new anti-cancer treatment prior to an event will be censored on the date of the last evaluable tumor assessment before anti-cancer therapy is given.
- Participants with an event after two or more missing tumor assessments will be censored on the date of the last evaluable tumor assessment.
- Participants who do not have a baseline tumor assessment or who do not have any post-baseline tumor assessments will be censored on the date of first dose unless death occurred on or before the time of the second planned tumor assessment in which case the death will be considered an event.

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

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

The time and duration of response per participant will also be displayed in a swimmer graph in participants with unconfirmed/confirmed CR/PR. A participant listing will provide the following information: cohort, participant identifier, age, sex, race, date of first response, duration of response, date of last tumor assessment, censoring reason.

## 14.3 Progression Free Survival

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

The analysis of PFS will be performed according to RECIST 1.1 assessed by the Investigator.

PFS will be censored in the following scenarios:

- Participants who do not experience an event (PD or death) will be right-censored on the date of the last evaluable (non missing and non "Not Evaluable") tumor assessment.
- If death without previously documented PD is observed after more than 84 days (12 weeks) during the first year of follow-up, or after more than 168 days (24 weeks) during the following years of follow-up of last tumor assessment, the patient will be right-censored at the date of the last evaluable tumor assessment.
- Participants who start new anti-cancer treatment prior to an event will be censored on the date of the last evaluable tumor assessment before anti-cancer therapy is given.
- Participants with an event after two or more subsequent missing response assessments (i.e. 2 times the scheduled time interval between two subsequent response assessments) will be censored on the date of the last evaluable tumor assessment.
- Participants who do not have a baseline tumor assessment or who do not have any post-baseline tumor assessments will be censored on the date of first dose unless death occurred on or before the time of the second planned tumor assessment in which case the death will be considered an event.

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

The date of PFS event / censoring will be defined in Table 1.

**Table 1 Progression-free Survival Event / Censoring**

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

PFS: progression-free survival

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

A participant listing will provide the following information: cohort, participant identifier, age, sex, race, date of first administration, date of last tumor assessment, date of event/censoring, event/censoring reason, time to event.

## 14.4 Overall Survival

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

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

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

The date of event / censoring is defined in Table 2.

**Table 2 Survival Event / Censoring**

| | Date of event / censoring | Censoring |
|---|---|---|
| Participants alive or lost to follow-up before or at cut-off date | Last date known to be alive (as defined in Section 9.6) | Yes |
| Participants who died before or at cut-off date | Date of death | No |

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

A participant listing will provide the following information: cohort, participant identifier, age, sex, race, date of first administration, date of last tumor assessment, date of event/censoring, event/censoring reason, time to event.

## 14.5 Immunogenicity

*Analysis Sets: Immunogenicity analysis set - 3 different sets of outputs:*

- *Participants included in Cervical Cancer cohorts*
- *Participants included in Dose Escalation/HCC 3 mg/kg cohorts*
- *Participants included in Dose Expansion cohorts*

Anti-drug antibody (ADA) will be assessed before the trial treatment start, and on Days 15, 29, 43, 85 after the start of infusion, 6-weekly up to/including Week 25, every 12 weeks after Week 25, and during safety follow-up (the term HAHA is used in the eCRF). Samples collected after the on-treatment period (e.g. safety follow-up) will be included in the analysis as well. If the sample

is positive for ADA, it will be re-analyzed to determine the titer. The ADA results will be derived based on the algorithm in Table 3.

**Table 3 Algorithm for the Derivation of ADA Results**

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

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

Negative, number, or positive-TNR are valid results while number and positive-TNR are considered as positive. Participants will be characterized into different categories based on the criteria in Table 4.

**Table 4 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 (a single positive evaluation, or duration between first and last positive result <16 weeks) and last assessment not positive. | Number of participants with at least one valid post-baseline result and without positive baseline result (including 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) |

Samples with a reportable ADA titer will also be tested in the two nAb assays, PD-L1 and TGF beta. The nAb results will be derived based on the algorithm in Table 5. Participants will be characterized into different nAb categories based on the criteria in Table 6.

**Table 5 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 6 Participants Characterized based on nAb Results**

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

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

### Start of Immunogenicity Response (Seroconversion)

For participants with any positive ADA/nAb 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.

For ADA/nAb results, the following analysis will be described in each cohort and in all cohorts combined (the titer summaries will be provided if the titer data is available):

- The frequency and percentage of each ADA/nAb 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

A listing of all individual ADA results from ever positive participants will be prepared by time point. The listing will include cohort, participant identifier, age, gender, race, ADA categories, date of assessment and results of screening, confirmatory and titer values.
A listing of nAb results from ever positive participants (in either assay) will be provided with the following: cohort, participant identifier, age, sex, race, assay (PD-L1 or TGFβ), nAb categories status, visit, date of assessment and results of screening, titer values.

### 14.5.1 Evaluation of Potential Effect of ADA/nAb on bintrafusp alfa Efficacy

Potential effect of ADA on confirmed BOR, PFS and OS will be evaluated on:

- ADA positive status: ever positive vs never positive

A listing of efficacy assessments for Investigator assessment, and immunology parameters for ADA positive participants will be prepared including: cohort, participant identifier, age, sex, ADA and nAb categories, date of first study treatment, treatment duration, cBOR, PFS, time to response, DR and OS. Participants with treatment ongoing or response ongoing will be flagged.

# 15 Safety Analyses

The safety evaluation will be performed by cohort and all cohorts combined (where applicable).

*Analysis Sets: Safety analysis set - 3 different sets of outputs:*

- *Participants included in Cervical Cancer cohorts*
- *Participants included in Dose Escalation/HCC 3 mg/kg cohorts*
- *Participants included in Dose Expansion cohorts*

## 15.1 Adverse Events

The safety evaluation will be performed by cohort and all cohorts combined (where applicable).

*Analysis Sets: Safety analysis set*

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 as defined in Section **Error! Reference source not found.**. Adverse events will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA) at the time of data cut-off.

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

All analyses described in Section **Error! Reference source not found.** will be based on TEAEs (started during the on-treatment period) unless otherwise specified. The AE listings will include all AEs (whether treatment-emergent or not). AEs outside the on-treatment period will be flagged in the listings. AE occurring during the reinitiation phase will be considered in the summary tables and will be flagged in the listings.

- **Related Adverse Events:** adverse events with relationship to study treatment (as recorded on the AE eCRF page, Relationship with study treatment = Related) reported by the investigator and those of unknown relationship (i.e. no answer to the question "Relationship with study treatment").
- **Serious Adverse Events (SAE):** serious adverse events (as recorded on the AE eCRF page, Serious Adverse Event = Yes).

- **Adverse Events Leading to Treatment Discontinuation:** adverse events leading to permanent discontinuation of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Drug withdrawn).
- **Adverse Event Leading to Death:** adverse event leading to death (as recorded on the AE eCRF page, Outcome = Fatal, as well as AEs of Grade 5).
- **Adverse Events of Special Interest (AESI):** Adverse events of special interest will be identified according to a pre-specified search list of MedDRA Preferred Terms. Categories of AESI include:
    - Infusion-Related Reactions including hypersensitivity: any infusion-related reaction or hypersensitivity (regardless of grade) (see description in 15.2.3.1)
    - Immune Related Adverse Events (irAE) including autoimmune disorders (see description in 15.2.3.2)
    - Skin AE possibly related to TGFβ inhibition (see description in 15.2.3.3)
    - Anemia
- **Bleeding events** are those AEs belonging to the MedDRA SMQ Haemorrhage terms (excluding laboratory terms).
- **Anemia events** are those AEs belonging to the MedDRA HLT Anaemias NEC, HLT Anaemias haemolytic immune, HLT Anaemias haemolytic NEC or PT = Haemoglobin decreased.

### 15.1.1 All Adverse Events

A table presenting the overall summary of AEs will be presented by cohort and all cohorts combined (where applicable) with the following information:

- TEAEs
- Related TEAEs
- Serious TEAEs
- Related serious TEAEs
- TEAEs, grade $\geq 3$
- TEAEs, grade $\geq 4$
- Related TEAEs, grade $\geq 3$
- Related TEAEs, grade $\geq 4$
- TEAEs leading to death
- Related TEAEs leading to death
- TEAEs of special interest (all categories listed)
- Related TEAEs of special interest (all categories listed)
- TEAEs leading to study termination

- Related TEAEs leading to study termination
- Bleeding TEAE
- Related bleeding TEAE

A listing including all adverse events, and a listing including adverse events with onset date after the CSR cutoff (i.e. between 15MAY 2020 and 14MAY 2021) will be provided and will contain the following information: cohort, participant identifier, age, gender, race, preferred term, reported term for the adverse event, start date, end date, duration of AE (in days), day relative to the first infusion, day relative to the most recent infusion prior to AE onset, relationship to study treatment, toxicity grade, action(s) taken, outcome, seriousness (Y/N), AESI infusion-related (Y/N), AESI immune-related (Y/N), medication administered. The listing will be sorted by cohort, participant identifier, and start date.

### 15.1.2 Adverse Events Leading to Discontinuation of Study Treatment

Will not be reported.

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

### 15.2.1 Deaths

Will not be reported.

### 15.2.2 Serious Adverse Events

Will not be reported.

## 15.2.3 Other Adverse Events of Interest

### 15.2.3.1 Infusion Related Reaction

Infusion Related Reactions (IRR) are defined as adverse events with PTs according to a pre-specified MedDRA search list, which started on the study drug dosing date (not prior to the infusion of study drug) or the following day of study drug infusion. IRR can be split into Reactions, and Signs and symptoms subgroups as follows:

**Table 7: Criteria for Infusion-Related Reactions**

| | |
|---|---|
| Infusion related reactions | Reactions – considered when onset is on the day of bintrafusp alfa infusion (during or after the infusion) or the day after then bintrafusp alfa infusion (irrespective of resolution date):<br>• Infusion related reaction<br>• Drug hypersensitivity<br>• Anaphylactic reaction<br>• Hypersensitivity<br>• Type I hypersensitivity<br><br>Signs and Symptoms – occurring on the day of bintrafusp alfa infusion (during or after the infusion) and resolved with end date the day of bintrafusp alfa or the day after:<br>• Pyrexia<br>• Chills<br>• Flushing<br>• Hypotension<br>• Dyspnea<br>• Wheezing<br>• Back pain<br>• Abdominal pain<br>• Urticaria |

### 15.2.3.2 Immune Related Adverse Events

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

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

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

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

4) On the imAE eCRF page, either:

a. The question "Does any of the following provide a clear etiology for the event?" has the answer "No" selected, indicating that the AE is not attributable to underlying cancer disease/PD, prior or concomitant medications/procedures, nor another medical condition such as an infection or pre-existing disease.

OR

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

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

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

#### 15.2.3.3 Potential TGFβ-mediated Skin TEAE

Skin AE possibly related to TGFβ inhibition will be selected based on MedDRA PTs according to a pre-specified MedDRA search list:

**Narrow definition:**

- Keratoacanthoma
- Squamous cell carcinoma of skin

**Broad definition** has additional PTs:

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

## 15.3 Clinical Laboratory Evaluation

### 15.3.1 Hematology and Chemistry Parameters

Baseline Laboratory Assessments are defined as the last non-missing observation prior to first dose of trial treatment, while Treatment Emergent Laboratory Assessments are any sample collected after 1st drug administration and within the on-treatment period as defined in Section 9.8. Laboratory values (including corresponding normal ranges) converted in standard unit will be used for listings. Laboratory results will be classified according to the NCI-CTCAE criteria version 4.03. Non-numerical qualifiers (with exception of fasting flags) will not be taken into consideration in the derivation of CTCAE criteria (e.g., hypokalemia Grade 1 and Grade 2 are only distinguished by a non-numerical qualifier and therefore Grade 2 will not be derived, similar for Grade 1 and Grade 3 hyperuricemia where Grade 3 will not be derived). Additional laboratory results that are not part of NCI-CTCAE will be categorized as follows: below normal limits, within normal limits, and above normal limits.

**Gradable parameters**

Gradable parameters to analyze with their respective NCI-CTC name and direction of abnormality are provided in Table 8. For parameters which are graded with both low and high values such as potassium (hypokalemia/hyperkalemia), the toxicities will be summarized separately. Low direction toxicity (e.g. hypokalemia) grades at baseline and post-baseline will be set to 0 when the variables are derived for summarizing high direction toxicity (e.g. hyperkalemia), and vice versa.

**Table 8 NCI-CTC Gradable parameters**

| Parameter (LBTEST) | Parameter code (LBTESTCD) | Name in NCI-CTC | Direction of abnormality |
|---|---|---|---|
| **Biochemistry** | | | |
| Alanine Aminotransferase | ALT | Alanine aminotransferase increased | High |
| Albumin | ALB | Hypoalbuminemia | Low |
| Alkaline Phosphatase | ALP | Alkaline phosphatase increased | High |
| Amylase | AMYLASE | Serum amylase increased | High |
| Aspartate Aminotransferase | AST | Aspartate aminotransferase increased | High |
| Bilirubin | BILI | Blood bilirubin increased | High |
| Calcium | CA | Hypercalcemia/Hypocalcemia | High/Low |
| Cholesterol | CHOL | Cholesterol high | High |
| Creatinine | CREAT | Creatinine increased | High |
| Gamma Glutamyl Transferase | GGT | GGT increased | High |
| Glucose | GLUC | Hyperglycemia/Hypoglycemia | High/Low |
| Haptoglobin[1] | HAPTOG | Haptoglobin decreased | Low |
| Lipase | LIPASET | Lipase increased | High |
| Magnesium | MG | Hypermagnesemia/Hypomagnesemia | High/Low |
| Phosphate | PHOS | Hypophosphatemia | Low |

| Parameter (LBTEST) | Parameter code (LBTESTCD) | Name in NCI-CTC | Direction of abnormality |
|---|---|---|---|
| Potassium | K | Hyperkalemia/Hypokalemia | High/Low |
| Sodium | SODIUM | Hypernatremia/Hyponatremia | High/Low |
| Triglycerides | TRIG | Hypertriglyceridemia | High |
| Urate | URATE | Hyperuricemia | High |
| **Hematology** | | | |
| Activated Partial Thromboplastin Time | APTT | Activated partial thromboplastin time prolonged | High |
| Hemoglobin | HGB | Anemia/Hemoglobin increased | Low/High |
| Leukocytes | WBC | Leukocytosis/White blood cell decreased | High/Low |
| Lymphocytes | LYM | Lymphocyte count decreased/Lymphocyte count increased | High/Low |
| Neutrophils | NEUT | Neutrophil count decreased | Low |
| Platelets | PLAT | Platelet count decreased | Low |
| Prothrombin Intl. Normalized Ratio | INR | INR increased | High |

[1] Haptoglobin is collected only in case of episode of anemia. It will be described in listings only.

For **WBC differential counts** (neutrophil, lymphocyte counts), the absolute value will be used when reported. When only percentages are available (this is mainly important for neutrophils and lymphocytes, because the CTCAE grading is based on the absolute counts), the absolute value is derived as follows:

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

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

- Lymphocyte count decreased:
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % Lower Limit of Normal (LLN) value, and
  - derived absolute count ≥ 800/mm3
- Neutrophil count decreased:
  - derived absolute count does not meet Grade 2-4 criteria, and
  - % value < % LLN value, and
  - derived absolute count ≥ 1500/mm3

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

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

**Non-gradable parameters**

Non-gradable parameters are provided in Table 9.

**Table 9 Non-NCI-CTC Gradable Parameters**

| Parameter (LBTEST) | Parameter code (LBTESTCD) |
|---|---|
| **Biochemistry** | |
| Blood Urea Nitrogen | BUN |
| C Reactive Protein | CRP |
| Chloride | CL |
| Creatine Kinase | CK |
| Creatinine Clearance[1] | CREATCLR |
| Indirect Bilirubin[2] | BILIND |
| Lactate Dehydrogenase | LDH |
| Protein | PROT |
| Urea[3] | UREA |
| **Hematology** | |
| Basophils/Leukocytes | BASOLE |
| Eosinophils/Leukocytes | EOSLE |
| Ery. Mean Corpuscular HGB Concentration | MCHC |
| Ery. Mean Corpuscular Hemoglobin | MCH |
| Ery. Mean Corpuscular Volume | MCV |
| Erythrocytes | RBC |
| Hematocrit | HCT |
| Monocytes/Leukocytes | MONOLE |
| Prothrombin Time | PT |
| Reticulocytes/Erythrocytes | RETIRBC |

[1] Creatinine clearance is only collected at screening to check inclusion criteria. It will be presented in listings only.

[2] Indirect bilirubin is not measured but obtained by site calculation (=Total bilirubin – Direct bilirubin). It will be presented in listings only.

[3] Total urea is collected in Europe only (in the US, Blood Urea Nitrogen is collected instead).

Separate listings of Hematology and Biochemistry with onset date after the CSR cutoff (i.e. between 15MAY 2020 and 14MAY 2021) will be created. Each listing will include assigned cohort, participant identifier, age, gender, race, first dose date, last dose date, dose, number of doses, laboratory parameter (units), visit, date, SI value, LLN, ULN, Indicator of Normal Range (Low, Normal, High), and toxicity grade. These listings will be sorted by cohort, participant identifier, and laboratory measurement date.

## 15.4 Vital Signs

Will not be reported.

## 15.5 Other Safety or Tolerability Evaluations

Will not be reported.

# 16 Analyses of Other Endpoints

## 16.1 Pharmacokinetics

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

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

### 16.1.1 Missing PK Data

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

Pharmacokinetic concentrations below the lower limit of quantification (<LLOQ) will be taken as zero for descriptive statistics.

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

Concentrations will be set to missing in summary tables if the value is reported as no result. Pharmacokinetic concentrations which are erroneous due to a protocol violation (as defined in the clinical trial 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 relevant listing/table and/or Clinical Trial Report (CTR).

**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/Pharmacodynamics Data Analyst must not be excluded from the analysis. Any implausible data will be documented in the relevant listing/table and/or CTR.

If a PK parameter cannot be derived from a subject'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 subject discontinues the treatment). For statistical analyses, PK parameters coded as NC will be set to missing.

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

### 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 SAF. Concentrations will be reported with the same precision as the source data.

Tables:

Pharmacokinetic concentrations will be presented in tables and descriptively summarized by treatment and cohort (as appropriate) separately, day, and nominal time using: n, arithmetic mean (Mean), StD, coefficient of variation (CV%), Min, median (Median), Max, n of log-transformed data (n,Geo), geometric mean (GeoMean), StD of log-transformed data (logStD), and geometric coefficient of variation (GeoCV%). Summaries will be based on the PKAS and separately PKASCC by dose.

Additional table(s) will summarize PK concentrations with further stratification by ADA subsets ever positive and never positive, based on PKADACC by dose. Additional table(s) will summarize PK concentrations with further sub-stratification by nAb subsets ever positive and never positive, based on PKNABCC by dose. For nAb subsets, summaries will be done for ever positive in either of 2 nAb assays (PD-L1 and TGFß receptor neutralization; PD-L1+/TGFß+, PD-L1+/TGFß–, PD-L1–/TGFß+) versus never positive (PD-L1–/TGFß–). Additional table(s) will summarize PK concentrations with stratification by Asian versus non-Asian, based on PKASCC by dose. Only subgroup sample size with a minimal 3 subjects will be displayed.

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

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

Subjects for whom the Week 1 dose amount differs from the subsequent dose will have the relevant Week 1 data excluded from statistics.

Figures:

No concentration figures are planned to be produced.

### 16.1.3 Pharmacokinetic Parameter Analysis

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

| | |
|---|---|
| $C_{trough}$ | The concentration observed at the end of the dosing interval, before next dosing (serum trough concentration) |
| $C_{EOI}$ | The concentration observed at the end of infusion |

For subjects in the PKASCC, accumulationAccumulation ratios for $C_{trough}$ ($C_{trough}$ Day X / $C_{trough}$ Day 15) and $C_{EOI}$ ($C_{EOI}$ Day X / $C_{EOI}$ Day 1) will be derived to assess time-dependent PK effects.

**Presentation of PK Parameter Data**

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

Tables:

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

Additional table(s) will summarize $C_{trough}$ and $C_{EOI}$ with further stratification by ADA subsets ever positive and never positive, based on PKADACC for 1200 mg dose. Additional table(s) will summarize $C_{trough}$ and $C_{EOI}$ with further sub-stratification by nAb subsets ever positive and never positive, based on PKNABCC for 1200 mg dose.. For nAb subsets, summaries will be done for ever positive in either of 2 nAb assays (PD-L1 and TGFß receptor neutralization; PD-L1+/TGFß+, PD-L1+/TGFß−, PD-L1−/TGFß+) versus never positive (PD-L1−/TGFß−). Additional table(s) will summarize $C_{trough}$ and $C_{EOI}$ with stratification by Asian versus non-Asian, based on PKASCC for 1200 mg dose. 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, Asian versus non-Asian, based on PKADACC for 1200 mg dose.. For nAb ever-positive subjects, serum bintrafusp alfa $C_{trough}$ will be descriptively summarized in additional table(s) for nAb status subgroups (positive in any of 2 nAb assays), based on PKNABCC 1200 mg dose group. Additional table(s) will summarize $C_{trough}$ of ADA Treatment-emergent subjects and nAb Treatment-emergent subjects by PK day relative to day of seroconversion, based on PKADACC and PKNABCC, respectively (all subjects with CC). Only subgroup sample size with a minimal 3 subjects will be

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

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

Subjects for whom the first dose amount differs from the subsequent dose will have the relevant first dose data excluded from statistics.

Figures:

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

Arithmetic mean $C_{trough}$ ±StD will be plotted versus nominal day on a linear scale, based on the PKAS and separately PKASCC by dose. Arithmetic mean $C_{EOI}$ ±StD will be plotted versus nominal day on a linear scale, based on the PKASCC by dose.. Additional figures will present Mean $C_{trough}$ and $C_{EOI}$ with further stratification by ADA subsets ever positive and never positive, based on PKADACC 1200 mg. Additional figures will present Mean $C_{trough}$ and $C_{EOI}$ with further sub-stratification by nAb subsets ever positive (either assay) and never positive, based on PKNABCC 1200 mg. Additional figures will present Mean $C_{trough}$ and $C_{EOI}$ with stratification by Asian versus non-Asian, based on PKASCC for 1200 mg. 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 relative PK day for the PKADACC by dose (for readability, split further into groups of 10 subjects or fewer as needed). Box plots will be prepared for $C_{trough}$ versus relative PK day for the PKADACC by dose.

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 relative PK day for the PKNABCC by dose (for readability, split further into groups of 10 subjects or fewer as needed). Box plots will be prepared for $C_{trough}$ versus relative PK day for the PKNABCC by dose. .

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

CCI [redacted]

[redacted]

## 16.3 Patient Reported Outcome

Will not be reported.

## 17 References

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

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

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

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

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

National Cancer Institute. NCI dictionary of cancer terms – second-line therapy. https://www.cancer.gov/publications/dictionaries/cancer-terms?CdrID=346513 (accessed 29 March 2017).

## 18 Appendices

Not applicable.

Signature Page for VV-CLIN-290447 v1.0

| | |
|---|---|
| Approval | PPD |

Signature Page for VV-CLIN-290447 v1.0