Integrated Analysis Plan

Clinical Study Protocol Identification No.	MS200647_0017
Title	A Phase II, Multicenter, Open Label Study of Bintrafusp alfa (M7824) Monotherapy in Participants with Advanced, Unresectable Cervical Cancer with Disease Progression During or After Platinum-Containing Chemotherapy
Study Phase	Phase II
Investigational Medicinal Product(s)	Bintrafusp Alfa (M7824)
Clinical Study Protocol Version	22 June 2021/ Version 3.0
Integrated Analysis Plan Author	PPD
Integrated Analysis Plan Date and Version	21 February2022/ Version 4.0
Integrated Analysis Plan Reviewers	**Function** **Name** PPD

Approval Page

Integrated Analysis Plan: MS200647_0017

A Phase II, Multicenter, Open Label Study of Bintrafusp alfa (M7824) Monotherapy in Participants with Advanced, Unresectable Cervical Cancer with Disease Progression During or After Platinum-Containing Chemotherapy

Approval of the IAP by all Merck 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

Approval Page 3
1 Table of Contents.....4
2 List of Abbreviations and Definition of Terms 8
3 Modification History 13
4 Purpose of the Integrated Analysis Plan.....15
5 Objectives and Endpoints 15
6 Overview of Planned Analyses.....16
6.1 Overview of Planned Analyses after discontinuation of NSCLC and BTC trials.....18
7 Changes to the Planned Analyses in the Clinical Study Protocol 20
7.1 COVID-19 Impact 20
8 Analysis Populations and Subgroups.....21
8.1 Definition of Analysis Populations.....21
8.2 Subgroup Definition and Parameterization 22
9 General Specifications for Data Analyses 26
9.1 Data Handling After Cut-off Date 27
9.2 Definition of Baseline and Change from Baseline 27
9.3 Study Day / Study Treatment Day.....27
9.4 Definition of Duration and 'time since' Variables 27
9.5 Conversion Factors 28
9.6 Date of Last Contact 28
9.7 Time Window 28
9.8 Definition of On-treatment Period.....29
9.9 Imputation of Missing Data 30
9.10 Scoring of HRQOL Data 35
CCI [redacted]
9.11 Other Derivations.....36

10 Study Participants 38
10.1 Disposition of Participants and Discontinuations 38
10.2 Protocol Deviations / Exclusion from Analysis Populations 39
10.2.1 Important Protocol Deviations 39
10.2.2 Reasons Leading to the Exclusion from an Analysis Population 40
11 Demographics and Other Baseline Characteristics 40
11.1 Demographics 40
11.2 Medical History 41
11.3 Other Baseline Characteristics 42
11.3.1 Disease History 42
11.3.2 HPV test 42
11.3.3 PD-L1 test 43
11.3.4 Vital signs at Baseline 43
11.3.5 ECOG Performance Status at Baseline 44
11.3.6 Skin status history 44
CCI [redacted]
12 Previous or Concomitant Medications/Procedures 45
12.1 Previous and concomitant medications 45
12.2 Premedications prior to bintrafusp alfa administration 45
12.3 Concurrent procedures 46
12.4 Previous anticancer treatments and procedures 46
12.5 Previous pre-cancerous procedures of the cervix 47
12.6 Subsequent anticancer treatments and procedures 47
13 Study Treatment: Compliance and Exposure 48
14 Efficacy Analyses 50
14.1 Primary Endpoint: Confirmed objective response according to RECIST 1.1 as adjudicated by IRC 50
14.1.1 Primary Objective: Objective Response Rate 50
14.2 Further Efficacy Endpoints 54
14.2.1 Confirmed objective response according to RECIST 1.1 as assessed by Investigator (Secondary endpoint) 54
CCI [redacted]

CCI [redacted]
14.2.4 Progression-Free Survival (Secondary Endpoint) 57
CCI [redacted]
15 Safety Analyses 64
15.1 Adverse Events 64
15.1.1 All Adverse Events 66
15.1.2 Adverse Events Leading to Discontinuation of Study Treatment 68
15.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events 69
15.2.1 Deaths 69
15.2.2 Serious Adverse Events 69
15.2.3 Adverse Events of Special Interest 70
15.2.3.1 Infusion-Related Reaction including Immediate Hypersensitivity.....70
CCI [redacted]
15.2.3.3 TGF-β inhibition mediated skin adverse events 72
15.2.3.4 Anemia.....73
15.2.3.5 Bleeding events.....73
15.3 Clinical Laboratory Evaluation.....73
15.4 Vital Signs 79
15.5 Other Safety or Tolerability Evaluations 79

16 Analyses of Other Endpoints 80
16.1 Pharmacokinetics 80
16.1.1 Missing/non-quantifiable Pharmacokinetic Data Handling....80
16.1.2 Descriptive Pharmacokinetic Analysis 81
CCI [redacted]
16.3.1.2 Summary Statistics 86
16.3.1.3 Longitudinal analysis of change from baseline 87
16.4 Immunogenicity 88
16.4.1.1 Antidrug Antibody 88
16.4.2 Neutralizing Anti-drug Antibody 91
CCI [redacted]
16.5.1 PD-L1 Expression (Secondary Endpoint) 93
CCI [redacted]
17 References....95
18 Appendices 96
18.1.1 Appendix 1: Definition of important protocol deviations 96
18.1.2 Appendix 2: Definition of NCI-CTCAE grading 96
CCI [redacted]

2 List of Abbreviations and Definition of Terms

Abbreviation	Definition
2L	Second-line
aCSR	abbreviated CSR
ADA	Antidrug Antibody
ADaM	Analysis Data Model
AEs	Adverse Events
AESI	Adverse Events of Special Interest
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ANC	Absolute Neutrophil Count
anti-PD-L1	Anti-programmed Death-ligand 1
aPTT	Activated Partial Thromboplastin Time
ART	Antiretroviral Therapy
AST	Aspartate Aminotransferase
CCI [redacted]	
BMI	Body Mass Index
BTC	Biliary Tract Cancer
CBC	Complete Blood Count
CCRT	Concurrent Chemoradiation Therapy
CDISC	Clinical Data Interchange Standards Consortium
C_{EOI}	Concentration observed immediately at the end of infusion
CI	Confidence Interval
COVID-19	Coronavirus disease 2019
CNS	Central Nervous System
CPS	Combined Positive Score
CR	Complete Response
CrCL	Creatinine Clearance
CROs	Contract Research Organizations
cSCC	Cutaneous Squamous Cell Carcinoma
CSR	Clinical Study Report

CCI [redacted]	
C_{trough}	Concentration observed immediately before next dosing (corresponding to pre-dose or trough concentration for multiple dosing)
CV%	Coefficient of Variation
D	Day
DNA	Deoxyribonucleic Acid
DOR	Duration of response
DRR	Durable Response Rate
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Forms
CCI [redacted]	
FDA	Food and Drug Administration
FA	Final Analysis
FAS	Full Analysis Set
FT4	Free Thyroxine
GCP	Good Clinical Practice
GeoMean	Geometric Mean
GeoCV%	Geometric Coefficient of Variation
GI	Gastrointestinal
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HgB	Hemoglobin

HIV	Human Immunodeficiency Virus
HPV	Human Papilloma Virus
CCI	[redacted]
IA	Interim Analysis
IAP	Integrated Analysis Plan
IB	Investigator's Brochure
ICF	Informed Consent Form
ICH	International Council for Harmonization
IEC	Independent Ethic Committee
IMM	Immunogenicity Analysis Set
CCI	[redacted]
IRB	Independent Review Board
IRC	Independent Review Committee
IRR	Infusion-related Reaction
irRECIST	Immune-related RECIST
IV	Intravenous
KA	Keratoacanthoma
LDH	Lactate Dehydrogenase
LFT	Liver Function Test
logStD	Standard deviation of log-transformed data
Max	Maximum
Mean	Arithmetic Mean
Med	Median
MedDRA	Medical Dictionary for Regulatory Activities
Min	Minimum
MRI	Magnetic Resonance Imaging
CCI	[redacted]
N	Number of Participants
n	Number of Participants with Non-missing Values
nAb	Neutralizing anti-drug antibody
NC	Not Calculated

NCCN	National Comprehensive Cancer Network
NCI-CTCAE	National Cancer Institute – Common Terminology Criteria for Adverse Event
NE	Not Evaluable
NSCLC	Non-Small Cell Lung Cancer
NSAID	Nonsteroidal Anti-inflammatory Drug
ORR	Objective Response Rate
OS	Overall Survival
PA	Primary Analysis
PAI-1	Plasminogen Activator Inhibitor-1
PD	Progressive Disease
PD-1	Programmed Death-1
PD-L1	Programmed Death-ligand 1
PFS	Progression-free Survival
CCI [redacted]	
PK	Pharmacokinetic
PKAS	Pharmacokinetic Analysis Set
PKADA	Pharmacokinetic ADA Analysis Set
PKNAB	Pharmacokinetic nAb Analysis Set
PR	Partial Response
CCI [redacted]	[redacted]
PS	Performance Status
PT	Prothrombin Time/Preferred Term
Q1	25th Percentile
Q3	75th Percentile
CCI [redacted]	
RECIST 1.1	Response Evaluation Criteria in Solid Tumors Version 1.1
RoW	Rest of the World
RP2D	Recommended Phase II Dose
SAE	Serious Adverse Event

SAF	Safety Analysis Set
SCR	Screening (Analysis Population)
SD	Stable Disease
SDTM	Study Data Tabulation Model
SI	International System of Units
SoA	Summary of Activities
SOC	System Organ Class
StD	Standard deviation
SUSARs	Suspected Unexpected Serious Adverse Reactions
TEAEs	Treatment-Emergent Adverse Events
CCI [redacted]	
TSH	Thyroid-stimulating Hormone
TTP	Thrombotic Thrombocytopenic Purpura
ULN	Upper Limit of Normal
V	Visit
VAS	Visual Analog Scale
vCPS	Visual element Combined Positive Score (Ventana)
W	Week
W1D1	Week 1 Day 1
W3	Week 3
WHO-DD	WHO Drug Dictionary
WOCBP	Woman of Childbearing Potential
β-HCG	β-human Chorionic Gonadotropin

3 Modification History

Unique Identifier for Version	Date of IAP Version	Author	Changes from the Previous Version
1.0	20May2020	PPD	NA
2.0	10December2020	PPD	CCI - Update in section 15.3: the shift tables are made on on-treatment toxicity grade (and not on-study toxicity grade) - Section 15.5: added listing for non-protocol Related Hospital Visits which was missing in V1 - Section 15.2.4.3: added two new terms to the Broad Definition PT list - Section 15: update Treatment-related Anemia to report all Anemia (not only treatment-related). - Sections 9.8 and 15.2.4.2 update time window for irAE - Section 14.1.1: the start of subsequent therapy is not expected to be displayed on the figure - Section 14.1: two spider plots are needed, one for all population and one for responders. - Section 15.3: add clarification on which data are displayed for boxplots. - All document: add Covid-19 analyses -CCI - Section 14: update wording for best overall response definition - eDISH plot: update on the left quadrant - Updated section 16.5: updated PD-L1 secondary efficacy analysis and PD-L1 expression in tumor tissue measured by 22C3 central lab testing - Updated Pharmacokinetics analyses - Update in section 8.2: Additional subgroup prior radiotherapy added - Update in section 14.2.2: Update definition for irRECIST definition - Update in section 6: Ad-hoc 40-patient analysis added

Unique Identifier for Version	Date of IAP Version	Author	Changes from the Previous Version
2.1	16December2020	PPD	- DI calculation updated - Appendix links 1 and 2 updated
3.0	1July2021	PPD	– Update in section 7.1: – listing of COVID-19 impact added – COVID-19 vaccination related outputs added CCI [redacted] – Update in section 8.2: – describe analyses for China and Japan authorities – Number of prior treatment lines, Prior anti-cancer RT and HPV status by central lab definition updated for clarification – ADA treatment-emergent status added – Section 9.1 Data Handling After Cut-off Date CCI [redacted] – Section 9.9 added some clarifications on Rules to define previous and/or concomitant medication table – Updated section 9.11 Moore criteria – Updated section 10.2.1: Important Protocol Deviations reporting – Section 11.3.2: added clarification for HPV calculation – Updated section 12.4 Previous anticancer treatments and procedures including CCRT description – Updated section 13 Cumulative dose and relative dose intensity formulas – Section 14 Non-CR/Non-PD options added for BOR calculations – Section 14.1.1 waterfall with subgroups added – Updated section 14.2.4 censoring reasons categories – Section 15.4 removed “except oxygen saturation” – Updated section 16.3 CCI [redacted] – Longitudinal analysis of change from baseline – Updated Section 16.4 Immunogenicity analyses CCI [redacted]
4.0	21February2022	PPD	– Section 6: Added “Overview of Planned Analyses after discontinuation of NSCLC and BTC trials”

4 Purpose of the Integrated Analysis Plan

The purpose of this integrated analysis plan (IAP) is to document technical and detailed specifications for interim, primary and final analyses of data collected for protocol MS200647_0017.

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 bintrafusp alfa based on ORR	Confirmed objective response according to RECIST 1.1 assessed by an IRC	14.1.1
Secondary		
To evaluate clinical efficacy of bintrafusp alfa based on DOR	DOR according to RECIST 1.1 assessed by an IRC	14.2.8
To evaluate clinical efficacy of bintrafusp alfa based on DRR	Durable response of at least 6 months according to RECIST 1.1 assessed by an IRC	14.2.10
To evaluate clinical safety of bintrafusp alfa	Occurrence of TEAEs and treatment-related AEs including AEs of special interest	15.1, 15.2
To evaluate clinical efficacy based on PFS	PFS according to RECIST 1.1 assessed by an IRC	14.2.1
To evaluate ORR, DOR, DRR and PFS by Investigator read	Confirmed objective response, DOR, DRR, and PFS according to RECIST 1.1 assessed by Investigator	14.2.1, 14.2.6, 14.2.8, 14.2.10
To evaluate clinical efficacy based on OS	OS	14.2.13
To characterize the pharmacokinetic (PK) profile of bintrafusp alfa	The concentration observed immediately at the end of infusion (C_{EOI}) of bintrafusp alfa The concentration observed immediately before next dosing (corresponding to pre-dose or trough concentration [C_{trough}] for multiple dosing) of bintrafusp alfa	16.1
To characterize the immunogenicity of bintrafusp alfa	Immunogenicity of bintrafusp alfa as measured by ADA assay from Screening through Safety Follow-up Visit (up to 28 days after last treatment)	16.4

Objectives	Endpoints (Outcome Measures)	IAP section
To evaluate clinical efficacy of bintrafusp alfa according to PD-L1 expression	Efficacy endpoints by PD-L1 expression in tumor	16.5.1
CCI		

6 Overview of Planned Analyses

This IAP addresses the interim, primary and final analyses. If the study continues beyond 67% of the participants died, subsequent analyses may be performed but are not described in this IAP. All statistical analyses will be performed based on Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) data using cleaned electronic Case Report Forms (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 defined in the following subsections.

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).

Table 2 displays an overview of the analyses to be performed for this study.

Table 2 Overview of Analyses

CCI [redacted]			
Interim analysis (IA)	CCI [redacted]	• 81 first dosed participants • All dosed participants (all analyses except efficacy analyses)	• Participant disposition • Demographics • Medical history • Previous and subsequent anticancer medications and other baseline characteristics (disease history, Eastern Cooperative Oncology Group Performance Status [ECOG PS]) • Concomitant medications • Treatment compliance and exposure • Confirmed objective response as adjudicated by IRC • Progression-free survival (PFS) as adjudicated by IRC • Duration of response (DOR) as adjudicated by IRC • Durable response rate (DRR) as adjudicated by IRC • ORR, DoR and PFS based on investigator assessment • Safety analyses (Adverse events (AEs), deaths, vital signs and clinical laboratory evaluations)
Primary analysis (PA)	CCI [redacted]	All participants	Full evaluation of all efficacy and safety endpoints
Final analysis (FA)	CCI [redacted]	All participants	Full evaluation of all efficacy and safety endpoints for the data that will be updated after PA

If key results presented during the IA are positive, meaning a lower bound of the exact 95% confidence interval (CI) for ORR is above 15% based on 81 participants, a full evaluation of all

efficacy and safety endpoints will be also provided at this time. In case the recruitment between the 81st and the last global participants is going fast and the timing between IA and PA is getting close, the possibility to not perform the IA will be discussed.

6.1 Overview of Planned Analyses after discontinuation of NSCLC and BTC trials

The data review outcome from the 3 randomized controlled studies in NSCLC (Non-Small Cell Lung Cancer) and BTC (Biliary Tract Cancer) (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 restrict the analyses on this study MS200647-017. The analyses described in Table 2 will not be performed as planned, but one analysis based on the primary analysis cutoff of 15 February 2022 will be performed for an abbreviated CSR (aCSR).

The analyses considered relevant for this purpose are listed below:

- Participant disposition
- Important protocol deviations
- Demographics, medical history and other baseline characteristics (disease history, HPV and PD-L1 test at screening, vital signs at baseline, skin status history)
- Vaccines for COVID 19
- Previous and concomitant medications, procedures, follow-up treatments
- Treatment exposure and compliance
- Efficacy analyses:
    - ORR per RECIST 1.1 as assessed by IRC
        - ORR and DCR per RECIST 1.1 as assessed by IRC
        - Reasons for Non-evaluable BOR based on Confirmed Responses according to RECIST 1.1 as Adjudicated by IRC
        - Percent Change from Baseline in Sum of Diameters according to RECIST 1.1 as Adjudicated by IRC
        - Percent Change in Sum of Diameters between Baseline and Best Post-baseline Assessment according to RECIST 1.1 as Adjudicated by IRC and selected subgroups
        - ORR per RECIST 1.1 as assessed by IRC - Forest plot of odds ratio by subgroups
        - ORR and DCR per RECIST 1.1 as assessed by Investigator

- Concordant and Discordant Objective Response based on confirmed responses according to RECIST 1.1 between IRC and Investigator

- PFS per RECIST 1.1 as assessed by IRC
  - PFS per RECIST 1.1 as assessed by IRC - Kaplan-Meier Curve
  - PFS per RECIST 1.1 as assessed by IRC by histology and PD-L1 - Kaplan-Meier Curve
  - PFS per RECIST 1.1 as assessed by IRC - Forest plot of hazard ratio and free survival rates at 6 and 12 months by subgroups
  - PFS per irRECIST 1.1 as assessed by IRC - Kaplan-Meier Curve
- Duration of response per RECIST 1.1 as assessed by IRC
  - Duration of response per RECIST 1.1 as assessed by IRC - Kaplan-Meier Curve
  - Time and Duration of Confirmed Response per Subject per RECIST 1.1 as assessed by IRC
- Overall survival (OS)
  - Overall survival - Kaplan-Meier Curve
  - Overall survival by histology and PD-L1 - Kaplan-Meier Curve
  - Overall Survival – Forest plot of hazard ratio by subgroups
- Follow-up time for Overall Survival (OS)

- Safety analyses (overview of treatment-emergent adverse events (TEAEs), serious TEAEs (SAE), treatment-related TEAEs, TEAEs leading to treatment discontinuation/interruption/modification, TEAEs Leading to death, adverse events of special interest (AESI), TEAEs associated to COVID-19, deaths, vital signs, ECOG, clinical laboratory evaluations).
- Pharmacokinetic Data

- Descriptive Statistics for Bintrafusp Alfa Serum CEOI and Ctrough (<unit>) by Nominal Day
- Descriptive Statistics for Bintrafusp Alfa Serum Accumulation Ratios (RCEOI and RCtrough) by Nominal Day
- Arithmetic Mean Bintrafusp Alpha Serum Trough Concentrations (Ctrough) versus Nominal Day on Linear (±StD) Scale
- Geometric Mean Bintrafusp Alpha Serum Trough Concentrations (Ctrough) versus Nominal Day on Linear (±logStD) Scale
- Median Bintrafusp Alpha Serum Trough Concentrations (Ctrough) versus Nominal Day on Linear Scale
- Arithmetic Mean Bintrafusp Alpha Serum End of Infusion Concentrations (CEOI) versus Nominal Day on Linear (±StD) Scale
- Geometric Mean Bintrafusp Alpha Serum End of Infusion Concentrations (CEOI) versus Nominal Day on Linear (±logStD) Scale
- Median Bintrafusp Alpha Serum End of Infusion Concentrations (CEOI) versus Nominal Day on Linear Scale
- Individual Bintrafusp Alpha Serum Trough Concentrations (Ctrough) versus Actual Day on Linear Scale
- Individual Bintrafusp Alpha Serum End of Infusion Concentrations (CEOI) versus Actual Day on Linear Scale

CCI [redacted]

- Immunogenicity
  - Overview of ADA Subgroups
  - Listing of Subjects with Ever Positive ADA Results

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

7 Changes to the Planned Analyses in the Clinical Study Protocol

CCI [redacted]

7.1 COVID-19 Impact

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

Instead, additional outputs will be generated to assess potential impacts of COVID-19 to this study including:

- Listing of participants potentially affected by pandemic and with any AEs or PDs related to COVID-19
- Listing of participants with any missed treatment administrations, tumor assessments or missed visits due to COVID-19
- Listing of PDs related to COVID-19
- Listing of AEs related to COVID-19
- Table of COVID-19 vaccinations
- Listing of COVID-19 vaccinations
- Table of TEAEs associated with the vaccination for COVID-19 (in case that AEs associated with COVID-19 vaccination will be observed for 10 subjects or more)
- Listing of AEs related to COVID-19 vaccination.

8 Analysis Populations and Subgroups

8.1 Definition of Analysis Populations

Screening Analysis Set (SCR)

All participants who signed the informed consent.

Full Analysis Set (FAS)/Safety Analysis Set (SAF)

The Full Analysis Set as well as the Safety Analysis Set include all participants who were administered at least 1 infusion of bintrafusp alfa.

The FAS terminology will be used for all analyses of demographics and baseline characteristics, efficacy, CCI [redacted]. The SAF terminology will be used for all exposure and safety analyses.

Pharmacokinetic Analysis Set (PKAS)

The Pharmacokinetic Analysis Set (PKAS) includes all participants who complete at least one infusion of bintrafusp alfa, and who provide at least one 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 10.2 and Section 16.1.1 for handling of protocol deviations relevant to PK.

Pharmacokinetic ADA Analysis Set (PKADA)

A subpopulation of the PKAS restricted to subjects who have in addition at least one valid result of antidrug antibody (ADA) at any time point.

Pharmacokinetic nAb Analysis Set (PKNAB)

A subpopulation of the PKAS 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 were administered at least one infusion of bintrafusp alfa and have at least one valid ADA result. All ADA 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	FAS	PK	PKADA	PKNAB	IMM
Disposition	✓						
Demographics			✓				
Baseline Assessments			✓				
Previous and Concomitant Therapies			✓				
Compliance and Exposure		✓					
Efficacy			✓				
Safety and Tolerability		✓					
Pharmacokinetics				✓	✓	✓	
CCI [redacted]							
Immunogenicity							✓
CCI [redacted]							

8.2 Subgroup Definition and Parameterization

Analysis of primary and key secondary efficacy endpoints may be performed on subgroups of interest as specified in this IAP Section 14. CCI [redacted]

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

In case of low number of participants within a category (< 7 participants, which is about 5% of the randomized population), categories will be pooled when meaningful.

The following subgroups will be defined and used for the subgroup efficacy analyses:

- Age
    - Age ≥ 18 - < 50 years
    - Age ≥ 50 - < 65 years
    - Age ≥ 65 years
- Moore criteria
    - Low
    - Medium
    - High
    - Unknown
- Histology
    - Squamous cell carcinoma
    - Adenocarcinoma
    - Adenosquamous cell carcinoma
- Number of prior treatment lines (metastatic/locally advanced prior anti-cancer drug therapies (see Section 12.4)
    - 0
    - 1
    - 2
    - 3
    - ≥4
- Pooled Region
    - North America
    - Europe
    - South America
    - Asia
    - Rest of the World (RoW)
- Prior Bevacizumab use

  - Yes
  - No
- Prior anti-cancer radiotherapy (see Section 12.4)
  - Yes
  - No
- ADA status
  - Ever positive
  - Never positive
- ADA treatment-emergent status
  - Treatment-emergent ALL
- Non-treatment-emergent nAb status
  - Ever positive (for either assay TGF-β or PD-L1)
  - Never positive

All subgroups are associated with demographic and clinical variables corresponding with cervical cancer prognosis except ADA and nAb status.

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

The subgroups coming from the individual items composing the Moore criteria will be defined and used only for the Objective Response Rate (ORR) and Duration of Response (DOR) subgroup analyses:

- ECOG PS at baseline
  - ECOG PS 0
  - ECOG PS 1
- RACE from Moore criteria
  - Black or African American
  - Not Black or African American
- RACE
  - White
  - Black or African American
  - Chinese
  - Japanese
  - Korean

  - Other Asian
  - Other
- Measurable disease in pelvis
  - Yes
  - No
- Disease progression-free interval from date of diagnosis < 1 year
  - Yes
  - No
- Prior platinum therapy exposure to radiotherapy
  - Yes
  - No

CCI [redacted]

- PD-L1 expression at baseline measured by central lab:
  - PD-L1 expression in tumor tissue measured by 22C3 (CPS >= 1, < 1, NE or NA)
  - PD-L1 expression in tumor tissue measured by SP263 (vCPS (Visual element Combined Positive Score (Ventana))>=5%, <5%, NE or NA)
- HPV status collected on the "Disease History" eCRF page
  - HPV16+ or HPV18+
  - HPV+ (all other subtypes)
  - HPV-
  - Not done
- HPV status at baseline measured by central lab (see section 11.3.2)
  - High risk HPV+
  - Low Risk HPV+
  - HPV-
  - Not done

For HPV status collected on the eCRF, as they are several possibilities of testing (tumor, pre-malignant cervical lesion and other), it will be enough for a participant to have one positive value in one of the three tests to be in the associated subgroup.

CCI [redacted]

To support submission purposes a subset of outputs may be repeated for the sub-populations described below:

- China: A Chinese participant is defined as a participant enrolled at a site in China
- Japan: A Japanese participant is defined as a participant enrolled at a site in Japan and of Japanese ethnicity as recorded in the case report form.

9 General Specifications for Data Analyses

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

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

The "start date" for this study is the start date of treatment.

Continuous (non-PK) variables will be summarized using descriptive statistics, i.e., the number of participants with non-missing values (n), arithmetic mean (Mean), standard deviation (StD), median (Med), 25th percentile (Q1) and 75th percentile (Q3), minimum (Min), and maximum (Max).

Qualitative variables will be summarized by frequency 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.

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

Pharmacokinetic variables (concentrations and parameters) will also be summarized as described in Section 16.1.

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

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 statistical analyses will be performed using SAS® Software Version 9.4 or higher in the SAS Grid environment (Statistical Analysis System, SAS-Institute, Cary NC, USA). The computer program Phoenix® WinNonlin® Version 8.0, or higher (Certara, L.P., Princeton, New Jersey, USA) could be used for PK data.

9.1 Data Handling After Cut-off Date

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

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

9.2 Definition of Baseline and Change from Baseline

The last non-missing measurement prior to the first study treatment administration will be used as the baseline measurement for safety and efficacy analyses.

If an assessment that is planned to be performed before treatment per protocol is performed on the same day as the start of treatment, respectively, but the assessment time is not available, it will be assumed that it was performed prior and will be considered as baseline. 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.

Absolute and percent changes from baseline are defined as

absolute change = visit value – baseline value

percent change = 100 * (visit value – baseline value) / baseline value

9.3 Study Day / Study Treatment Day

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

9.4 Definition of Duration and ‘time since’ Variables

Durations in days will be calculated by the difference of start and stop date + 1 (e.g. survival time (days) = date of death – date of first study intervention + 1) if not otherwise specified.

Date of first study intervention is defined as the date of the first administration of bintrafusp alfa during the study.

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 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.

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 cut-off using the latest complete date prior to or at the data cut-off date among the following:

- All participant assessment dates (blood draws (laboratory, PK), vital signs, performance status, electrocardiogram (ECG), tumor assessments, quality of life assessments)
- Start and end dates of anti-cancer therapies administered after study treatment discontinuation.
- AE start and end dates
- Last known to be alive date collected on the 'Survival Follow-up' eCRF (do not use follow-up date)
- Study drug start and end dates (including reinitiation of treatment)
- Date of discontinuation on disposition eCRF pages (do not use if reason for discontinuation is lost to follow-up)

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. Assessment dates after the cut-off date will not be applied to derive the last contact date.

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

CCI

9.8 Definition of On-treatment Period

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 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.

Any systemic anticancer therapy, any anticancer surgery and any anticancer radiotherapy as documented in the "Anti-cancer treatment after discontinuation details", "Radiotherapy after discontinuation details" and "Surgery after discontinuation details" eCRF pages 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 the 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.

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.

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.

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 (StD) cannot be computed and should be presented as "nd".

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

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

For identification of previous or concomitant medications/procedures, rules presented in Table 5 and Table 6 below will be used to define if a medication/procedure is considered as a previous, concomitant or both previous and concomitant medication/procedure.

Table 5 Stopping rules for medication/procedure end dates

End date of medication/procedure			Stopping rule
Day	**Month**	**Year**	
UNK	UNK	UNK	After treatment start (ongoing)
UNK	UNK	< Treatment start (year)	Before treatment start

End date of medication/procedure			Stopping rule
Day	Month	Year	
UNK	UNK	>= Treatment start (year)	After treatment start
UNK	< Treatment start (month and year)		Before treatment start
UNK	>= Treatment start (month and year)		After treatment start
< Treatment start (complete date)			Before treatment start
>= Treatment start (complete date)			After treatment start

UNK = Unknown.

Table 6 Rules to define previous and/or concomitant medication

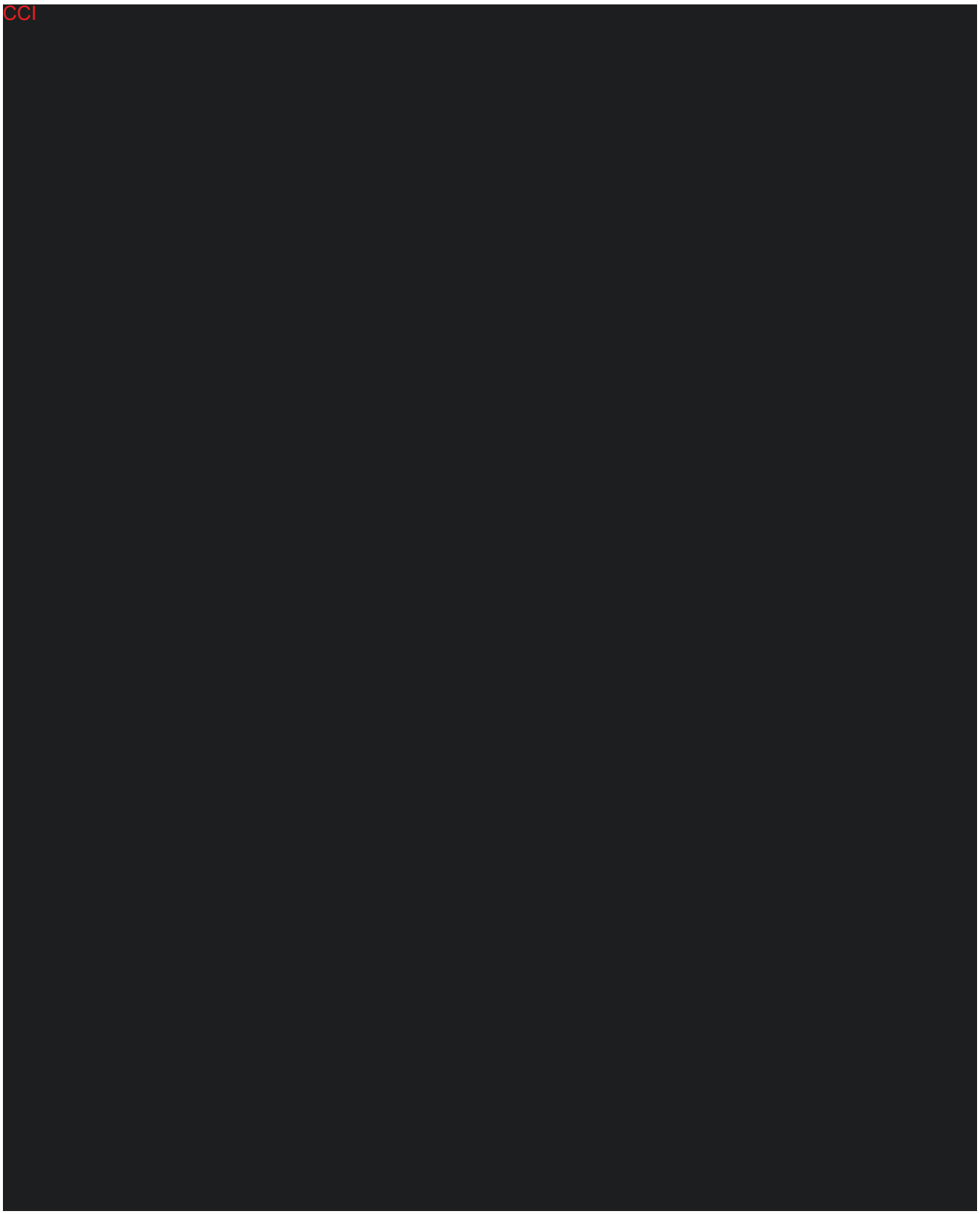
Start date of medication/procedure			Stopping rule (see Table 5)	Medication/procedure
Day	Month	Year		
UNK	UNK	UNK	Before treatment start	Previous
UNK	UNK	UNK	After treatment start	Previous and concomitant
UNK	UNK	<= Treatment start (year)	Before treatment start	Previous
UNK	UNK	<= Treatment start (year)	After treatment start	Previous and concomitant
UNK	UNK	> Treatment start (year) and <= Treatment end + 30 days (year) OR the earliest date of subsequent anticancer drug therapies minus 1 day, whichever occurs first.	After treatment start	Concomitant
UNK	<= Treatment start (month and year)		Before treatment start	Previous
UNK	<= Treatment start (month and year)		After treatment start	Previous and concomitant
UNK	> Treatment start (month and year) and <= Treatment end + 30 days (month and year) OR the earliest date of subsequent anticancer drug therapies minus 1 day, whichever occurs first.		After treatment start	Concomitant
<= Treatment start (date)			Before treatment start	Previous
<= Treatment start (date)			After treatment start	Previous and concomitant

Start date of medication/procedure			Stopping rule (see Table 5)	Medication/procedure
Day	Month	Year		
> Treatment start (date) and <= Treatment end + 30 days (date) OR the earliest date of subsequent anticancer drug therapies minus 1 day, whichever occurs first.			After treatment start	Concomitant

UNK = Unknown.

Dates of study treatment	Start date of study treatments: • No imputation will be done. End date of study treatments: In case the last date of study drug is missing or incomplete the date of last administration of study drug will be taken from the treatment termination eCRF pages. • If the last date of study drug is completely missing and there is no End of Treatment eCRF page and no death date the participant should be considered to be ongoing and use the cut-off date for the analysis as the last dosing date • If the last date of study drug is completely or partially missing and there is EITHER an End of Treatment eCRF page OR a death date available (within the cut-off date) then imputed last dose date is: = 31DECYYYY, if only Year is available and Year < Year of min (EOT date, death date) = Last day of the month, if both Year and Month are available and Year = Year of min (EOT date, death date) and Month < the month of min (EOT date, death date) = min (EOT date, death date), for all other case
Death date	For the purpose of survival analyses (PFS and OS) 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 contact and the 15th day of the month. Otherwise it will not be imputed. In listings, the imputed date of death will be presented with a flag indicating the level of imputation.

CCI



CCI

CCI

CCI

[redacted]

9.11 Other Derivations

Data collected after reinitiated treatment	Data collected after reinitiation of treatment will be included in the summary statistics. A data listing will include AE data for all reinitiated patients.
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).
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 CCI [redacted]. 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.
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.
Moore criteria	The Moore criteria is defined using the following negative risk factors: • ECOG > 0 at baseline • Measurable disease in pelvis if there is presence of target lesions in the pelvis coming from "Disease History" eCRF page • Race is "BLACK OR AFRICAN AMERICAN" • Disease progression-free interval from date of diagnosis < 1 year. The negative risk is met if the following date is strictly earlier than a year after the date of initial cancer diagnosis coming from "Disease History" eCRF page: o The first documented progression disease date from "Prior Anti-Cancer Drug Therapies Details" eCRF page. Date is imputed as the disease history detailed in Section 9.9. • Overlapping of prior platinum therapy and exposure to radiotherapy. Platinum-based treatment are defined as treatment with PTs according to a pre-specified WHODrug search list which will be finalized before database lock. If the answer to the question "Was Radiotherapy administered as part of the therapy?" is Yes on the "Prior Anti-Cancer Drug Therapies Details" eCRF page , then it is a negative factor. Dates are imputed as the previous and concomitant medication/procedure detailed in Section 9.9. Risk categories are defined as low-risk (0–1 factor met), mid-risk (2–3 factors met), and high-risk (4–5 factors met).
Categorization of participant for COVID-19 impact assessment	For the assessment of COVID-19 impact on this study, participants will be categorized as being potentially affected by COVID-19 based on the COVID-19 pandemic start date, defined as the minimum of the first COVID-19 death date per country and 11 March 2020 (WHO-start of world-wide pandemic). First death from COVID-19 occurred per country is determined according to the published data

	by European Centre for Disease Prevention and Control (status of 26th June 2020).

10 Study Participants

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

10.1 Disposition of Participants and Discontinuations

The number and percentage of participants in each of the below disposition categories will be presented based on the electronic case report form (eCRF) data. Percentages will be presented with respect to the number of treated participants

- 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 (FAS/SAF analysis set)
- Number and percentage of participants with treatment ongoing at the data cut-off date
- Number and percentage 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 and percentage of participants who reinitiated the study treatment and number of participants who discontinued the study intervention after reinitiation
- Number and percentage of participants with treatment ongoing at the end of study
- Number and percentage 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)

In addition, the number of participants screened, and enrolled in each analysis population defined in Section 8.2 will be summarized, overall and by region (North America, Europe, South America, Asia, RoW), by country within region and by site. Subject disposition and analysis sets may be repeated for China and Japan sub-populations.

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 study treatment 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 reinitiation study treatment administration date and status at end of treatment reinitiation 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, a listing will include the following information (as applicable): participant identifier, date of first and last dose, pandemic start date in the country, if treatment period was potentially affected by pandemic, if any COVID-19 related AE or PD occurred during the pandemic period.

10.2 Protocol Deviations / Exclusion from Analysis Populations

10.2.1 Important Protocol Deviations

Analysis Set: FAS

A full list of potential protocol deviations including definition and categorization is maintained in Appendix 1: Definition of important protocol deviations.

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 MS200647-0017 will be evaluated as described on Section 7.1. Table on important protocol deviations may be repeated for China and Japan sub-populations.

Considerations for PK:

Protocol deviations or events will be reviewed by the study pharmacokineticist and biostatistician to identify deviations or events which have the potential to affect the PK results. Deviations resulting in no evaluable PK results for a subject, and thus exclusion from PK summaries, will be documented as reasons for exclusion from these summaries.

Changes to the procedures which may impact the quality of the PK data will be considered protocol deviations and will be described within the CSR body text. Other events which may impact the quality of the PK data will be described within the CSR body text. These changes or events will include any circumstances that will alter the evaluation of the PK. Examples of important events for PK in terms of this study may include, but may not be limited to, the following:

Actual dosing time not recorded

- Dose change, missed dose, or incomplete/inaccurate 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 PKAS. Other changes to the procedures or events which do not impact the quality of the PK data will not be considered important protocol deviations for PK. Common examples of minor (non-important) protocol deviations are a missed sample or minor deviations from sample collection times/windows.

Refer to Section 16.1.1 for more details of protocol deviations and handling relevant to PK.

10.2.2 Reasons Leading to the Exclusion from an Analysis Population

Reasons for exclusion from the PKAS will be provided in table(s)/listing(s).

11 Demographics and Other Baseline Characteristics

Analysis Set: FAS

Demographics medical history and other baseline characteristics may be repeated for China and Japan sub-populations.

11.1 Demographics

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

The following demographic characteristics will be included:

- Sex: female
- Ethnicity: Hispanic or Latino, not Hispanic or Latino
- Race:
    - For participants reporting one race only: White, Black or African American, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian, American

Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Not collected at the site, Other.

- For participants reporting multiple races, all combinations will be reported under 'More than one race" category.

- Age (years)
- Age categories
  - < 65 years, ≥ 65 years
  - 18-49 years, 50-64 years, 65-74 years, 75-84 years, ≥ 85 years
- Pooled Region
  - North America
  - Europe
  - South America
  - Asia
  - RoW

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.
- Site codes will be used for the determination of the participant's geographic region.

This summary table will also be repeated to evaluate subjects split by prior Bevacizumab used or not, and may be repeated for Japanese status (Japanese, non-Japanese, all).

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 and category), height, weight and body mass index (BMI) at baseline 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) at the time of the database lock, 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. Categories for Human immunodeficiency virus (HIV) infection, Multi-drug resistance preventing effective antiretroviral therapy, Multi-drug resistance not preventing effective antiretroviral therapy and Aids defining opportunistic infection will be also displayed in a separate table.

Listing of medical history including participant identifier, age, sex, race, PT, 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:

- Tumor histology: Squamous cell carcinoma, Adenocarcinoma, Adenosquamous cell carcinoma, Other
- Time since initial cancer diagnosis (months)
- Time since documented, locally advanced, inoperable or metastatic disease (months)
- Time since last progression of disease prior to study entry (months)
- TNM classification at initial diagnosis: each T, N, M category will be described (TX, T0, N1, etc.)
- MSI status: High, Low, MSS, Unknown. MSI status will be centrally determined by means of the Tumor tissue sample at Baseline and will be provided as external data

This summary table will also be repeated to evaluate subjects split by prior Bevacizumab used or not, and may be repeated for Japanese status (Japanese, non-Japanese, all).

Listing will also be provided with the following information:

- Participant identifier, age, sex, race, tumor histology, date of initial cancer diagnosis (months), date of documented, locally advanced, inoperable or metastatic disease (months), date of last disease progression (months), TNM classification at initial diagnosis and MSI status

11.3.2 HPV test

HPV testing history as collected on the "Disease History" eCRF page and HPV status retrospectively measured from central testing will be summarized respectively by the frequency and percentage of participants having the following history of:

- HPV16+ or HPV18+, HPV+ (all other subtypes), HPV-, or NA
- High risk HPV+ (if HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, or 68 is positive), Low risk HPV+ (if subject is positive for any other HPV subtype than the ones listed on high risk), HPV- (if none of HPV subtypes are detected), or Not done (if result is not reported).

For data coming from eCRF, as they are several possibilities of testing (tumor, pre-malignant cervical lesion and other), it will be enough for a participant to have one positive value, HPV16, HPV18 or in another subtype, in one of the three possible tests to be in the associated subgroup.

This summary table may be repeated for Japanese status (Japanese, non-Japanese, all).

Listing will also be provided and include: participant identifier, age, sex, race, visit, testing date, location, test source, kit used and HPV status of tumor and pre-malignant lesion or other history of HPV positive testing. Tissue sample collection date will be also provided for data coming from central lab.

11.3.3 PD-L1 test

Programmed Death-ligand 1 (PD-L1) expression will be collected from two different sources, eCRF page and external data coming from tests conducted by central lab.

All tests collected on the "PD-L1 Tumor Testing" eCRF page will be summarized. This includes 22C3, SP142, SP263, Dako 28-8, Dako 73-10 and other testing. All tests will be described using the following categories: positive, negative, NE or NA except Dako 28-8 which will use the following categories: < 1%; >= 1%; >= 5%; >= 10%, NE or NA.

In addition, PD-L1 expression as collected from central lab testing will be described using the following categories:

- PD-L1 expression in tumor tissue measured by 22C3 (CPS >= 1, < 1, NE or NA)
- PD-L1 expression in tumor tissue measured by SP263 (vCPS>=5%, <5%, NE or NA)

This summary table will also be repeated to evaluate subjects split by prior Bevacizumab used or not, and may be repeated for Japanese status (Japanese, non-Japanese, all)

Listing will also be provided and include: participant identifier, age, sex, race, unique sample identifier, visit, sample collection date, date of PD-L1 testing, method of testing, manufacturer/qualified pathologist reported result, raw value for the central lab testing and PD-L1 expression status.

11.3.4 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

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

Other vital signs parameters such as systolic and diastolic blood pressure, pulse rate and respiration rate at baseline will be included in the summary tables by timepoint.

11.3.5 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 house work, office work.
- 2: Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours.
- 3: Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours.
- 4: Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair.
- 5: Dead.

11.3.6 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.

CCI [redacted]

CCI

12 Previous or Concomitant Medications/Procedures

Analysis Set: FAS

The tables in this section may be repeated for China and Japan sub-populations.

12.1 Previous and concomitant medications

Concomitant treatments are medications, other than study treatment, which are taken by participants any time during the on-treatment period, see Section 9.8. All medications starting the same day as study drug will be considered as concomitant. Medications starting 30 days after the last dose of study treatment will also be considered as concomitant.

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

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

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

Specific rules will be used to define if a medication/procedure is considered as a previous, concomitant or both previous and concomitant medication/procedure as detailed in Section 9.9.

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

12.2 Premedications prior to bintrafusp alfa administration

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, premedication prior to bintrafusp alfa administration for the first two infusions is optional and at the discretion of the investigator. If Grade 2 or more infusion reaction(s) are seen during the first 2 infusions, premedication should be continued/implemented for future infusions.

If at least 10 participants received premedication during the study, the number of participants receiving premedication will be summarized for each treatment visit based on "Premedication Bintrafusp Alfa Details" eCRF page. Percentages will be calculated on the number of participants who received an infusion at the associated visit.

Listing will 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.

Listing will be provided including: 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 radiotherapy
- Participants with at least one previous anticancer surgery
- Participants with at least one previous Concurrent Chemoradiation Therapy (CCRT)

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 prior lines of therapy for metastatic/locally advanced disease: 0 / 1 / 2 / 3 / ≥ 4
- Number of prior lines of therapy for metastatic/locally advanced disease excluding CCRT: 0 / 1 / 2 / 3 / ≥ 4
- 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 Evaluable / Unknown.

Please note that radiotherapies taken alone will not be considered as a line of therapy

Previous anticancer drug therapies and radiotherapies and previous anticancer surgery will be presented in separate listings:

- The previous anticancer drug therapies and radiotherapies listing will contain participant identifier, age, sex, race, regimen ID, PT, medication name, start date, end date, intent of therapy, best response, date of progression. Radiotherapy administered as part of previous drug therapy with start date, end date, total dose, number of fractions and location will also be displayed.
- 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 Previous pre-cancerous procedures of the cervix

The previous pre-cancerous procedures of the cervix are collected under the “Previous Pre-cancerous Procedures of the Cervix Details” eCRF page.

Previous pre-cancerous 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 and reason for procedure.

12.6 Subsequent anticancer treatments and procedures

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

If collected in the eCRF, the best response of first subsequent anticancer drug therapy will also be described (Complete Response (CR) / Partial Response (PR) / Stable Disease (SD) / Progressive

Disease (PD) / Non-Complete Response/Non-Progressive Disease (non-CR/non-PD) / Not Evaluable / Unknown).

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

- For medication: participant identifier, age, sex, race, PT/medication name, regimen name, start date, end date, route and best response (if collected in the eCRF)
- For radiotherapy and surgery: participant identifier, age, sex, race, start date, end date, radiotherapy site or name of surgery/location, if surgery outcome and was the surgery curative in intent (Y/N)

13 Study Treatment: Compliance and Exposure

Analysis set: SAF

The tables in this section may be repeated for China and Japan sub-populations.

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.

All dosing calculations below and summaries will be based on "Bintrafusp Alfa Administration Details" eCRFs page. Data collected during the treatment reinitiation phase will be included in the summary statistics and described in listings.

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

Imputation for incomplete start and end dates of study treatments are described in Section 9.9.

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** overall of bintrafusp alfa per participant is the sum of the actual dose that the participants received during the study (i.e., total dose administered (mg)).

The **dose intensity** of treatment (DI) (mg/cycle) of bintrafusp alfa during the study is defined as

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

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

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

Where the planned dose is 1200 mg/cycle

The following summary tables will be provided:

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

Duration of therapy and cumulative dose, dose intensity and relative dose intensity tables may be repeated by Japanese status (Japanese, non-Japanese, all).

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 reinitiation 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/week), and relative dose intensity (%).

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 Set: FAS

14.1 Primary Endpoint: Confirmed objective response according to RECIST 1.1 as adjudicated by IRC

Primary endpoint is the confirmed objective response according to Response Evaluation Criteria in Solid Tumors Version 1.1 (RECIST 1.1) assessed by an IRC.

14.1.1 Primary Objective: Objective Response Rate

Objective Response Rate (ORR)

The ORR is the proportion of participants having a confirmed Objective Response (OR) in the analysis set.

The confirmed objective response according to RECIST 1.1 as adjudicated by the IRC is the primary endpoint of the study. The study aims to estimate the Objective Response Rate (ORR) with a sufficient level of precision and that the associated lower bound of the two-sided 95% CI is above a minimal threshold of 15%.

Confirmed Objective Response (OR)

The Confirmed Objective Response (OR) is defined as a best overall response of confirmed complete response (CR) or partial response (PR) according to RECIST v1.1 as adjudicated by IRC.

Best overall response will be assessed based on reported overall responses at different evaluation time points from the treatment start date until documented disease progression in accordance to RECIST v1.1, taking requirements for confirmation into account as detailed below. Only tumor assessments performed before the start of any subsequent anti-cancer therapies will be considered in the assessment of best overall response. If a tumor assessment was performed on the same day as start of new anti-cancer therapy, it will be assumed that the tumor assessment was performed prior to the start of the new anti-cancer therapy, therefore the tumor assessment will be included in the assessment of best overall response. Clinical deterioration will not be considered as documented disease progression. The date of the overall response assessment is the earliest date for imaging of target, non-target and new lesions of images taken at that response assessment.

Best Overall Response based on confirmed responses according to RECIST 1.1 will be derived as follows

- CR = at least two determinations of CR at least 4 weeks apart (with no PD in between)
- PR = at least two determinations of PR or better (PR followed by PR or PR followed by CR) at least 4 weeks apart (and not qualifying for a CR), with no PD in between
- 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 to Investigator assessment)
- PD = progression ≤ 16 weeks after first date of study intervention (and not qualifying for CR, PR, non-CR/non-PD or SD).
- Not Evaluable (NE): all other cases.

Disease Control Rate (DCR)

Disease Control Rate (DCR) is defined as the proportion of participants with best overall response according to evaluation criteria of confirmed CR, PR, or SD out of the total number of participants in the analysis set.

Best Overall Response based on unconfirmed responses

Best Overall Response can be assessed based on unconfirmed responses according to RECIST 1.1 and derived as follows:

- 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 first date of study intervention 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 to Investigator assessment)
- PD = progression ≤ 16 weeks after first date of study intervention (and does not meet criteria of CR, PR, SD)
- NE = all other cases.

Details of analyses

Separate tables will be provided for Best Overall Response based on confirmed and unconfirmed responses. Best Overall Response based on confirmed responses of CR, PR, SD, PD and NE as well as the ORR and the DCR will be provided with number and percentage of participants.

The number and percentage of participants will be provided for Best Overall Response of CR or PR based on unconfirmed responses within 8 weeks before the cutoff. The participants will be taken into account only if their first CR or PR occurred within 8 weeks before the cutoff, they don't have confirmed response of CR or PR and they continue to have tumor assessments (they didn't discontinue the study for any reason and didn't discontinue treatment because of progressive disease). The denominator will be all participants in the analysis set.

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).

For all participants, 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) for these participants will be displayed 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). Same spider plot will be performed only on participants with a response, confirmed, unconfirmed or delayed.

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 will 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.

The best percent change in sum of target lesion diameters will also be presented in a waterfall plot with confirmed Best Overall Response per IRC, histology, prior Bevacizumab use, prior anti-cancer radiotherapy, Moore criteria, number of prior treatment lines (0, 1 and ≥2) and PD-L1 expression at baseline measured by central lab (see section 8.2). A spider plot including all subjects with at least one target lesion previously irradiated will be provided. This plot will display the individual percentage of change in the sum of diameter since baseline for the target lesions previously irradiated and the curves will be color-coded to differentiate responder vs non-responder subjects.

Two listings of tumor assessment will be provided with the following information. First listing will display participant identifier, age, sex, race, first and last date of treatment, date of start of subsequent anticancer therapy, date of death when death occurs, visit, date(s) of imaging, description of target lesions (size, site, type, method), non-target lesions (status, site, type, method), and new lesions (site, type, method) and if the lesion split or merged. Second listing will present participant identifier, age, sex, race, first and last date of treatment, Best Overall Response based on confirmed responses, Best Overall Response based on unconfirmed responses, date of start of subsequent anticancer therapy, date of death when death occurs, target response, non-target response, new lesion, sum of diameters, % of change in sum of diameters since baseline and overall response.

In addition, a summary table of the reasons for non-evaluable Best Overall Response based on confirmed responses 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 Best Overall Response is NE due to both early SD and late PD will be classified into this category
- Non-CR/Non-PD of insufficient duration (<6 weeks after 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 Best Overall Response based on confirmed responses 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 Best Overall Response based on confirmed responses non-evaluable.

Clinical efficacy of Bintrafusp alfa will be evaluated according to PD-L1 expression as a Secondary Endpoint, see section 16.5.1.

The following SAS code will be used to analyse the primary endpoint, ORR with its 95% CI:

```
proc freq data = <DATASET>;
        table <VARIABLE> / binomial(exact) alpha=.05 out = <OUTDATASET>;
run;
```

14.2 Further Efficacy Endpoints

14.2.1 Confirmed objective response according to RECIST 1.1 as assessed by Investigator (Secondary endpoint)

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 14.1.1 will be repeated based on Investigator assessment, i.e.:

- Number and percentage of participants with Best Overall Response based on confirmed responses of CR, PR, SD, PD and NE and corresponding listing.
- The number and percentage of participants will be provided for Best Overall Response of CR or PR based on unconfirmed responses within 8 weeks before the cutoff. The participants will be taken into account only if their first CR or PR occurred within 8 weeks before the cutoff, they don't have confirmed response of CR or PR and they continue to have tumor assessments (they didn't discontinue the study for any reason and didn't discontinue treatment because of progressive disease). The denominator will be all participants in the analysis set.
- ORR and DCR will be provided with number and percentage of participants with their 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). Analyses will be repeated for unconfirmed responses.
- Spider and Waterfall plots
- A summary table of the reasons for non-evaluable Best Overall Response based on confirmed responses and related listing, with 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 Best Overall Response BOR is NE due to both early SD and late PD will be classified into this category
- 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

- Listing of reasons for non-evaluable Best Overall Response based on confirmed responses
- Listing of individual tumor assessments as well as listing of time point tumor response and overall response

In addition, a summary of the Best Overall Response based on confirmed responses and unconfirmed responses as adjudicated by IRC versus Investigator assessment will be provided including numbers of concordant and discordant assessments, and a listing of inconsistencies will be provided.

CCI

CCI

CCI

[redacted]

14.2.4 Progression-Free Survival (Secondary Endpoint)

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 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 prior to the event during the first 12 months i.e. 365 days of follow-up or 168 days prior to the event 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.

Censoring rules are also summarized in Table 7.

Table 7 Censoring Rules for Analyses of PFS

Situation		Rule 1	Rule 2
No PD and no death	New anticancer therapy is not initiated	Censored at last tumor assessment*	Censored at last tumor assessment*
	New anticancer therapy is initiated	Censored at last tumor assessment* before new anticancer therapy	Censored at last tumor assessment*
No baseline assessment or no evaluable post-baseline assessment	No death or death > 16 weeks (112 days) after start of study treatment	Censored at date of first administration of study treatment	Censored at date of first administration of study treatment
	Death ≤ 16 weeks (112 days) after start of study treatment	Progressed at date of death	Progressed at date of death
PD or death	After ≤ 1 subsequent missing response assessment[a]	Progressed at date of documented PD or death, whichever came first	Progressed at date of documented PD or death, whichever came first
	After ≥ 2 subsequent missing response assessment[a]	Censored at last tumor assessment* before missing assessments.	Progressed at date of documented PD or death, whichever came first
	Before new anticancer therapy has started	Progressed at date of document PD or death, whichever came first	Progressed at date of documented PD or death, whichever came first
	After new anticancer therapy has started	Censored at last tumor assessment* before new anticancer therapy	Progressed at date of documented PD or death, whichever came first

Rule 1 used for PFS as assessed by the Investigator, by the IRC and for irPFS.

Rule 2 used for 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.

* 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 prior to the event during the first 12 months of follow-up or 168 days prior to the event after the first 12 months of follow-up.

The analysis of PFS as measured by the IRC (secondary endpoint) 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 Med 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 Med, Min, and Max. Number of participants at risk, failed and PFS rates with their CI at 3, 6, 9, 12, 18 and 24 months will also be provided. Censoring reasons will also be described. Censoring reasons are as follows:

- Ongoing in the study without an event
- No baseline assessment
- No adequate 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:

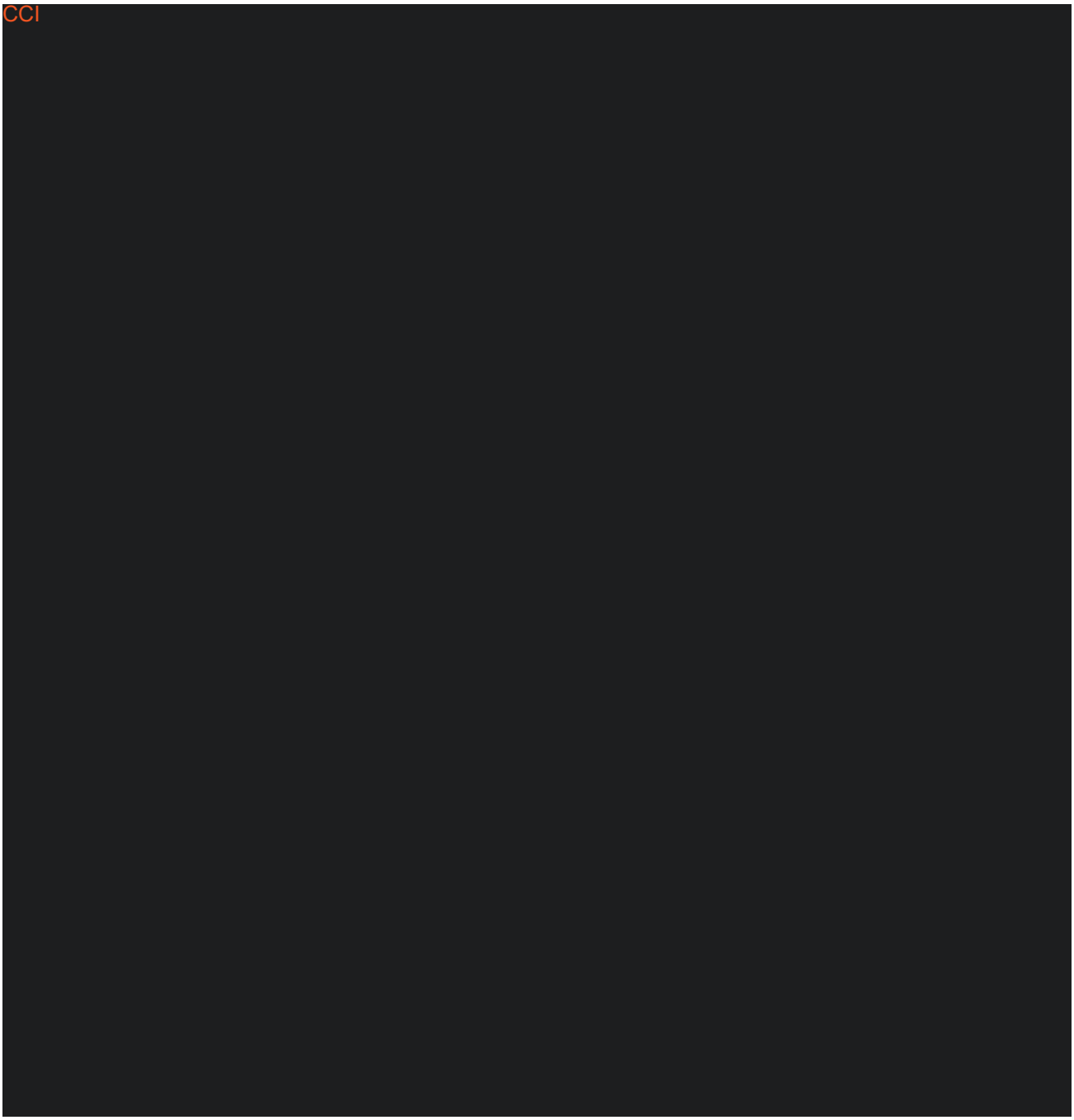
- Lost to follow-up status is collected on the eCRF treatment termination page or eCRF study termination page prior to the analysis cut-off
- Participants with the last alive date > 14 weeks prior to the analysis cut-off date (duration of 14 weeks is based on the assessment schedule of every 3 months for survival follow-up interval + 2-weeks window)
- Participants who discontinued/completed the study without an event

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

Listing will be provided 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.

CCI

[redacted]



14.2.8 Duration of Response (Secondary Endpoint)

Duration of Response (DOR) is defined for participants with a confirmed objective response as the time from first documentation of a confirmed objective response (CR or PR) according to RECIST 1.1 to the date of first documentation of objective progression of disease (PD) or death due to any cause, whichever occurs first. The censoring rules for DOR are as described for PFS (Table 7).

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

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.

Frequency (number and percentage) of participants with an event, with ongoing response at data cut-off date and without event will be presented as well as the DOR time Med, Min, and Max. Number of participants at risk, failed and DOR rates with their CI at 3, 6, 9, 12, 18 and 24 months will also be provided.

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 Med 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, 18 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 14.2.12 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, duration of response and if the response is ongoing (administrative censoring for subject ongoing in the study without an event).

CCI

14.2.10 Durable Response Rate (Secondary Endpoint)

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. Participants for whom the DOR is censored will be treated as failures (successes) in the analysis of durable response if the censored DOR is below (at least) 6 and 12 months.

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 of at least 12 months will also be provided.

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

CCI

[redacted]

14.2.13 Overall Survival (Secondary Endpoint)

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 8.

Table 8 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

Last date known to be alive will be derived as specified in Section 9.6.

The analysis of OS time will be performed with a Kaplan-Meier method with the same approach as for PFS described in section 14.2.4. 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, 18 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.

CCI

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.

Additional analyses may be performed on safety events to explore specific risk factors.

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. Such entries reporting the same event in such immediately consecutive periods will be considered as one event in the analysis (using the field "AE id for new grade" in the eCRF). 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 during on-treatment period will be flagged in listings. AE occurring during the reinitiation phase will be considered as TEAE and will be reported in the summary tables and will be flagged in the listings.

Incomplete AE-related dates will be handled as stated in Section 9.9.

Related Adverse Events are those events with relationship to study treatment (as recorded on the "Adverse Events Details" eCRF page, Relationship with Bintrafusp alfa= Related) reported by the Investigator and those of missing or unknown relationship.

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 Bintrafusp alfa" = "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 Bintrafusp alfa" = "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) which will be finalized before the database lock. Categories of AESI include:

- Infusion-Related Reactions (IRRs)
- CCI [redacted]
- TGF-β inhibition mediated Skin Reaction Anemia
- Bleeding events

The tables in this section may be repeated for China sub-population.

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 available version of MedDRA PT at the time of the database lock as event category and MedDRA primary 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 of special interest:
    - Infusion-related reactions (IRRs)
    - CCI [redacted]
    - TGF-β inhibition mediated skin adverse events
    - Anemia
    - Bleeding events
- Related AEs of special interest:
    - Infusion-related reactions (IRRs)

- TGF-β- inhibition mediated skin adverse events
- Anemia
- 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 and TEAE with NCI-CTCAE severity grade ≥3 or ≥4).

TEAEs and related TEAEs by worst grade will also be summarized, and the most frequent PTs (at least 5%) will be presented 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, PT, reported term for the AE, start date, end date, duration of AE (in days), day relative to the first infusion, day relative to the most recent infusion prior to AE onset, relationship to study treatment, toxicity grade, action(s) taken, outcome, seriousness (Y/N), AESI infusion-related (Y/N), AESI immune-related (Y/N), TGF-β inhibition mediated skin adverse events (Y/N), bleeding (Y/N), anemia AESI (Y/N), AE occurring after COVID-19 pandemic start (Y/N). TEAEs and AEs occurring during the reinitiation period specifically will be flagged.

Following listings will be provided with the relevant information:

- Listing of all AEs (whether treatment-emergent or not) (TEAEs will be flagged) (AEs occurring during reinitiation phase will be flagged)
- Listing of TEAEs
- Listing of non-TEAEs for AEs starting and not worsening 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 AEs with onset or worsening after the on-treatment period (AEs occurring during reinitiation phase will be flagged)
- Listing of all AEs for reinitiated participants

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 15.2.3.2)
- IRRs (see definition in Section 15.2.3.1)

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).

Evaluation of COVID-19 effects on AEs

The direct effect of COVID-19 for AEs will be assessed via listing of COVID-19 related AEs. The following listing will be generated using the 'COVID-19 related terms using the latest available version of MedDRA 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'. Same information as for the listing of all AEs will be provided.

15.1.2 Adverse Events Leading to Discontinuation of Study Treatment

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

- TEAEs leading to temporary treatment interruption
- Related TEAEs leading to temporary treatment 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 reinitiated participants, the last dose will be the last dose of the reinitiation 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 reinitiation 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, primary reason, autopsy (Y/N/U), AEs with fatal outcome (list PTs 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.

The tables and listings in this section may be repeated for China sub-population.

15.2.2 Serious Adverse Events

The following overall frequency tables will be prepared for serious adverse events (SAEs):

- Incidence of serious AEs by SOC and PT
- Incidence of related serious AEs by SOC and PT

The listings of SAEs will also be provided with the relevant information with a flag for TEAEs and AEs occurring during reinitiation phase (see description of listing in Section 15.1.1).

This may be repeated for China sub-population.

15.2.3 Adverse Events of Special Interest

The tables in this section may be repeated for China sub-population.

15.2.3.1 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 which will be finalized before database lock, and are divided into two subcategories: "Reactions" and "signs and symptoms" based on criteria on the timely relationship as detailed below:

Reactions of IRR: will 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: will 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.
- Time related to first onset (infusion 1/ infusion 2/ infusion 3/ infusion 4 or later). The events will be assigned to the actual drug infusions that the participant received, not to the planned dates. An IRR is assigned to a drug infusion if its onset is at the same date and time is missing or same or after dosing (but not before dosing when time is recorded) or the following day of drug infusion.

The frequency table of IRR and drug related 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.

15.2.3.2 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 PT matches a PT on the list of pre-selected MedDRA terms.

2) The AE onset or worsening occurs 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.

3) On the "Adverse Events Details" eCRF page, the question "Were Corticosteroids, Immunosuppressants, or hormonal therapy (e.g. Thyroid) or Insulin 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.

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 myositis, Other immune-mediated adverse events. PTs belonging to the category "Immune related endocrinopathies – Thyroid disorders" will also be compiled into sub-subcategories: Hyperthyroidism, Hypothyroidism, Thyroiditis.

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).

15.2.3.3 TGF-β inhibition mediated skin adverse events

To identify TGFβ inhibition mediated skin adverse events, 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. Further details (e.g. MedDRA PT queries) are regularly updated based on the current MedDRA version.

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

The overall summary of TGF-β inhibition mediated 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

Table 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. This listing will also provide, for the AE identified from the PT list, the number of lesions, if a biopsy or an excision was done and if it confirmed the diagnosis and the lesion location coming from "TGFβ MEDIATED SKIN REACTION" eCRF page.

15.2.3.4 Anemia

To identify potential anemia AEs, MedDRA PT queries will be used to search for anemia AEs of interest in the clinical database. Further details (e.g. MedDRA PT queries) are regularly updated based on the current MedDRA version.

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

A frequency tables of treatment-emergent anemia AEs and treatment-emergent related anemia AEs by worst grade (any grade, Grade ≥ 3, Grade ≥ 4, Grade 5), SOC and PT will be provided.

A listing of anemia AEs will also be provided with the relevant information (see description of listing in Section 15.1.1). Bintrafusp alfa related anemia will be flagged.

15.2.3.5 Bleeding events

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

Bleeding events will be summarized by the following variables:

- Any bleeding events
- Bleeding events by the worst grade
- Bleeding events leading to permanent treatment discontinuation
- Serious bleeding events

15.3 Clinical Laboratory Evaluation

Baseline and on-treatment laboratory values (including corresponding normal ranges), converted in standard unit, will be used for boxplots and shift tables.

Laboratory results will be classified according to the NCI-CTCAE criteria version 5.0 and as specified in Appendix 2: Definition of NCI-CTCAE grading. 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).

The following figures will be provided for hematology and biochemistry parameters:

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

Boxplots will display scheduled visits per protocol. In case two values are related to the same visit, the first result will be used.

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 2: Definition of NCI-CTCAE grading of this IAP.

Table 10 NCI-CTC Gradable parameters

Parameter	Parameter code	Name in NCI-CTC	Direction of abnormality
Biochemistry			
Alanine Aminotransferase [c]	ALT	Alanine aminotransferase increased	High
Albumin	ALB	Hypoalbuminemia	Low
Alkaline Phosphatase [c]	ALP	Alkaline phosphatase increased	High
Amylase	AMYLASE	Serum amylase increased	High
Aspartate Aminotransferase [c]	AST	Aspartate aminotransferase increased	High
Bilirubin total [c]	BILI	Blood bilirubin increased	High
Corrected Calcium[a]	CA	Hypercalcemia/Hypocalcemia[a]	High/Low
Creatinine [c]	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

Parameter	Parameter code	Name in NCI-CTC	Direction of abnormality
Hematology			
Absolute eosinophils[c]	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
Prothrombin International Normalized Ratio[b]	INR	INR increased	High

[a] based on corrected calcium (see Appendix 2: Definition of NCI-CTCAE grading).

[b] reported on the "Coagulation" eCRF page.

[c] baseline will be presented as Normal/Abnormal. In high direction low values will be accounted as normal and for low direction high values will be accounted as normal. 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 (mg/dL) = Calcium (mg/dL) – 0.8 [Albumin (g/dL)-4], or
- Corrected calcium (mmol/L) = measured total calcium (mmol/L) + 0.02 (40 – serum albumin [g/L])

Laboratory parameters with NCI-CTC grades not available

Table 11 Non-NCI-CTC Gradable Parameters

Parameter (LBTEST)	Parameter code
Biochemistry	
Bilirubin direct	BILDIR
Bilirubin Indirect	BILIND
Chloride	CL
C-Reactive Protein	CRP
Total Protein	PROT
Urea	UREA
Urea Nitrogen	BUN
Hematology	
Absolute Basophils	BASO
Absolute Eosinophils	EOS
Absolute Monocytes	MONO

Parameter (LBTEST)	Parameter code
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

* reported on the "Coagulation" eCRF page.
Calcium will be listed but not summarized

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)

In case the indirect bilirubin is not provided in the database but, at the same time point, data are provided for both total bilirubin and direct bilirubin, the missing value will be estimated using the formula: indirect bilirubin = total serum bilirubin - direct bilirubin

Separate listings of hematology (including coagulation) and biochemistry will be created. Each listing will include: participant identifier, age, sex, race, first dose date, last dose date, laboratory parameter (units), visit, date, SI value, change from baseline value, lower 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.

The laboratory shift tables may be repeated for China sub-population.

Liver function tests

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

The number and percentage of participants within each of the following liver function categories during 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 left quadrant indicates 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×ULN but not satisfying Hy's Law). Same plot will be provided for AST.

Liver function elevation may be repeated for China sub-population.

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, tuberculosis test, HIV test, serology

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

- Tuberculosis parameters: M. tuberculosis IFN Gamma Response by T-SPOT ELISPOT, Tuberculosis skin test (TST), M. tuberculosis IFN Gamma Response by QuantiFERON TB Gold ELISA
- Hormonal parameters: thyroxine free (Free T4), thyrotropin (Thyroid-Stimulating Hormone; TSH)
- HIV parameters: HIV test (HIV), HIV RNA viral load, CD4 Lymphocyte count (CD4LY)
- 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" eCRF page 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

All vital sign parameters from the on-treatment period, will be summarized using descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from baseline for each visit over time. End of treatment visit will be summarized separately. The changes computed will be the differences from baseline.

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 pulse rate
- ≥ 100 beats/min and increase from baseline ≥ 20 beats/min in pulse 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 decrease
- ≥ 10% weight increase

The tables in this section may be repeated for China sub-population.

A listing of vital signs including results for oxygen saturation 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.

15.5 Other Safety or Tolerability Evaluations

ECG

Single 12-lead ECGs will be obtained at screening and repeated during the study if clinically indicated using an ECG machine that automatically calculates the heart rate and measures PR, QRS, QT, and QTc Intervals, and ECG parameters will be derived from the data collected on the "Electrocardiogram" eCRF page.

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 (ECOG PS)

The ECOG PS 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 PS 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, visit, 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 SAF. Summaries and statistical analyses will be based on the PKAS. Only subgroup sample size with a minimum of 3 subjects will be displayed.

Pharmacokinetic samples to be collected were specified in Section 1.3 of the protocol. Pharmacokinetic concentrations/parameters refer to bintrafusp alfa concentrations/PK parameters.

16.1.1 Missing/non-quantifiable Pharmacokinetic 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 violation (as defined in the clinical study protocol), documented handling error, or analytical error (as documented in the bioanalytical report) may be excluded from the PK analysis if agreed upon prior to performing a statistical analysis. In this case the rationale for exclusion must be provided in the CSR and/or relevant listing/table.

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 (Ctrough) values in case samples are taken at least 7 days late or early

Any other PK concentrations that appear implausible to the Pharmacokineticist/PK Data Analyst must not be excluded from the analysis. Any implausible data will be documented in the CSR and/or 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 (NC)]. 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 only.

Relevant decisions on participant inclusion in the PKAS will be made before database lock in the Database Review Meeting.

16.1.2 Descriptive Pharmacokinetic 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.

Presentation of PK Parameter Data

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

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

RC_{EOI}	Accumulation ratio for C_{EOI}. Calculated as C_{EOI} Day 29 / C_{EOI} Day 1.
RC_{trough}	Accumulation ratio for C_{trough}. Calculated as C_{trough} Day X / C_{trough} Day 15. All applicable visits.

Individual PK C_{trough} and C_{EOI} will be listed by nominal study day based on the SAF. Individual PK C_{trough} and C_{EOI} will be reported with the same precision as the source data. Accumulation ratios will be reported with 3 significant figures.

Pharmacokinetic parameter data will be presented in tables and descriptively summarized by nominal study day using: n, Mean, StD, coefficient of variation (CV%), Min, Med, Max, geometric n, 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 PKAS. At least 3 values for a summary category will be required for descriptive statistics to be calculated. If n = 2, these will be presented as Min and Max. If n <2, only n will be presented. Only subgroup sample size with a minimum of 3 subjects will be displayed.

Additional table(s) will summarize PK concentrations with further stratification by ethnicity (e.g., Japanese, non-Japanese Asian, all Asian, non-Asian, Chinese mainland, non-Chinese mainland Asian), based on the PKAS. Additional table(s) will summarize C_{trough} and C_{EOI} with further stratification by ADA subsets ever positive and never positive, based on the PKADA and . 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 (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} of ADA "Treatment emergent (ALL)" versus "ADA Non-Treatment-emergent", based on the PKADA. Additional table(s) will summarize C_{trough} of ADA Treatment-emergent (ALL) subjects and nAb Treatment-emergent subjects by PK day relative to seroconversion, based on the PKADA and PKNAB, respectively. Additional table(s) may summarize C_{trough} and C_{EOI} with (further) stratification by race (Asian versus nonAsian) and ethnicity (e.g. Japanese, Chinese, etc.) subgroups or prior Bevacizumab use (Bevacizumab used versus Bevacizumab not used), as appropriate. Only subgroup sample size with a minimum of 3 subjects will be displayed.

In export datasets, as well as in the SDTM PP domain, PK parameters will be provided with full precision (i.e., as much as allowed by software) 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 SAF.

Arithmetic mean (±StD), 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 PKAS.

Additional figures will present Mean, GeoMean, and Median PK concentration-time profiles with further stratification by ethnicity (e.g., Japanese, non-Japanese Asian, all Asian, non-Asian, Chinese mainland, non-Chinese mainland Asian), based on the PKAS. Additional figure(s) will present Mean, GeoMean, and Med C_{trough} and C_{EOI} 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) may summarize C_{trough} and C_{EOI} with (further) stratification by ethnicity (e.g., Japanese, Chinese, etc.) subgroups, as appropriate. Only subgroup sample size with a minimum of 3 subjects will be displayed.

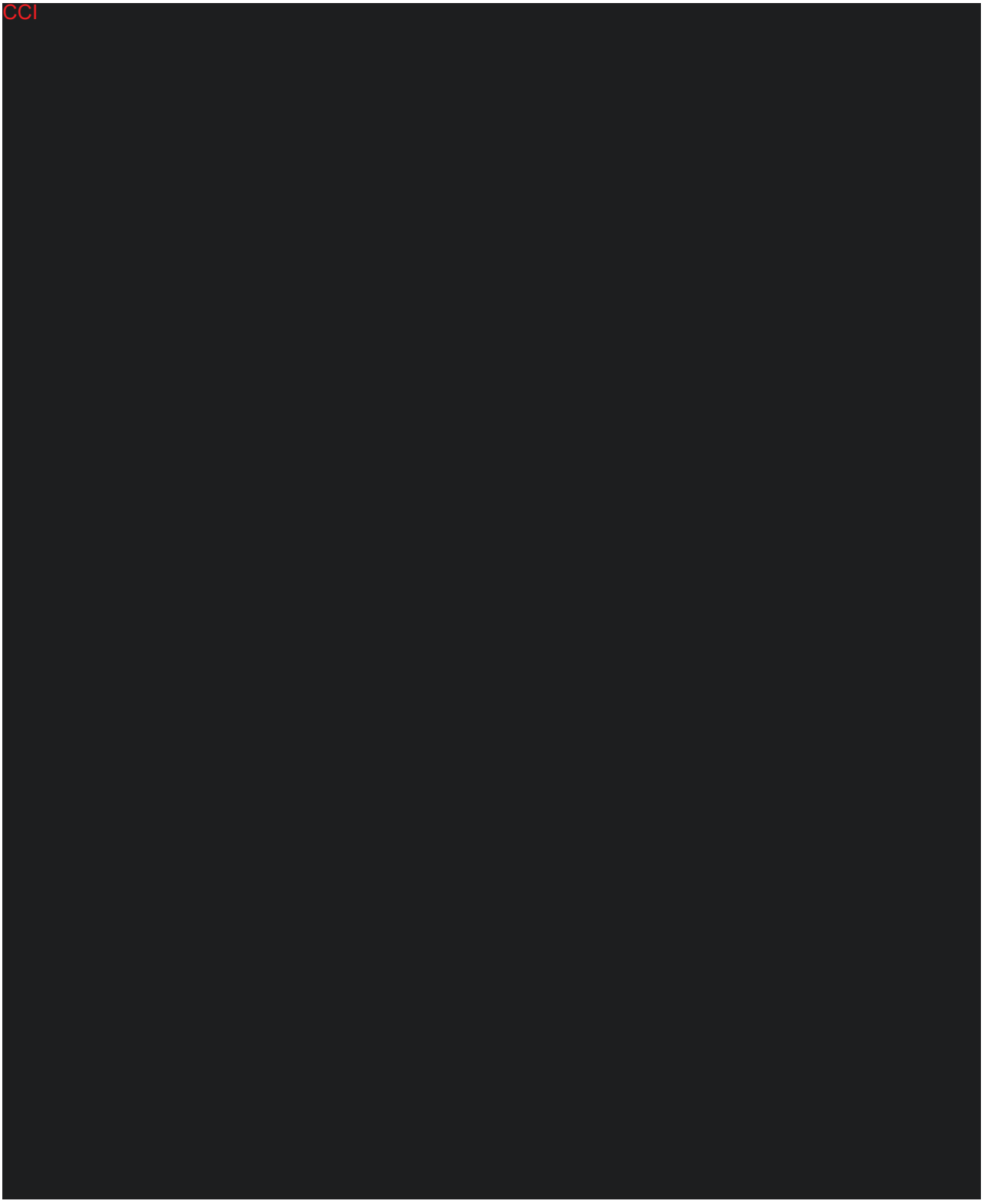
For ADA treatment-emergent (ALL) subjects with at least one C_{trough} measurement before and on/after ADA seroconversion, individual C_{trough} will be plotted versus PK day relative to seroconversion (for readability, split further into groups of 10 subjects or fewer as needed), based on the SAF. Box plots will be prepared for C_{trough} versus PK day relative to seroconversion, based on the PKADA. Additional figure(s) may summarize C_{trough} with (further) stratification by race (Asian versus non-Asian) and ethnicity (e.g. Japanese, Chinese, etc.) subgroups or prior Bevacizumab use (Bevacizumab used versus Bevacizumab not used), as appropriate. Only subgroup sample size with a minimum of 3 subjects will be displayed.

For nAb treatment-emergent subjects with at least one C_{trough} measurement before and on/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), based on the SAF. Box plots will be prepared for C_{trough} versus PK day relative to seroconversion, based on the PKNAB. Additional figure(s) may summarize C_{trough} with (further) stratification by ethnicity (e.g., Japanese, Chinese, etc.) subgroups, as appropriate. Only subgroup sample size with a minimum of 3 subjects will be displayed.

CCI

CCI

CCI


CCI

CCI

CCI

CCI

[redacted]

16.4 Immunogenicity

16.4.1.1 Antidrug Antibody

Analysis Sets: IMM

The ADA results will be derived for each visit based on the algorithm in Table 14.

Table 14 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 reinitiation 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 15.

Table 15 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 at least one post-baseline value with ADA titer result ≥ 8*baseline ADA titer result	Number of participants with valid baseline result and at least one valid post-baseline result
Treatment emergent (induced)	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)
Treatment emergent (ALL)	Treatment-emergent plus treatment-boosted	Number of participants with at least one valid post-baseline result at any timepoint
ADA Non-Treatment-emergent	All subjects minus treatment-emergent All	Number of participants with at least one valid post-baseline result at any time point
Transient positive	If treatment emergent participants have - a single positive evaluation and last assessment not positive 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)

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 for treatment-emergent ALL

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

- (Date of last positive ADA assessment – date of first positive ADA assessment + 1) / 7 for treatment-emergent participants

(Date of last positive ADA assessment - date of first boosted (titer ≥ 8×baseline titer) ADA assessment + 1) / 7 for treatment boosted participantsThe following analysis will be described:

- The frequency and percentage of each ADA category, as defined in Table 15, 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 onset of response for treatment-emergent ALL participants will be summarized
- The duration of ADA immunogenicity response for treatment-emergent ALL participants will be summarized

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 14.2.1)
    - for duration of confirmed response according to RECIST 1.1 as adjudicated by IRC (see Section 14.2.8)
    - for durable confirmed response according to RECIST 1.1 as adjudicated by IRC (see Section 14.2.10)
    - for overall survival (see Section 14.2.13)
- Evaluation of potential effect of ADA on bintrafusp alfa Pharmacokinetics (see Section 16.1)

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: IMM

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 16. Treatment boosted is not defined for nAb.

Table 16 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 17 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 16.4.1.1 for the derivation of time to first positive response and duration of response.

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 14.2.4)
    - for duration of confirmed response according to RECIST 1.1 as adjudicated by IRC (see section 14.2.8)
    - for durable confirmed response according to RECIST 1.1 as adjudicated by IRC (see section 14.2.10)
    - for overall survival (see section 14.2.13)
- 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).

CCI

[redacted]

16.5.1 PD-L1 Expression (Secondary Endpoint)

Clinical efficacy of bintrafusp alfa will be evaluated according to PD-L1 expression measured by SP263 central lab testing as a Secondary Endpoint.

- PD-L1 expression in tumor tissue by SP263 is defined as: vCPS>=5%, <5%, NE or NA

CCI

[redacted]


CCI

17 References

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D.,Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer, National Cancer Institute of Canada-Clinical Trials Group, 10 Stuart Street, Queen's University, Kingston, Ontario, Canada. eeisenhauer@ctg.queensu.ca, 2009;45:228-47

Fayers, P.M., Aaronson, N.K., Bjordal, K., Groenvold, M., Curran, D., Bottomley, A., on behalf of the EORTC Quality of Life Group. *EORTC QLQ-C30 Scoring Manual*. 3rd edn. EORTC, Brussels; 2001.

Greimel ER, Kuljanic Vlasic K, Waldenstrom AC, et al; European Organization for Research and Treatment of Cancer Quality-of-Life Group. The European Organization for Research and Treatment of Cancer (EORTC) Quality-of-Life questionnaire cervical cancer module: EORTC QLQ-CX24. Cancer. 2006;107(8):1812–1822.

Hua, C. H. et al. Validation of the European Organization for Research and Treatment of Cancer cervical cancer module for Chinese patients with cervical cancer. Patient Prefer. Adherence 7, 1061–1066 (2013).

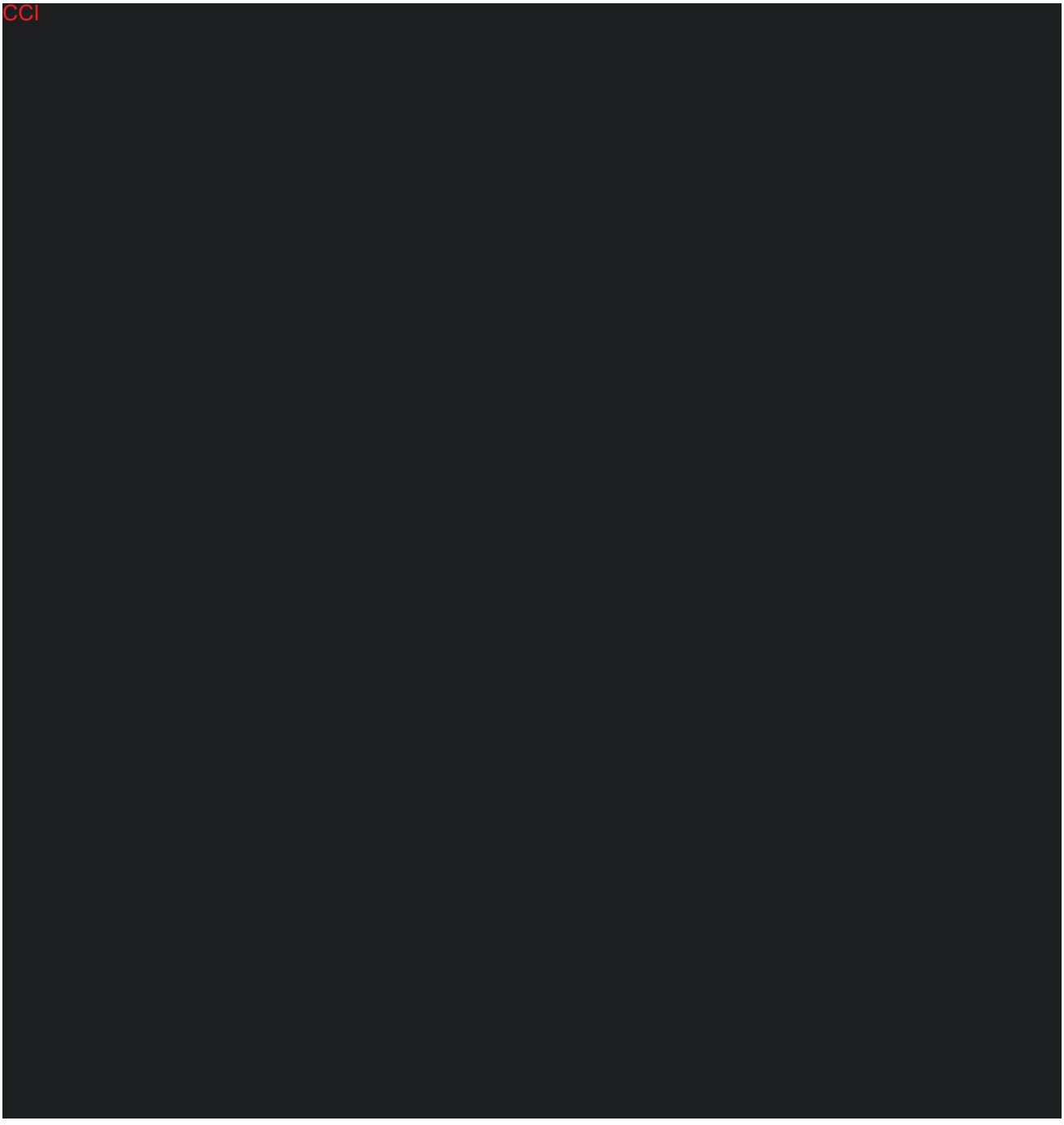
18 Appendices

18.1.1 Appendix 1: Definition of important protocol deviations

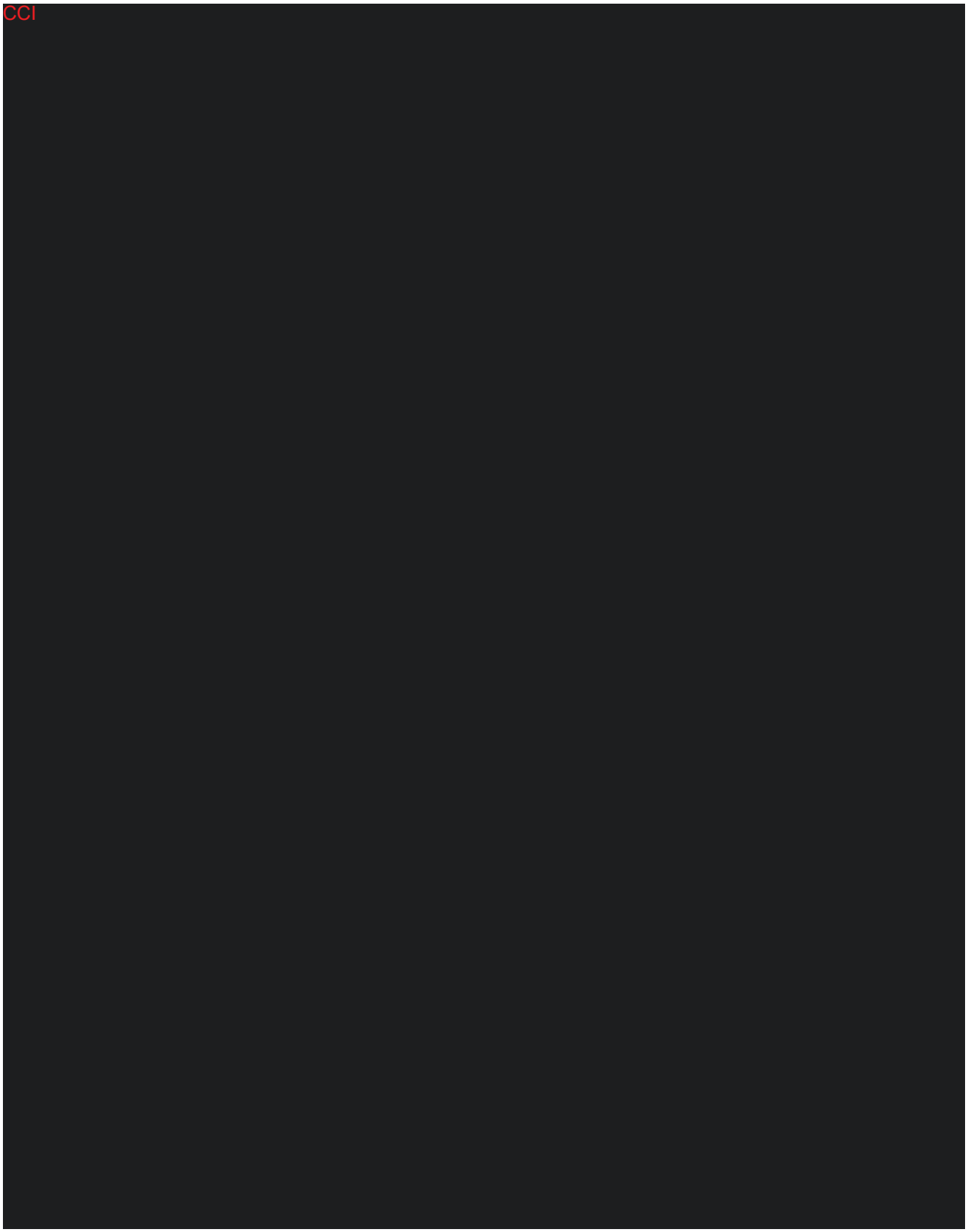
Refers to: csp-ms200647-0017-iap-appendix-protocol-deviation-classification-v1.pdf

18.1.2 Appendix 2: Definition of NCI-CTCAE grading

Refer to: csp-ms200647-0017-iap-appendix-nci-ctcae-grading-v1.xlsx



CCI



CCI

Signature Page for VV-CLIN-287969 v2.0

Approval	PPD

Signature Page for VV-CLIN-287969 v2.0