Integrated Analysis Plan

Clinical Study Protocol Identification No.	MS200647-0037
Title	An Adaptive Phase III, Multicenter, Randomized, Open-Label, Controlled Study of M7824 (bintrafusp alfa) versus Pembrolizumab as a First-line Treatment in Patients with PD-L1 Expressing Advanced Non-small Cell Lung Cancer
Trial Phase	Phase III
Investigational Medicinal Product(s)	Bintrafusp alfa
Clinical Study Protocol Version	10 February 2020/Version 3.0
Integrated Analysis Plan Author	**Coordinating Author** PPD, Merck PPD **Function** — **Author(s) / Data Analyst(s)** PPD
Integrated Analysis Plan Date and Version	14 July 2021 / Version 5.0
Integrated Analysis Plan Reviewers	**Function** — **Name** PPD — PPD PPD, Merck PPD, Merck PPD, Merck PPD, Merck PPD, Merck PPD, Merck PPD, Merck PPD, Merck PPD, Merck PPD, Merck PPD, Merck

Approval Page

Integrated Analysis Plan: MS200647-0037

An Adaptive Phase III, Multicenter, Randomized, Open-Label, Controlled Study of M7824 (bintrafusp alfa) versus Pembrolizumab as a First-line Treatment in Patients with PD-L1 Expressing Advanced Non-small Cell Lung Cancer

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

Merck Data Analysis Responsible	Date	Signature
PPD	Via CARA approval process	

1 Table of Contents

2 List of Abbreviations and Definition of Terms

CCI	
AE	Adverse event
AESI	Adverse events of special interest
ALK	Anaplastic lymphoma kinase
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
aPTT	Activated partial thromboplastin time
AST	Aspartate aminotransferase
BOR	Best overall response
C_{eoi}	The concentration observed immediately at the end of infusion
CI	Confidence Interval
COVID-19	Coronavirus Disease 2019
CSR	Clinical Study Report
CR	Complete response
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
C_{trough}	The concentration observed immediately before next dosing
DCR	Disease Control Rate
DOR	Duration of Response
DI	Dose Intensity
ECG	Electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	Electronic case report form
EGFR	Epidermal growth factor receptor
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer-Quality of Life Questionnaire Core-30
EQ-5D-5L	European Quality of Life 5-dimensions 5-level questionnaire
FAS	Full analysis set
FT4	Free thyroxine
1L	First-line
GeoMean	Geometric Mean

HBV	Hepatitis B virus
HCV	Hepatitis C virus
HLT	High-level Term
HLGT	High-level Group Term
HR	Hazard ratio
IAP	Integrated Analysis Plan
ICH	International Council for Harmonization
IHC	Immunohistochemical
IDMC	Independent Data Monitoring Committee
IL-29	Item List-29
CCI [redacted]	
IMP	Investigational Medicinal Product
INR	International normalized ratio
irAE	Immune-related AE
CCI [redacted]	
IRC	Independent Review Committee
CCI [redacted]	
IRR	Infusion-related reactions
CCI [redacted]	
IRT	Interactive Response Technology
MedDRA	Medical Dictionary for Regulatory Activities
NC	Not Calculated
NCI	National Cancer Institute
NSCLC	Non-small cell lung cancer
NSCLC-SAQ	Non-small cell lung cancer symptom assessment questionnaire
OR	Objective Response
ORR	Objective response rate
OS	Overall survival
PD	Progressive disease
PD-L1	Programmed death-ligand 1
PFS	Progression-free survival

PFS2	Progression-free survival after next anticancer treatment
CCI	[redacted]
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PK	Pharmacokinetic
PR	Partial response
PRESCR	Prescreening analysis set
PRO	Patient-reported outcomes (it may refer also to patient-reported outcomes analysis set)
PT	Preferred Term
RDI	Relative Dose Intensity
RECIST 1.1	Response Evaluation Criteria in Solid Tumors Version 1.1
SAE	Serious adverse event
SAF	Safety analysis set
SCR	Screening analysis set
SD	Stable disease
SDTM	Study Data Tabulation Model
SMQ	Standardised MedDRA Queries
SOC	System Organ Class
StD	Standard Deviation
2L	Second-line
TEAE	Treatment Emergent Adverse Events
TGFβ	Transforming growth factor β
TSH	Thyroid-stimulating hormone
CCI	[redacted]
ULN	Upper limit of normal
VAS	Visual analog scale

3 Modification History

Unique Identifier for Version	Date of IAP Version	Author	Changes from the Previous Version
1.0	06 February 2019	PPD	NA
2.0	30 March 2020		• Changed drug name from M7824 to bintrafusp alfa • Added PROs: IL-29, PGIS and PGIC in the list of abbreviations, "Objectives and Endpoints" and "Overview of Planned Analyses" sections • Study objectives (according to protocol v. 3.0) • Changes in overview of planned analyses and efficacy analyses (according to global protocol amendment, protocol v. 3.0) • Replaced subject with participant • Renamed ITT analysis set as FAS analysis set • [redacted] • Merged IDMC SAP v.1.1 with the current IAP • Added clarification in the definition of baseline for safety analyses • Added clarification in the definition of on-treatment period • Added imputation rules for previous and concomitant medications, start date of subsequent anticancer therapies, death dates • Added handling rules for tumor assessments • Updates in participant disposition table • Renamed IVRS with IRT • Added listing of participants for which study treatment has been reinitiated • ECOG Performance status moved in "Other Baseline Characteristics" section • Changes in Nicotine Consumption table • Changes in PD-L1 table • Added details regarding handling of data after treatment reinitiation • Renamed administration with intervention • Added analysis "Follow-up Time since Randomization" • Added analyses for the evaluation of potential effect of ADA on bintrafusp alfa safety, efficacy and PK • Added listing of AEs occurring after treatment reinitiation • Clarifications on safety section and changes in Immune-related Adverse Events definition • Added "Bleeding events" section • Added clarification on treatment duration period regarding reinitiation

Unique Identifier for Version	Date of IAP Version	Author	Changes from the Previous Version
			• Added eDISH plot for AST values • Removed section on PRO, it will be provided in a separate Appendix of the IAP • Changed testing strategy: Holm group sequential design provided • [redacted] • Provided new wording for censoring reasons (administrative censoring) • [redacted] • Provided new wording for Three-Tier approach analysis • Added details for WBC differential counts and calcium in the "Clinical Laboratory Evaluation" section • Specified type of analyses performed on coagulation parameters
3.0	23 November 2020	PPD	• Added COVID-19 abbreviation • Removed secondary objectives of ORR and PFS assessed by Investigator as per protocol v.3.0 and added these analyses as sensitivity analyses for BOR and PFS in the Efficacy Analyses section • Updated overview of planned analyses for PFS IA as per Merck request: added forest plot of HR by subgroups, removed analyses on PRO questionnaires, removed OR by PD-L1 status and AESI by PD-L1 status • Added section 7.1 COVID-19 Impact • Added more detailed definition of baseline and added definition of study day/study treatment day • Section General Specifications for Data Analyses: added definition of pre, during, post pandemic period for COVID-19 impact assessment. • Updated sections "Disposition of participants and Discontinuations", "Protocol Deviations" "Adverse Events" including COVID-19 analyses details • Updated lost to follow-up definition replacing "censoring" date with "last known alive" date • Updated definition of Immune-related Adverse Events • Added "Lip squamous cell carcinoma" and "Bowen's disease" in the broad definition of potential TGFbeta-mediated skin AEs • Added section "Anemia adverse events"

Unique Identifier for Version	Date of IAP Version	Author	Changes from the Previous Version
			• Removed unnecessary formula for calcium derivation • Added Creatinine Clearance derivation
4.0	22 April 2021	PPD	• Added section 16.2 "Pharmacokinetics" • [redacted] • Added the computation description for the duration of nicotine consumption • Nicotine consumption: years since quitting, clarifications added in the computation description • Added imputation rules for dates of nicotine consumption • Added summary tables for important/non-important COVID-19 related PDs • Added listing of concomitant COVID-19 vaccinations • Removed previous medications text as this section has been removed from e-crf as it was out of study objectives • Added analyses of potential effect of n-AB on safety CCI [redacted] • Updates applied to section "Three-Tier Approach to Summarizing and Analyzing AEs" • Added section "Overview of Planned Analyses after Trial Discontinuation" • Definition of on-treatment period: added reference to expanded on-treatment period for irAEs • Added flag variable for bleeding events in AEs listings • Added flag variable of possible Hy's law cases in the listing of total bilirubin, ALT, AST and ALP values during on-treatment period • Removed Appendix "Rules for the identification of previous, concomitant and both previous and concomitant medications/procedures", not applicable due to the previous medications section removal
5.0	14 July 2021	PPD	CCI [redacted]

Unique Identifier for Version	Date of IAP Version	Author	Changes from the Previous Version
			• ADA: changed the category name in Table 11 from "Treatment emergent (baseline non-positive)" to "Treatment emergent (baseline induced)" • ADA and Nab: removed sentence regarding censoring for subjects still on treatment at data cut-off and positive assessment at their last assessment before cut-off • n-AB: modified "Subject at Risk (Denominator for Incidence)" in Table 13. • n-AB: removed analyses on titer values

4 Purpose of the Integrated Analysis Plan

The purpose of this Integrated Analysis Plan (IAP) is to document technical and detailed specifications for the analyses at periodic safety evaluation by IDMC, interim and primary analyses of data collected for protocol MS200647-0037. Results of the analyses described in this IAP will be included in the Clinical Study Report (CSR). Additionally, the planned interim and primary analyses identified in this IAP will be included in regulatory submissions or future manuscripts. Any post-hoc, or unplanned analyses performed to provide results for inclusion in 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.

5 Objectives and Endpoints

The objectives and endpoints are defined in Table 1.

Table 1 Study Objectives and Endpoints

Objectives	Endpoints (Outcome Measures)	IAP Section
Primary		
To demonstrate improvement in PFS with bintrafusp alfa compared with pembrolizumab To demonstrate improvement in OS with bintrafusp alfa compared with pembrolizumab	PFS according to RECIST 1.1 assessed by IRC OS	Efficacy Analyses, sections 14.4, 14.11
Secondary		
• Safety To evaluate the safety and tolerability of bintrafusp alfa compared with pembrolizumab	• Occurrence of TEAEs and treatment-related AEs	Safety Analyses, section 15
• Efficacy To evaluate the efficacy in objective response of bintrafusp alfa compared with pembrolizumab	• Objective response according to RECIST 1.1 assessed by IRC	Efficacy Analyses, section 14.1
- DOR	• DOR assessed from CR or PR according to RECIST 1.1 assessed by IRC until PD, death, or last tumor assessment	Efficacy Analyses, section 14.8
CCI		
CCI		


CCI

Objectives	Endpoints (Outcome Measures)	IAP Section
CCI		
Long-term follow-up of survival for 5 years after the last dose of bintrafusp alfa or pembrolizumab in a participant unless reported as lost to follow-up, had died, or the study is terminated.		

ADA = anti-drug antibodies; AE = adverse event; C_{eoi} = concentration at end of infusion; C_{trough} = concentration at the end of the dosing interval; CR = complete response; DOR = duration of response; EORTC QLQ-C30 = European Organization for Research and Treatment of Cancer-Quality of Life Questionnaire Core-30; EQ-5D-5L = European Quality Of Life 5-dimensions 5-level questionnaire; IHC = immunohistochemistry; IL-29 = Item List-29;IRC = Independent Review Committee; CCI NSCLC = non-small cell lung cancer; NSCLC-SAQ = non-small cell lung cancer symptom assessment questionnaire; OR = objective response; ORR = objective response rate; OS = overall survival; PA = primary analysis; PD = progressive disease; CCI 1; PFS = progression-free survival; survival; CCI PGIC = Patient Global Impression of Change; PGIS = Patient Global Impression of Severity; PK = pharmacokinetic; PR = partial response; PRO = patient-reported outcome; RECIST = Response Evaluation Criteria in Solid Tumors; TEAE = treatment-emergent adverse event; TMB = tumor mutation burden.

6 Overview of Planned Analyses

The following analyses are described in this IAP:

- Periodic Independent Data Monitoring Committee (IDMC) Safety Evaluations
- Objective Response Interim analysis (OR IA)
- Progression-Free Survival Interim analysis (PFS IA)
- PFS Primary analysis (PFS PA)
- Overall Survival Interim Analysis (OS IA)
- Overall Survival Primary Analysis (OS PA)

Details of outputs belonging to each of the analyses listed above are provided in a separate document ("IAP_MS200647-0037 List of outputs") provided in Appendix to this IAP.

Analysis for periodic IDMC safety evaluations

The IDMC will periodically review accumulating safety data. The IDMC will evaluate quality and completeness of safety data based on the provided study data. Moreover, the IDMC will provide recommendations regarding modification and/or continuation of the study, including any additional safety monitoring or risk mitigation procedures to ensure participant safety and the validity and scientific merit of the study.

The first IDMC safety data review meeting took place when the first 60 participants have completed 6 weeks of follow up post randomization. The following IDMC safety meetings will be held every 6 months following the initial IDMC meeting. Further ad-hoc meetings may be scheduled by the IDMC as needed. For each IDMC meeting, all data will be included up to an administrative cutoff date.

Adaptive Phase III Trial Design and analyses before/after adaptation decision

In protocol version 3.0, the study design was modified to an adaptive trial design to permit sample size adaptation based on prespecified adaptation rules evaluating the objective response at a pre-planned interim analysis.

Analysis for adaptation decision (OR IA)

The study's IDMC will evaluate the results from the pre-defined interim analysis (OR IA) based on objective response as assessed by the IRC on the 73-10 PD-L1 high analysis set restricted to participants randomized at least 6 months before the data cutoff date which is planned at 6 months after the randomization of the 100th participant tested 73-10 PD-L1 high. In this analysis, the IDMC will decide, based on ORR difference between treatment arms per study design, to either keep the study running at the initial sample size (N=300), or expand the sample size (N=584), or stop for

futility. The ORR criterion for expansion is prespecified in the corresponding confidential IDMC charter appendix. The study is stopped for futility (non-binding) at the OR IA, if $\Delta^{ORR} \leq 0$ (i.e. the response rate in the bintrafusp alfa arm is not higher than in the pembrolizumab arm).

Analyses after adaptation decision (PFS and OS IA and PA)

If IDMC will decide to keep the study running, either at the initial sample size or at an expanded sample size, the PFS and OS interim and primary analyses will be provided at the cut-off dates determined by a pre-specified number of observed events. The number of events triggering the analyses differ according to the total sample size and are displayed in the table below (Table 2).

Table 2 Cutoff dates for the planned PFS and OS IA and PA analyses

Analysis	Data cutoff date	
	Initial Sample Size (N=300)	Expanded Sample Size (N=584)
PFS IA	When 140 Progression-free survival (PFS) events have been reached	When 140 Progression-free survival (PFS) events have been reached
PFS PA*	When 224 Progression-free survival (PFS) events have been reached	When 386 Progression-free survival (PFS) events have been reached
OS IA*	When 169 Overall survival (OS) events have been reached	When 270 Overall survival (OS) events have been reached
OS PA	When 197 Overall survival (OS) events have been reached	When 339 Overall survival (OS) events have been reached

*On discretion of the sponsor the same cutoff can be used for OS IA and PFS PA for operational reasons if the number of events for PFS PA is reached and at the same time the number of OS events is in the range of ± 5% of the planned event size for OS IA. A Lan-DeMets O'Brien-Fleming approximation spending function will be used for calculation of efficacy bounds for PFS and OS at corresponding analyses. Efficacy boundaries will be recalculated on the basis of the actual number of events reached at cut-off date.

Testing strategy for PFS and OS

The 2 primary endpoints of this study are PFS according to RECIST 1.1 assessed by IRC and OS. The study will be considered positive if either the OS analysis results and/or the PFS analysis results are statistically significant.

The following efficacy analyses are planned: PFS IA, PFS PA, OS IA, OS PA.

- **Primary endpoint PFS**

The following null hypothesis will be tested:

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

Where$\lambda^{PFS}(t)$ represents the hazard at time t and θ the unknown constant of proportionality of hazards in the treatment groups *M* (M7824/bintrafusp alfa) and *P* (pembrolizumab).

- **Primary endpoint OS**

The following null hypothesis will be tested:

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

Where $\lambda^{OS}(t)$ represents the hazard at time t and θ the unknown constant of proportionality of hazards in the treatment groups *M* (M7824/bintrafusp alfa) and *P* (pembrolizumab).

The type I error is strictly controlled at a level of 2.5% one-sided based on work by Chen et al. (Chen 2018). Bonferroni method is used to split the alpha of 2.5% and use 0.5% for testing H_0^{PFS} and 2% for testing H_0^{OS} (all one-sided).
Group sequential boundaries for PFS and OS endpoints will be derived from the alpha-spending function according to Lan-DeMets (O'Brian-Fleming boundaries) at local significance levels α^{PFS} for PFS and α^{OS} for OS, respectively (all one sided).
The Holm group sequential design will be used (Ye et al 2013): monitor PFS endpoint using its local α^{PFS} and its specific group sequential boundary c^{PFS} and monitor OS endpoint using its local α^{OS} and its specific group sequential boundary c^{OS}. If either of the two endpoints (PFS or OS) crossed its corresponding boundary at some specific look (either PFS IA or PFS PA or OS IA or OS PA) then its local type I error can be reallocated to the other endpoint so that the other endpoint can be tested using the full level alpha boundary (i.e α = 2.5%) taking into account the group-sequential testing procedure.
The family-wise error rate is controlled at 2.5% (one-sided) for the described testing procedure. An appendix to this IAP, "Adaptation Plan", includes simulation results, which show that this holds true under broad variation of assumptions.
The multiple primary results will be considered statistically significant only if either of the two individual hypotheses H_0^{PFS} or H_0^{OS} are rejected by the closed test at any look. Operating characteristics of the study design are detailed in Table 3 below:

Table 3 Operating Study Characteristics for Phase II/III Design

Analysis		Initial Sample Size (N=300)				Expanded Sample Size (N=584)			
		Planned Events	Critical Values		Local Power	Planned Events	Critical Values		Local Power
			HR	p-value			HR	p-value	
PFS	IA	140	0.57	0.00038	17%	140	0.47	$3.15*10^{-6}$	2%
	PA	224	0.71	0.00487	50%	386	0.77	0.00500	90%

Analysis		Initial Sample Size (N=300)				Expanded Sample Size (N=584)			
		Planned Events	Critical Values		Local Power	Planned Events	Critical Values		Local Power
			HR	p-value			HR	p-value	
OS	IA	169	0.71	0.01201	52%	270	0.75	0.00914	71%
	PA	197	0.74	0.01658	14%	339	0.80	0.01731	18%

HR = hazard ratio; IA = interim analysis, OS = overall survival, PA = primary analysis, PFS = progression-free survival. Hazard ratio is an expected one based on the assumption of exponential distribution. The decision will be based on the p-value. Alpha is distributed to the PFS and OS analyses based on Bonferroni's method.

Unblinding

The study team is blinded for all analyses conducted by treatment arm until the IDMC recommends unblinding.

Until the unblinding is recommended, the analyses will be performed by an independent statistical provider for the IDMC and they will be transmitted from this group to the IDMC only.

After the prospectively determined data cutoff date is reached (displayed in Table 2) the data center statistician will prepare the outputs (using programs prepared by the blinded team based on dummy treatment arms) in agreement with the IDMC charter and transmit the analyses, tabulations, and listings to the IDMC for the meeting. The data center statistician will be available at the IDMC meeting should any questions from the IDMC members arise regarding the data and/or analyses.

6.1 Periodic IDMC Safety Evaluations

Periodic IDMC safety evaluations will be taken every 6 months following the initial IDMC meeting (performed when the first 60 participants have completed 6 weeks post randomization) and will evaluate:

- Participant disposition
- Important protocol deviations
- Enrollment details (participants randomized by region and country, randomized by strata)
- Demographics and other baseline characteristics (height, weight and BMI at baseline, disease history)
- Treatment compliance and exposure (duration of therapy)
- Safety analyses (overview of treatment-emergent adverse events (TEAEs), serious TEAEs (SAE), treatment-related TEAEs, adverse events of special interest (AESI), Bleeding events, clinical laboratory evaluations, vital signs)

Further details can be found in the IDMC Charter.

6.2 OR Interim Analysis

The OR interim analysis will be triggered at 6 months after the randomization of the 100th participant tested 73-10 PD-L1 high and will evaluate:

- Unconfirmed best overall response per RECIST 1.1 as assessed by IRC (Difference in ORR, see Section 14 for details) and reasons for non-evaluable unconfirmed best overall response.
- Duration of response

Participant disposition, important protocol deviations, enrollment details (participants randomized by region and country, randomized by strata), demographics and other baseline characteristics (height, weight and BMI at baseline, disease history, PD-L1 assays concordance), duration of therapy and safety analyses (overview of treatment-emergent adverse events (TEAEs), serious TEAEs (SAE), treatment-related TEAEs, adverse events of special interest (AESI), Bleeding events, clinical laboratory evaluations, vital signs) will also be described at the time of BOR interim analysis. Further evaluations on PD-L1 assays may be included in this analysis (on request of the IDMC) e.g. unconfirmed OR by PD-L1 assays status and/or TEAEs and AESI by PD-L1 assays status (as described in sections 14.1.1 and 15.5). Additionally descriptive statistics of demographics and baseline characteristics in subjects with and without bleeding events may be included on IDMC request (as described in section 15.2.4).

6.3 PFS Interim Analysis

The PFS Interim analysis will be triggered by the data cutoff date displayed in Table 2 and will describe:

- Best overall response per RECIST 1.1 as assessed by IRC
- Best overall response per RECIST 1.1 as assessed by Investigator
- PFS per RECIST 1.1 as assessed by IRC (H_0^{PFS} tested see Section 14 for details)
- PFS per RECIST 1.1 as assessed by Investigator
- PFS per RECIST 1.1 as assessed by IRC subgroup analyses: forest plot of HR and its corresponding 95% confidence interval (CI) by subgroups
- Duration of response

Participant disposition, important protocol deviations, enrollment details (participants randomized by region and country, randomized by strata), demographics and other baseline characteristics (height, weight and BMI at baseline, disease history, PD-L1 assays concordance), duration of therapy, cumulative dose, dose intensity and relative dose intensity, safety analyses (overview of treatment-emergent adverse events (TEAEs), serious TEAEs (SAE), treatment-related TEAEs, adverse events of special interest (AESI), bleeding events, clinical laboratory evaluations, vital signs) will also be described at the time of PFS interim analysis. Additional descriptive statistics

of demographics and baseline characteristics in subjects with and without bleeding events may be included on IDMC request (as described in section 15.2.4).

6.4 PFS Primary Analysis and OS Interim Analysis

The PFS primary analysis and OS Interim analysis will be triggered by the data cutoff dates displayed in Table 2 and will include the following sections of this IAP:

- Participant disposition
- Important protocol deviations
- Enrollment details (participants randomized by region and country, randomized by strata)
- Demographics, medical history and other baseline characteristics (height, weight and BMI at baseline, disease history, skin status history, nicotine consumption, tumor biopsy, PD-L1 assays concordance)
- Previous and concomitant medications and procedures
- Treatment exposure and compliance (e.g. duration of treatment, number of infusions and dose intensity)
- Efficacy analyses:
    - PFS per RECIST 1.1 as assessed by IRC (H_0^{PFS} tested, see Section 14 for details)
    - PFS per RECIST 1.1 as assessed by Investigator
    - PFS per irRECIST as assessed by IRC
    - BOR per RECIST 1.1 as assessed by IRC
    - BOR per RECIST 1.1 as assessed by Investigator
    - BOR per irRECIST as assessed by IRC
    - Change from baseline and Maximum shrinkage in sum of diameter of target lesions according to RECIST 1.1 assessed by IRC
    - Time to response
    - Duration of response
    - Overall survival (OS) (H_0^{OS} tested, see Section 14 for details). Alpha re-allocation is possible if H_0^{PFS} is rejected previously

- Safety analyses (overview of treatment-emergent adverse events (TEAEs), serious TEAEs (SAE), treatment-related TEAEs, adverse events of special interest (AESI), Bleeding events, ECOG PS, vital signs, clinical laboratory evaluations, and electrocardiograms [ECG])

- Analyses on Patient-Reported Outcomes (PRO) (EORTC-QLQ-C30, NSCLC-SAQ, EQ-5D-5L, PGIS, PGIC and IL-29). See section 16.4

CCI [redacted]

- Further evaluations on PD-L1 assays may be included in this analysis (on request of the IDMC) e.g. OR by PD-L1 status and/or TEAEs and AESI by PD-L1 assays status (as described in sections 14.1.1 and 15.5)

- Additionally descriptive statistics of demographics and baseline characteristics in subjects with and without bleeding events may be included on IDMC request (as described in section 15.2.4).

6.5 OS Primary Analysis (Final analysis)

The following analyses will be described in the OS primary analysis, triggered by the data cutoff point displayed in Table 2:

- Participant disposition

- Important protocol deviations

- Enrollment details (participants randomized by region and country, randomized by strata)

- Demographics, medical history and other baseline characteristics (height, weight and BMI at baseline, disease history, skin status history, nicotine consumption, tumor biopsy, PD-L1 assays concordance)

- Previous and concomitant medications, procedures, follow-up treatments

- Treatment exposure and compliance (e.g. duration of treatment, number of infusions, and dose intensity)

- Efficacy analyses:

    - PFS per RECIST 1.1 as assessed by IRC (H_0^{PFS} tested, see Section 14 for details). Hypothesis tested only if not rejected in previous analyses and alpha re-allocation is possible if H_0^{OS} is rejected

    - Overall survival (OS) Overall survival (OS) (H_0^{OS} tested, see Section 14 for details). Alpha re-allocation is possible if H_0^{PFS} is rejected previously

    - BOR on the subsequent line of therapy, CCI [redacted]

- Safety analyses (overview of treatment-emergent adverse events (TEAEs), serious TEAEs (SAE), treatment-related TEAEs, adverse events of special interest (AESI), Bleeding events, ECOG PS, vital signs, clinical laboratory evaluations, and ECG)

- Analyses on PRO (EORTC-QLQ-C30, NSCLC-SAQ,EQ-5D-5L, PGIS, PGIC and IL-29). See section 16.4

CCI [redacted]

- Further evaluations on PD-L1 assays may be included in this analysis e.g. TEAEs and AESI by PD-L1 assays status (as described in section 15.5)

- Additionally descriptive statistics of demographics and baseline characteristics in subjects with and without bleeding events may be included (as described in section 15.2.4).

6.6 Overview of Planned Analyses after Trial Discontinuation

At the time of the PFS IA the IDMC recommended to terminate the study. Based on this recommendation, the Sponsor decided to discontinue the clinical study MS200647-0037 as the study was unlikely to meet the primary endpoints. Therefore, the analyses described in Sections 6.1 - 6.5 will not be performed as planned, but one primary analysis will be performed for reporting purpose. The analyses considered relevant for this purpose are listed below:

- Participant disposition

- Important protocol deviations

- Enrollment details (participants randomized by region and country, randomized by strata)

- Demographics, medical history and other baseline characteristics (height, weight and BMI at baseline, disease history, ECOG at baseline, skin status history, nicotine consumption)

- Demographics and baseline characteristics in subjects with and without bleeding events

- Previous and concomitant medications, procedures, follow-up treatments

- Treatment exposure and compliance (e.g. duration of treatment, number of infusions and dose intensity)

- Efficacy analyses:

    - PFS per RECIST 1.1 as assessed by IRC

    - PFS per RECIST 1.1 as assessed by Investigator

- BOR per RECIST 1.1 as assessed by IRC
- Duration of response
- Overall survival (OS)

- Safety analyses (overview of treatment-emergent adverse events (TEAEs), serious TEAEs (SAE), treatment-related TEAEs, adverse events of special interest (AESI), Bleeding events, ECOG PS, vital signs, clinical laboratory evaluations)

CCI [redacted]

7 Changes to the Planned Analyses in the Clinical Study Protocol

The statistical methods as described in the protocol will be adopted. There are no changes to the planned analyses, except additional outputs that will be generated to assess the potential impact of COVID-19 pandemic and described in the Section 7.1 below.

7.1 COVID-19 Impact

No changes to the planned analysis of the efficacy endpoints will be performed due to the impact of COVID-19 outbreak. Instead, additional outputs (summary tables and listings) will be generated to assess potential impacts of COVID-19 to this study.

- An overview table of the impact by COVID-19
- An overview table of participants who started the treatment pre/during COVID-19 study period
- A listing of COVID-19 impact
- A listing of COVID-19 related protocol deviations
- A table of treatment-emergent adverse events associated to COVID-19
- A listing of AEs related to COVID-19.
- A listing of COVID-19 vaccinations.

Details of the categorization of participants for COVID-19 impact assessment is provided in Section 9 and details of the analyses are provided in Sections 10.1, 10.2, 15.1.1, 12.1.

8 Protocol Deviations and Analysis Sets

8.1 Definition of Protocol Deviations and Analysis Sets

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

Important protocol deviations will be identified and confirmed prior to or at each Cross-Functional Data Review Meetings and will include:

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

Important protocol deviations are listed in a specific document (Merck template CS_TP_MS_PD001), deviations to be checked programmatically are identified in this list. Important protocol deviations list is provided in Appendix 18.3 of this IAP.

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

Further considerations for PK analysis:

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

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

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

Refer to Section 16.2 for more details of protocol deviations and handling relevant to PK.

8.2 Definition of Analysis Sets and Subgroups

Prescreening Analysis Set (PRESCR)

The prescreening analysis set includes all participants who signed the prescreening informed consent.

Screening Analysis Set (SCR)

The screening analysis set includes all participants who signed the main informed consent and who are not pre-screening failures, regardless of the participant's randomization and study intervention status in the study.

Full Analysis Set (FAS)

The full analysis set (FAS) will include all participants who were randomized to either bintrafusp alfa or pembrolizumab arm. Analyses performed on the FAS set will take into account participants' allocation to treatment arms as randomized.

Full Analysis Set with 6 months follow-up (FAS with 6m f-up)

The FAS with 6m f-up analysis set will include all participants randomized at least 6 months before cut-off date which is planned at 6 months after enrolling the 100th participant tested 73-10 PD-L1 high.

Safety Analysis Set (SAF)

The safety set will include all participants who were administered any dose of any Investigational Medicinal Product (IMP, bintrafusp alfa or pembrolizumab). Analyses performed on the safety set will consider participants as treated. Participants will be classified according to the treatment received at randomization unless the incorrect treatment(s) was/were received throughout the dosing period in which case participants will be classified according to the first IMP intervention received.

CCI

CCI

Table 4 displays the use of the analysis sets in the different analyses:

Table 4 Overview of the Analysis Set Used in the Analyses

Analyses	ALL	FAS	CCI	FAS with 6m f-up	SAF	PRO	CCI	PKAS
Participant disposition	✓							
Important Protocol Deviations		✓						
Enrollment details		✓						
Demographics		✓			✓			
Baseline Assessments		✓			✓*			
Past and Concomitant Therapies		✓						
Compliance and Exposure					✓			
Efficacy: Primary		✓						
Efficacy: Secondary, OR IA								
Efficacy: Secondary, OR IA sensitivity analyses				✓				
Efficacy: Secondary		✓						
Safety and Tolerability					✓			
PRO questionnaires						✓		
CCI								
CCI								

*Only the following tables: Disease history; Height, Weight and Body Mass Index.

For including baseline variables into Cox's proportional hazards model and logistic models the following parameterization is to be used. For variables with more than two categories, an indicator variable will be defined for each category except for the first category, which always defines the reference. In case of low number of participants within a category (i.e. 20 participants), categories may be pooled. CCI [redacted]

The following subgroups will be defined:

Age

- age < 65 years (Ref)
- age ≥ 65 years

Age 2:

- age < 75 years (Ref)
- age ≥ 75 years

Sex

- Male (Ref)
- Female

Race

- Caucasian / White (Ref)
- Asian
- Other

Ethnicity 1:

- Japanese
- Not Japanese (Ref)

Ethnicity 2:

- Hispanic or Latino
- Not Hispanic or Latino (Ref)

ECOG PS at baseline

- ECOG PS 0 (Ref)
- ECOG PS 1

NSCLC Histology:

- Squamous cell carcinoma (Ref)
- Non-squamous cell carcinoma

Smoking status:

- Never smoker (Ref)
- Ever smoker

PD-L1 test used for participant enrollment:

- 22-C3[1]
- 73-10 (Ref)

Prior radiotherapy:

- Yes
- No (Ref)

Prior cytotoxic therapy:

- Yes
- No (Ref)

Brain lesions at baseline:

- Yes
- No (Ref)

Pooled Region:

- North America (Ref)
- Europe
- Asia
- Latin America

ADA status:

- Ever positive (Ref)
- Never positive

[1] Enrollment by local CCI [redacted] is possible only if the pre-screening informed consent was signed up until and including 23 September. From 24 September 2019 onwards only enrollment by Central 73-10 PD-L1 (DAKO) assay was allowed.

9 General Specifications for Data Analyses

Unless otherwise indicated all analyses will be presented separately for the two treatment arms (bintrafusp alfa and pembrolizumab).
For the IDMC meetings, open session tables will be presented with treatment arms combined (only total column will be displayed); closed session tables will include separate columns for the two treatment arms and a total column when appropriate.
The summary of outputs to be produced for open and/or closed session for IDMC meetings is presented in Table 5.

Table 5 Outputs to be produced for open and/or closed sessions of the IDMC meetings

Analyses	IDMC meeting open session	IDMC meeting closed session
Participant disposition status	✓	✓
Important Protocol Deviations	✓	✓
Enrollment details	✓	
Summary of timeliness/cleanliness of Data	✓	
Demographics and other baseline characteristics	✓	✓
Treatment duration		✓
Efficacy analyses*		✓
Safety analyses		✓

* Efficacy analysis will be provided in addition to outputs for regular IDMC safety meetings only for efficacy assessments IDMC meetings.

Data handling after cutoff date:

Data after the cutoff will not be displayed in any listings or used for summary statistics, e.g. laboratory values of samples taken after data cutoff, AEs with onset date after data cutoff, etc. will not be included in any analysis or listing.
Stop dates are not affected by this rule, e.g. a stop date of AEs, which starts prior to the cutoff, but stopped after date of cutoff, will not be changed.
Data cut will occur before the analysis database creation process, i.e. on the SDTM in input to the process for the ADAM creation.

Pooling of centers:

Because of the high number of participating centers and the anticipated small number of participants randomized in each center data will be pooled across centers, and the factor center will not be considered in statistical models or for subgroup analyses.

Unscheduled assessments

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

- For shift tables, they will be taken into account in the definition of the worst assessment during the study
- For description at each time post-baseline point, the first available result (in chronological order) will be taken into account in the analysis in case of multiple values.

For description at baseline, the last available result before randomization will be taken into account in the analysis in case of multiple values.

Presentation of continuous and qualitative variables:

Continuous variables will be summarized using descriptive statistics, i.e.

- number of participants (N), number of participants with missing values
- mean, standard deviation (StD)
- median, 25th Percentile - 75th Percentile (Q1-Q3),
- minimum and maximum.

If there are no missing values this should be indicated by a 0.

Qualitative variables will be summarized by counts and percentages.

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

Definition of baseline and change from baseline:

The last non-missing measurement prior to randomization or prior to the first dose of IMP will serve as the baseline measurement, depending of the type of analysis. Details are given in the table below:

Table 6 Baseline date to be used in the analyses

Analysis	Date of randomization*	Date of start of first dose of IMP
Disposition	✓	
Protocol deviations	✓	
Demographics and others baseline characteristics	✓	
Medical history	✓	
Concomitant medications, Previous/ Subsequent anticancer therapy		✓
Treatment exposure and compliance		✓
Efficacy (OS, PFS, BOR…)	✓	
Safety (TEAE, SAE, AESI…)		✓
Laboratory evaluation		✓
Vital signs		✓
Other safety (ECG/ECOG)		✓
PRO	✓	
CCI		

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

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

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

Absolute and percent changes from baseline are defined as:

absolute change = timepoint value – baseline value

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

Study Day/Treatment Day

Study day 1 is relative to randomization and treatment day 1 is relative to first dose of IMP, the use of study day 1 or treatment day 1 depend on the type of analysis, as detailed in Table 6 of this document.

Definition of on-treatment period:

The on-treatment period is defined as the time from the first IMP intervention to the last IMP intervention date + 30 days OR the earliest date of subsequent anticancer therapy minus 1 day, whichever occurs first, unless otherwise stated. Dates of subsequent anticancer therapy include dates of subsequent anticancer drug therapy, dates of subsequent anticancer radiotherapy, dates of subsequent anticancer surgery. Note that the on-treatment period will include the initial treatment period as well as the reinitiation of treatment period, as applicable. Whether a participant reinitiates treatment (following the rules as outlined in the protocol) or not, the on-treatment period is defined as the time from the first IMP intervention administration to the last IMP 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 analyses on immune-related adverse events an expanded on-treatment period will be used. See section 15.2.3.2 for details.

Definition of duration:

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

The time since an event (e.g. time since first diagnosis) will be calculated as reference date minus date of event.

Time since initial cancer diagnosis

Time since initial cancer diagnosis (months) = (date of randomization – date of initial cancer diagnosis)/30.4375

Time since documented, locally advanced, inoperable or metastatic disease

Time since documented, locally advanced, inoperable or metastatic disease (months) = (date of randomization – date of documented, locally advanced, inoperable or metastatic disease)/30.4375

Nicotine consumption: duration of consumption

The duration of nicotine consumption is computed as follows:

- If a subject is a former smoker, the difference in years between end date and start date of nicotine consumption is computed (end date of nicotine consumption – start date of nicotine consumption+1) / 365.25

- If a subject is a current smoker, the difference in years between nicotine consumption collection date and start date of nicotine consumption is computed. (nicotine consumption collection date – start date of nicotine consumption+1) / 365.25

Nicotine consumption: years since quitting

Years since quitting is computed for former smokers only as (collection of nicotine consumption date - end date of smoking+1)/365.25.

Conversion factors:

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

Handling of missing data:

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

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

Handling of incomplete dates:

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

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

Incomplete dates for disease history (date of initial cancer diagnosis, date of documented, locally advanced, inoperable or metastatic disease diagnosis) and dates of nicotine consumption:

- 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.

Incomplete dates for concomitant medications will be imputed as follows:

For start date of medication:

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

For end date medication:

- If the day is missing, it will be imputed to the last day of the month.
- If both day and month are missing, the month and day will be imputed as December 31st
- If the date is completely missing, no imputation will be performed.

Note: In case the imputation results in a date later than the date of patient's death, then the date of death will be used to impute the incomplete stop date.

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

- If only day is missing, it will be imputed as the last day of the month unless the end date of subsequent anticancer therapy is before that date. In that case, the incomplete anticancer therapy start date will be imputed as the end date of the anticancer therapy.
- If both day and month are missing, no imputation will be performed.

Incomplete subsequent anticancer therapy stop dates will not be imputed.

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

Incomplete date of death:

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 between:
  - The day after the last known alive date (non-imputed date) and
  - 15th day of the month
- Otherwise it will not be imputed

Outside survival analyses no death imputation is done.

Incomplete date of first dose of IMPs:

In case the date of first dose of IMP is missing, it is assumed that the first dose of IMP is given at the randomization date. The randomization date will replace incomplete dates of the first dose of IMPs. 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.

Handling rules for tumor assessments

If there are multiple scan dates associated with an evaluation, the earliest of the scan dates associated with the evaluation will be used as the date of assessment.

Preferred term for analysis of WHO-DD coded data:

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

Data collected after reinitiated treatment:

Data collected after reinitiation of treatment will be included in the summary statistics. Data listings will include both data collected during the first period of treatment, and during the reinitiation of treatment (except for listings displaying parameters derived during the first treatment period only). Data collected during the reinitiation of treatment will be flagged.

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). If a participant is re-screened several times, then he will be counted only once in the disposition table in the number of re-screened participants.

Definition of Pre-, During-, Post-pandemic Time Periods

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

- Pre-pandemic time period: subjects who started the treatment before the start of COVID-19 pandemic.
- During-pandemic time period: subjects who started the treatment on the same date or after the start of COVID-19 pandemic
- Post-pandemic time period: the end of pandemic is considered not yet reached anywhere at the time of this IAP version redaction, consequently no participant will be grouped into the post COVID-19 study period for this study and no post-pandemic date will be defined.

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

distribution-covid-19-cases-worldwide) or 11th March 2020 (when the WHO declared COVID-19 pandemic).

Software:

All analyses will be performed using SAS 9.4 or higher in the SAS Grid environment.

10 Study Participants

The subsections in this section include specifications for reporting participant disposition and treatment/study discontinuations. Additionally, procedures for reporting protocol deviations are provided (see Section 10.2).

10.1 Disposition of Participants and Discontinuations

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

The following information will be reported:

- Total number of participants prescreened
- Number of participants who did not continue beyond prescreening overall and grouped by reason (reason related to central testing, reason related to local testing, reason independent of local/central testing)
- Number of participants who completed the pre-screening (pre-screening ICF signed), but have not signed the main ICF yet
- Total number of screened participants
- Total number of re-screened participants
- Number of participants who did not continue beyond screening overall and grouped by reason (participant did not meet all eligibility criteria, withdrew informed consent, progressive disease, adverse event, lost to follow-up, death, other)
- Number of participants who completed the screening (main ICF signed), but have not been randomized yet
- Number of randomized participants overall and by treatment arm
- Number of randomized participants who did not receive any treatment
- Number of randomized participants with treatment on-going in each treatment arm
- Number of randomized participants off treatment, grouped by treatment arm and main reason
- Number of randomized participants with treatment reinitiated
- Number of randomized participants with treatment on-going after reinitiation
- Number of randomized participants who discontinued treatment after reinitiation

- Number of randomized participants who discontinued the treatment but are still in study for follow-up, grouped by treatment arm
- Number of randomized participants who discontinued the study, grouped by treatment arm and main reason (as reported on "Study termination" page of the eCRF).

In addition, the number of participants in each analysis set defined in Section 8.2 will be summarized. The percentage of participants with percentage based on the number of participants in the FAS population will be presented. The disposition of participants may also be presented with some PD-L1 assay information in a CONSORT diagram (e.g. PD-L1 assay test performed for the inclusion; Dako 73-10, CCI results). The discrepancies between planned and treated arms will be cross-tabulated and provided in a listing. The number of randomized participants as per Interactive Response Technology (IRT) by region and country will be provided, as well as the number of participants by randomization strata (as per IRT).

For the IDMC meetings, timeliness and Cleanliness of the data will be summarized as follows:

For participants who are still on treatment:

- Number of participants with last on-study safety information (AEs, Labs, VS, Other safety evaluations: ECG, ECOG) entered into the data extract
  - Within 1 month prior to the data cutoff date
  - Between 1 and 2 months prior to the data cutoff date
  - Between 2 and 3 months prior to the data cutoff date
  - More than 3 months prior to the data cutoff date
  - No information

For participants who have had their end of treatment visit:

- Number of participants in the follow-up phase with and without follow-up information
- Among participants with follow-up information, number of participants without a death report in the follow-up phase
- Among participants with follow-up information and no death report in the follow-up phase, number of participants with last follow-up date (from the Participant Status/Survival Follow-Up eCRF page)
  - Within 1 month prior to the data cutoff date
  - Between 1 and 2 months prior to the data cutoff date
  - Between 2 and 3 months prior to the data cutoff date
  - More than 3 months prior to the data cutoff date

The listing of participant disposition will include all participants (i.e. pre-screened participants, pre-screening failures, participants included in screening, screening failures, but not re-screened participants). The listing will include the following information (if applicable): planned arm (participants having actual arm different from planned arm will be flagged and displayed in a separate listing), participant identifier, date of prescreening informed consent, included in the screening phase and reason for exclusion, date of main informed consent, included in the study, reason for exclusion, randomization date, first/last treatment date, reason off-treatment, date and reason off-study, population flags. When the reasons, such as the reason off-treatment is categorized as "Other, specify" or "Withdrew consent from treatment, specify", the verbatim text as entered in the eCRF will be presented in the listing.

If any re-screened participants will be observed, the listing of re-screened participants will be provided and will include the following information: planned arm, participant identifier, date of informed consent, randomization date, participant identifier at screen failure, date of informed consent at screen failure, date of screening failure, reason of screening failure.

In addition, a listing of participants for which study treatment has been reinitiated will be provided with the following information: participant identifier, date of randomization, planned treatment, date of first IMP intervention, date of last IMP intervention, date and reason for treatment termination (if treatment is discontinued), as applicable.

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

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

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

The frequency distribution of participants who started the treatment before or during the pandemic per country-specific start of COVID-19 study period (see Section 9 for details on the COVID-19 categorization by study period) will be also displayed.

A listing of COVID-19 impact will be provided with the following information: treatment arm, participant identifier, first and last date of treatment administration, date of the event, visit, category, event, event description/reason

10.2 Protocol Deviations

Important protocol deviations will be determined for all participants by either medical review processes or programming based on the inclusion/exclusion criteria or other criteria presented in the protocol and documented in SDTM. The following summary tables and listings of important protocol deviations will be provided (separately for pre-/post inclusion deviations):

- Frequency table per reason of important protocol deviations
- Listing of important protocol deviations.

Potential impact of COVID-19 pandemic in MS200647-0037 will be evaluated by an analysis of protocol deviations:

- The summary of important COVID-19 related PDs by category and type of PD will be provided
- The summary of non-important COVID-19 related PDs by category will be provided
- The listing of important protocol deviations will be enriched by the variable "COVID-19 Related Protocol Deviations" flagging all protocol deviations due to COVID-19.

11 Demographics and Other Baseline Characteristics

Summaries will be presented for the FAS analysis set.

11.1 Demographics

Demographic characteristics will be summarized as follows:

Demographic characteristics (from "Demographics" e-CRF page)

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

- Age categories:
  - < 65 years, ≥ 65 years
  - 65-74, 75-84 and ≥85 years
- Geographic Region:
  - North America
  - Europe
  - Asia and Pacific
  - Latin America

Specifications for computation:

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

Handling of incomplete dates of birth or informed consent is specified in Section 9 of this IAP, paragraph "Handling of incomplete dates".

Investigator site codes will be used for the determination of the participant's geographic region.

A listing of individual demographics will be also provided.

11.2 Medical History

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

Medical history will be displayed in terms of frequency tables: ordered by primary SOC and preferred term (PT) in alphabetical order. Each participant will be counted only once within each PT or SOC.

The related listing will also be provided.

11.3 Other Baseline Characteristics

Analyses of baseline characteristics with respect to ECGs and clinical laboratory evaluations are discussed in Section 15 (Safety Analyses).

11.3.1 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).

BMI (kg/m2) will be computed as weight(kg)/[height(m)]^2

11.3.2 Disease History

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

- Tumor histology: adenocarcinoma, adenosquamous carcinoma, squamous cell carcinoma, large cell carcinoma, sarcomatoid carcinoma, neuroendocrine carcinoma, diffuse idiopatic carcinoma, other
- Time since initial cancer diagnosis (months), computation details provided in Section 9
- Time since documented, locally advanced, inoperable or metastatic disease (months), computation details provided in Section 9
- TNM classification at initial diagnosis and at study entry: each T, N, M category will be described (TX, T0, N1, etc.)
- Molecular abnormalities:
    - EGFR (mutated, wild)
    - ALK translocation (yes, no)
    - ROS1 rearrangement (yes, no)
    - BRAF V600E mutation (yes, no).

Disease history will also be presented in a dedicated listing.

11.3.3 ECOG Performance status

The ECOG Performance Status will be derived 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.

11.3.4 Skin Status History

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

- Personal history of frequent sunburn (Yes, No, Unknown)
- Personal history of easy sunburn (Yes, No, Unknown)
- Personal history of skin cancer (Yes, No, Unknown)
- Personal history of significant UV exposure (Yes, No, Unknown)
- Personal history of photosensitivity due to skin disorder (Yes, No, Unknown)
- Personal history of 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 conditions, 1 condition, 2 conditions, 3 or more conditions).

11.3.5 Nicotine Consumption

The nicotine consumption information will be collected from the "Nicotine consumption" e-CRF page and it will be described as follows:

- Nicotine use: Never user, Ever user (including further breakdown: former / current) as collected in e-CRF
- Smoking exposure (pack-years): 0, <20, 20-<40, ≥40 and summary statistics
- Years since quitting: never smoker, current smoker, <5, 5-<10, ≥10 and summary statistics

A listing of nicotine consumption, including treatment arm, participant identifier, age, sex, race, will also be produced with the following data: nicotine use status, start/end date of nicotine consumption, and duration of consumption (years).

12 Previous or Concomitant Medications/Procedures

Summaries will be presented for the FAS analysis set

12.1 Concomitant Medications

Concomitant medications are medications, other than IMPs and premedications for IMP, which are taken by participants any time on-treatment (see on-treatment definition in Section 9), medications taken on the first day of IMP will be considered on-treatment. In case the date values will not allow to unequivocally allocate a medication to concomitant medication the medication will be considered as concomitant medication. Medications started before the signing of the main informed consent and continue past the time of the signing will be categorized as concomitant medications.

Concomitant medications are reported on the "Concomitant medications details" eCRF page.

Summaries on concomitant medications will present the number and percentage of participants by drug class and preferred term according to the most recent available version of the WHO-DD dictionary.

A participant will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. The summary will be sorted by Drug Class and Preferred Term in alphabetical order.

Concomitant medications will be presented in a listing. The listing will include treatment arm, participant identifier, age, sex, race, preferred term and medication name as provided by the investigator, start date, end date, dose, dose units, frequency, route, reason for the medication. The listing will be sorted by treatment arm, participant identifier, start date, end date (note that missing end date are considered as ongoing and will be displayed after non-missing date in case of same start date) and preferred term.

A separate listing of concomitant COVID-19 vaccinations will be provided, this listing will include the same information as concomitant medications listing.

12.2 Premedications

Premedications are medications administered per protocol on the same day as, but prior to, bintrafusp alfa intervention. In the protocol v.3.0 bintrafusp alfa premedication is an optional procedure.

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

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

Premedications will be presented in a listing from "Premedication Details" eCRF page. The listing will include participant identifier, age, sex, race, medication name, visit, date/time of intervention, dose, dose units, and route. The listing will be sorted by participant identifier, premedication start date and medication name.

12.3 Concomitant Procedures

All concomitant procedures, which were undertaken any time during on-treatment period are collected into the CRF page "Concomitant Procedures Details".

Concomitant procedures will be presented in a listing including treatment arm, participant identifier, age, sex, race, name of procedure (as provided by the investigator), start date, end date, indication, and reason for procedure. A flag will be added to identify each procedure as prior to treatment and on-treatment. The listing will be sorted by treatment arm, participant indentifier, start date, end date (note that missing end date are considered as ongoing and will be displayed after non-missing date in case of same start date) and procedure name.

12.4 Previous Anticancer Treatments and Procedures

The previous anticancer treatments and procedures are collected under the "Prior Anticancer Drug Therapies Details", "Prior Anticancer Radiotherapy Details", and the "Prior Anticancer Surgeries Details" eCRF pages.

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

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

Previous anticancer drug therapy will be summarized as follows:

- Type of previous anticancer therapy: Cytotoxic therapy/Endocrine therapy/Monoclonal antibodies therapy/Small molecules/Immunotherapy/ Other.
- Intent of therapy: Neoadjuvant /Adjuvant
- Best response of last treatment regimen: complete response (CR)/partial response (PR)/stable disease (SD)/non-complete response/non-progressive disease (Non CR/Non PD), progressive disease (PD)/not evaluable/unknown.

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

The previous anticancer drug listing will contain treatment arm, participant identifier, age, sex, race, preferred term, medication name, start date, end date, type of therapy, intent of therapy, best response and criteria on the last treatment regimen, and date of progression on the regimen. This listing will be sorted by treatment arm, participant identifier, anticancer drug start date, end date and preferred term

The previous radiotherapy listing will contain treatment arm, participant identifier, age, sex, race, start date, end date, prior radiotherapy to bone for palliation only, site of therapy, total dose, number of fractions and best response on the treatment regimen. This listing will be sorted by treatment arm, participant identifier, radiotherapy start date, radiotherapy end date and site of radiotherapy.

The previous anticancer surgery listing will contain treatment arm, participant identifier, age, sex, race, date of surgery, name and site of surgery, curative intent of surgery (Y/N), and outcome of surgery. This listing will be sorted by treatment arm, participant identifier, surgery date and name of surgery.

12.5 Anticancer Treatment after Discontinuation

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

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

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

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

be described, as well as the intent of therapy (Neoadjuvant/Adjuvant) and the best response (CR, PR, stable disease, PD, non-CR/non-PD, not evaluable, unknown).

In addition, the anticancer treatment after discontinuation of IMP will be provided in a data listing including type of therapy, preferred term/medication name, medication type, regimen name, best response, start date, end date, radiotherapy site and name of surgery/location.

13 Treatment Compliance and Exposure

Summaries will be performed on SAF analysis set.

Participants will be treated with bintrafusp alfa or pembrolizumab at the following doses:

- Bintrafusp alfa at a dose of 1200 mg per infusion once every 2 weeks
- Pembrolizumab at a dose of 200 mg per infusion once every 3 weeks.

These doses will be administered as "flat" doses independent of the body weight. Both treatments will be administered until PD per RECIST 1.1, unacceptable toxicity, or for up to 24 months.

All dosing calculations and summaries will be based on "Study Treatment Administration Details" eCRF page. Data collected during the treatment reinitiation phase will be included in the summary statistics.

A dose is regarded to be administered, if the actual dose received is > 0

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

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

with "last dose" being the last dose of reinitiation period, if applicable.

The duration of pembrolizumab treatment (in weeks) during the study is defined as:

$$\text{duration}=\left(\frac{\text{date of last dose-date of first dose+21}}{7}\right)$$

with "last dose" being the last dose of reinitiation period, if applicable.

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

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

$$\text{dose intensity}=\left(\frac{\text{Cumulative dose}}{(\text{duration of therapy (in weeks)})/2}\right)$$

The dose intensity (DI) (mg/cycle) of pembrolizumab per 3-week period is defined as:

$$\text{dose intensity}=\left(\frac{\text{Cumulative dose}}{(\text{duration of therapy (in weeks)})/3}\right)$$

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

RDI (%) = 100 * [DI (mg/cycle)/(planned dose per cycle) (mg)]

For participants who reinitiate the treatment, the varying duration of cycle will be taken into account.

The following summary tables will be provided:

- Duration of therapy (weeks)
- Total number of infusions received overall
- Cumulative dose (mg)
- Dose intensity (mg/cycle)
- Relative dose intensity (%).

A listing of treatment exposure will include treatment arm, participant identifier, age, sex, race, visit, infusion start date and time, infusion end date and time, infusion rate (mL/hr), actual dose (mg), administration location, administration modification and reason for modification, change in administration detail, treatment delay (days). This listing will be sorted by treatment arm, participant identifier, infusion start date and infusion end date.

An additional listing of treatment exposure will include treatment arm, participant identifier, age, sex, race, duration of therapy (weeks), total number of infusions received, cumulative dose of therapy (mg), dose intensity (mg/cycle), and relative dose intensity (%). This listing will be sorted by treatment arm and participant identifier.

Therapy Delays

Delays of therapy will be derived by infusion as follows:

- For bintrafusp alfa interventions as the no. of days since last infusion - 14
- For pembrolizumab interventions as the no. of days since the start of last infusion – 21.

If the value above is >0 days, 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 delays
- Longest delay per participant (no delay, 1-2 days, 3-8 days, 9-15 days, ≥16 days)
- Number of delays per participant (0 delay, 1 delay, 2 delays, 3 delays, ≥ 4 delays).

Infusion Temporary Interruptions

IMP infusion 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 IMP interruption, reason for infusion temporary interruptions, as well as a categorization of the number of IMP interruptions (1 / 2 / ≥3) will be summarized.

Infusion Rate Reductions

Infusion rate reductions as recorded on the "Study Treatment Administration Details" page of the eCRF will be used for analysis. Number of participants with at least one infusion rate reduced to 50%, reasons for infusion rate reduction, as well as a categorization of the number of infusion rate reductions (1 / 2 / ≥3) will be summarized.

14 Efficacy Analyses

Best Overall Response efficacy analyses at OR interim analysis will be performed on 73-10 PD-L1 High analysis set, if any Best Overall Response sensitivity analyses will be performed at OR interim analysis it will be based on FAS with 6m f-up analysis set. Remaining efficacy analyses will be presented for the FAS analysis set.

A group sequential testing procedure will be used to test primary endpoints. CCI [redacted]

14.1 Best Overall Response According to RECIST 1.1 as Assessed by IRC (Secondary Endpoint)

Best overall response (BOR) will be assessed based on the tumor response at different evaluation time points from randomization until the first documented disease progression for RECIST 1.1 criterion. Only tumor assessments performed before the start of any further anticancer treatment will be considered in the assessment of BOR. Clinical deterioration will not be considered as documented disease progression.

The BOR according to RECIST 1.1 criteria as assessed by IRC can be CR, PR, Non-CR/Non-PD, SD, PD, or NE.

The following rules will be applied to the derivation of BOR:

- CR = at least one determination of CR before disease progression
- PR = at least one determination of PR before disease progression (and not qualifying for a CR)
- SD = at least one SD assessment ≥ 6 weeks after start date and before disease progression (and not qualifying for CR or PR)
- As tumor lesions are evaluated by the IRC, it may happen that the independent reviewer disagrees with the Investigator and does not assess any tumors as "measurable" at screening.

In that case, the BOR could be rated as "non-CR/non-PD" (if no "CR" or "PD" are previously reported and if the "non-CR/non-PD" assessment is ≥ 6 weeks after start date).

- PD = progression ≤ 12 weeks after randomization (and not qualifying for CR, PR, SD or non-CR/non-PD)
- If a participant has a missing baseline tumor assessment and/or no (or NE) tumor assessments on-treatment, BOR will be Not Evaluable (NE).
- In the case the single response is SD, SD must have been assessed no less than 6 weeks (at least 42 days) after randomization, otherwise, the best response will be NE.

The number and percentage of participants with unconfirmed BOR of CR, PR, Non-CR/non-PD, SD, PD, and NE will be tabulated by treatment arm.

A listing will be provided with the following information: treatment arm, participant identifier, age, sex, race, date of randomization, first/last dose, unconfirmed BOR, date of subsequent anticancer therapy (i.e. drug therapy, radiotherapy or surgery), date of death, visit, date(s) of imaging, lesion type (target lesion, non-target lesion, new lesion), description of target lesions (size, site, type, method, response), non-target lesions (status, site, type, method, response) and new lesions (site, type, method), sum of lesion diameters, percent change in target lesions from baseline, and overall response, sorted by treatment arm, participant identifier, visit, date of imaging and lesion ID.

A summary table of the reasons for non-evaluable BOR as adjudicated by the IRC by treatment arm will be provided:

- No baseline assessment
- No post-baseline assessments due to death within 6 weeks after randomization
- 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 randomization without further evaluable tumor assessment). Note: Special and rare cases where BOR is NE due to both early SD and late PD will be classified into this category.
- Non-CR/non-PD of insufficient duration (<6 weeks after randomization without further evaluable tumor assessment)
- PD too late (i.e. tumor assessment of PD was >12 weeks after randomization and there was no evaluable tumor assessment in between)
- No IRC review may also be added if required by the data
- If the IRC is not able to identify any disease at baseline (target or non-target lesions), the BOR may be rated as "No Disease" (ND), "No Disease" may category may also be added if required by the data.

A listing of reasons for non-evaluable BOR will also be created including: treatment arm, participant identifier, age, sex, race, date of randomization, date of first and last dose, date(s) of imaging, overall response, date of subsequent anticancer therapy (i.e. drug therapy, radiotherapy or surgery) and the reason for unconfirmed BOR NE.

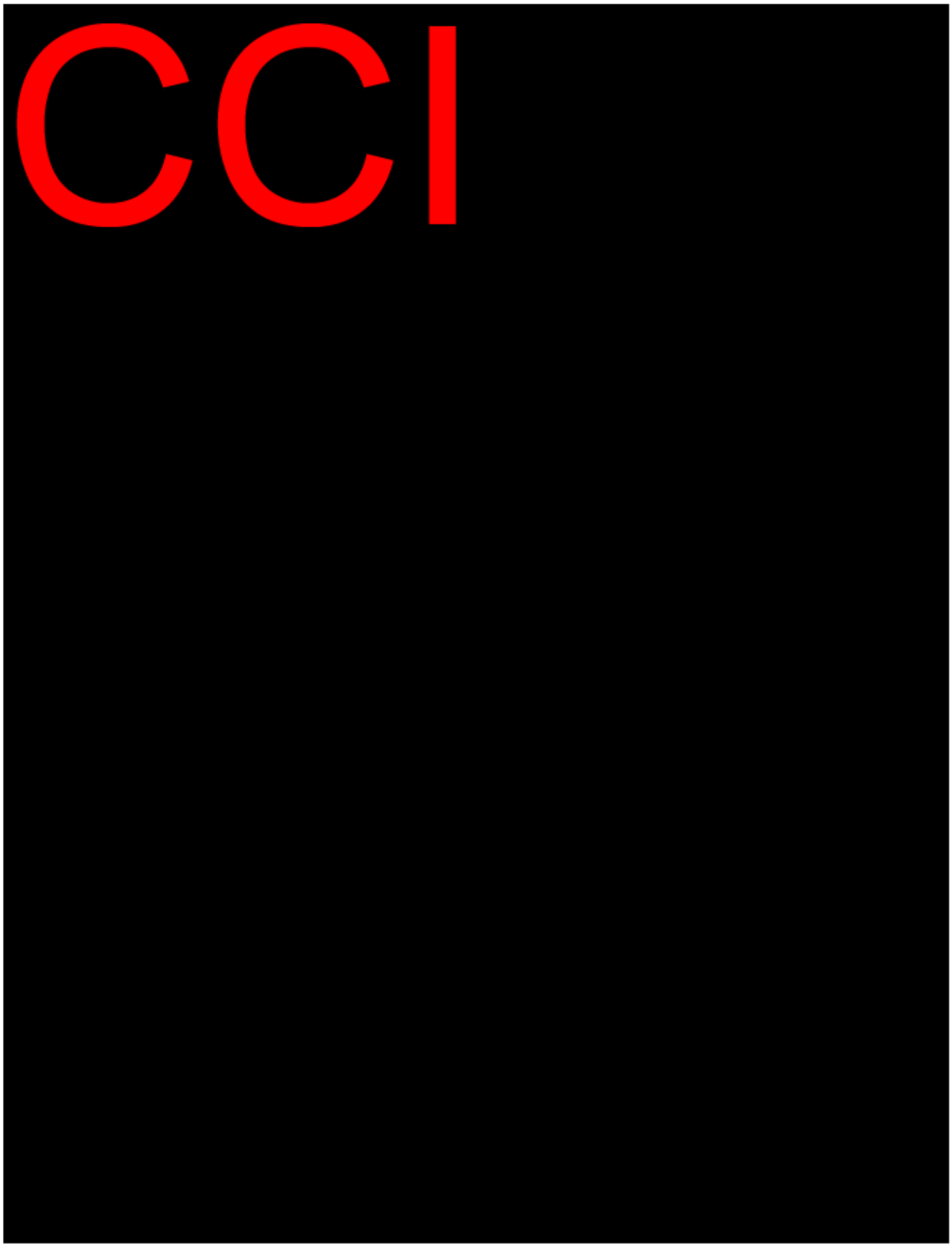
The **Objective Response Rate (ORR)** is defined as the number of participants having a BOR assessment of CR or PR, out of the total number of participants belonging to the analysis set of interest.

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

The unconfirmed ORR will be calculated for each treatment arm with a two-sided 95% CI using the Clopper-Pearson method (exact CI for a binomial proportion as computed by default by the FREQ procedure using the EXACT option). At OR IA, the difference Δ^{ORR} will be estimated by the crude rate difference between bintrafusp alfa and pembrolizumab arms and complemented by a 95% Miettinen-Nurminen CI (computed by the FREQ procedure using the RISKDIFF(CL=MN) option).

At PFS IA and following analyses, the difference in ORR will be estimated based on Cochran-Mantel-Haenszel method (taking into account the randomization strata) and the test statistics will be presented.

The odds ratio adjusted by randomization strata with associated 95% CI will also be presented. Additionally, DCR will be calculated along with the corresponding two-sided exact Clopper-Pearson 95% CI per treatment arm.

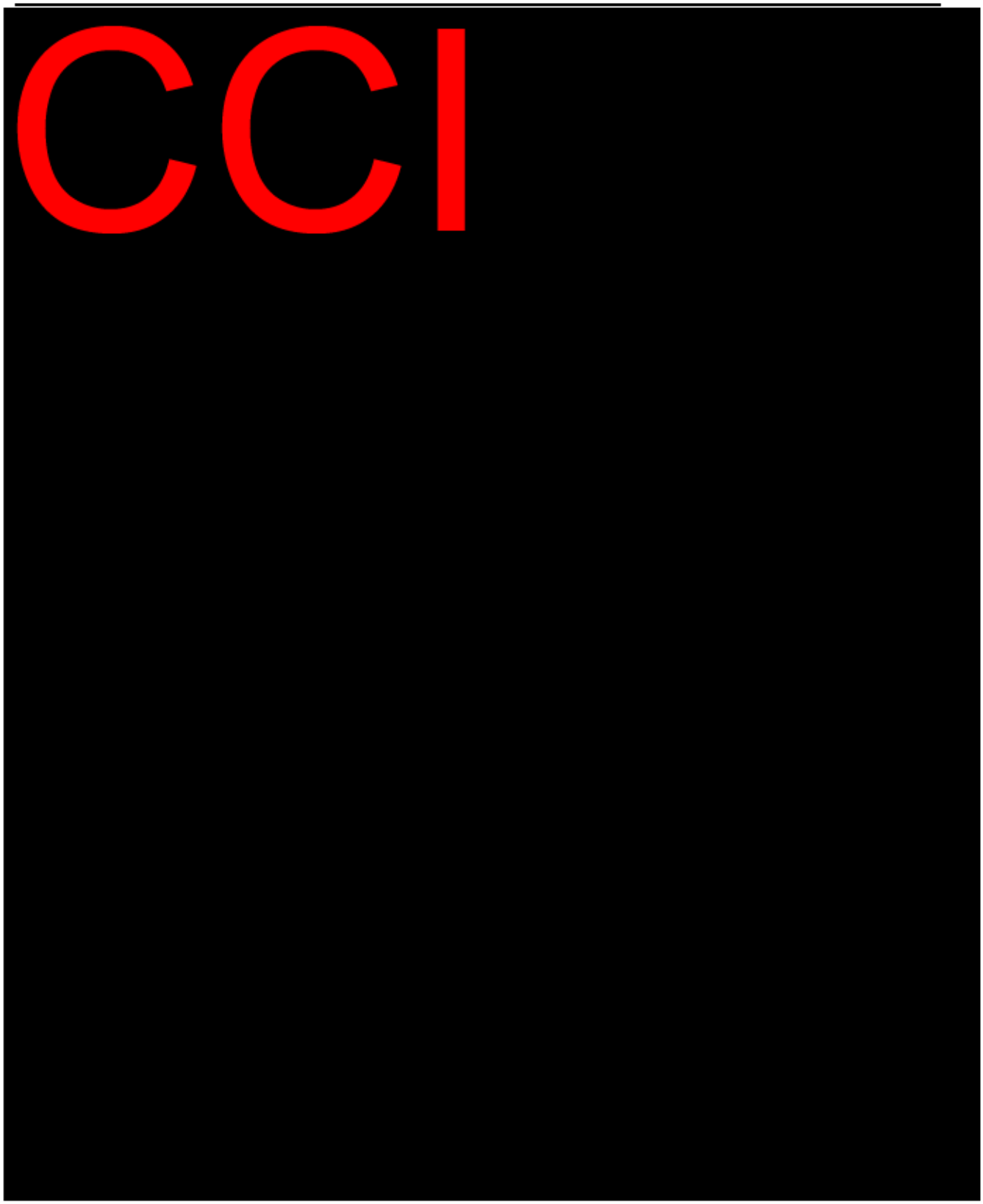

CCI

CCI

CCI

CCI

CCI


CCI

14.4 Progression Free Survival According to RECIST 1.1 as assessed by IRC (Primary Endpoint)

A primary endpoint of this study is PFS time. The disease progression will be based on IRC assessment on imaging and classified according to RECIST version 1.1 criteria.

Progression Free Survival (PFS) time is defined as the time from randomization to the date of the first documentation of objective PD as assessed by IRC or death due to any cause, whichever occurs first. Death will be considered as event if reported within:

- 12 weeks (84 days) after the last tumor assessment without progression or the randomization (whichever occurs later) – during the first 18 months of follow-up
- 24 weeks (168 days) after the last tumor assessment without progression or the randomization (whichever occurs later) - after 18 months of follow-up.

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

- Any participant with neither assessment of tumor progression, nor death date within 12 weeks after last tumor assessment (or 24 weeks after 18 months of follow-up) or randomization will be censored on the date of last evaluable tumor assessment or randomization whatever is later
- If death without previously documented PD is observed after more than 84 days during the first 18 months of follow-up (or 168 days after the first 18 months of follow-up) of last tumor assessment, the participant will be censored at the date of the last evaluable tumor assessment
- Participants who do not have a baseline tumor assessment or who do not have any post-baseline tumor assessments will be censored at randomization unless death occurred on or before the time of the second planned tumor assessment (i.e. 12 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
- Participants with an event after two or more subsequent missing response assessments (i.e. no assessments in 84 days during the first 18 months of follow-up or 168 days after the first 18months of follow-up) will be censored on the date of the last evaluable tumor assessment.

Censoring rules are also summarized in Table 7.

Table 7 Censoring Rules

PFS Event Status		Censoring	Date of event / censoring
Progressed or died	Within two subsequent scheduled tumor assessments after last response assessment of CR, PR or SD	Event	Minimum (Date of PD, Date of death)
	Otherwise	Censored	Date of last evaluable tumor assessment with outcome CR, PR or SD or date of randomization, whatever is later

PFS Event Status	Censoring	Date of event / censoring
Neither progressed nor died	Censored	Date of last evaluable tumor assessment with outcome CR, PR or SD or date of randomization, whatever is later

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

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

Kaplan-Meier analysis

The analysis of PFS will be presented by treatment arm and will be performed with a Kaplan-Meier method (product-limit estimates) and a summary of associated statistics will be presented including corresponding two-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and CIs for the survival function estimates at Month 3, 6, 9, 12 and 24 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 an event, overall and per event type (PD or death) will be presented by treatment arm. Censoring reasons will be described by treatment arm, as follows:

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

  Lost to follow-up will include the following participants:

  - Lost to follow-up status is collected on the eCRF treatment termination page or eCRF study termination page prior to the analysis cutoff;
  - Participants with last known alive date > 14 weeks prior to the analysis cutoff date (duration of 14 weeks is based on the assessment schedule of every 3 months for survival follow-up interval + 1 week window).

A participant listing will be also provided with the following information: treatment arm, participant identifier, age, sex, race, randomization date, date of last tumor assessment, date of event/censoring, event/censoring reason, time to event.

Log-rank test

The primary efficacy analysis of PFS will be performed based on the FAS using the strata assigned at randomization.

The primary efficacy analysis will compare the PFS time between the two treatment arms, and will be performed using a one-sided stratified log-rank test, according to testing strategy defined in Section 6.

The stratification factor applied are identical to those used for randomization (as captured via the IRT at randomization):

- Squamous histology
- Nonsquamous histology and never smoked
- Nonsquamous histology with a smoking history.

The following null hypothesis will be tested:

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

Where $\lambda^{PFS}(t)$ represents the hazard at time t and θ the unknown constant of proportionality of hazards in the treatment arms M (M7824/bintrafusp alfa) and P (pembrolizumab).

Cox's regression model

The treatment effect will be estimated using a Cox's Proportional Hazard model stratified by the randomization strata to calculate the hazard ratio. Randomization strata will be taken as specified and documented in the IRT. Each stratum will define a separate baseline hazard function (using the 'STRATA' statement in SAS PROC PHREG), i.e. for the i-th stratum the hazard function is expressed as: h(i;t) = h(i,0;t) exp(βX), where h(i,0;t) defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0=pembrolizumab, 1= M7824 (bintrafusp alfa)) and beta is the unknown regression parameter. Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

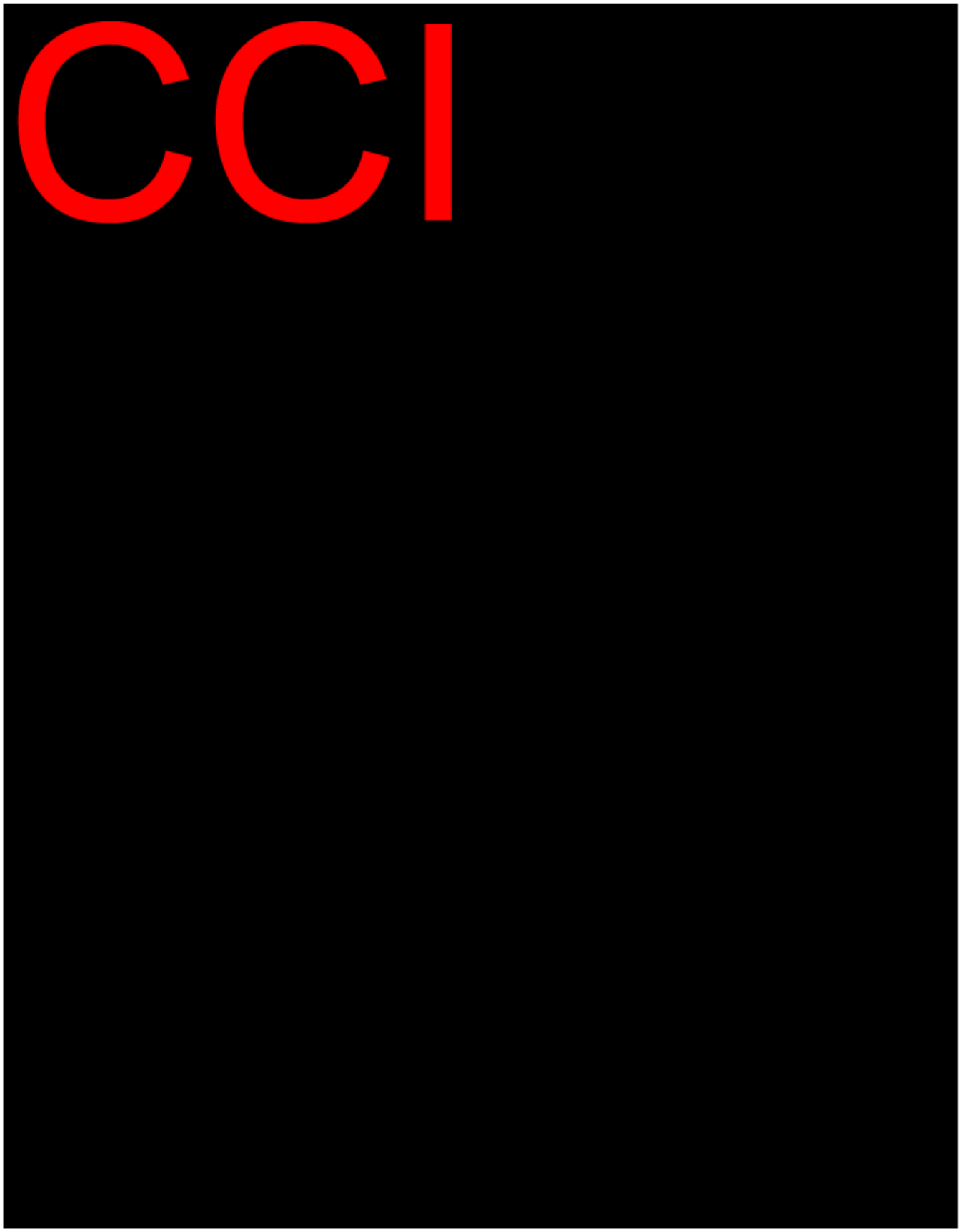
CCI [redacted]

CCI

14.4.3 Time to Follow-up for PFS

A Kaplan-Meier analysis for PFS follow-up duration will be generated to assess the follow-up time in the treatment arms reversing the PFS censoring and event indicators.

Kaplan-Meier estimates will be presented by treatment arm together with the median time of follow-up for PFS. In particular, the follow-up rate at 3, 6, 9, 12 and 24 months will be estimated with corresponding two-sided 95% CIs.


CCI

14.8 Duration of Response (Secondary Endpoint)

Duration of response (DOR) is defined as the time from first documentation of objective response (CR or PR) to the date of first documentation of objective progression of disease (PD) or death due to any cause whichever occurs first. The analysis of DOR will be performed on unconfirmed objective response of CR or PR according to RECIST 1.1 as adjudicated by IRC. The censoring rules for DOR are as described above for PFS. The number of participants with ongoing response will be described as the number of censored participants with administrative censoring.

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

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

The time and duration of response per participant may also be displayed in a swimmer graph in participants with unconfirmed CR/PR (delayed response is also considered).

A participant listing will provide the following information: treatment arm, participant identifier, age, sex, race, date of randomization, date of first response, time to response, date of last tumor assessment, censored (Y/N), date of event/censoring, event/censoring reason, DOR.

14.8.1 Sensitivity Analyses for DOR

The analysis of DOR will be repeated on confirmed CR/PR according to RECIST 1.1 as adjudicated by IRC.

The analyses for DOR based on unconfirmed responses will be repeated for DOR based on confirmed responses (i.e. Kaplan-Meier analysis, listing of DOR, swimmer graph of time and duration of response in participants with confirmed CR/PR).

CCI

CCI

14.11 Overall Survival (Primary Endpoint)

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

OS = (date of event/censoring – date of randomization + 1)/30.4375 (months).

For participants alive at the time of data cutoff 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 cutoff date	Last date known to be alive	Yes
Participants who died before or at cutoff date	Date of death	No

The following complete dates will be used to determine the last date known to be alive. Only the ones among them that are before or at data cutoff and which are not imputed shall be used in the derivation. Dates past the data cutoff will be ignored by the derivation:

- All participant assessment dates (blood draws (laboratory, PK), vital signs, performance status, ECG, tumor assessments, PRO assessments)
- Start and end dates of anticancer therapies administered after IMP discontinuation.
- AE start and end dates
- Last known alive date collected on the 'Survival Follow-up' eCRF (do not use follow-up date)
- IMPs start and end dates (including reinitiation of treatment)
- Randomization date
- Completion/Discontinuation date from the "Study Termination" eCRF page (do not use if reason for discontinuation is lost to follow-up or death)

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

The analysis of OS time will be performed with a Kaplan-Meier method with the same approach described for PFS in Section 14.4. Kaplan-Meier estimates (product-limit estimates) will be presented together with a summary of associated statistics including corresponding two-sided 95% CIs. They will be estimated at 12, 18, 24, 30, 36, 42, and 48 months.

A Cox's regression model will be used with the same approach as for PFS described in Section 14.4.

The following null hypothesis will be tested:

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

Where $\lambda^{OS}(t)$ represents the hazard at time t and θ the unknown constant of proportionality of hazards in the treatment arms M (M7824/bintrafusp alfa) and P (pembrolizumab).

H_0^{OS} will be tested according to the testing strategy described in Section 6.

Lan-DeMets alpha spending with O'Brien-Fleming-like boundaries will be adopted for the currently observed event size at cutoff date.

A participant listing will provide the following information: treatment arm, participant identifier, age, sex, race, date of randomization, date of event/censoring, event/censoring reason, and time to event.

14.11.2 Follow-up time since randomization

A Kaplan-Meier analysis will be performed on follow-up time using OS data by reverting the censoring flag (participants still alive or lost to follow-up are counted as events, deceased participants are counted as censored). Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with the median time of follow-up. In particular, the follow-up rates at 3, 6, 9, 12 and 24 months will be estimated with 95% CIs.

15 Safety Analyses

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. Baseline values for safety analysis are defined in Section 9 of this IAP "Definition of baseline and change from baseline".

15.1 Adverse Events

Treatment emergent adverse events (TEAEs) are those events with onset dates occurring during the on-treatment period (as defined in Section 9 of this IAP) or if the worsening of an event is during the on-treatment period.

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

date of the entire event. The overall outcome of the adverse event is the outcome of the last event in the sequence. Duration of the AE and the TEAE flag will be adjusted accordingly in the analysis.

All analyses described in this section will be based on TEAEs if not otherwise specified and will be described by treatment arm. The AE listings will include all AEs (whether treatment-emergent or not). AEs outside the on-treatment period will be flagged in the listings.

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

Related Adverse Events are those AEs with relationship to IMP reported by the investigator as related (i.e. answer to the questions "Relationship with study treatment" = "Related" on "Adverse Event Details" eCRF page) and those of unknown relationship (i.e. no answer to the question "Relationship with study treatment").

Serious Adverse Events (SAEs) are those events reported on the "Adverse Event 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 study treatment" includes "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 study treatment" includes "Drug withdrawn" on "Adverse Event Details" eCRF page).

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

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

Adverse Events of Special Interest (AESI): AESI will be identified according to a pre-specified search list of MedDRA Preferred Terms.

Categories of AESIs include (see details in Section 15.2.3):

- Infusion-related reactions (IRRs)
- Immune-related adverse events (irAEs)
- Skin AE possibly related to TGFβ inhibition
- Anemia.

15.1.1 All Adverse Events

Adverse events will be summarized by worst severity (according to NCI-CTCAE version 5.0) per participant, using MedDRA (latest version) preferred term as event category and MedDRA (latest version) primary system organ class (SOC) body term as Body System category.

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

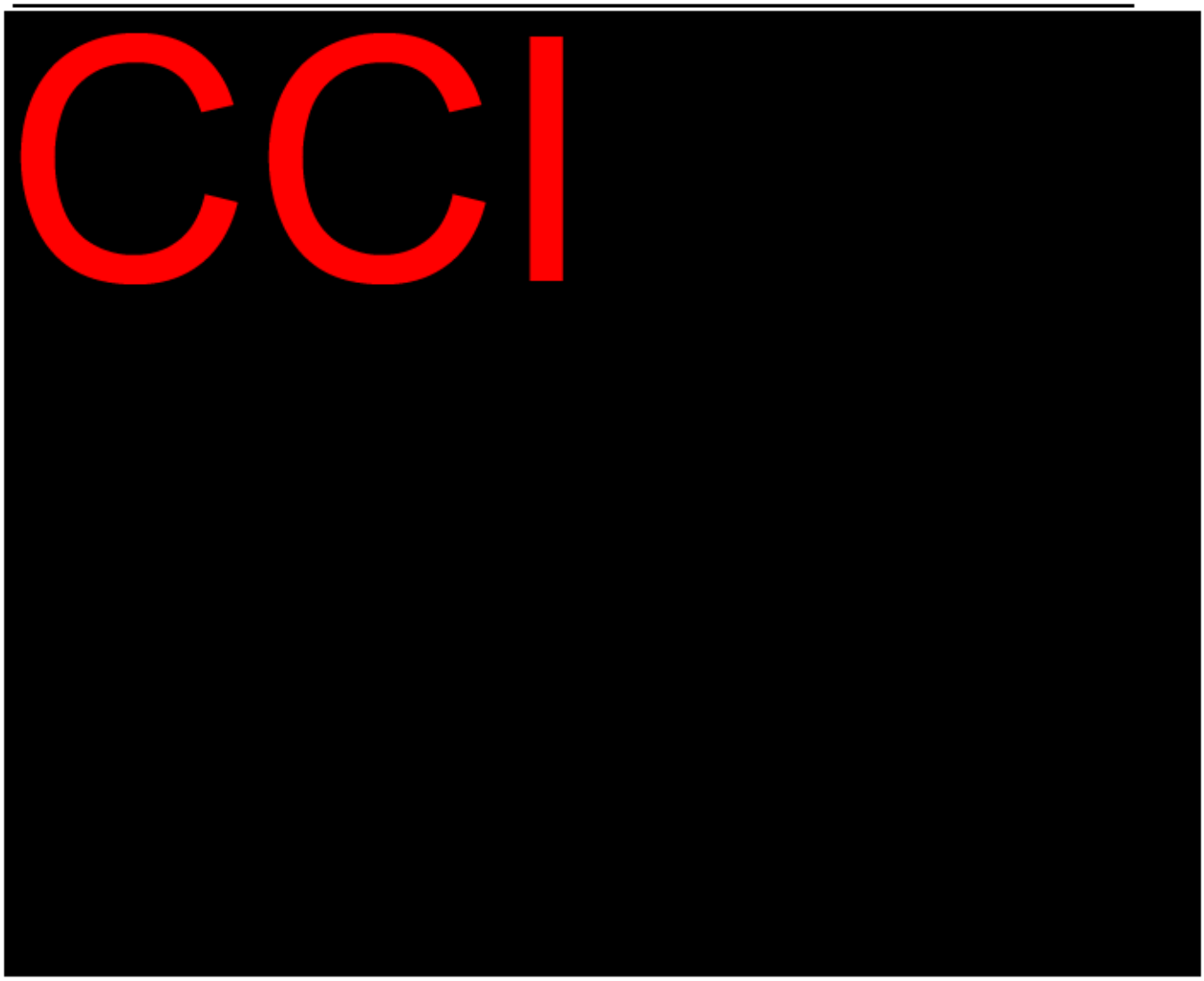
If an AE is reported for a given participant more than once during treatment, the worst severity and the worst relationship to IMP will be tabulated. In case a participant had events with missing and non-missing grades, the maximum of the non-missing grades will be displayed.

A table presenting the overall summary of AEs will be presented by treatment arm with the following information:

- 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
- TEAEs and related TEAEs of special interest:
    - Infusion-related reactions (IRRs)
    - Immune-related AEs (irAEs)
    - TGF-β inhibition mediated skin TEAEs
    - Anemia
- Bleeding TEAEs
- Related bleeding TEAEs

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) and by treatment arms. Each participant will be counted only once within each PT or SOC. Categories for TEAEs of special interest will be described as detailed in Section 15.2.3 of this IAP.

TEAEs and related TEAEs by worst grade will also be summarized, and PTs that differs by at least 5% between treatment arms will be presented graphically by worst grade and PT with bar chart figures.



Clinicaltrials.gov and EudraCT -requirements

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

The following listings will be provided:

- Listing of all AEs (whether treatment-emergent or not): TEAEs and AE occurring during reinitiation phase will be flagged

- Listing of non-TEAE for AEs occurring after enrollment (date of first signature of informed consent/date of first signature of first informed consent) but prior to the first dose of study intervention

Listing of AEs with onset or worsening after the on-treatment period (AEs occurring during reinitiation phase will be flagged). The listing of AEs will contain the following information: treatment arm, participant identifier, age, sex, race, first and last date of treatment administration, preferred term, reported term for the AE, start date, end date, duration of AE (in days), day relative

to the first infusion, day relative to the most recent infusion prior to AE onset, relationship to IMP, toxicity grade, action(s) taken, outcome, seriousness (Y/N), AESI infusion-related (Y/N), AESI immune-related (Y/N), TGF-β inhibition mediated skin AESI (Y/N), Treatment-related anemia AESI (Y/N), Bleeding events (Y/N). The listing will be sorted by treatment arm, participant identifier, preferred term, start date and end date.

15.1.2 Adverse Events Leading to Treatment Discontinuation

Frequency tables summarizing the following actions taken with IMP in response to TEAEs will be prepared and presented by PT and primary SOC in alphabetical order:

- 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
- TEAEs leading to temporary IMP interruption
- Related TEAEs leading to temporary IMP interruption.

The listing of AEs 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 IMP, death within 60 days after first dose as well as primary reason for death, will be tabulated based on information from the “Death” eCRF page.

The following summaries will be provided by treatment arm:

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

In addition, date and cause of death will be provided in an individual participant data listing together with selected dosing information (IMP received, date of first/last administration, number of infusions, day relative to first and last infusion) and will include columns for

- Participant identifier, analysis set, age, sex, race
- AEs with fatal outcome (list preferred terms of AEs with outcome=fatal, Grade 5 or Serious resulting in death)
- Autopsy performed (Yes/No/Unknown)
- Flag for death within 30 days of last IMP intervention
- Flag for death within 60 days of first IMP intervention.

15.2.2 Serious Adverse Events

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

- Serious TEAEs
- Related serious TEAEs

Please refer to Section 15.1.1. A participant listing will be provided including all SAEs.

15.2.3 Adverse Events of Special Interest

15.2.3.1 Infusion-related Reactions Including Immediate Hypersensitivity

Infusion-related reactions (IRR) are defined as adverse events with PTs on a pre-specified list of MedDRA PTs and divided into reactions versus signs and symptoms.

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

Signs and symptoms of IRRs and hypersensitivity/allergic reactions: should be considered when onset is on the day of infusion (during or after the infusion) and resolved completely with the end date on the same day of the infusion or the day after.

IRR will be summarized by the following variables per subgroup of IRR and overall:

- 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 should be assigned to the actual drug infusions that the participant received, not to the planned dates. An IRR is assigned to a drug infusion if its onset is at the same date (but not before dosing) or the following day of drug infusion.

Moreover, the following frequency tables will be provided by worst grade, SOC, and PT:

- IRR by worst grade, SOC, and PT

The listing of IRRs will also be provided with the relevant information (see description of listing in Section 15.1.1). A listing displaying IMP intervention together with IRR will also be displayed, including administration date (day) /time, IRR AE PT, IRR AE grade, IRR AE start day/stop day, and 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 preferred term matches a preferred term on the list of pre-selected MedDRA terms.

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

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

4) On the imAE eCRF page, either:

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

OR

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

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

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

PTs will be compiled into categories on the basis of the list of pre-selected MedDRA terms: 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.

Immune-related adverse events (irAEs) will be described in the following tables:

- irAEs
- irAEs by the worst grade
- irAEs leading to permanent treatment discontinuation
- Serious irAEs
- Moreover, the frequency table of irAEs by worst grade, category, subcategory (for Immune-mediated endocrinopathies), sub-subcategory (for Immune-mediated endocrinopathies – thyroid disorders) and PT will be provided

The listing of irAEs will also be provided with the relevant information including additional interventions for irAEs (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

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

Narrow definition:

- Keratoacanthoma
- Squamous cell carcinoma of skin

Broad definition has additional PTs:

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

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

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

Tables for skin TEAEs frequency will be provided by MedDRA Preferred Terms (PTs) including both narrow and broad definition PTs.

A listing of skin TEAEs 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 Adverse Events

The following HLTs and PTs (using the latest version of MedDRA available) will be used to select the anemia AEs:

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

A listing of anemia adverse events will be also provided with the relevant information (see description of listing in Section 15.1.1). A table displaying the overall summary of Anemia AEs by treatment arm may be also presented.

15.2.4 Bleeding events

Bleeding events are defined as adverse events with PTs according to the MedDRA Standardised MedDRA Queries (SMQ) for Haemorrhage terms (excl laboratory terms). Treatment-emergent bleeding events and treatment-emergent 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:

- Any grade (including missing grade)
- Grade 1
- Grade 2
- Grade 3
- Grade 4
- Grade 5

On request of IDMC, baseline participants characteristics like demographic characteristics, height, weight and BMI, disease history, presence of any bleeding events in medical history, use of antithrombotic agents (in medical history and ongoing at the start of bleeding event) may be summarized in participants with and without bleeding events according to the definition above. Antithrombotic agents categories belonging to ATC code B01 are considered.

15.2.5 Three-Tier Approach to Summarizing and Analyzing AEs

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

Risk Measures

The following statistics for summarizing safety data will be provided:

Crude Rate: is calculated as number of participants with specific AE out of the total number of participants at risk expressed as percentage.

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

Tier 1, 2, and 3 identifications

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

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

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

  - Other immune-related adverse events
- Infusion-related reactions
  - Reactions
  - Signs and Symptoms
- TGF-β inhibition mediated skin reactions
- Anemia
- Bleeding adverse events
- Impaired wound healing

Summary statistics

The following summaries will be presented:

For the pre-specified terms::

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

For any TEAEs:

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

No multiplicity adjustment will be applied for Tier 1 and 2 TEAEs.

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

15.3 Clinical Laboratory Evaluation

Treatment emergent laboratory assessments are any sample collected during on-treatment period.

Laboratory values (including corresponding normal ranges) converted in standard units will be used for summary statistics and shift tables.

Laboratory results will be classified according to the NCI-CTCAE Version 5.0. Additional laboratory results that are not part of NCI-CTCAE will be presented according to the categories: low, normal and high (according to the laboratory normal ranges). Quantitative data will be examined for trends using descriptive statistics (mean, StD, median, Q1, Q3, minimum, and maximum) of actual baseline values, on-treatment values and changes from baseline to each on-treatment visit over time. The changes computed will be the differences from baseline. Participants without post baseline laboratory samples will be excluded from analyses with respect to values after the baseline. Qualitative data based on reference ranges will be described according to the categories (i.e. low, normal, and high).

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, the toxicities will be summarized separately. Low direction toxicity grades at baseline and post-baseline will be set to 0 when the variables are derived for summarizing high direction toxicity, and vice versa.

For gradable parameters, the following summaries will be displayed by treatment arm:

- 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 following figures will also be provided by treatment arm:

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

The definitions of toxicity grading for each parameter are available in the NCI-CTCAE toxicity grading version 5.0 (in Appendix 18.1 of this IAP).

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

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

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

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

Creatinine Clearance (CrCl) will be derived through the Cockcroft-Gault formula (Cockcroft and Gault, 1976), from the serum creatinine measurements in the "FULL CHEMISTRY" and "CORE CHEMISTRY" e-crf pages, the weight in the "VITAL SIGNS" e-crf page and the age re-calculated at the sampling date as follows:

$$CrCl = FACTOR * (140 - AGE) * \frac{WT}{S_{CREA} * 0.8136}$$

where FACTOR is gender-dependent with value 1 (male) or 0.85 (female), AGE is the subject age in years, WT is the subject weight in kg and S_{CREA} is the observed serum creatinine in umol/l.

Toxicity grading for creatinine clearance will be defined as per NCI-CTCAE toxicity grading version 5.0, the following normal ranges will be considered: 110 to 150mL/min in males and 100 to 130mL/min in females.

Laboratory parameters with NCI-CTC grades not available

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

- Shift from baseline to lowest on-treatment value (classified as normal, high, low)
- Shift from baseline to the highest on-treatment value (classified as normal, high, low)

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), bilirubin, and creatinine.

ALP, ALT, AST, bilirubin and creatinine

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 ≥ 5xULN, AST ≥ 10×ULN, AST ≥ 20×ULN
- (ALT and AST) <3×ULN, (ALT or AST) ≥ 3×ULN, (ALT or AST) ≥ 5×ULN, (ALT or AST) ≥ 10×ULN, (ALT or AST) ≥ 20×ULN
- Total bilirubin < 2×ULN, total bilirubin ≥ 2×ULN
- Concurrent ALT ≥ 3×ULN and total bilirubin ≥ 2×ULN
- Concurrent AST ≥ 3×ULN and total bilirubin ≥ 2×ULN

- Concurrent (ALT or AST) ≥ 3×ULN and total bilirubin ≥ 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and total bilirubin ≥ 2×ULN and ALP > 2×ULN
- Concurrent (ALT or AST) ≥ 3×ULN and total bilirubin ≥ 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 highest values of ALT versus highest values of 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 patients with possible Gilbert's cholestasis; the right upper quadrant are the possible Hy's Law patients; the right lower quadrant is possible Temple's Corollary (patients with ALT ≥ 3 x ULN but not satisfying Hy's Law).

A plot of highest values of AST versus highest values of total bilirubin, both relative to the ULN will be also provided.

In addition, a listing of total bilirubin, ALT, AST and ALP values during on-treatment period for participants with a post-baseline total bilirubin ≥ 2×ULN, ALT≥ 3×ULN or AST ≥ 3×ULN will be provided, including values expressed as multiples of ULN. Possible Hy's Law patients will be flagged in this listing.

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

Pregnancy test

All test results for pregnancy test as collected on the "Pregnancy Test" eCRF page will be listed.

Coagulation parameters

Descriptive statistics of coagulation parameters results by visits will be performed, shift tables and listings will not be provided.

15.4 Vital Signs

Vital signs data during on-treatment period will be summarized by treatment arm for the Safety analysis set.

All vital sign parameters will be summarized using descriptive statistics (mean, StD, median, Q1, Q3, minimum, maximum) of actual baseline values, on-treatment values and changes from baseline to each visit over time.

In addition, the maximum changes of vital sign measurements baseline to maximum changes after start of 1st treatment will be grouped as follows:

Table 9 Categories of Maximum Change from Baseline in Vital Signs

Body temperature increase from baseline < 37 °C, 37-<38°C, 38-<39°C, 39-<40°C, ≥40°C	< 1°C , 1-<2°C , 2-<3°C, ≥ 3 °C
Body weight increase	< 10%, ≥ 10%
Body weight decrease	< 10%, ≥ 10%
Heart rate increase from baseline <100 bpm ; ≥ 100 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
Heart rate decrease from baseline ≤50 bpm ; > 50 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
SBP increase from baseline <140 mmHg; ≥ 140 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
SBP decrease from baseline ≤95 mmHg; > 95 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP increase from baseline <90 mmHg; ≥ 90 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP decrease from baseline ≤45 mmHg; > 45 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg

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

- Maximal shifts (changes in categories)

The following potentially clinically significant abnormalities in vital signs will be also 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 bpm and decrease from baseline ≥ 20 bpm in heart rate
- ≥ 100 bpm and increase from baseline ≥ 20 bpm in heart rate
- ≥ 10% weight increase
- ≥ 10% weight decrease

All vital signs will also be listed, baseline values and post-baseline values collected after the on-treatment period will be flagged in the listing. The listing will include treatment arm, participant identifier, age, sex, race, vital sign parameter, visit, date, on-treatment flag, value, unit, baseline value, and change from baseline. Listings will be sorted by treatment arm, participant identifier, vital sign parameter and vital sign measurement date.

15.5 Other Safety or Tolerability Evaluations

ECG

A listing of ECG values will be provided with the following information: treatment arm, participant identifier, age, sex, race, ECG parameter and unit, ECG date and value. The listing will be sorted by treatment arm, participant identifier, ECG parameter, and ECG date. It will include ECG values collected at screening and at safety follow-up visits.

ECOG Performance Status

The ECOG shift from baseline to the highest score during the on-treatment period will be summarized by treatment arm. ECOG performance status will also be presented in a listing at each time point.

Safety analyses by PD-L1 statuses

Further analyses on TEAEs and AESI by PD-L1 assays statuses may be provided on IDMC request.

The following safety tables may be provided by PD-L1 High/Not High statuses for PD-L1 assays of interest (e.g. SP263 assay):

- TEAEs by Worst Grade, SOC and PT
- Overview of Infusion-Related Reactions
- Overview of Immune-Related AE
- Overview of TGF-β inhibition mediated skin TEAE
- Overview of Treatment-Related Anemia

Subset of participants with Haemoptysis, Gastrointestinal Haemorrhages or TEAE leading to Death

Specific listings may be provided for AEs of participants meeting at least one of the following criteria:

- haemoptysis : PT code = Haemoptysis
- Gastrointestinal haemorrhage: SMQ=20000108
- TEAE leading to death within 30 days from first IMP intervention

The following listings are considered:

Medical history listing showing the following information: participant identifier, treatment arm, haemoptysis (Y/N), gastrointestinal haemorrhage(Y/N), TEAE leading to death (Y/N), date of first infusion of trial drug, date of last infusion of trial drug, country, age, sex, race, reported medical history, start date of medical history, end date of medical history, relationship to study condition, ongoing, grade, medical history of bleeding events.

Concomitant medications listing including: participant identifier, treatment arm, haemoptysis (Y/N), gastrointestinal haemorrhage(Y/N), TEAE leading to death (Y/N), date of first infusion of trial drug, date of last infusion of trial drug, country, age, sex, race, preferred term of concomitant medication, medication name, start date of concomitant medication, end date of concomitant medication, dose (units), frequency, route, reason for medication (indication), reason for medication, use of antithrombotic agents (Y/N), concomitant medication (taken any time during on-treatment).

Adverse Events listing including: participant identifier, treatment arm, haemoptysis (Y/N), gastrointestinal haemorrhage(Y/N), TEAE leading to death (Y/N), Date of first infusion of trial drug, Date of last infusion of trial drug, preferred term, investigator term, start or change date of AE, end date of AE, treatment-emergent AE (Y/N), duration of AE (days), day since first infusion of trial drug (days), day since most recent infusion prior to AE onset (days), relationship to trial drug, grade, causality factor, action taken on trial drug, outcome, seriousness, country, age, sex, race, date of initial cancer diagnosis, TNM classification at initial diagnosis, TNM classification at study entry, prior anticancer therapy, tumor histology, tumor location, bleeding events in medical history (Y/N), PT of bleeding events in medical history, antithrombotic agents (any prior or concomitant medication of ATC code B01), last available collection date of hematology test, hemoglobin value (g/L), Platelets value (10^9/L), last available collection date of coagulation tests, PT value (sec), aPTT value (sec), INR value (ratio).

Subset of participants with ALP increase TEAE or Elevated ALP of Grade >=1 during On-Treatment

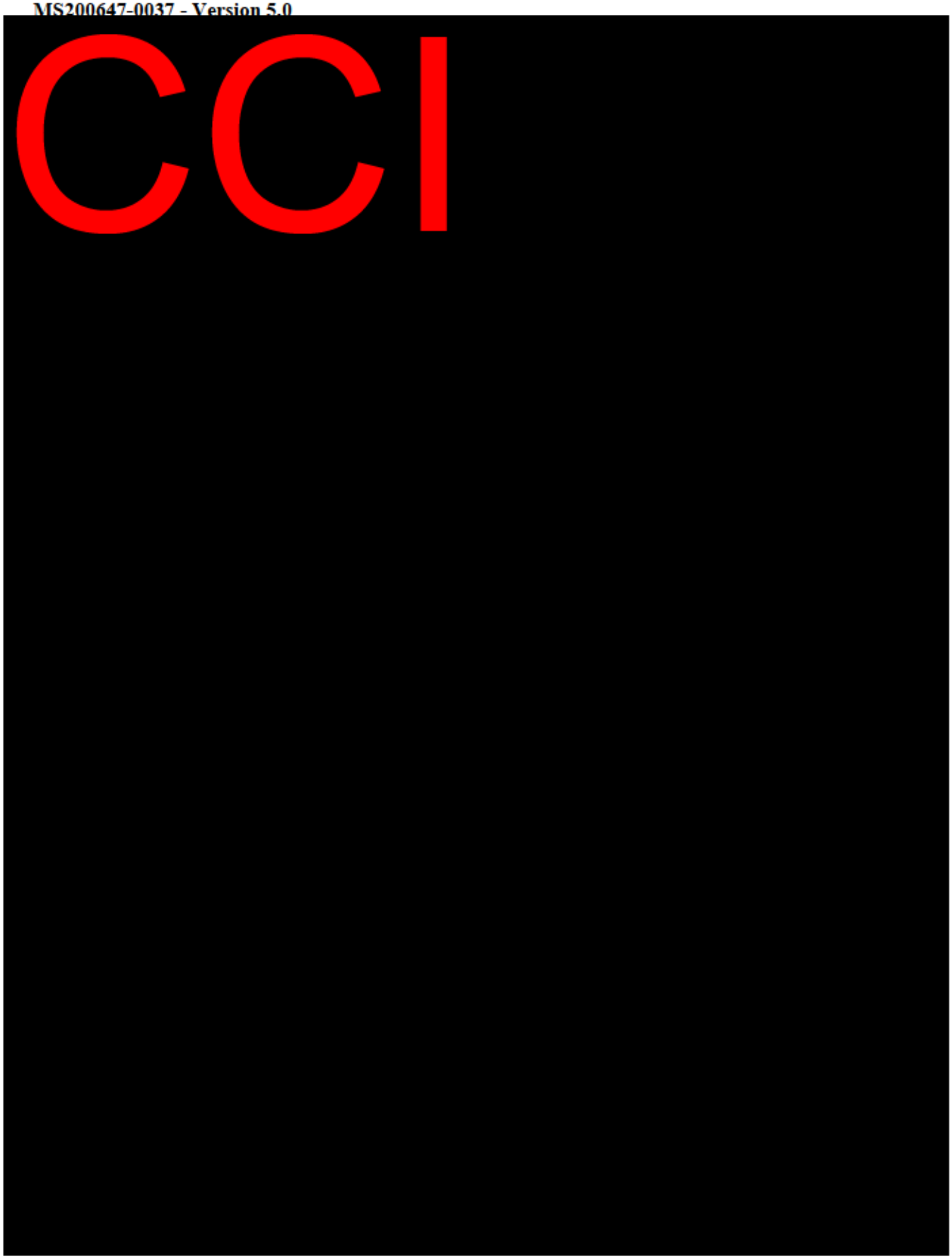
On IDMC request, a specific listing may be provided for participants meeting at least one of the following criteria:

- increased ALP TEAE
- elevated ALP of grade >=1 on treatment

This listing will include, subject with increased ALP TEAE (Y/N) , Subject with elevated ALP of grade >=1 on treatment (Y/N), Analysis Visit, Analysis Date, Analysis Relative Day, On-Treatment Flag, ALP value, ALP ULN, Ratio ALP value/ULN, ALP toxicity grade, ALT value, ALT ULN, Ratio ALT value/ULN, ALT toxicity grade, AST value, AST ULN, Ratio AST value/ULN, AST toxicity grade, Bilirubin value, Bilirubin ULN, Ratio Bilirubin value/ULN, Bilirubin toxicity grade, GGT value, GGT ULN, ratio GGT value/ULN, GGT toxicity grade.

16 Analyses of Other Endpoints

CCI


CCI

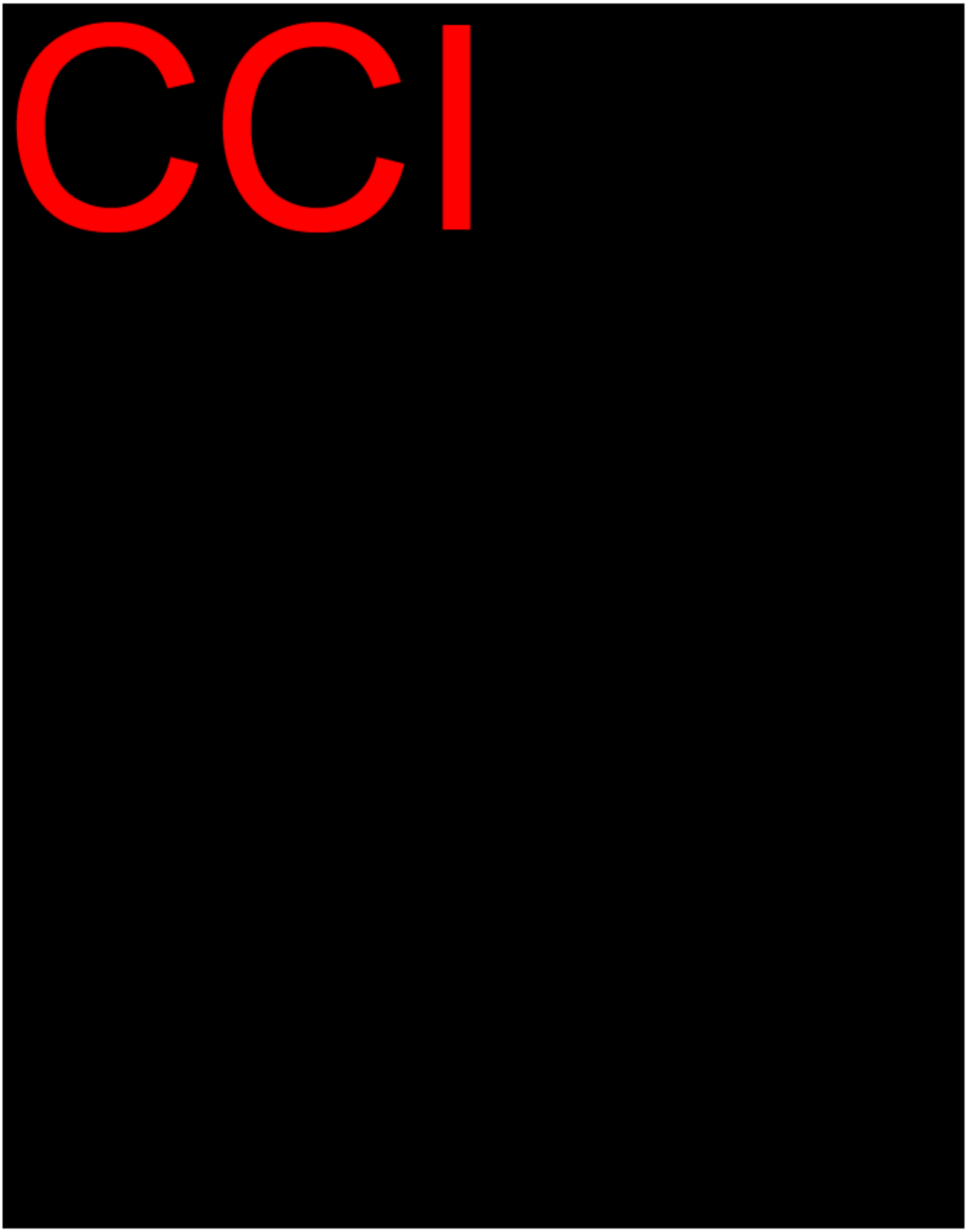

CCI


CCI

CCI

CCI

CCI


CCI


CCI
CCI

16.4 Patient-Reported Outcomes

Patient-reported outcomes will be collected using the European Organization for Research and Treatment of Cancer-Quality of Life Questionnaire Core-30 (EORTC-QLQ-C30), NSCLC symptom assessment questionnaire (NSCLC-SAQ) and European Quality of Life 5-dimensions 5-level questionnaire (EQ-5D-5L). The PROs should be performed in the same sequence at each visit.

EORTC-QLQ-C30

The EORTC-QLQ-C30 is a cancer specific health-related quality of life questionnaire that has been widely used in clinical studies and investigations using PROs for individual participant management. It includes 5 function domains (physical, emotional, social, role, cognitive), 8 symptoms domains (fatigue, pain, nausea/vomiting, constipation, diarrhea, insomnia, dyspnea, and appetite loss), as well as global health status. The EORTC also includes an item library from which items can be selected to supplement the core instrument. Items measuring rash, pruritus, and trouble by side effects will be added to the items on the C30.

A score can be computed for each domain by taking the following steps:

1) For each domain the average of the items that contribute to the domain should be estimated, this is the row score (i.e Row Score = Sum of the items responses that contribute to the domain/ number of non-missing items that contribute to the domain).

2) Subtract 1 from the row score obtained at step one

3) Divide the result obtained at step 2 by the range of item score. The range of item score is the difference between the maximum possible value of row score and the minimum possible value (e.g. most items are scored 1 to 4, giving range = 3)

4) If functional domain, subtract 1 from the result obtained at step 3) and multiply by 100

5) If symptom domain/Global health status, multiply the result obtained at step 3 by 100

The domain score will range from 0 to 100; a higher score represents a higher (“better”) level of functioning or a higher (“worse”) level of symptoms.

NSCLC-SAQ
The NSCLC-SAQ is a 7-item patient-reported outcome questionnaire created to measure the cardinal symptoms of NSCLC (McCarrier, 2016). Items measure 5 domains:

- Cough (item 1. “How would you rate your coughing at its worst…?”)

- Pain (item 2. “How would you rate the worst pain in your chest over the last 7 days?” and item 3. “How would you rate the worst pain in areas other than your chest over the past 7 days?”)

- Dyspnea (item 4. "How often did you feel short of breath during usual activities…?")
- Fatigue (item 5. "How often did you have low energy over the last 7 days?" and item 6. "How often did you tire easily over the last 7 days?"),
- Appetite loss (item 7. "How often did you have a poor appetite over the last 7 days?").

Each of these items is rated in terms of symptom severity or frequency on a 5-point response scale (1 = none or never; 5 = very severe or always).

The scoring algorithm of the NSCLC-SAQ total score is as follows:

- Cough Domain Score: score of the cough item, or missing if skipped
- Fatigue Domain Score: if both items present, compute mean; or use score from 1 item if the other is missing; or set to missing if both are skipped
- Pain Domain Score: if both items present, use most severe of both; or use score from 1 item if the other is missing; or set to missing if both are skipped
- Dyspnea Domain Score: score of the shortness of breath item, or missing if skipped
- Appetite Domain Score: score of the poor appetite item, or missing if skipped

NSCLC-SAQ Total Score: sum all five domain scores; if any are missing, a total score is not computed. This creates a total score ranging between 0 and 20 with higher scores indicating more severe symptomatology.

EQ-5D-5L

The EQ-5D-5L is a 6-item instrument developed by the EuroQol Group as a measure of health-related quality of life that can be used in a wide range of health conditions and treatments. The EQ-5D-5L consists of a descriptive system and a single item EQ Visual Analog Scale (VAS). The descriptive system comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each EQ-5D-5L dimension has 5 response categories: no problems, slight problems, moderate problems, severe problems and extreme problems. The participant is asked to indicate his/her health state by ticking the box next to the most appropriate statement in each of the 5 dimensions. The responses (ranging from 1 to 5) to the 5 dimensions are summarized into a 5-digit profile.

The EQ VAS records the patient's self-rated health on a vertical visual analogue scale ranging from 0 to 100.

Analyses on Patient-Reported Outcomes will not be included in the CSR

17 References

Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993 Mar 3;85(5):365-76

Agresti, A. (2007). An Introduction to Categorical Data Analysis (2 ed.). New York: John Wiley & Sons.

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

Chen C, Anderson K, Mehrotra DV, et al. A 2-in-1 adaptive phase 2/3 design for expedited oncology drug development. Contemp Clin Trials. 2018;64:238-42.

Clopper, C. J., & Pearson, E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika, 26, 404–413.

Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? J Clin Oncol. 2008 Aug 1;26(22):3791-6.

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

EMA. (13Dec2012). Guideline on the evaluation of anticancer medicinal products in man. European Medicines Agency. Retrieved from http://www.ema.europa.eu/docs/en_GB/ document_ library/Scientific_guideline/2013/01/WC500137128.pdf.

FDA. (2007). Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Washington DC: US Food and Drug Administration.

Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011 Dec;20(10):1727-36.

Hommel G1, Bretz F, Maurer W. Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. Stat Med. 2007 Sep 30;26(22):4063-73.

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

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

McCarrier KP, Atkinson TM, DeBusk KP, et al. Qualitative Development and Content Validity of the Non-small Cell Lung Cancer Symptom Assessment Questionnaire (NSCLC-SAQ), A Patient-reported Outcome Instrument. Clin Ther. 2016 Apr;38(4):794-810.

Tang, P. A., Pond, G. R., & Chen, E. X. (2010, Jan). Influence of an independent review committee on assessment of response rate and progression-free survival in phase III clinical trials. Ann Oncol, 21(1), 19-26.

Ye Y, Li A, Liu L, Yao B. A group sequential Holm procedure with multiple primary endpoints. Stat Med. 2013 Mar 30;32(7):1112-24

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

Shahbaz H, Gupta M. Creatinine Clearance. 2020 Sep 2. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing.

18 Appendices

18.1 IAP_MS200647-0037 List of outputs

The list of outputs will be provided in a separate Appendix of this IAP.

18.2 Toxicity grading for laboratory parameters according to NCI-CTCAE v. 5

Toxicity gradings will be described in a separate Appendix of this IAP.

18.3 Important Protocol Deviations

Important protocol deviations list is provided as a separate appendix of this IAP.

18.4 Adaptation Plan

A report of simulation results (adaptation plan) showing that the family-wise error rate is controlled under broad variation of assumptions is provided as a separate appendix of this IAP.

18.5 Literature (SAP template)

References for study specific literature are provided in Section 17.

GBS internal guidance documents:

- Standard analyses for Time to event endpoints In Oncology Trials from GBS Perspective
- Definition and Analysis of Best Overall Response in Oncology Trials from GBS Perspective
- Proposal for Primary Definition of Progression Free Survival (PFS) endpoint In Oncology Trials from GBS Perspective
- Guidance on subgroup analyses, interpretations and its consequences for future trial planning from GBS Perspective
- Standards for implementation of 3-tier Approach to Summarizing and Analyzing AEs in Clinical Trials and Safety summaries