agenus

STATISTICAL ANALYSIS PLAN PHASE 2

A PHASE 1/2, OPEN-LABEL, MULTIPLE ASCENDING DOSE TRIAL TO INVESTIGATE THE SAFETY, TOLERABILITY, PHARMACOKINETICS, BIOLOGICAL, AND CLINICAL ACTIVITY OF AGEN2034 IN SUBJECTS WITH METASTATIC OR LOCALLY ADVANCED SOLID TUMORS, WITH EXPANSION TO SECOND LINE CERVICAL CANCER

VERSION: 4

DATE OF PLAN: 8 FEBRUARY 2021

BASED ON:

Amendment 5, 27 September 2019

Amendment 4, 24 January 2019

Amendment 3, 26 September 2017

Amendment 2, 11 January 2017

Amendment 1, 07 December 2016

Original Version, 26 October 2016

STUDY DRUG: BALSTILIMAB (AGEN2034)

PROTOCOL NUMBER: C-700-01

SPONSOR:

Agenus Inc.
3 Forbes Road
Lexington, MA 02421-7305
[redacted]

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

APPROVALS

By signing this document, I acknowledge that I have read the document and approve of the planned statistical analyses described herein. I agree that the planned statistical analyses are appropriate for this study, are in accordance with the study objectives, and are consistent with the statistical methodology described in the study, and all applicable regulatory guidance and guidelines.

This document has been reviewed and accepted by:

[redacted] Agenus, Inc.	Signature [redacted]	Date [redacted]
[redacted] Agenus, Inc.	Signature [redacted]	Date [redacted]
[redacted] Agenus, Inc.	Signature [redacted]	Date [redacted]
[redacted] Agenus, Inc.	Signature [redacted]	Date [redacted]

TABLE OF CONTENTS

LIST OF TABLES

1. LIST OF ACRONYMS AND ABBREVIATIONS

Table 1: List of Acronyms and Abbreviations

Abbreviation	Term
ADA	Anti-Drug Antibody
AE	Adverse Event
AESI	Adverse Event of Special Interest
ATC	Anatomical Therapeutic Chemical
AUC	Area under the drug concentration time curve
BOR	Best Overall Response
CI	Confidence Interval
C_{max}	Maximum concentration
CPS	Combined Positive Score
CR	Complete Response
CTCAE	Common Terminology Criteria for Adverse Events
DCR	Disease Control Rate
DOR	Duration of Response
EAS	Evaluable Analysis Set
eCRF	electronic Case Report Form
EU	European Union
FAS	Full Analysis Set
HLGT	High-level Group Term
HLT	High-level Term
ICH	International Conference on Harmonization
IERC	Independent Endpoint Review Committee
irAE	Immune-related Adverse Event
IRAE	Infusion-related Adverse Event
ITT	Intention-to-Treat
MedDRA	Medical Dictionary for Regulatory Activities
Msec	Milliseconds
NCI	National Cancer Institute
NEC	Not Elsewhere Classified
ORR	Objective Response Rate
OS	Overall Survival

PD	Progressive Disease
PD-L1	Programmed Death-Ligand 1
PFS	Progression Free Survival
PK	Pharmacokinetic(s)
PP	Per Protocol
PR	Partial Response
PT	Preferred Term
QTc	QT Interval Corrected
RECIST	Response Evaluation Criteria in Solid Tumors
SAP	Statistical Analysis Plan
SD	Stable Disease
SMQ	Standardised MedDRA Queries
SOC	System Organ Class
TCR	Tumor Control Rate
TEAE	Treatment Emergent Adverse Event
ULN	Upper Limit of Normal
USA	United States of America

AMENDMENT FROM PREVIOUS VERSION

- Updated timing of filing analysis cutoff date, to provide minimum 12 months median follow-up and 6-months follow up on objective responses, as requested by FDA.
- Removed from SAP the exploratory analysis of AEs using sponsor broad and narrow definition of immune-mediated AEs — due to FDA request not to use a definition based on use of steroids.
- Divided cluster of immune-mediated Skin Disorders into Rash AEs and Pruritis AEs
- Flattened the hierarchical structure of clusters of immune-mediated AEs
- Specified that clusters of immune-mediated AEs will be evaluated during extended on-treatment period.
- Editorial changes/corrections

2. INTRODUCTION

The objective of this Statistical Analysis Plan (SAP) is to provide a comprehensive and detailed description of the strategy, rationale, and statistical techniques that will be used for analysis of Phase 2 data in order to address the primary and secondary objectives of the dose expansion part of the protocol, titled: "A Phase 1/2, Open-Label, Multiple Ascending Dose Trial to Investigate the Safety, Tolerability, Pharmacokinetics, Biological, and Clinical Activity of AGEN2034 in Subjects with Metastatic or Locally Advanced Solid Tumors, with Expansion to Second Line Cervical Cancer."

This document has been prepared based on protocol version 6 (amendment 5), dated 27 Sep 2019 and Case Report Form dated 18 Mar 2020. The statistical analyses adhere to principles specified in E9 guidance "Statistical Principles for Clinical Trials" of International Conference on Harmonisation (ICH).

2.1. STUDY DESIGN OVERVIEW

This study is conducted in 2 parts: a Phase 1, open-label, dose-escalation study in subjects with metastatic or locally advanced solid tumors, with a consecutive Phase 2 dose expansion to evaluate safety and efficacy in subjects with recurrent, unresectable, or metastatic (advanced) cervical cancer that has progressed on or after a platinum-based treatment regimen. The subsequent summary of design and objectives addresses only study design aspects relevant for Phase 2.

2.2. OVERALL STUDY DESIGN

Phase 2: Dose Expansion Phase

To further characterize safety and efficacy, subjects with recurrent, unresectable, or metastatic cervical cancer that has progressed after a platinum-based treatment regimen will be enrolled in Phase 2 and receive the recommended Phase 2 dose (RP2D) of balstilimab (3 mg/kg every 2 weeks) for a maximum of 24 months or until progression, unacceptable toxicity, or any criterion for stopping the study drug or withdrawal from the trial occurs. (…) an Independent Data Monitoring Committee (IDMC) will evaluate safety and efficacy.

The details of the investigational plan for Phase 2 are in the study protocol, including inclusion and exclusion criteria (Sections 4.2 and 4.3), criteria for subject withdrawal (Section 4.4), dosage and administration (Section 5.4), concomitant treatments (Section 5.5), management of immune-related adverse events (irAEs, Section 5.6), and schedule of trial assessments and procedures (Section 6).

3. STUDY OBJECTIVES AND ENDPOINTS

3.1. PRIMARY OBJECTIVE OF PHASE 2

- To assess the objective response rate (ORR) according to Response Evaluation Criteria in Solid Tumors (RECIST) 1.1, as determined by an Independent Endpoint Review Committee (IERC)

3.2. SECONDARY OBJECTIVES OF PHASE 2

- To assess the safety and tolerability of balstilimab in subjects with metastatic, locally advanced, and/or unresectable cervical cancer
- To characterize the balstilimab pharmacokinetic (PK) profile
- To evaluate the immunogenicity of balstilimab and correlate it to exposure
- To assess ORR according to RECIST 1.1 as determined by investigator
- To assess duration of response (DOR), disease control rate (DCR), duration of stable disease (SD), time to response, and progression-free survival (PFS) time per RECIST 1.1
- To assess overall survival (OS) rate
- To assess OS time

3.3. EXPLORATORY OBJECTIVES OF PHASE 2

- [redacted]
- Baseline expression of programmed death-ligand 1 (PD-L1) association with clinical outcomes

3.4. PRIMARY ENDPOINTS OF PHASE 2

- Confirmed ORR per RECIST 1.1, as determined by an IERC, in the analysis population.

Note: The population for primary efficacy analysis, specified in the study protocol, consists of "All subjects in Phase 2 who have received ≥ 1 dose of trial treatment and have measurable disease at baseline, according to the IERC assessment."

3.5. SECONDARY ENDPOINTS OF PHASE 2

- Frequency, severity, and duration of treatment-emergent adverse events (TEAEs) and laboratory abnormalities, using NCI-CTCAE v4.03
- Balstilimab PK parameters which may include (but are not limited to) maximum drug concentration observed postdose at steady-state ($C_{max\text{-}ss}$), minimum observed concentration at steady-state ($C_{min\text{-}ss}$), area under the drug concentration-time curve within time span t_1 to t_2 at steady-state ($AUC_{(t1\text{-}t2)\text{-}ss}$), area under the drug concentration-

time curve from time zero to time t ($AUC_{(0-t)}$), area under the drug concentration-time curve from time zero to infinity ($AUC_{(0-\infty)}$), time to maximum observed concentration (t_{max}), terminal disposition rate constant (λz), terminal elimination half-life ($t_{½}$), systemic clearance (CL), and volume of distribution (Vd)

Note: Analysis of PK parameters and correlation with anti-drug antibody concentrations is not covered in this SAP, it will be conducted according to a separate PK analysis plan.

- Anti-drug antibody (ADA) concentrations and correlation with AGEN2034 PK exposure metrics
- Confirmed ORR per RECIST 1.1, as determined by an investigator
- DOR per RECIST 1.1, as determined by an IERC and investigator, defined as time from first observation of response to first observation of documented disease progression (or death within 12 weeks after last tumor assessment). Subjects without an event at analysis cutoff date will be censored on date of last tumor assessment.
- DCR, defined as proportion of subjects with complete response (CR), partial response (PR), or stable disease (SD) for at least 12 weeks
- Tumor control rate (TCR), defined as proportion of subjects with CR, PR, or SD lasting at least 6 weeks

  Note: endpoint mot originally in the protocol, added in SAP
- Time to response, defined as the time from the first dose date to first observation of confirmed response.
- PFS time, defined as time from first treatment administration to first observation of documented disease progression (or death within 12 weeks after last tumor assessment), per RECIST 1.1, as determined by an IERC and investigator. Subjects without an event at analysis cutoff date will be censored on date of last tumor assessment.
- Median OS and OS rate
- OS time, defined as time from start of treatment to death. For subjects who are still alive at time of data cutoff for trial analysis or who are lost to follow-up, survival will be censored at the last recorded date that the subject is known to be alive as of the cutoff date for analysis.

3.6. EXPLORATORY ENDPOINTS OF PHASE 2

- [redacted]
- [redacted]
- [redacted]

Note: The analysis of exploratory endpoints is not covered in this SAP. If applicable, these analyses will be performed according to a separate Exploratory Statistical Analysis Plan.

4. PLANNED ANALYSIS

4.1. SUMMARY OF CHANGES TO THE PLANNED STATISTICAL ANALYSIS DESCRIBED IN THE PROTOCOL

1. Change of the name of analysis set:
   The name of the primary efficacy analysis set has been modified. Evaluable Efficacy Analysis Set is now referred as the Intention-to-Treat Efficacy Analysis Set (ITT EAS) to emphasize that inclusion in this analysis set is only based on the presence of measurable disease at baseline (per IERC) and no subject may be excluded from this set based on information collected after study treatment initiation.
   Note: The original term might be incorrectly interpreted as excluding subjects based on non-evaluable post-baseline data.
2. Definition of ITT Prior Line of Therapy (ITT PLT) Analysis Set. The SAP introduces the ITT PLT set, which includes all subjects with measurable disease and a prior platinum-based line of treatment in a recurrent/persistent/metastatic disease setting as determined by an Independent Review Committee. The analysis set will be used for secondary analysis of efficacy.
3. Definition of Per Protocol Efficacy Analysis Set (PP EAS): PP EAS is defined as a subset of ITT EAS with exclusion of subjects with major protocol deviations deemed likely to affect the efficacy outcomes; it will be used for sensitivity analysis of efficacy.
4. Removal the secondary endpoint of duration of stable disease specified in the protocol due to its similarity to progression-free survival (another secondary endpoint).
5. Second interim analysis:
   The first interim analysis was conducted as specified in the protocol. Then, study protocol prescribed the subsequent interim analysis to be "performed 3 months after approximately 100 subjects have been dosed." The intended goal of this analysis was to support potential Biologics License Application (BLA) submission under subpart E for accelerated approval. However, due to fast accrual in the final stages of study enrollment and feedback received from FDA, the plans for this analysis were modified: the interim analysis was replaced by the BLA Filing analysis which is to be performed using data of all subjects dosed for safety analysis and all subjects dosed with measurable disease for efficacy analysis to support regulatory filing (see Section 4.2 for planned timing of these analyses.
6. Inclusion of a new secondary endpoint (tumor control rate: TCR) defined as the rate of best overall response with either a confirmed response or stable disease at 6 weeks (see Section 7.7.5).

4.2. INTERIM ANALYSES AND DATA MONITORING

The Protocol specified that Phase 2 is designed to have at least 2 interim analyses of the primary endpoint and selected secondary efficacy and safety endpoints. No early stopping for efficacy will be performed, as a more complete assessment of safety and durability of responses will be required for evaluation of risks and benefits. Consequently, the overall type I error will not be affected by the interim analyses.

The first pre-specified interim analysis was performed based on the prospectively defined interim analysis SAP, when data were available for approximately 30 subjects treated for at least 3 months. The analysis included, for efficacy, all subjects dosed as of 15 Jul 2019 (N = 44), including N = 42 consecutive subjects with measurable disease at baseline for the primary analysis of efficacy. For safety, interim analysis included all subjects dosed by 16 Oct 2019 (the date of the data cutoff), N = 60 such subjects available at the time of the original analysis (however, it was discovered later during re-analysis of the same cohort, using data cutoff of 27 Mar 2020, that N=61 subjects had been actually dosed by 16 Oct 2019; one subject was not accounted for in the original safety analysis due to delayed data entry; the missing subject was included later in an updated analysis). The analysis was conducted according to a standalone Interim Analysis SAP. A non-binding futility assessment consisted of evaluation of conditional and predicted probability (He et al 2012) of excluding the ORR of 5% by the lower limit of the Wilson's confidence interval for observed ORR, if the study continued enrollment to 100 and 150 patients.

The second pre-specified analysis was originally planned 3 months after approximately 100 subjects have been dosed and was to include, as the primary analysis of safety and efficacy, all subjects in the evaluable efficacy analysis set among the first approximately 100 subjects dosed. The goal of this analysis was to support regulatory BLA filing. However, due to fast accrual at the final stage of enrollment (consequently short follow-up at the time of the planned analysis) and in consideration of feedback received from FDA, the plans for this analysis were modified: the analysis was replaced by BLA Filing analysis below described below.

BLA Filing analysis is planned to be performed on all subjects dosed using data cutoff of 11 Feb 2021, for at least 12-month median follow-up duration (counting from time of the first dose to data cutoff date) and 6-month follow-up on confirmed objective responses (counting from the start of the last objective response to data cutoff date). All subjects dosed will be used in safety analysis and all subjects dosed with measurable disease at baseline (per IERC) will be used for efficacy analysis.

4.3. FINAL ANALYSIS AND REPORTING

The final analysis will be performed after study completion and it will supplement the BLA Filing analysis.

All final planned analyses per protocol and this analysis plan must be performed after either locking records for relevant subjects or visits included in the analysis or the full database lock. See Section 7 for discussion of requirements for data freeze, locking, and source data verification in the context of the COVID-19 pandemic.

5. DEFINITIONS AND CONVENTIONS FOR DATA HANDLING

5.1. DEFINITION OF BASELINE

For all evaluations, unless otherwise noted, baseline is defined as the most recent non-missing measurement prior to the first administration of study drug.

5.2. DEFINITION OF STUDY DAYS

Unless otherwise noted, study days of an evaluation are defined as the number of days relative to the first dose date which is designated as Day 1, and the preceding day is Day -1, the day before that is Day -2, etc.

Study days are calculated as

- (Date of assessment – first dose date + 1) for assessments on/after first dose date
- (Date of assessment – first dose date) for assessments before first dose date

5.3. DEFINITION OF ON-TREATMENT PERIOD

On-treatment period is defined as the time from the first dose of study treatment until the end of treatment safety follow-up (4 weeks ± 7 days from the last dose of study drug, i.e., until 35 days from the last dose) or end of study, whichever occurs first.

5.4. DEFINITION OF AN EXTENDED ON-TREATMENT PERIOD

Extended on-treatment period (follow-up) is defined as the time from the first dose of study treatment through 90 days after the last dose of study treatment or end of study, whichever occurs first. This is used for reporting immune-related AEs and also as a sensitivity analysis for reporting all treatment emergent adverse events.

5.5. ANALYSIS VISIT WINDOW

Data will be analyzed by scheduled visits, where appropriate. Data collected out of the prescribed timing window will be analyzed as scheduled (i.e., at intended visits). Data that are collected from an unscheduled visit will not be included in the by-visit summary tables but will be presented in the listings. However, data collected at an unscheduled visit will also be considered for toxicity grading, in particular, for laboratory assessments. Additionally, both unscheduled and out of window imaging assessments will be included in efficacy analysis.

5.6. HANDLING OF PARTIAL DATES FOR ADVERSE EVENTS

When determining TEAEs, partial dates will be handled as follows.

- If the day of the month is missing, the onset day will be set to the first day of the month unless it is the same month and year as study treatment. In this case, the onset date will be assumed to be the first dose date of treatment.

- If the onset day and month are both missing, the day and month will be assumed to be 1 January, unless the event occurred in the same year as the study treatment. In this case, the event onset will be coded to the first dose date of treatment to conservatively report the event as treatment emergent.
- A missing onset date will be coded as the first dose date of treatment. If the resulting onset date is after a reported date of resolution, the onset date will be set equal to the date of resolution.
- A partially missing resolution date will be conservatively imputed as the latest date in the range of missing date element(s) (for example 2020-10-31 if the date is 2020-10-UNK), date of last dose + 90 days, day of study discontinuation, or death, whichever comes first.

Data listings will present the partial date as recorded in the electronic case report form (eCRF).

5.7. HANDLING OF PARTIAL DATES FOR MEDICATIONS

When determining prior or concomitant medications, partial start dates of prior and concomitant medications will be assumed to be the earliest possible date consistent with the partial date.

- Medication start dates with a missing day and non-missing month will be assumed to have occurred on the first day of the non-missing month, except for medications occurring in the first month of dosing, in which case the date will be one day before the first day of dosing.
- Medication start dates with missing day and month will be assumed to have occurred on the first day of the non-missing year (i.e., 1 January), except for medications occurring in the first year of dosing, in which case the date will be the informed consent date.
- Medications that are not ongoing and have a medication stop date with a missing day and non-missing month will be assumed to have stopped on the last day of the non-missing month.
- Medications that are not ongoing and have a medication stop date with a missing month will be assumed to have stopped on the last day of the non-missing year (i.e., 31 December).

5.8. HANDLING OF PARTIAL DATES FOR PROCEDURES

For records with a missing procedure date, the following procedure will be employed to determine whether the procedure is prior or concomitant:

- For a procedure with a missing or partially missing date performed on a scheduled visit, the date of the scheduled visit will be used.
- If a procedure was not scheduled, the missing part of the date will be imputed as discussed in Section 5.9.

5.9. HANDLING OF PARTIAL DATES FOR OTHER DATA

Other data (such as disease history) with partial dates will be listed as collected. For calculation of time intervals, the dates will be imputed with the first day of the month (if the day is missing) or with 1 January (if both day and month are missing).

5.10. HANDLING OF MISSING DATA

Missing data will not be imputed, except for:

- Missing date parts, i.e., partial dates, discussed above
- Missing AEs grade severity and relationship to study drug of AEs, discussed in Section 7.6.2
- Missing efficacy assessments discussed in Section 7.7.

6. ANALYSIS SETS

The following analysis sets will be used for analysis of Phase 2 data:

Table 2: Definitions of Analysis Sets

Analysis Set	**Definition**	**Analysis to Be Conducted**	**Interim Analysis**
All subjects analysis set	Includes all subjects who signed informed consent form (ICF), irrespective if they passed or failed screening for Inclusion and Exclusion criteria.	Will be used only for listing of all subjects who signed informed consent who either received study treatment or were excluded from the study during the screening period due to failure to meet eligibility criteria.	All patients who signed informed consent by the cut-off date for data.
Safety (Note: All dosed subjects)			
Safety analysis set (SAS)	Includes all subjects who received ≥ 1 dose of study treatment.	It will be used for safety evaluation, as the primary analysis in the final analysis of data and as a secondary set in in any interim analysis.	All patients dosed by the cut-off date for data.
ITT (Note: No exclusions based on information obtained after study treatment initiation)			
ITT efficacy analysis set (ITT EAS)	Includes all subjects who received ≥ 1 dose of study treatment, with measurable disease at baseline (per IERC).	Primary analysis set for efficacy.	All patients with measurable disease at baseline (per IERC) dosed by the cutoff-date for enrollment.

ITT Prior Line of Therapy analysis set (ITT PLT)	Includes all subjects who received ≥ 1 dose of study treatment, with measurable disease at baseline (per IERC), and had prior line of platinum-based treatment in metastatic, persistent or recurrent setting, as determined by Independent Review Committee	Secondary analysis set for efficacy	N/A
Per Protocol (Note: Possible exclusions based on information obtained after study treatment initiation)			
Per Protocol (PP) Efficacy Analysis Set (PP EAS)	Includes all subjects who received ≥ 1 dose of study treatment, had measurable disease at baseline by IERC, and were not excluded due to important protocol deviations.	The PP EAS set will be used for sensitivity analysis of the primary efficacy analysis.	All patients dosed by the cutoff-date for enrollment, had measurable disease at baseline by IERC, and were not excluded due to important protocol deviations (exclusions will be decided case by case).
Pharmacokinetic analysis set (PK analysis set)	Includes all subjects who have completed ≥ 1 infusion of study drug, and who have sufficient evaluable drug concentration measurements prior to and after treatment.	Pharmacokinetics Analysis Set will be used for pharmacokinetic analysis that will be conducted according to a separate pharmacokinetic analysis plan.	

6.1. ANALYSIS TO BE CONDUCTED FOR DIFFERENT ANALYSIS SETS

Table 3 below shows analysis sets used for different data omitting All Subjects Analysis Set (as this is only used for listing of screened and enrolled subjects) and PK Analysis Set (as this analysis will be conducted according to a separate PK analysis plan).

Table 3: Applications of Analysis Sets

Analysis	ITT EAS	ITT PLT	PP EAS	SAS
Subject Disposition	2	2		1

Protocol Deviations				1
Demographic and Baseline Characteristics	2	2		1
Medical History				1
Ongoing Medical History				1
Cancer History	1	1		2
Prior Cancer Therapies and Procedures	1	1		
Prior and Concomitant Medications				1
Drug Exposure				1
Adverse Events				1
Anti-Drug Antibodies				1
Clinical Laboratory Tests				1
Body Weight and Vital Signs				1
12-lead ECG				1
Pregnancy tests				1
ORR per RECIST 1.1	1	1	2	
Time to Response	1	1	2	
Duration of Response	1	1	2	
Disease Control Rate	1	1	2	
Tumor Control Rate	1	1	2	
Progression-Free Survival	1	1	2	
Overall Survival	2	2		1

Legend: 1 = Primary Analysis, 2 = Secondary Analysis

6.2. PROTOCOL DEVIATIONS AND EXCLUSIONS FROM ANALYSIS SETS

A protocol deviation is any change, divergence, or departure from the study design or procedures defined in the protocol. Important protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being (ICH E3 Questions and Answers R1).

Protocol deviations will be reviewed by study team to designate important protocol deviations. These important protocol deviations will be summarized by category, and all protocol deviations will be listed.

Important protocol deviations may lead to exclusions from Per Protocol analysis sets such as PP EAS. The exclusions will be finalized before clinical database is available for analysis.

7. STATISTICAL CONSIDERATIONS

All analyses described in this plan are considered a priori analyses as they have been defined prior to database availability for analysis. The COVID-19 pandemic may impact Agenus ability to complete data monitoring, collection, and cleaning to achieve the data lock as originally planned for our clinical programs. As a result, a priori analysis of Agenus clinical trial data may be completed prior to all data locked; however, all analysis will occur using locked and frozen clinical database records and ensuring trial data integrity is maintained as locked or frozen on the subject and visit levels. Clinical data management will notify the team when data are available (all locked or frozen) for download and the planned analysis. The impact of lack of source data verification or principal investigator (PI) signature will be performed at the time of data download. Only significant impact/risks to analysis of data, with the planned mitigation, will be discussed in the clinical study report (CSR). Clinical data management will keep a log of requested data changes after the data download for the analysis, result of completion of Source Data Verification (SDV) or new information provided. Requested changes will be reviewed by the team before changes are made to the database. An additional analysis may be performed if data changes would affect study conclusions; additional analysis will be described in the applicable SAP. All other analyses, if any, designed after completing analysis specified in this SAP and additional analysis plans referred by the SAP, will be considered post-hoc analyses and will be described in a separate exploratory analysis plan (if applicable).

All summaries and statistical analysis will be performed using SAS Software (SAS Institute Inc., Cary, NC, USA), version 9.4 or later, or R (R Foundation for Statistical Computing, Vienna, Austria), version 3.6.3 or later.

7.1. SAMPLE SIZE

No formal hypothesis testing is being planned for this Phase 2 study, rather the goal is to estimate the primary endpoint (ORR per IERC based on RECIST 1.1) with confidence intervals (CI), interpreted as a plausible range for a true (unobserved) ORR that could be interpreted considering historical data.

The sample size was planned to allow for exclusion of the ORR of 5% by the lower confidence limit Wilson score CI and to provide adequate safety data. With 150 subjects originally planned for the final analysis, the power to exclude an ORR of 5% by the lower limit of the 95% Wilson score interval is 92.2% and 96.2%, assuming a true ORR of 12% and 13%, respectively. The sample size for 150 subjects also provides ≥ 77% probability to observe an AE with an underlying rate of ≥ 1%.

7.2. GENERAL STATISTICAL PROCEDURES

Frequency distributions for categorical variables will be provided as number of subjects in the category and the percentages of the total number of subjects in the given population as noted. Counts of missing observations will be included in the denominator and presented as a separate category. Percentages will be reported to one decimal place.

A 2-sided 95% Wilson score CI for categorical variables without multiplicity adjustment will be provided where appropriate for efficacy analysis.

The descriptive statistics for continuous variables will be number of subjects, mean, standard deviation, median, minimum, and maximum. Mean and median will be reported to 1 more decimal place than the raw data, while the standard deviation will be reported to 2 more decimal places than the raw data. Minimum and maximum will be reported the same as the original data.

Summary tables may be replaced with listings when appropriate. For instance, an AE frequency table may be replaced with a listing if it only contains a few unique preferred terms (PTs) reported on relatively few subjects.

All listings (with the exception of All Subject listings) will be performed using the Safety Analysis Set and sorted by subject ID, ITT EAS, ITT PLT analysis set, and PP EAS membership (first subjects included in these sets) and, if applicable, by visit date unless otherwise specified.

The study drug name will be abbreviated in the tables, listings, and figures (TLFs) as “Bal” for balstilimab (AGEN2034).

Time intervals (in days), unless specified otherwise, will be calculated as difference between ending and start day + 1 day. For conversion to months, interval days will be divided by 30.4375.

7.3. SUBJECT ENROLLMENT

The number of subjects per region (USA, EU, other) and country in each analysis set will be summarized. In addition, a listing will be provided for All Subjects Analysis Set, including country, informed consent date, screening status (screen failure/enrolled), date of first dose, and analysis sets membership (Yes/No).

7.4. SUBJECTS DISPOSITION

Subject disposition will be summarized including

- Number of subjects treated
- Number of subjects dosed at least 12 months before data cutoff
- Median time and range (in months) from first dose until data cutoff
- Number of subjects who completed study treatment as per protocol
- Number of subjects with treatment ongoing at the time of analysis
- Number of subjects who discontinued treatment and primary reasons of discontinuation (with “Progression of disease (compared to baseline)” listed as “Progressive disease”).
- Number of subjects who completed the study as per protocol
- Number of subjects ongoing on study
- Number of subjects who discontinued study and primary reasons of discontinuation

7.5. DEMOGRAPHICS AND BASELINE CHARACTERISTICS

7.5.1. Demographic and Baseline Characteristics

Demographic and baseline characteristics will be tabulated using descriptive statistics. The following variables will be included:

- Age at screening (years)
- Age category (age < 55 vs age ≥ 55, age < 65 vs age ≥ 65, age < 75 vs age ≥ 75)
- Race (since race was not reported for patients treated in France, these patients will be summarized under separate "Not reported (France)" category)
- Ethnicity
- Height (cm)
- Weight (kg)
- BMI (body mass index, kg/m^2), defined as weight (kg) / height $(m)^2$
- Eastern Cooperative Oncology Group (ECOG) performance status

Conversions for height and weight are as follows:

- Height (cm) = Height (inches) x 2.54
- Weight (kg) = Weight (lb) x 0.4536

Demographics and baseline characteristics will be listed in Safety Analysis Set.

7.5.2. Medical History

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 22.1 or higher.

Conditions and symptoms ongoing immediately prior to first dose will be summarized as "ongoing medical history at baseline". Previous medical history data will be summarized separately as "medical history".

The frequency count and percentage of subjects experiencing any medical conditions will be tabulated by system organ classifications (SOC) and PT. If a PT or SOC was reported more than once for a subject, the subject would only be counted once in the incidence for that PT or SOC.

7.5.3. Cancer History and Characteristics

Subjects' cancer history prior to study entry will be summarized, including:

- Histological diagnosis at diagnosis classified as squamous, adenocarcinoma, adenosquamous, and other histology types.
- FIGO 2008 stage at diagnosis (In some cases FIGO data was not provided. To address this missing data point, for patients for whom AJCC-8 TNM staging was entered, but

FIGO is missing, TNM class will be mapped to FIGO using AJCC-8 algorithm described at the end of the section, both original and derived FIGO will be presented in listings).

- Time from original diagnosis to first dose date (years), calculated as (first dose date – initial diagnosis date)/365.25.
- PD-L1 tumor expression status in the most recent available pre-treatment biopsy (frequency and percentage of subjects classified as positive if combined positive score (CPS) ≥ 1%, and negative if CPS < 1%).
- Tumor burden (sum of baseline target lesions by IERC assessment) using descriptive statistics and frequency ≤ 50 mm vs > 50 mm.

The mapping of AJCC-8 TNM stage to FIGO will be performed as prescribed by American Joint Committee on Cancer:

Table 4: Mapping from AJCC-8 TNM to FIGO 2008 Stage

Stage 0	Tis	N0	M0
Stage I	T1	N0	M0
Stage IA	T1a	N0	M0
Stage IA1	T1a1	N0	M0
Stage IA2	T1a2	N0	M0
Stage IB	T1b	N0	M0
Stage IB1	T1b1	N0	M0
Stage IB2	T1b2	N0	M0
Stage II	T2	N0	M0
Stage IIA	T2a	N0	M0
Stage IIA1	T2a1	N0	M0
Stage IIA2	T2a2	N0	M0
Stage IIB	T2b	N0	M0
Stage III	T3	N0	M0
Stage IIIA	T3a	N0	M0
Stage IIIB	T1-T3	N1	M0
	T3b	Any N	M0
Stage IVA	T4	Any N	M0
Stage IVB	Any T	Any N	M1

Note: Any Nx, or Mx staging (that is, with unknown status in nearby lymph nodes or distant metastases) will be treated for the purpose of mapping as N0 or M0, respectively.

7.5.4. Prior Cancer Therapies and Procedures

Prior cancer systemic therapies will be summarized and listed including:

- Prior line of platinum-based systemic treatment in recurrent /persistent/metastatic setting as determined by Independent Review Committee
- Prior platinum-based treatment (string “plat” in reported or coded terms)

- Prior bevacizumab use (strings “bev” or “avastin” in reported or coded terms)
- BOR on the last treatment/therapy before enrollment
- Time from last day of last treatment to first dose date (months).

Listing will provide original term and PT coded using WHO Drug Sep 2019, the dates of start and end of therapy, location, intent, dose, BOR, and relapse date for each therapy recorded, as well as prior line of platinum-based systemic treatment in recurrent /persistent/metastatic setting as determined by Independent Review Committee.

Prior radiation therapies will be listed including the following information:

- Description
- Dose of radiation received
- BOR
- Time from last day of last treatment to first dose date (months).

Prior surgical procedures will be listed for Safety Analysis Set. Surgical procedures will be listed using both original entry as well as PT coded using WHO Drug Sep 2109.

7.5.5. Prior and Concomitant Medications

Prior and concomitant medications will be summarized and listed in Safety Analysis Set. Prior medication is defined as any non-study medication started before the first dose of balstilimab. Concomitant medication is defined as any non-study medication that was dosed on or after first study drug administration.

A prior medication could also be classified as “both prior and concomitant medication” if the end date is on or after first dose of balstilimab. In the listing, it will be indicated whether a medication is prior-only, concomitant-only, or both prior and concomitant medication.

7.6. SAFETY ANALYSIS

7.6.1. Drug Exposure

Balstilimab is administered intravenously (IV) every 2 weeks (Q2W).

Overall drug exposure will be summarized in Safety Analysis Set by the following parameters:

- Duration of treatment (week), calculated as: (last dose date – first dose date + 1)/7
- Number of balstilimab infusions received
- Total cumulative dose (mg), defined as the sum of the actual doses (mg) taken
- Dose compliance, defined as percentage of patients with any actual dose higher than 10%, 15%, or 20% or lower than 10%, 15%, or 20% relative to planned dose as entered on eCRF (Note: pharmacy manual specified that dose does not need to be adjusted for weight changes within 10%)

- Number of doses with infusion related reactions (as reported on drug administration eCRFs)
- Occurrences and number of actual doses exceeding adjusted planned dose by more than 5% (adjusting for up to 10 % weight loss.)

  Per pharmacy manual, no adjustment in dose was required for body weight changes within 10%, therefore dose exceeding planned dose by > 5% will be interpreted as any dose 5% greater than the highest daily dose prescribed by the clinical trial protocol, after accounting for up to 10% changes due to decrease in body weight. Adjusted planned dose is calculated by multiplying prescribed dose level by adjusted weight W_{adj}. Denoting by W_0 and W the baseline and the most recent weights, adjusted weight is: $W_{adj} = W$ for $W \geq W_0$ and $W_{adj} = \min(W_0, 1.1 \times W)$ for $W < W_0$.
- Subjects with any dose interrupted/not completed
- Number of doses interrupted/not completed

Missing weight measurements in drug exposure analysis will be imputed by last observation carried forward, multiple measurements for the same timepoint will be averaged.

Details of study treatment exposure will be listed for Safety Analysis Set.

7.6.2. Adverse Events

Adverse events (AEs) will be coded using MedDRA v22.1 or later and will be classified by SOC and PT.

Severity of AEs will be assessed by investigators according to NCI-CTCAE v4.03, and AEs will be classified by the investigator as either related or not related to study drug

Pre-existing conditions are those that started before and continued after signing of informed consent. Prior AEs (non-treatment emergent) are those occurring after the subject signed the informed consent and before the administration of the first dose of study treatment. Immune-related AEs will be evaluated in the extended on-treatment period to account for late toxicities. A separate listing will be provided for AEs incurred after on-treatment period.

TEAEs are AEs with onset during or after administration of first dose, or the worsening of a pre-existing condition during the on-treatment period. Immune-related AEs will be considered TEAEs during the extended on-treatment period. In a sensitivity analysis, TEAEs will be defined as AEs with onset during or after administration of first dose, or the worsening of a pre-existing condition during the extended on-treatment period. A separate listing will be provided of all AEs with onset date, or worsening date between end of “On-treatment period” (35 days) and end of “Extended on-treatment period” (90 days).

Any missing onset date, causality, or severity must be queried for resolution. Unresolved / missing causality and severity will be handled according to the following rules:

- An unresolved causality will be considered treatment-related
- An unresolved severity will be identified as an unknown severity

AEs with partial dates will be considered TEAEs unless the AE can unequivocally be determined as not treatment emergent.

The following AEs are considered as AEs of special interest (AESIs):

- Infusion-related reactions per definition in NCI-CTCAE v4.03 specification *(per investigator assessment)*
- Lab-based hepatic toxicity: An elevated aspartate transaminase (AST) or alanine transaminase (ALT) lab value ≥ to 3 × upper limit of normal (ULN) with a corresponding elevated total bilirubin value ≥ 2 × ULN, at the same time, an alkaline phosphatase lab value < 2 × ULN, as determined by way of protocol-specified laboratory testing or unscheduled laboratory testing
- Any AE suspected to be immune-mediated (i.e. an irAE) by the investigator

Unless otherwise stated, the number of subjects with an event (subject-wise) and the number of events (event-wise) will be provided for each summary.

Immune-related AEs will be tabulated using investigator-reported irAEs. .

Infusion related AEs will be tabulated using infusion-related reactions as reported by the investigators.

The following summaries by either SOC and PT or by PT and by toxicity grade will be provided. Summaries will be sorted by decreasing incidence of PT (within SOC, if applicable).

- TEAEs by SOC and PT, and by PT
- TEAEs by relationship to study drug, toxicity grade, SOC, and PT
- Drug-related TEAEs by SOC and PT, and by PT
- TEAEs leading to permanent treatment discontinuation by PT
- TEAEs leading to death by PT
- Drug-related TEAEs leading to permanent treatment discontinuation by PT
- Drug-related TEAEs leading to death by PT
- Serious TEAEs by SOC and PT, and by PT
- Serious TEAEs by relationship to study drug and toxicity grade, by SOC and PT
- Drug related serious TEAEs by SOC and PT, and by PT
- irAEs by SOC and PT, and by PT
- irAEs by relationship to study drug, toxicity grade, SOC, and PT
- Infusion-related TEAEs by SOC and PT, and by PT
- Infusion-related TEAEs by relationship to study drug, toxicity grade, SOC, and PT
- Laboratory abnormalities worsening (in at least one on-study measurement, relative to baseline) by NCI-CTCAE v4.03 criteria.
- ALT, AST, or bilirubin normal at baseline but with elevations ≥ 2 × ULN.

The overall summary of will be reported by age (age < 55 vs age ≥ 55, age < 65 vs age ≥ 65, and age < 75 vs age ≥ 75), by race group including at least 5 subjects (groups < 5 subjects will be reported summarily as “Other”, patients treated in France who did not report race will be summarized as “Not reported (France)”), region (US vs EU vs other), and by normal vs abnormal creatinine and bilirubin levels at baseline.

In addition, the overall summary of AE incidence per month of drug exposure will be presented. Such a summary will report #patients/(total time of exposure in months) and #AEs/(total time of exposure in months.)

The analysis of irAEs in the above summaries will be performed using data as reported by investigator (the primary analysis). In addition, as an exploratory analysis, TEAEs using sponsor's broad and narrow definitions of irAEs, will be summarized. For each irAE, the summary will include median time (and range) to onset, median duration (and range) of the first instance of such AE per patient. Finally, descriptive statistics will be provided for time of onset, cycle of onset, and cumulative dose at onset of first irAE of any type.

If an SOC or PT was reported more than once for a subject, the subject would only be counted once in the incidence for that SOC or PT.

In tabulation by severity grade:

- For a given SOC, only the most severe SOC for each subject will be included
- For a given PT, only the most severe PT for each subject will be included

In addition, TEAEs (evaluated within extended treatment period) will be reported for the following clusters of immune-mediated AEs: pneumonitis, colitis and diarrhea, hepatitis, and liver function abnormalities, endocrinopathies, nephritis and renal dysfunction, skin disorders, other immune-related AEs. These clusters will be defined by the following MedDRA version 22.1 Preferred Terms:

- **Pneumonitis**: interstitial lung disease; pneumonitis; acute interstitial pneumonitis; immune-mediated pneumonitis
- **Colitis:** allergic colitis, autoimmune colitis, colitis, colitis erosive, colitis microscopic, enterocolitis hemorrhagic, eosinophilic colitis, necrotizing colitis, neutropenic colitis, enterocolitis, immune-mediated enterocolitis,
- **Diarrhea:** diarrhoea, diarrhoea hemorrhagic
- **Hepatitis:** acute hepatic failure, autoimmune hepatitis, hepatic failure, hepatitis, hepatitis acute, hepatoxicity, liver injury, hepatocellular injury, immune-mediated hepatitis, immune-mediated hepatic disorder, drug-induced liver injury
- **Hypothyroidism:** autoimmune hypothyroidism, hypothyroid goiter, hypothyroidism, immune-mediated hypothyroidism, myxoedema, thyroid atrophy
- **Hyperthyroidism**: Basedow's disease, hyperthyroidism, Marine Lenhart syndrome, primary hyperthyroidism, immune-mediated hyperthyroidism, secondary hyperthyroidism, tertiary hyperthyroidism, thyroid dermatopathy, thyroid crisis, thyrotoxic periodic paralysis, toxic goiter, toxic nodular goiter,
- **Thyroiditis**: autoimmune thyroiditis, thyroiditis, thyroid acute, thyroiditis chronic, thyroiditis fibrous chronic, thyroiditis subacute, immune-mediated thyroiditis

- **Neuroendocrine Disorders (Pituitary-Adrenal insufficiency):** Addison's disease, adrenal androgen deficiency, adrenal insufficiency, adrenal suppression, adrenocortical insufficiency acute, glucocorticoid deficiency, hypoaldosteronism, mineralocorticoid deficiency, hypophysitis, lymphocytic hypophysitis, hypopituitarism
- **Diabetes** without prior history: type 1 diabetes mellitus, diabetes mellitus, latent autoimmune diabetes in adults, diabetic ketoacidosis.
- **Nephritis**: autoimmune nephritis, lupus nephritis, nephritis, nephritis hemorrhagic, immune-mediated nephritis
- **Rash**: pemphigoid, rash, rash erythematous, rash generalized, rash macular, rash maculo-papular, rash pruritic, rash papular, immune-mediated dermatitis
- **Pruritus**: pruritus, pruritus allergic, pruritus generalized

  Other Immune-related AEs: myocarditis, immune-mediated myocarditis, uveitis, immune-mediated uveitis, iritis, vitiligo, psoriasis, myositis, immune-mediated myositis, rheumatoid arthritis, systemic inflammatory response syndrome, sarcoidosis, autoimmune disorder, meningitis, encephalitis, encephalitis autoimmune, immune-mediated encephalitis, encephalopathy, autoimmune encephalopathy, immune-mediated encephalopathy, Guillain-Barre syndrome, myasthenia gravis, demyelination, stomatitis, pancreatitis, autoimmune pancreatitis, pancreatitis acute, immune-mediated pancreatitis, immune-mediated gastritis, lichen planus, colitis ulcerative, Crohn's disease, inflammatory bowel disease, hemolytic anemia, immune-thrombocytopenia, immune thrombocytopenic purpura, aplastic anemia, bicytopenia, pancytopenia, immune-mediated pancytopenia

For each cluster, the analysis will summarize: the total incidence and number of all TEAEs included in the cluster and by severity grade, % of pts who received treatment with pituitary and hypothalamic hormones and analogues, with corticosteroids for systemic use, or with thyroid therapy (concomitant medications in, respectively, H01, H02 and H03 ATC2 groups indicated for AE and started following AE onset but no later than the date of AE resolution or 21 days from AE onset, whichever comes earlier), % leading to dose hold, % leading to drug discontinuation, % resolved, time to onset of first TEAE and duration of all TEAEs and TEAEs of grade ≥ 3.

The median duration of TEAE (in days) will be calculated using Kaplan Meier analysis for all patients and AE included in a cluster. Unresolved AEs or AEs with missing resolution information and date will be treated as censored data with the end of extended on-treatment period as censoring date.

Death summary table will present reasons for deaths on treatment period, in survival follow-up, and overall.

Death details will also be listed (if applicable) including:

- Number of doses of balstilimab
- Number of cycles

- Age at death
- Study days relative to first dose date, calculated as (death date – first dose date + 1)
- Number of days to death from last dose, calculated as (death date – last dose date + 1)
- Precipitating AE (if death is an outcome of AE)
- Immune-related AE (if death is an outcome of irAE per investigator)
- Cause of deaths

The following listings will be provided in Safety Analysis Set:

- All AEs (with TEAEs flagged)
- Related TEAEs
- Serious TEAEs
- TEAEs leading to treatment discontinuation
- TEAEs leading to death
- Death details (including age, study day, and days from last dose, AEs, and death cause).

Adverse events observed between signing the informed consent and first dose of study drug will be listed with negative study days.

7.6.3. Anti-Drug Antibodies

Results of the ADA assay including titers will be listed and the number of subjects with a confirmed positive sample will be summarized. Neutralizing antibody assays results will be listed for positive ADA samples.

7.6.4. Clinical Laboratory Tests

For clinical laboratory parameters, descriptive summary tables for observed values and changes from baseline will be provided by visit.

Laboratory values will also be categorized according to their NCI-CTCAE v4.03 toxicity grade, for the following parameters: ALT, AST, Bilirubin, Creatinine, Hemoglobin, Platelets, Neutrophils, and Lymphocytes. For applicable parameters, shift tables will be presented from the baseline toxicity grade to the worst post-baseline visit value (scheduled or unscheduled). All test results and associated normal ranges from local laboratories will be reported in SI units. All tests with numeric values will have a unique unit per test. Any laboratory test results and associated normal ranges from local laboratories will be converted to SI units. For the limited number of cases where the associated normal ranges from a local laboratory cannot be obtained despite due diligence, if query is unsuccessful at resolving issue and analysis is mandatory then the clinical scientist and medical monitor can provide a suitable normal range to be used in determining CTC grading and flags for above and below normal and the inferred normal range will be flagged in listings.

If a test value is reported with "<" or ">" sign (as below or above some detection limit), the test value will be imputed in summaries at the limit of detection (that is ignoring "<" or ">" signs), however the measurement will be listed as originally reported.

A standard 'eDISH' (evaluation of Drug-Induced Serious Hepatotoxicity) plot will be presented to display on the log-log scale the correlation between peak total bilirubin and alanine aminotransferase (ALT), both in multiples of ULN, with horizontal and vertical lines for Hy's law thresholds, i.e., 3×ULN for ALT and 2×ULN for total bilirubin.

Listing will be provided for Safety Analysis Set.

7.6.5. Body Weight and Vital Signs

The observed value and percentage of changes from baseline will be summarized and listed by visit for Safety Analysis Set. If there are multiple measurements per visit, these measurements will be averaged.

Maximum weight loss from baseline will be summarized as 5 to <10% loss from baseline; 10 – <20% loss from baseline; and ≥ 20% loss from baseline.

7.6.6. 12-lead Electrocardiogram

Number of subjects with abnormal clinically not significant and abnormal clinically significant electrocardiogram (ECG) results, as well as number of subjects with QTc_F (QT interval corrected using Fridericia's formula) greater than 450, 480, or 500 milliseconds will be summarized by visit and timing with respect to drug dose. In addition, QTc_F changes from pre-dose will be summarized for number of patients with a change exceeding a threshold of 30 msec or 60 msec.

Other ECG parameters will be reported and summarized. Hart rate (HR) will be summarized as for number of patients ≤ 50 bpm and ≥ 120 bpm, PR interval for number of patients with PR ≥ 220 msec, and duration of QRS complex for number of patients with QRS ≥ 120 msec. RR intervals, when missing, will be calculated from HR using the formula: RR = 60000/HR (in milliseconds, assuming HR in beats per minute). Then QT interval corrected using Fridericia's formula will be calculated as $QTc_F = QT/(RR/1000)^{1/3}$ (i.e, QT interval in msec divided by cube root of RR interval in seconds).

In all these summaries, QT, PR, QRS durations, and HR will be averaged for ECG taken in triplicate.

7.6.7. Pregnancy tests

Dates and results of pregnancy tests will be listed within Safety Analysis Set.

7.7. EFFICACY ANALYSIS

7.7.1. Best Overall Response and Objective Response per RECIST 1.1

Best overall response (BOR) will be derived from time-point RECIST 1.1 assessments, either by IERC or by investigator. For the primary endpoint, tumor responses will be evaluated using IERC assessments. In the primary analysis, BOR will be evaluated over all on-study time-point assessments recorded before start of a new anti-cancer therapy, in consideration of RECIST 1.1 definition: "The best overall response is the best response recorded from the start of the study treatment until the end of treatment taking into account any requirement for confirmation. On occasion a response may not be documented until after the end of therapy so protocols should be clear if post-treatment assessments are to be considered in determination of best overall response" (Eisenhauer et al 2009). Since all assessments are considered for BOR, the confirmed objective response or stable disease (SD) may occur after a transient time-point progressive disease (PD) assessment(s).

In the secondary analysis, conducted for sensitivity evaluation, an alternative, more conservative approach will be used, evaluating BOR only over time-point assessments until first observed PD, as initially proposed in RECIST 1.0 (Therasse et al 2000). Such a definition was implemented in the Charter of IERC, therefore BOR evaluation derived by IERC will be used as sensitivity analysis.

Confirmation of objective response (complete response [CR] or partial response [PR]) is required for respective BOR of CR or PR. Confirmation must be obtained at a subsequent assessment at least 4 weeks apart. In case of CR, the confirmatory assessment must be performed at a consecutive assessment, except for intercurrent non-evaluable assessments.

In case of PR, confirmatory assessment after initial CR/PR may be obtained a) in the primary analysis any time at least 4 weeks apart from the initial assessment, without restrictions on intercurrent assessment b) in the IERC-derived BOR evaluation, any time after 4 weeks and after at most 1 intercurrent assessment of non-evaluable (NE) or SD.

For patients who did not fulfill requirements for BOR of CR or PR as described above, BOR of SD is defined, in the primary analysis, as i) initial time point assessment(s) of SD, PR, CR without PD at a minimum interval of 39 days (6 weeks-3 days), disregarding NE assessments and/or ii) unconfirmed PR or CR any time at initial evaluation, without PD at a minimum interval of 39 days and/or iii) durable SD following transient PD: at least 2 assessment(s) of SD, PR, CR (ignoring intercurrent NE assessments) at least 39 days (6 weeks-3 days) apart and without new PD within 84 days, counting time from the last transient PD assessment. Subjects who have a response of PR or CR with no subsequent tumor assessments will be assigned a BOR of SD and will be flagged in listings. In the sensitivity analysis performed using IERC-derived BOR evaluation of SD is only considered based on only criterion i) above.

For patients with time point assessment(s) of PD, excluding any transient PDs (PD followed by confirmed objective response or durable SD) and who do not satisfy criteria for BOR of CR, PR, and SD described above, the BOR assessments in the primary analysis will be PD. In the

sensitivity analysis, using IERC-derived BOR evaluation, a similar definition will be used but without excluding transient PDs.

Finally, BOR of NE will be assigned where there are either no evaluable time point assessments at all or first time point assessment(s) of SD was before minimum interval of 39 days without evaluable assessments preformed after 39 days.

The objective response rate (ORR) is the percentage of subjects with a BOR of PR or CR (i.e., objective response). The primary analysis of ORR will include only confirmed responses, with sensitivity analyses based on unconfirmed responses, derived from timepoint assessments according to RECIST 1.1 guidance for studies not requiring confirmation of response (i.e., occurrence of time-point assessment of CR or PR anytime on study). Subjects with missing tumor response status assessment will be counted as non-response (that is, subjects without any follow-up assessment) will be counted in the denominator for ORR.

BOR determined by investigator and by IERC will be compared by cross-tabulating categories of both BOR evaluations in a 2-way table.

BOR and ORR will be listed for Safety Analysis Set as available using primary and alternative definitions. For confirmed response, duration of response will be listed (see Section 7.7.3). The listings of time-point response will include assessments of target lesion, non-target lesions, and new lesion data, as well as percentage change from baseline in the sum of diameters of target lesions.

7.7.2. Time to Response

Time from treatment initiation to start of objective response (in days) will be summarized using descriptive statistics in subjects with confirmed CR or PR.

7.7.3. Duration of Response

Duration of response (DOR) will be analyzed only for subjects who responded to the study treatment. DOR is defined as the time from first observation of CR or PR which was subsequently confirmed until the first time of disease progression or death by any cause within 12 weeks of last tumor assessment. All assessments performed before initiation of a new anti-cancer therapy will considered for evaluation of DOR. Similar to evaluation of BOR, in the primary analysis any transient time-point PD(s) will be disregarded for determination of DOR when followed by confirmed objective response or durable stable disease, as defined in section 7.1.1. In the alternative (sensitivity) analysis, DOR will be counted using the date of first observed PD, whether transient or not.

The starting event for DOR will be the date of first objective response. The terminating event will be date of PD or death within 12 weeks of last tumor assessment, whichever occurs first. The censoring event will be the earliest of the dates of last evaluable assessment, last evaluable assessment performed before initiation of new anti-cancer therapy (if applicable), or before 2 or more missed imaging assessments (i.e., no assessment within 90 days) followed by PD or death.

DOR will be reported in months and summarized by quartiles estimated from Kaplan-Meier curve, as well as Kaplan-Meier estimates at 6, 9, 12, 18 months. Number and percent of censored and non-censored observations and a plot Kaplan-Meier curve will be provided.

In addition, duration of observation of DOR (whether censored or not) will be summarized as "Observed DOR", including the shortest and longest observation will be provided with censoring indicator.

7.7.4. Disease Control Rate

Disease control rate (DCR) is defined as proportion of subjects who have a confirmed response (CR or PR) or SD without PD within 81 days (12 weeks – 3 days) at study start, or durable SD following PD. In the alternative analysis, used for sensitivity analysis in parallel with alternative algorithm for BOR, described previously it is defined as proportion of subjects who have a confirmed response (CR or PR) or SD without PD within 81 days at study start.

DCR and number of subjects with disease control will be tabulated using IERC and investigator assessments.

7.7.5. Tumor Control Rate

Tumor control rate (TCR) is defined as proportion of subjects who have BOR of either SD (as defined previously) or a confirmed objective response (CR or PR).

TCR and number of subjects with tumor control will be tabulated using IERC and investigator assessments.

7.7.6. Efficacy Analysis Plots

The following plots will be provided for Safety and Efficacy Analysis Set, respectively:

- Waterfall plots of the best percentage change in tumor size (i.e., maximum tumor reduction, or minimum increase in the absence of any reduction) (Gillesp et al 2012).
- Swim lane plot for responders in ITT EAS and ITT PLT sets.

7.7.7. Sensitivity Analyses of Efficacy

The analysis of efficacy per IERC in ITT EAS and ITT PLT, including summary of BOR, ORR, DCR, TCR, DOR will be performed:

- Using alternative definition of BOR, ORR, DOR, and PFS based on evaluation per IERC
- Using a subset of patients with minimum expected follow up (i.e., time from first dose until data cutoff) of at least 12 months (analysis of BOR, ORR, DCR, and TCR only)

Sensitivity analysis will be also performed by using PP EAS.

7.7.8. Subgroup Analyses of Efficacy

Summaries of BOR, ORR, DOR, DCR, and TCR and the waterfall plot, for IERC assessment will be reported in the following subgroups:

- by PD-L1 status (positive, negative, unknown)
- by histology (squamous, adenocarcinoma, other)
- by prior use of bevacizumab (yes vs no)
- by region/country (regions: USA, EU, other)
- by age groups (age < 55 vs ≥ 55, age < 65 vs age ≥ 65, age < 75 vs age ≥ 75)
- by race group including at least 5 subjects and grouping together all categories with < 5 subjects as "Other" (since race was not reported for patients treated in France due to local regulations these patients will be summarized as "Not reported (France)" category).
- by tumor burden (sum of baseline target lesions by investigator assessment) ≤ 50 mm vs > 50 mm

Sensitivity analysis by PD-L1 status in ITT AES and ITT PLT analysis sets will be performed both for primary and alternative definitions of BOR, ORR, DOR, and PFS.

These subgroup evaluations will apply to primary analyses in ITT EAS and to secondary analysis in ITT PLT analysis sets.

7.7.9. Progression-Free Survival

Progression-free survival (PFS) is defined as the interval from the date of first dose of investigational agent until the earliest date of PD, as determined by IERC and investigator assessment of objective radiographic disease assessments per RECIST 1.1, or death due to any cause if occurring sooner than progression.

The primary analysis will be performed ignoring transient PDs, using the same definition of transient PDs as in section 7.7.1. Thus, the terminating event will be a non-transient PD or death, whichever occurs first. Patients who did not experience a non-transient PD or death will be censored on the date of last evaluable tumor assessment or the date of initiation of new anti-cancer therapy (if applicable), whichever comes earlier.

The alternative (sensitivity) analysis will be performed for terminating event of death or first observed time-point PD, whichever occurs first. Censoring will be on the date of last evaluable tumor assessment, the date of the last evaluable assessment before two or more consecutive missed imaging visits, or the date of initiation of new anti-cancer therapy (if applicable), whichever occurs first.

For ITT EAS and ITT PLT analysis sets, PFS will be summarized using PD assessments by IERC and by Investigator.

PFS will be reported in months. Median PFS will be provided (estimated using K-M estimate, with 95% CI) and Kaplan-Meier curves of PFS will be plotted using ITT EAS and ITT PLT.

Listing will be provided for Safety Analysis Set.

7.7.10. Overall Survival

Overall survival (OS) is defined as the interval from the date of first dose of investigational agent until the date of death.

OS in months is calculated as:

- (date of death, or censoring – date of first dosing + 1) / 30.4375

Where the censoring rules are specified as following:

- For subjects who have not died, OS will be censored at last contact date.

Median OS will be provided (estimated using K-M estimate, with 95% CI) and Kaplan-Meier curve of OS will be plotted for Safety Analysis Set.

Listing will be provided for Safety Analysis Set.

7.8. PD-L1 Expression Status

PD-L1 tumor expression status at baseline, an exploratory endpoint in this study, will be summarized as counts and percentage of PD-L1 positive and negative tumors (classified as positive if combined positive score [CPS] ≥ 1%, negative if CPS < 1%, unknown if not available or not evaluable). PD-L1 expression status will be used for subset analysis of BOR, ORR, DOR, and DCR, as described in Section 7.7.8. The summaries of PD-L1 expression status will be included as part of tumor baseline characteristic.

Listing of CPS and derived PD-L1 expression status will be provided for Safety Analysis Set. The listing will also include data about baseline biopsy sample collection date, whether the sample was FFPE block, test date, sample stability status, and the reason for non-evaluable samples (tumor samples are considered to be within stability period for PD-L1 assay if the time interval between sample collection and testing does not exceed 5 years for formalin-fixed paraffin-embedded (FFPE) blocks and 6 months for slides).

7.9. Pharmacokinetic Parameters

PK analyses will be conducted according to separate Pharmacokinetic Analysis Plan (PAP).

8. REFERENCES

He P, Lai TL, and Liao OY. “Futility Stopping in Clinical Trials.” Statistics and Its Interface 5, no. 4 (2012): 415–23. https://doi.org/10.4310/SII.2012.v5.n4.a4.

Eisenhauer E.A., Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, et al. “New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1).” *European Journal of Cancer* 45, no. 2 (January 2009): 228–47.

Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, Van Oosterom AT, and Christian MC. “New Guidelines to Evaluate the Response to Treatment in Solid Tumors.” Journal of the National Cancer Institute 92, no. 3 (2000): 205–216. https://doi.org/10.1093/jnci/92.3.205.

Gillespie TW. “Understanding Waterfall Plots.” Journal of the Advanced Practitioner in Oncology 3, no. 2 (2012): 106–11.

U.S. Department of Health and Human Services Food and Drug Administration: Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) (2015) Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics Guidance for Industry